ОБРАЗОВАНИЕ, НАУЧНИ ИЗСЛЕДВАНИЯ И ИНОВАЦИИ

Научно списание
Година IV, книжка 2, 2026

EDUCATION, SCIENTIFIC RESEARCH AND INNOVATIONS

Scientific journal
Vol. IV, Issue II, 2026

ЗА БУКВИТЕ О ПИСМЕНЕХЪ

Профил на списанието

„Образование, научни изследвания и иновации“ е двуезично (на английски и български език) научно списание, което се издава от Академично издателство „За буквите – О писменехъ“ – гр. София, България. Изданието цели да популяризира висококачествени теоретични и приложни изследвания. Тематични рубрики на списанието са:

- История и археология;
- Обществени комуникации и информационни науки;
- Информатика и компютърни науки;
- Национална сигурност;
- Актуално.

Списанието отразява иновативни идеи, научни резултати, съвременни тенденции и перспективи за развитие в изброените научни области. Публикуват се само оригинални научни статии, преминали двойно сляпо рецензиране.

Списание „Образование, научни изследвания и иновации“ е със свободен достъп.

https://e-journal.unibit.bg/
ISSN 2815-4630
https://doi.org/10.70300/WTRT9015

Journal Scope

“Education, Scientific Research and Innovations” is a bilingual (in English and in Bulgarian) scientific journal, which is published by the Academic Publisher “Za Bukvite – O Pismeneh” – Sofia, Bulgaria. The publication aims to promote high-quality theoretical and applied research works. The thematic sections of the journal are:

- History and archaeology;
- Public communications and information sciences;
- Informatics and computer sciences;
- National Security;
- Current topics.

The journal reflects innovative ideas, scientific results, modern trends and prospects for development in the listed scientific fields. Only original scientific articles that have passed double-blind peer review are published.

The “Education, scientific research and innovation” journal is open access.

https://e-journal.unibit.bg/
ISSN 2815-4630
https://doi.org/10.70300/WTRT9015

ГЛАВЕН РЕДАКТОР

доц. д-р Елена Савова
Факултет по библиотекознание
и културно наследство
Университет по библиотекознание
и информационни технологии
бул. „Цариградско шосе“ № 119,
София 1784, България
тел.: +359 878 97 03 40
е-поща: e.ignatova@unibit.bg

EDITOR-IN-CHIEF

Assoc. Prof. Elena Savova, PhD
Faculty of Library Studies
and Cultural Heritage
University of Library Studies
and Information Technologies
119, Tsarigradsko Shosse Blvd.
Sofia 1784, Bulgaria
tel. +359 878 97 03 40
E-mail: e.ignatova@unibit.bg

ИЗДАТЕЛ

Академично издателство
„За буквите – О писменехъ“
Университет по библиотекознание
и информационни технологии

PUBLISHER

Academic Publisher
“Za Bukvite – O Pismeneh”
University of Library Studies and Information
Technologies

ДИРЕКТОР

доц. д-р Диана Стоянова
бул. „Цариградско шосе“ № 119,
ет. 2, стая 213
София 1784, България
тел.: +359 879 14 83 85
е-поща: d.stoyanova@unibit.bg

DIRECTOR

Assoc. Prof. Diana Stoyanova, PhD
119, Tsarigradsko Shosse Blvd.
fl. 2, room 213
Sofia 1784, Bulgaria
tel.: +359 879 14 83 85
E-mail: d.stoyanova@unibit.bg

Списание „Образование, научни изследвания и иновации“ излиза четири пъти годишно:
книжка 1 – януари – март;
книжка 2 – април – юни;
книжка 3 – юли – септември;
книжка 4 – октомври – декември.

Journal “Education, Scientific Research and Innovations” is published four times a year:
book 1 – January – March;
book 2 – April – June;
book 3 – July – September;
book 4 – October – December.

С изпращането на текст и илюстрации до Академично издателство „За буквите – О писменехъ“ авторът се съгласява да преотстъпи правото за публикуването, анонсирането и разпространението им за нуждите на всички издания на Академично издателство „За буквите – О писменехъ“. Материали, които не са одобрени за публикуване, не се редактират и не се връщат на авторите.

By sending texts or illustrations to the Academic Publisher “Za Bukvite – O Pismeneh” the author agrees to submit the copyright for publishing, dissemination and announcing in all Academic Publisher “Za Bukvite – O Pismeneh” editions. Materials that are not approved for publication are not edited and are not returned to the authors.

СЪДЪРЖАНИЕ

Обществени комуникации и информационни науки

CONTENTS

Public Communications and Information Sciences

ОБЩЕСТВЕНИ КОМУНИКАЦИИ И ИНФОРМАЦИОННИ НАУКИ
PUBLIC COMMUNICATIONS AND INFORMATION SCIENCES

THE VOPA+ MODEL AS A GUIDELINE FOR TRANSFORMATIVE LEADERSHIP

Alexander Manuel Boschuk
University of Library Studies and Information Technologies
https://doi.org/10.70300/ABLT2026

Abstract: *Digital transformation and increasing volatility are presenting new challenges for managers and leaders. Traditional leadership approaches are reaching their limits in VUCA environments. This article analyzes Buhse and Petry's VOPA+ model as a framework for transformative leadership and examines its relevance for change management processes. Based on a systematic literature analysis of German and international publications on digital leadership and VUCA environments, the five principles of the model are examined: networking as the basis for digital collaboration, openness as the key to innovation, participation to involve employees, agility for flexibility in dynamic environments, and trust as a fundamental element. The analysis shows that the five dimensions are closely interlinked and that trust serves as a foundation without which the other principles are ineffective. Implementation requires a comprehensive cultural change that goes beyond merely introducing digital tools. Managers must change their role from that of the sole decision-maker to that of facilitator and enabler. The VOPA+ model offers a valuable framework for transformation processes, although the specific design must be context-dependent, depending on the industry, size and organizational culture. Further empirical research on the measurability of success is necessary.*
Keywords: *VOPA+; VUCA; Digital Leadership; Transformation; Change Management*

INTRODUCTION

The current working world is characterized by increasing uncertainty and volatility. A global survey of 4,701 CEOs conducted by PricewaterhouseCoopers clearly shows that macroeconomic fluctuations (35%), geopolitical tensions (27%) and cyber risks (24%) are perceived as key threats for 2025 (PricewaterhouseCoopers 2025). German CEOs also see these risks as particularly relevant, with the shortage of skilled workers being highlighted at 21 per cent (ibid.).

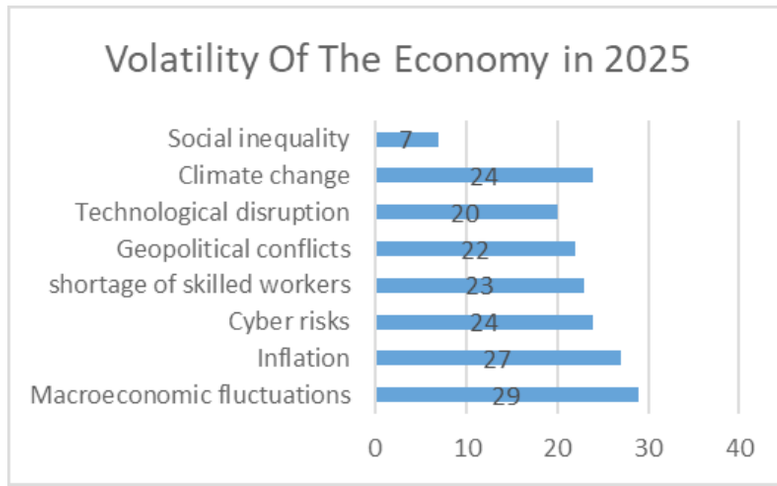


Fig. 1. CEOs concerned about economic volatility
Source: PwC Annual Global CEO Survey 2025

Not only global CEOs, but also German CEOs are feeling a growing sense of uncertainty. In January 2026, 84% of the German population stated that the current situation was cause for concern (Infratest dimap 2026). This mood and these developments are reflected in companies. In

2016, over 58% of executives surveyed in a study stated that the complexity of collaboration was a particular challenge (Hays 2017, p. 15). Increasing digitalization and flexible working models, generally referred to as "new work", also pose new tasks for managers.

Current studies show that today's working world and its environment can no longer be described as purely linear or predictable, and this is also how those affected and involved perceive it. The working world is characterized by volatility, uncertainty, complexity and ambiguity. The VUCA framework provides a systematic approach to understanding the challenges of modern leadership (Lippold 2019, p. 25). These dilemmas can be exacerbated by the challenges of digital transformation. According to McKinsey, 90 per cent of companies worldwide are now implementing digital transformation in some form (Rachytski and Runets 2025). In Germany, the DIHK Digitalization Survey 2025 shows that many companies are actively engaged in transformation processes, with 54 per cent citing complexity and 60 per cent citing a lack of resources as key challenges (DIHK 2026). This is where traditional management approaches, which originated in other eras such as the second or third industrial revolution, reach their limits and fade into the background.

New leadership approaches are replacing outdated models because they appear better suited to today's requirements. These new leadership models, such as digital leadership, combine technological, strategic and emotional skills, as they are particularly relevant (Boschuk 2024, p. 7). The VOPA+ model, introduced by Willms Buhse and expanded by Thomas Petry, offers a systematic framework for modern leadership. The aim of this article is to analyze the model's principles and highlight their relevance to the design of change management processes.

This article will first provide a rough overview of new leadership approaches and describe the VOPA+ model, as well as substantiate its relevance for the current challenges faced by managers. Recommendations for action round off this work.

RESEARCH METHODOLOGY

This article is based on a systematic analysis of literature and publications from German-speaking and Anglo-Saxon countries on digital leadership, the VUCA environment, and VOPA+. The acronym VOPA is only found in German-language literature, which is why the analysis was supplemented with the term's ambidextrous leadership and ambidexterity. The theoretical foundations are supported by selected sources from the current scientific discussion, such as specialist journals and online articles.

RESULTS

New leadership approaches in the context of digital transformation

The demands of digital transformation have triggered a fundamental reassessment of traditional leadership approaches. Leadership research distinguishes between various models, each with a different focus, and understanding these models makes it easier to classify the VOPA+ model. The transactional leadership approach is based on the principle of exchange. Performance is rewarded and deviations from this are corrected (Bass and Riggio 2006, p. 8). It is suitable for stable, clearly structured environments, but reaches its limits in dynamic transformation processes. The transformational leadership approach goes beyond this. It aims to change employees' values and attitudes, emphasizing vision and inspiration. Individual development is also anchored in this leadership approach (ibid., p. 6). Situational leadership models, such as that of Hersey and Blanchard (1969), emphasize the need to adapt the leadership style to employees' maturity level (Lippold 2019, p. 21). Servant leadership views the manager primarily as an enabler who puts employees' needs first (Simsa, p. 140).

These approaches form the basis for digital leadership as an integrative approach to the present. Digital leadership is not an entirely new theoretical construct, but rather a further development of proven leadership concepts for the digital age. It combines technological competence and strategic foresight and is thus specifically geared towards volatile, networked and rapidly changing organizational environments (Petry 2024, p. 56). The VOPA+ model operationalizes digital leadership by identifying five specific leadership dimensions that managers can use to guide their actions in the digital transformation process.

The principles of the VOPA+ model

The VOPA+ model is a framework developed in response to the new challenges of a world characterized by VUCA or BANI factors. The original VOPA model was presented by Willms Buhse in his 2014 book "Management by Internet: New Leadership Models for Companies in Times of Digital Transformation". It describes the key principles highly relevant to managers in an increasingly digitalized and networked world. The acronym VOPA stands for networking, openness, participation and agility (Buhse 2014):

Networking as the basis for digital collaboration in the context of leadership

Networking is a central component of digital transformation. It encompasses interaction between people as well as connections between humans and machines or between machines themselves, for example in the Internet of Things (Buhse 2014, p. 24). Comprehensive internet-based networking, now a global infrastructure for information exchange, serves as the basis for countless other technologies (Baltra et al. 2023, p. 2). Social media platforms such as Instagram, LinkedIn and X build on the basic technology of the Internet and not only enable private and professional communication, but also promote the establishment of networks, knowledge transfer and the organization of global communities (Lippold 2019, p. 27). Cloud technology and collaboration software, such as Confluence, are also noteworthy in a professional context as tools for networking and promoting digital collaboration. This networking, accelerated by digitalization, now enables global interaction, collaboration, and knowledge integration. Knowledge can thus be used today regardless of location. In addition, young technologies such as blockchain enable highly networked storage of data and information. They are increasingly influencing both the world of work and the private sphere. This technological dimension of networking is therefore crucial to modern leadership approaches. Nevertheless, people remain at the heart of successful leadership. An organizational culture is predominantly shaped by consistent interpersonal relationships (Schein and Schein 2017, p. 7), and trust between managers and employees is rooted in interpersonal interaction (Mayer, Davis and Schoorman 1995, p. 711). In the context of this article, it should be emphasized that these forms of networking among managers must be used strategically to enable not only technological but also interpersonal collaboration.

Openness of leaders as the key to innovation

In this context, openness, or "opening up," refers to transparency and a willingness to share knowledge and embrace new ideas. In earlier times, when processes and structures were less complex, it was not necessary to share knowledge to ensure a company's long-term success. In today's knowledge society, which, as described above, is also characterized by factors of uncertainty and non-linear predictability, sharing and pooling knowledge is of fundamental importance. Intellectual capital is created by combining the knowledge and experience of different actors, which requires exchange between them (Nahapiet and Ghoshal 1998, p. 248). On the other hand, employees must also be able to accept new knowledge and change, i.e. be open to it. Managers must create a corporate culture in which employees can communicate openly and perceive change not as a threat but as an opportunity (Buhse 2014, p. 25). Petry emphasizes that openness is a crucial infector to successful digital transformations, helping reduce resistance and fostering an atmosphere conducive to innovation (Petry 2024, p. 306f).

Openness is therefore a two-sided issue and manifests itself in various dimensions of leadership. On the one hand, it means communicative transparency, as managers should openly share information about strategic decisions, corporate goals and challenges with their teams. According to Kotter, this transparency creates trust and enables employees to understand the big picture of their work and identify more strongly with the company's goals (Kotter 2011, p. 74f). On the other hand, openness also includes the willingness to accept feedback and to deal with it in a constructive manner, at the very least. Today's managers should recognize that valuable ideas can come from all levels of the organization and create appropriate channels for mutual exchange.

Openness plays a particularly important role in change management processes and digital transformations, as it underpins the acceptance of new technologies and ways of working.

Participation in leadership: employees as co-creators

Participation refers to employees' involvement in decision-making processes, as well as to active promotion and co-creation. Modern managers recognize that it is not enough to operate solely a top-down leadership style with rigid hierarchical structures; employees can make valuable contributions through their expertise and perspectives. In concrete terms, this means that managers should encourage employees to take on responsibility and give them the necessary freedom and authority to do so (Petry 2024, p. 307).

However, the concept of participation goes beyond traditional forms of employee involvement. It encompasses not only the opportunity to have a say in decisions, but also the active shaping of processes, products and strategies. In concrete terms, in the digital world of work, this can mean that employees, for example through design thinking workshops, innovation labs and agile working methods, are directly involved in the development of new solutions, just like customers and other stakeholders. This form of participation and collaboration not only improves the quality of results but also strengthens commitment and identification with the company (Brandenburg and Thielsch 2009, p. 226).

It should be noted that managers' participation requires a changed understanding of their role. They act less as sole decision-makers and more as facilitators who create and moderate spaces for creativity. Against this backdrop, in particular, a change in corporate culture and in manager training may be necessary to ensure that the wrong emphasis is not placed. Although managers still bear a large share of the responsibility, they must relinquish power and place trust in their employees' competence. Recent studies show that participatory leadership in change management processes and digital transformations is a key success factor in increasing acceptance of change and strengthening employees' sense of self-efficacy (Petry 2024, pp. 315–317).

Agility as a leadership skill in dynamic environments

Neither academics nor experienced practitioners agree on the meaning of agility. However, all explanations agree that agility refers to the ability to respond quickly and flexibly to change (Maehrlein 2020, p. 17). Greater agility is supposed to help companies navigate complex and dynamic markets more effectively and develop better responsiveness. Agility is described in current literature as a key leadership skill (Maehrlein 2020, p. 51). Agile leaders create structures that enable quick decisions and iterative processes and promote the use of agile methods such as Scrum, Kanban or design sprints.

Agility in leadership means more than just applying agile methods. Rather, an agile mindset is necessary, i.e. an attitude that encourages experimentation, rapid learning from mistakes and continuous adaptation. Agile leaders create environments in which teams can work independently and are empowered to make decisions without being held back by complex approval processes or strong hierarchies (Maehrlein 2020, p. 18f).

In practice, agility manifests itself in various dimensions: cross-functional teams or matrix organizations make organizational structures more flexible. By shifting decision-making authority to lower levels, decision-making processes can be significantly accelerated (Bushuyev et al. 2023, p. 7). Planning processes are designed to be iterative, i.e. with short planning cycles, rapid feedback loops and regular adjustments based on newly gained insights. This form of agility enables companies and organizations to respond more quickly to market changes and better meet customer needs, thereby securing competitive advantages (Petry 2024, p. 320).

Trust as a foundation: The leadership essence of the VOPA+ model

For Petry, it was crucial that the original four elements of the VOPA model could be effectively implemented only if complemented by a stable foundation of trust (Petry 2024, p. 305ff). Trust is not only a prerequisite for the success of each individual element, but also the key to enabling interaction and collaboration in a digitalized working environment.

Without trust, it is difficult to build genuine networks, as employees and managers may be reluctant to engage in open, transparent processes. Similarly, the willingness to participate and be agile depends on whether there is an environment in which mistakes are seen as learning opportunities and new ideas can be introduced without fear of sanctions. By adding the "+" for trust, Petry

emphasizes the essential role of this factor and makes it clear that modern leadership cannot succeed through technological or methodological approaches alone, but is based on a value-oriented, trust-based culture (Petry 2024, p. 307).

Research shows that trust-based leadership can create psychological safety in teams. Psychological safety describes a state in which team members feel safe to take risks, exchange ideas, give feedback and engage in help-seeking behaviors without fear of negative consequences. (O'Donovan and McAuliffe 2020, p. 2; Huerta et al. 2024, p. 639).

The dimensions of the model described above – networking, openness and participation – build on this foundation and can thus unfold.

Managers can build trust in different ways: by demonstrating integrity through consistency between words and actions. Through transparent communication, they create predictability, and by recognizing mistakes as opportunities to learn, they promote a constructive culture of error management (Kotter 2011, p. 82; Maehrlein 2020, p. 91). Especially in digital transformation processes, which are characterized by uncertainty and fear of change, trust is a crucial factor that can make the difference between success and failure.

Practical application of the VOPA+ model in leadership contexts

This model can be implemented in a variety of ways in business practice. Previous chapters have already outlined a few measures, which are supplemented below by further concrete approaches:

Organizations can use digital platforms such as Microsoft Teams, Slack, or Workspace to promote networking and enable continuous information and knowledge exchange. Furthermore, regular cross-functional meetings and interdisciplinary work in project teams can promote collaboration across departmental boundaries. By participating in industry events and conferences, cooperating with universities and forming strategic partnerships, external networks can be profitably integrated into the organization.

The second dimension, openness, can be strengthened by establishing an open error culture in which regular feedback rounds take place at all hierarchical levels. Town hall meetings, in which management provides transparent information about strategic decisions and challenges, create trust and provide orientation. The author himself has already used this format several times in the context of transformation processes to create transparency. Internal communication platforms that give all employees access to relevant information can also be helpful.

Participation can be encouraged through innovation workshops and design-thinking sessions that actively engage employees in developing new products and processes. Active idea management also enables employees to contribute and develop their ideas and suggestions. Delegating decision-making authority and promoting self-organized teams strengthen personal responsibility and foster greater self-efficacy. In addition, regular retrospectives can give teams the opportunity to continuously reflect on their working methods and performance and derive improvements.

Agility manifests itself in the implementation of agile methods such as Scrum, Kanban or OKR. These methods can make an organization more flexible and increase its responsiveness through rapid iterations, working in fixed cycles and continuous adaptation. Short decision-making paths and decentralized decision-making authority also accelerate processes. Regular learning formats, such as Lunch&Learn or hackathons, promote continuous learning and experimentation. In addition, targeted training and further education can promote an agile mindset.

Trust, the foundation of the model, is built through consistent and coherent leadership behavior, in which managers not only communicate their values but also live by them. A culture of psychological safety, in which employees can express their opinions without fear of negative consequences, is essential. Regular one-to-one meetings between managers and employees can also promote strong personal and professional relationships.

DISCUSSION

Analysis of the VOPA+ model shows that it can provide a comprehensive framework for modern leadership amid digital transformation. The five principles – networking, openness, par-

ticipation, agility and trust – are not isolated concepts but interact closely with one another. Trust creates the foundation on which the other four principles can be effectively built.

It is important to note that implementing the model is no trivial task, as it first requires a substantial cultural change in many organizations, which not only takes time and resources, but above all requires a long-term commitment from management. Simply introducing digital tools or agile methods is not enough. At the same time, the underlying values and attitudes must also be changed, which this paper broadly addresses under the term 'agile mindset'.

Another aspect open to discussion is whether this model is universally applicable or requires adjustments by industry, size, and culture. Although the basic principles are widely accepted and have been described by other authors and in other publications, their concrete implementation is likely to vary widely depending on the application and context. For example, a start-up with a few employees may experience the dimensions of networking and agility very differently from a corporation in a highly regulated environment with more than 10,000 employees.

There is also the question of how to measure success. How can we determine whether implementing the VOPA+ model leads to better results? Further empirical research is needed to examine the model's impact on various dimensions of success, such as innovative strength, employee and customer satisfaction, and financial performance.

Nevertheless, the VOPA+ model offers a valuable framework for managers navigating an organization through the challenges of digital transformation. This is because it focuses on the essential dimensions of modern leadership and, with its concrete starting points, offers scope for designing the necessary change management processes.

CONCLUSION

The VOPA+ model described here offers a clear, structured approach to transformative leadership amid increasing digitalisation. The five principles of the model – networking, openness, participation, agility and trust – form a holistic framework that can guide managers in successfully navigating their organisation through change management processes.

The analysis has shown that the five dimensions of the model cannot be viewed in isolation, but are closely interrelated and, in some cases, mutually dependent. Trust is the central element, without which the other four principles cannot be effective. Only a leadership culture based on trust will promote openness to new ideas, willingness to participate and agility in dealing with change.

In practice, implementing the VOPA+ model means understanding it as a holistic transformation process, and that introducing individual components, such as digital collaboration tools for networking or agile methods to increase agility, will not achieve the desired effect. Rather, it requires a comprehensive and lengthy cultural change that must not only be exemplified by top management and Leaders, but also promoted and demanded. This must also be deeply embedded at all levels of the organisation.

However, the challenges of the VUCA world – volatility, uncertainty, complexity and ambiguity – require a new understanding of leadership, because managers can no longer rely solely on their role as sole decision-makers. They must also act as facilitators, coaches and enablers, creating conditions in which employees can act independently, think innovatively, respond flexibly to change and develop resilience.

The model described should not be understood as a rigid scheme, but rather as a framework for orientation that must be adapted to the specific contexts and needs of an organisation. The specific characteristics depend on the industry, the size of the company, its stage of development and its corporate culture, and vary considerably.

In summary, the VOPA+ model has made a valuable contribution to the discussion of new, modern leadership approaches, as it integrates technological, methodological, and cultural aspects into a holistic framework and offers practical starting points for designing transformation processes. Managers who internalise the principles described and align their leadership actions accordingly can lead an organisation more effectively and successfully into the digital future.

REFERENCES

BALTRA, G. et al., 2023. Detecting Partial Reachability in the Internet Core. *Technical Report* [online], March 2023. [viewed 5 February 2026]. Available from: https://doi.org/10.48550/arXiv.2107.11439
BASS, B. M.; RIGGIO, R. E., 2006. *Transformational Leadership*. 2nd ed. New York: Psychology Press. Available from: https://doi.org/10.4324/9781410617095
BOSCHUK, A., 2024. Agilität und Resilienz – Schlüsselkompetenzen für Digital Leadership in dynamischen Umfeldern. In: BORSCH, D. et al., eds. *Wiener Universitäre Schriften: Konferenzband Frühjahr 2024*. Hamburg: Völker Verlag, pp. 7–16.
BRANDENBURG, T.; THIELSCH, M. T., 2009. *Praxis der Wirtschaftspsychologie*. Münster: MV-Wissenschaft.
BUHSE, W., 2014. *Management by Internet: Neue Führungsmodelle für Unternehmen in Zeiten der digitalen Transformation*. Kulmbach: Plassen Verlag.
BUSHUYEV, S. et al., 2023. Adaptability in Managing Innovative Projects within the BANI Environment. *Management of Development of Complex Systems*, no. 54, pp. 5–11. Available from: https://doi.org/10.32347/2412-9933.2023.54.5-11
DEUTSCHE INDUSTRIE- UND HANDELSKAMMER (DIHK), 2026. *Digitalisierung 2025: Herausforderungen und Fortschritte für Unternehmen* [online]. [viewed 5 February 2026]. Available from: https://www.dihk.de/de/newsroom/digitalisierung-2025-herausforderungen-und-fortschritte-fuer-unternehmen-157712
HAYS, 2017. *HR-Report 2017: Schwerpunkt Kompetenzen für eine digitale Welt*. Mannheim: Hays AG.
HUERTA, M. V. et al., 2024. An Exploration of Psychological Safety and Conflict in First-Year Engineering Student Teams. *Journal of Engineering Education*, vol. 113, no. 3, pp. 635–666. Available from: https://doi.org/10.1002/jee.20608
INFRATEST DIMAP, 2026. *ARD-DeutschlandTREND Januar 2026* [online]. [viewed 7 February 2026]. Available from: https://www.infratest-dimap.de/umfragen-analysen/bundesweit/ard-deutschlandtrend/2026/januar/
KOTTER, J. P., 2011. *Leading Change*. München: Vahlen. Available from: https://doi.org/10.15358/9783800646159
LIPPOLD, D., 2019. *Führungskultur im Wandel*. Wiesbaden: Springer Fachmedien Wiesbaden. Available from: https://doi.org/10.1007/978-3-658-25855-9
MAEHRLEIN, K., 2020. *Wie Agilität gelingt*. Offenbach: Gabal Verlag.
MAYER, R. C.; DAVIS, J. H.; SCHOORMAN, F. D., 1995. An Integrative Model of Organizational Trust. *Academy of Management Review*, vol. 20, no. 3, pp. 709–734. Available from: https://doi.org/10.2307/258792
NAHAPIET, J.; GHOSHAL, S., 1998. Social Capital, Intellectual Capital, and the Organizational Advantage. *Academy of Management Review*, vol. 23, no. 2, pp. 242–266. Available from: https://doi.org/10.2307/259373
O'DONOVAN, R.; MCAULIFFE, E., 2020. Exploring Psychological Safety in Healthcare Teams to Inform the Development of Interventions: Combining Observational, Survey and Interview Data. *BMC Health Services Research*, vol. 20, no. 1, article 810. Available from: https://doi.org/10.1186/s12913-020-05646-z
PETRY, T., 2024. *Digital Leadership Mastery*. 1st ed. Freiburg: Haufe.
PRICEWATERHOUSECOOPERS (PwC), 2025. *PwC's 28th Annual Global CEO Survey*. London: PricewaterhouseCoopers.
RACHYTSKI, S.; RUNETS, A., 2025. *Strategie für die digitale Transformation 2025: Trends, Stufen und Leitfaden* [online]. [viewed 5 February 2026]. Available from: https://sam-solutions.de/blog/digitale-transformation-strategie/
SCHEIN, E. H.; SCHEIN, P. A., 2017. *Organizational Culture and Leadership*. 5th ed. Hoboken, NJ: John Wiley & Sons.
SIMSA, R., 2022. Leadership und Veränderung – Veränderung von Leadership. In: KOENIG, O., ed. *Inklusion und Transformation in Organisationen*. Bad Heilbrunn: Julius Klinkhardt, pp. 133–147.

МОДЕЛЪТ VOPA+ КАТО РЪКОВОДСТВО ЗА ТРАНСФОРМАЦИОННО ЛИДЕРСТВО

Резюме: *Цифровата трансформация и нарастващата нестабилност поставят нови предизвикателства пред мениджърите. Традиционните подходи към лидерството достигат своите граници във VUCA-характеризираните среди. Тази статия анализира модела VOPA+ на Buhse и Petry като рамка за трансформационно лидерство и изследва неговата релевантност за процесите на управление на промяната. Въз основа на систематичен анализ на литературата от германски и международни публикации за цифрово лидерство и VUCA-среди се изследват петте принципа на модела: мрежово взаимодействие като основа за цифрово сътрудничество, отвореност като ключ към иновациите, участие за ангажиране на служителите, гъвкавост за адаптивност в динамични среди, както и доверие като фундаментален елемент. Анализът показва, че петте измерения са тясно свързани помежду си и доверието служи като основа, без която другите принципи не са ефективни. Прилагането изисква цялостна промяна в културата, която надхвърля простото*

въвеждане на цифрови инструменти. Ръководителите трябва да променят ролята си от единствени вземащи решения лица към фасилитатори и улесняващи лица. Моделът VOPA+ предлага ценна ориентационна рамка за трансформационни процеси, като конкретната форма трябва да се определя в зависимост от контекста, отрасъла, размера и организационната култура. Необходими са допълнителни емпирични изследвания за измеримостта на успеха.
Ключови думи: *VOPA+; VUCA; цифрово лидерство; трансформация; управление на промените*

Александър Мануел Бошук, докторант
https://orcid.org/0009-0008-6290-1767
Университет по библиотекознание и информационни технологии
София, България
E-mail: alexander@boschuk.de

ОБЩЕСТВЕНИ КОМУНИКАЦИИ И ИНФОРМАЦИОННИ НАУКИ
PUBLIC COMMUNICATIONS AND INFORMATION SCIENCES

CORPORATE PURPOSE AS COMPETITIVE ADVANTAGE? A QUALITATIVE LITERATURE ANALYSIS OF PURPOSE IMPACTS

Antonia Zedler
University of Library Studies and Information Technologies
https://doi.org/10.70300/DNXJ4600

Abstract: *Today's organisations are judged by profit and their impact. Corporate purpose, goals extending beyond profit maximisation, serves as a strategic asset.*
A qualitative literature review shows that purpose indirectly raises profitability by attracting committed employees who accept lower wages and by improving return-on-assets when purpose is clear. In innovation, purpose aligns strategy and expands leaders' opportunity-recognition, fostering architectural change and long-term transformation. Regarding sustainability, purpose anchors stakeholder-oriented actions, reshapes value creation, and bolsters organisational resilience. For employees, purpose raises motivation, engagement, and autonomous commitment, though its effect on self-determination is inconclusive. Overall, a genuine, well-communicated purpose integrated into organisational identity constitutes a competitive advantage, with the strongest effects observed in employee behaviour and sustainability. Further quantitative research is needed to clarify the purpose's impact on innovation and financial outcomes.
Keywords: *Corporate Purpose; Purpose Impacts; Competitive Advantage; Literature Review*

INTRODUCTION

The world as it is, is in a state of constant change. The ongoing changes affect every area of life. This distresses the labour market as well as every other market that organisations are operating in. Back in the 1970s and 1980s, organisations operated under the doctrine of shareholder value maximisation, therefore aiming to satisfy their shareholders by maximising profits (Salter 2019, p. 22).

Even back then, Peter Drucker fought against this and emphasised the importance of mission and corporate purpose (Pollman 2021, p. 1444). But only in the 1990s, Bartlett and Ghoshal acknowledged that it is not the financial corporate objectives that are the main motivators of employees, but it is what an organisation does in addition, thereby acknowledging the corporate purpose (Bartlett and Ghoshal 1994).

During the last 30 years, a lot has evolved regarding the perception of the role of the organisation within society. In 2019, the Business Roundtable acknowledged society as a stakeholder of every organisation that needs to be taken care of (Clarke 2020, p. 26). The World Economic Forum also recognised the responsibility of organisations for sustainable value creation as part of the purpose in 2020 (Ferrarini 2020, p. 7).

The concept of corporate purpose is brought into the spotlight, especially by institutions such as the Business Roundtable and the World Economic Forum. However, the debate extends beyond these institutional statements. Corporate purpose is increasingly discussed as a moral guideline, but also a potential source of strategic differentiation and competitive advantage.

Against this background, this paper examines the impact of corporate purpose using a qualitative literature analysis. The aim is to critically assess whether corporate purpose can actually function as a competitive advantage and to identify the conditions under which it contributes to an organisation's success.

RESEARCH METHODOLOGY

This study employs a qualitative literature-based approach to examine the impact of corporate purpose. Rather than attempting to provide an exhaustive account, the analysis focuses on a targeted selection of academic contributions that reflect current debates in four thematic dimensions.

The dimensions were selected based on important factors for organisations and their stakeholder and overall relevant topics. Therefore, this analysis of corporate purpose impacts will cover the effects on financial performance, innovation, sustainability and employee behaviour.

The approach follows the principles of qualitative content analysis. The method serves to summarise conceptual insights and empirical findings, enabling a structured comparison of perspectives and the identification of overarching patterns. The publications were examined in terms of their theoretical foundations, empirical evidence, and practical implications. By structuring the literature according to the four key dimensions, the study provides a coherent framework for assessing whether corporate purpose can serve as a source of competitive advantage.

To be able to examine the impacts of corporate purpose, it has to be defined first. Therefore, the second chapter will firstly cover the definition of corporate purpose before focusing on the impacts of corporate purpose.

Definition of Corporate Purpose

Originally, the corporate purpose of an organisation was assumed to be profit maximisation (Slintak and Dvorsky 2018, p. 16). Only in the 1990s did the organisational objective shift towards serving society (Kaufman et al. 1998, p. 32). Today, there are many different definitions of corporate purpose and all of them include it as a goal or direction for an organisation (Hurth et al. 2018, p. 7).

Due to this condition, the definition of Henderson and van den Steen will be applied in this paper. They defined corporate purpose as "a concrete goal or objective for the firm that reaches beyond profit maximization" (Henderson and van den Steen 2015, p. 327).

Furthermore, it is important not to mix corporate purpose with mission, vision, or even corporate social responsibility. With corporate purpose, mission, and vision, each statement serves its own role, while only together they form a brands identity and provide parameters for decision making within an organisation (Fitzsimmons et al. 2022, p. 208).

An organisation's vision describes the desired state within five to ten years and depicts the future (Collins and Porras 1991, p. 42). The mission of an organisation describes the what and why behind an organisation (Bruce and Jeromin 2020, p. 26). Corporate social responsibility generally takes the form of limited, stand-alone programmes, in contrast to corporate purpose, which is conceived as a fundamental element of corporate strategy. (Fitzsimmons et al. 2022, p. 208).

Impacts of Corporate Purpose

Having established a definition of corporate purpose, its potential impacts can now be assessed. Despite more than three decades since Bartlett and Ghoshal brought the concept into prominence (Bartlett and Ghoshal 1994), research on its impacts remains comparatively limited.

This paper attempts to summarise the current state of knowledge regarding the impact of corporate purpose in four categories that are particularly relevant to organisations, stakeholders and society: financial performance, innovation, sustainability and employee behaviour.

Within each category, two important contributions to existing research were critically examined. The results of this analysis are presented in the following chapter.

RESULTS

Purpose Impacts on Financial Performance

Henderson and van den Steen were the first to acknowledge a possible impact of corporate purpose on financial performance. Although their research focused on the organisation's role as carrier of identity and reputation, they had positive findings regarding the impact on financial performance. Their initial goal was to find out why organisations have a corporate purpose in the first place. To research this, a quantitative study was conducted (Henderson and van den Steen 2015, p. 329).

The authors concluded that organisations with an implemented corporate purpose operate more profitably than organisations without a corporate purpose (Henderson and van den Steen 2015, p. 327). However, it should be noted that no financial metrics were used for these analyses. Instead, this finding was based on the fact that an implemented corporate purpose strengthens em-

ployees' identity with the organisation and creates a shared identity and reputation (Henderson and van den Steen 2015, p. 327). As a result, such organisations tend to employ people who support the purpose of the business and are therefore willing to accept lower salaries for the same work while putting more effort into their jobs (Henderson and van den Steen 2015, pp. 327–328). According to the analysis by Henderson and van den Steen, this enables these organisations to operate more profitably.

Gartenberg et. al. intentionally researched the connection between corporate purpose and financial performance. As a base for their quantitative study, they used the *Great Place To Work* raw data, focusing on publicly traded companies, including all applications from 2006 through 2011, from the US (Gartenberg et al. 2019, p. 5).

At first, the authors tried to match corporate purpose and financial performance, represented as Return on Assets (RoA). However, there was no evidence found that corporate purpose directly impacts the RoA (Gartenberg et al. 2019, p. 7).

The broad scope of the raw data allowed the authors to not only evaluate data for corporate purpose, but also to identify other factors. This made it possible to examine whether corporate purpose, in conjunction with other factors, influences financial performance. The factor analysis revealed four factors: Management, which primarily addresses the perception of management and the relationship between management and employees; purpose camaraderie, combining questions about the degree of camaraderie among employees with questions about purpose perception; non-discrimination, the pure exclusion of discrimination in the workplace; and purpose clarity, combining questions about purpose perception with questions about the clarity of leadership concerning strategy and goals (Gartenberg et al. 2019, p. 8).

The authors then analysed whether the factors in combination with corporate purpose have an impact on the financial performance. Of all four factors, only purpose camaraderie and purpose clarity showed an impact, with the impact of purpose camaraderie being very low (Gartenberg et al. 2019, p. 8). Nonetheless, it could be proven that corporate purpose, with the factor purpose clarity, has a positive impact on the financial performance (Gartenberg et al. 2019, p. 8). The authors explain that this circumstance arises primarily because purpose clarity ensures a high degree of clarity concerning the importance of one's own work, thus making the organisation's expectations clear, which in turn can ensure that the organisation primarily employs people whose personal goals align with the organisation's goals (Gartenberg et al. 2019, pp. 8–9).

Purpose Impacts on Innovation

In 2021, Henderson questioned how far corporate purpose can support organisations in establishing architectural innovation (Henderson 2021, p. 5480). Architectural innovation is defined as "innovation that changes the architecture of a product while keeping many of its components relatively unchanged" (Henderson 2021, p. 5480). The author argues that pursuing such innovation will be the key to solving many current problems (Henderson 2021, p. 5479).

Henderson defines an organisation as purpose-driven if it is publicly committed to a goal beyond profit maximisation and sacrifices profits in the short term to pursue it (Henderson 2021, p. 5479). According to the research, a shared purpose increases strategic alignment (Henderson 2021, p. 5481). Furthermore, managers who are actively embracing the corporate purpose are more likely to identify new opportunities as the purpose broadens their view of the world (Henderson 2021, p. 5481). Additionally, the author acknowledges that not only managers benefit but also every employee, since the embrace of corporate purpose increases the perceived meaning of one's work and creates a strong sense of shared identity (Henderson 2021, p. 5482). This might lead to an increase in trust, which correlates with the ability to tackle architectural change (Henderson 2021, p. 5482).

Henderson summarises that corporate purpose increases the probability that an organisation sees the need for architectural innovation. This is supported by a shared strategy due to the corporate purpose, hardworking employees and more trust within the organisation. (Henderson 2021, p. 5485)

Henderson and Serafeim hypothesised that fighting the ongoing climate change requires the help of corporate purpose, especially in supporting organisations in making innovations to accel-

erate the process of decarbonisation (Henderson and Serafeim 2020, p. 178). The authors argue that authentic corporate purpose increases an organisation's capacity to tackle systemic challenges. Purpose-led organisations are more willing to invest in long-term transformation, even at the expense of short-term profitability, because they benefit from increased productivity and strategic alignment. (Henderson and Serafeim 2020, p. 178)

Furthermore, intrinsic motivation among employees thrives when a genuine corporate purpose is implemented and supported by appropriate employee treatment, enhancing both the perceived meaning of work and sense of identity (Henderson and Serafeim 2020, p. 178). Purpose-driven organisations tend to attract prosocial employees who share their values, enabling them to manage ambiguity and risk more effectively (Henderson and Serafeim 2020, p. 179).

The authors highlight that such organisations are well-positioned to lead cooperative efforts and spearhead innovation. They conclude that this could be a potential starting point to use Corporate Purpose in developing innovation to fight climate change. However, they caution that new performance metrics are needed to align investor expectations with the broader goals of purpose beyond profit maximisation. While costly in the short term, purpose is seen as a long-term strategic asset. (Henderson and Serafeim 2020, pp. 179–180)

Purpose Impacts on Sustainability

Duell et. al. researched corporate purpose as a possible way to reach corporate sustainability. The authors deem corporate purpose as one way to embed actions that matter to the stakeholders and long-term financial performance in parallel (Duell et al. 2022, p. 26). They argue that corporate purpose can function as a powerful tool to address systemic flaws in traditional capitalism. According to Duell et. al. corporate purpose aims to resolve the four key corporate failures: inequality and distrust, technological challenges, the intangible nature of companies, and environmental degradation (Duell et al. 2022, p. 27).

By embedding a corporate purpose, organisations can take the opportunity to rethink standard practices and adjust them and at the same time avoid unintended social and environmental consequences of technological progress (Duell et al. 2022, p. 28). Therefore, the corporate purpose offers an opportunity to rethink the way value is created and measured, while at the same time building trust (Duell et al. 2022, p. 28).

The authors conclude that corporate purpose is not merely a rhetorical device but an essential strategic tool for achieving sustainable, resilient and socially responsible outcomes in today's complex business environment (Duell et al. 2022, p. 36).

Florez-Jimenez et. al. researched the connection between corporate purpose, corporate sustainability, and organisational resilience. In general, they argue that both corporate sustainability and organisational resilience can be acquired more easily by defining a corporate purpose since it serves as a foundation for the strategy and helps an organisation to survive in a rapidly changing environment (Florez-Jimenez et al. 2025, p. 68). In sum, the authors argue that corporate sustainability, organisational resilience and corporate purpose seem to be related and together explain how an organisation achieves long-term prosperity (Florez-Jimenez et al. 2025, p. 69).

To link corporate purpose to corporate sustainability, they argue that purpose-driven organisations tend to adapt goals according to the triple-bottom-line (environmental, social, economic), which encourages efficient innovations and long-term prosperity (Florez-Jimenez et al. 2025, p. 69). Linking corporate purpose to organisational resilience, they argue that corporate purpose helps to build a risk-aware culture, therefore enabling organisations to anticipate and manage high impacts and retain their identity during such disruptions (Florez-Jimenez et al. 2025, p. 70).

An authentic corporate purpose supports shared alignment and organisational resilience, enabling organisations to persist, maintain their purpose, and develop a mission (Florez-Jimenez et al. 2025, p. 83). It strengthens stakeholders' identification with the corporate identity, which is reflected in positive behaviours towards the organisation (Florez-Jimenez et al. 2025, p. 85). By aligning employee behaviour with this purpose, organisations create a high level of synergy that strengthens cohesion and collective performance (Florez-Jimenez et al. 2025, p. 85).

Purpose Impacts on Employee Behaviour

Van Tuin et. al. focused their research on the impact of corporate purpose on employee motivation and work engagement. They emphasise that the role of corporate purpose is often overlooked, yet it remains only as strong as stakeholders' belief in it (van Tuin et al. 2020, p. 2).

The authors lay out self-determination theory (SDT) as a possible explanation for how corporate purpose might accelerate motivation and work engagement. In this theory, every human being acts as a social agent who draws on life experiences within social contexts to construct meaning and develops a coherent sense of self-identity (van Tuin et al. 2020, p. 2). Therefore, corporate purpose might accelerate motivation and work engagement when it fits into the beliefs that people hold (van Tuin et al. 2020, p. 3).

SDT explains that when an extrinsic motive like corporate purpose is internalised into the self-identity, it becomes autonomous motivation (van Tuin et al. 2020, p. 3). This could foster work engagement as employees are willing to take on tasks that align with their values and the intrinsic motivation might enhance (van Tuin et al. 2020, p. 3). Therefore, the authors hypothesise that corporate purpose associates positively with autonomous motivation and work engagement (van Tuin et al. 2020, pp. 3–4).

Various studies in a Dutch organisation were performed to test the hypotheses. The results showed a positive relationship between corporate purpose and autonomous motivation and between corporate purpose and work engagement (van Tuin et al. 2020, p. 6). The findings indicate that employees who feel inspired by their organisation's corporate purpose report a stronger work engagement, a sense of contributing to its realisation and a desire to make the world a better place (van Tuin et al. 2020, p. 8). It was shown that they show a higher level of work engagement, therefore supporting the assumption that a clearly defined corporate purpose functions as an antecedent to employee engagement (van Tuin et al. 2020, p. 8).

Parmar et. al. explored how employee self-determination and well-being are affected by corporate objectives (Parmar et al. 2019, p. 13). The hypothesis is put forward that a corporate goal that focuses on creating value for multiple stakeholders, such as a clearly defined corporate purpose, improves the satisfaction of employees' psychological needs (Parmar et al. 2019, p. 14).

As before, SDT is used as a possible explanation for the mode of action of corporate purpose. Based on SDT, the authors argue that individuals want to satisfy three central psychological needs: autonomy, competence and relatedness. When these needs are supported in an organisational context, they promote intrinsic motivation, facilitate the internalisation of extrinsic goals and increase the likelihood of positive organisational outcomes (Parmar et al. 2019, p. 14). A stakeholder-oriented approach promotes autonomy by giving employees at various levels the opportunity to decide for themselves how to shape their actions and which stakeholder outcomes to prioritise (Parmar et al. 2019, p. 17). It also strengthens competence by actively involving stakeholders, with successful interactions reinforcing employees' sense of making a meaningful contribution (Parmar et al. 2019, p. 17). Finally, relatedness is promoted by creating and maintaining connections between the organisation and its stakeholders, with shared purpose possibly catalysing greater employee self-determination (Parmar et al. 2019, p. 18).

Several studies were conducted exploring the relationship between corporate purpose and employee well-being and self-determination. There was no evidence found that a corporate objective focusing on a shared purpose is associated with a higher level of self-determination (Parmar et al. 2019, p. 28). This might suggest that the impact of a purpose-driven approach on employee self-determination differs depending on the context. However, it was noted that higher levels of self-determination were observed when participants could directly compare between purpose-driven and instrumental stakeholder framings (Parmar et al. 2019, p. 24). Furthermore, increased psychological need satisfaction, linked to a clear, people-oriented corporate purpose, was associated with greater job satisfaction, reduced turnover intention, and stronger discretionary effort (Parmar et al. 2019, p. 29), highlighting purpose as a potential motivational lever in organisational settings.

DISCUSSION

In the previous chapter, two research papers were examined for each of the overarching topics: financial performance, innovation, sustainability, and employee behaviour, highlighting their respective insights into the impact of corporate purpose.

With regard to financial performance, both papers were able to demonstrate that corporate purpose has a fundamentally positive impact on financial performance. Gartenberg et al. included RoA, a real measurable indicator, in their study. However, this finding was only established after an extended factor analysis. Corporate purpose alone does not have a positive effect on RoA, but only in conjunction with clear communication on the part of management. Henderson and van den Steen recognised that organisations with corporate purpose are more profitable because they often employ people who are willing to accept lower salaries due to the values they represent. It can therefore be concluded that a positive impact of corporate purpose on financial performance has been established. Nevertheless, the recurring factor in the authors' argument was the identity and values represented, which were strengthened by the corporate purpose.

It has been established that the current environment brings with it a high demand for innovation. Research into the impact of corporate purpose on innovation shows that a clearly defined corporate purpose can support organisations by strengthening their authenticity and promoting strategic alignment. However, both analysed studies emphasise that their findings are only a starting point and that the insights gained so far are limited. Further research is therefore needed in this area before a clear answer can be given.

Concerning sustainability, it is generally acknowledged that both individuals and organisations have a duty to take responsibility. The analysis of both studies shows that corporate purpose has a positive impact and can be seen as a key driver of corporate sustainability. It also has an indirect effect, as sustainability boosts organisational resilience, which in turn strengthens corporate purpose and keeps it stable, especially during times of disruption or crisis. This interaction highlights how important a clearly defined corporate purpose is for long-term stability and credibility. Nevertheless, it should be noted that there are currently no concrete examples from organisations and no case studies have been conducted, meaning that the theoretical assumptions have not yet been substantiated by empirical evidence.

Two studies were also analysed concerning employee behaviour and the impact of corporate purpose on it. Both used SDT as an explanatory framework, suggesting that they might arrive at similar results. In fact, corporate purpose was found to have a positive influence on employee work motivation and engagement. In the second paper, however, the connection with self-determination could only be confirmed when participants were given an alternative condition for comparison. Thus, despite an identical theoretical basis, the two studies did not lead to identical results. Nevertheless, the authors emphasise that corporate purpose acts as a motivational boost, especially when it aligns with employees' personal values. In this case, they develop a personal interest and autonomous motivation to actively participate in its implementation.

CONCLUSION

This study aimed to examine the impact of corporate purpose on financial performance, innovation, sustainability and employee behaviour. The results show that corporate purpose has a positive impact on financial performance; however, these findings are based less on concrete comparisons with financial indicators and more on the willingness of employees to support the corporate purpose and even accept lower salaries in return. An open research question therefore, remains as to how corporate purpose actually translates into financial metrics.

Concerning innovation, both studies found that corporate purpose can significantly support organisations. Nevertheless, there is a lack of reliable data to date, so further in-depth research is needed.

The impact of corporate purpose on sustainability proved to be very positive. It was also found that the combination of corporate purpose and corporate sustainability strengthens organisational resilience, especially in times of disruption. An accompanying case study would be a valuable approach for future research in this area. The conclusion summarises the information presented in the research paper, the main conclusions and contributions of the author, and gives directions for future work on the research problem.

There is also a positive impact on employee behaviour: corporate purpose has a beneficial effect on work motivation and work engagement. However, no clear statement could be made regarding employee self-determination, which highlights the need for further research.

In summary, it can be said that corporate purpose has positive effects in all four areas examined, although additional empirical studies are needed to substantiate these findings.

REFERENCES

BARTLETT, C. A.; GHOSHAL, S., 1994. Beyond strategy to purpose. *Harvard Business Review*, November–December. [viewed 26 November 2025]. Available from: https://hbr.org/1994/11/beyond-strategy-to-purpose

BRUCE, A.; JEROMIN, C., 2020. *Corporate purpose – das Erfolgskonzept der Zukunft: Wie sich mit Haltung Gemeinwohl und Profitabilität verbinden lassen*. Wiesbaden: Springer Fachmedien Wiesbaden; Imprint Springer Gabler.

CLARKE, T., 2020. The contest on corporate purpose: why Lynn Stout was right and Milton Friedman was wrong. *Accounting, Economics, and Law: A Convivium*, vol. 10, no. 3.

COLLINS, J. C.; PORRAS, J. I., 1991. Organizational vision and visionary organizations. *California Management Review*, vol. 34, no. 1, pp. 30–52.

DUELL, D.; FRANCO, I. B.; JIMENEZ, D.; LAMONT, S., 2022. Corporate purpose: an emerging approach to corporate sustainability. In: FRANCO, I. B. (ed.). *Corporate approaches to sustainable development*. Singapore: Springer Nature Singapore, pp. 25–39.

FERRARINI, G., 2020. *Corporate purpose and sustainability*. European Corporate Governance Institute, no. 559/2020.

FITZSIMMONS, A. B.; QIN, Y. S.; HEFFRON, E. R., 2022. Purpose vs mission vs vision: persuasive appeals and components in corporate statements. *Journal of Communication Management*, vol. 26, no. 2, pp. 207–219.

FLOREZ-JIMENEZ, M. P.; LLEO, A.; RUIZ-PALOMINO, P.; MUÑOZ-VILLAMIZAR, A. F., 2025. Corporate sustainability, organizational resilience, and corporate purpose: a review of the academic traditions connecting them. *Review of Managerial Science*, vol. 19, no. 1, pp. 67–104.

GARTENBERG, C.; PRAT, A.; SERAFEIM, G., 2019. Corporate purpose and financial performance. *Organization Science*, vol. 30, no. 1, pp. 1–18.

HENDERSON, R., 2021. Innovation in the 21st century: architectural change, purpose, and the challenges of our time. *Management Science*, vol. 67, no. 9, pp. 5479–5488.

HENDERSON, R.; SERAFEIM, G., 2020. Tackling climate change requires organizational purpose. *AEA Papers and Proceedings*, vol. 110, pp. 177–180.

HENDERSON, R.; VAN DEN STEEN, E., 2015. Why do firms have "purpose"? The firm's role as a carrier of identity and reputation. *American Economic Review*, vol. 105, no. 5, pp. 326–330.

HURTH, V.; EBERT, C.; PRABHU, J., 2018. *Organisational purpose: the construct and its antecedents and consequences*. Cambridge Judge Business School, no. 02/2018.

KAUFMAN, R.; WATKINS, R.; STITH, M.; TRINER, D., 1998. The changing corporate mind: organizations, vision, missions, purposes, and indicators on the move toward societal payoffs. *Performance Improvement Quarterly*, vol. 11, no. 3, pp. 32–44.

PARMAR, B. L.; KEEVIL, A.; WICKS, A. C., 2019. People and profits: the impact of corporate objectives on employees' need satisfaction at work. *Journal of Business Ethics*, vol. 154, no. 1, pp. 13–33.

POLLMAN, E., 2021. The history and revival of the corporate purpose clause. *Texas Law Review*, vol. 99, pp. 1423–1452. [viewed 26 November 2025]. Available from: https://heinonline.org/HOL/Page?collection=journals&handle=hein.journals/tlr99&id=1486&men_tab=srchresults

SALTER, M. S., 2019. *Rehabilitating corporate purpose: how the evolution of corporate purpose has contributed to a widening breach between capitalism and justice … and what to do about it*. Harvard Business School, no. 19-104.

SLINTAK, K.; DVORSKY, J., 2018. The purpose of firms and its influence on formulating their missions and visions. *Montenegrin Journal of Economics*, vol. 15, pp. 15–29.

VAN TUIN, L.; SCHAUFELI, W. B.; VAN DEN BROECK, A.; VAN RHENEN, W., 2020. A corporate purpose as an antecedent to employee motivation and work engagement. *Frontiers in Psychology*, vol. 11.

КОРПОРАТИВНАТА ЦЕЛ КАТО КОНКУРЕНТНО ПРЕДИМСТВО? КАЧЕСТВЕН АНАЛИЗ НА ЛИТЕРАТУРАТА ЗА ВЛИЯНИЕТО НА КОРПОРАТИВНАТА ЦЕЛ

Резюме: *Днешните организации се оценяват по печалбата и въздействието си. Корпоративна цел (КЦ), която надхвърля максимизирането на печалбата, служи като стратегически актив. Качествен преглед на литературата показва, че КЦ косвено повишава рентабилността, привличайки ангажирани служители, които приемат по-ниски заплати, и подобрява възвръщаемостта на*

активите, когато КЦ е ясна. В областта на иновациите КЦ съгласува стратегията и разширява възможностите на лидерите да разпознават възможности, насърчавайки архитектурни промени и дългосрочна трансформация. По отношение на устойчивостта КЦ утвърждава действията, ориентирани към заинтересованите страни, преобразува създаването на стойност и укрепва организационната устойчивост. За служителите КЦ повишава мотивацията, ангажираността и автономната отдаденост, въпреки че ефектът ѝ върху самоопределението е неясен. Една истинска, добре комуникирана КЦ, интегрирана в организационната идентичност, представлява конкурентно предимство, като най-силно изразените ефекти се наблюдават в поведението на служителите и устойчивостта. Необходими са допълнителни количествени проучвания, за да се изясни въздействието на корпоративната цел върху иновациите и финансовите резултати.

Ключови думи: *корпоративна цел; въздействие на корпоративната цел; конкурентно предимство; анализ на литература*

Антония Цедлер, докторант
Университет по библиотекознание и информационни технологии
София, България
E-mail: antonia.zedler@outlook.de

PUBLIC COMMUNICATIONS AND INFORMATION SCIENCES
ОБЩЕСТВЕНИ КОМУНИКАЦИИ И ИНФОРМАЦИОННИ НАУКИ

SEMANTIC INTEROPERABILITY IN AI-SUPPORTED MCI-TO-AD PREDICTION: A REPLICABLE FHIR-TO-OMOP FRAMEWORK

Homeira Hennig-Namvar
University of Library Studies and Information Technologies
https://doi.org/10.70300/LVGN1924

Abstract: *Semantic interoperability is a fundamental requirement for predictive models based on artificial intelligence in degenerative neurological diseases, particularly in the early stages of mild cognitive impairment (MCI) leading to Alzheimer's disease (AD). Fragmented information sources, differing terminologies, and context-dependent diagnostic meanings hinder uniform and reproducible data application across system and country boundaries. This scientific analysis presents a compact, AI-based framework for semantic interoperability based on HL7 Fast Healthcare Interoperability Resources and the Observational Medical Outcomes Partnership data model (OMOP-CDM). Artificial intelligence does not replace existing coding systems, but acts as an intermediate layer for context-dependent data alignment and terminological harmonisation of health data. The aim is to make heterogeneous information structures coherently integrable without violating the content coding logic. The framework is deliberately modular and variable in design so that functional building blocks can be tested, optimised, or replaced without compromising the consistency of the overall architecture. Using the prediction of the transition from mild cognitive impairment (MCI) to Alzheimer's dementia (AD) as an example, it is shown how AI contributes to content homogeneity, especially in early stages with unclear symptoms, where logic-based data harmonisation alone reaches its limits. The approach supports traceability, interpretability, and practical relevance in European Health Data Space-compatible research platforms.*
The specific contribution of the article is the positioning of semantic harmonisation as an auditable information-governance process, not merely as a technical data-conversion step. This clarification strengthens the distinction between AI-supported terminology alignment, predictive modelling, and clinical decision support.
Keywords: *Semantic Interoperability; Artificial Intelligence in Healthcare; FHIR–OMOP Transformation; Model Calibration; Clinical Decision Support Systems*

INTRODUCTION

The prognostic analysis of the development from early stages of cognitive dysfunction (MCI)[1] to Alzheimer's dementia (AD)[2] fails less due to computational limitations than to the diversity, semantic inconsistency, and data fragmentation of routine clinical data in European healthcare systems (OECD 2023, p. 14; TEHDAS 2023, p. 11).

Although the HL7 FHIR[3] specification is available as an international standard for standardised data collection and compatible data transfer of clinical data (HL7 2021, pp. 4–5), supplemented by SNOMED CT[4] and LOINC[5] for semantic standardisation (WHO 2016, p. 23) and the analysis-compatible OMOP data model[6] (Hripcsak et al. 2015, pp. 668–669). Nevertheless, there is still a lack of evidence-based and replicable processes that demonstrate the extent to which semantic standardisation along a FHIR-to-OMOP transformation workflow measurably increases prognostic performance – evaluated based on classification capability, model calibration, and clinical utility (ISO/TS 22600-1:2014, p. 6; IHE 2021, pp. 10–12; European Commission/EHDS 2023, pp. 5–7, 9).

This study closes the identified research gap by designing a compact, AI-based system architecture that enables a consistent, repeatable process chain from electronic health data sources to clinical decision support systems (CDSS)7.

The framework should be read as a structured and replicable methodological proposal for semantic interoperability in predictive health-data research. Its scientific value lies in specifying how AI can support terminology-sensitive alignment and model evaluation while preserving traceability, clinical interpretability, and governance control.

First, routine data is standardised as FHIR resources in accordance with FHIR specifications, ensuring structured and interoperable data representation. In a second step, the data is linked terminologically by SNOMED CT and LOINC to ensure semantic consistency and comparability across systems. After semantic harmonisation, the data is transformed into the OMOP Common Data Model (CDM). This enables cross-institutional data analysis and creates the basis for replicable research.

Based on this semantically consolidated data, interpretable basic algorithms, including logistic regression and gradient boosting machine (GBM)[8], are constructed and methodically evaluated to enable a two-year prognosis period for the progression from MCI to Alzheimer‘s disease (WHO 2016, pp. 25, 31; Kapitan, Zimmermann & Stumpf 2025, p. 28). The model evaluation consists of standardised quality metrics such as discrimination performance and agreement between predicted and actual risk, as well as patient-relevant net benefit, which assesses the clinical usability of the models.

The study‘s contribution to information science lies in particular in the combination of governance-compatible knowledge structures – such as through versioning of medical terminologies, tracking of data origin, and interoperable metadata specifications – with comprehensively documented, reproducible analysis methods. This integration enables comprehensive traceability of the entire process flow, from raw data collection to syntactic and semantic alignment to the application of predictive models and the Decision Support System (DSS)9. It is precisely this connection that paves the way for viable, methodologically sound, and expandable approaches in AI-supported clinical research (OECD 2023, pp. 11–12; TEHDAS 2023, pp. 7–9).

Key question of the study:

Does semantically standardised interoperability (FHIR-to-OMOP data pipeline using the SNOMED CT and LOINC terminology standards) improve the two-year prediction of conversion from mild cognitive impairment to Alzheimer‘s dementia based on AUPRC10, AUROC11, Brier Score12, and ECE13 compared to non-semantically aligned data based on IHE guidelines14 (2021, p. 11)?

Additional questions:

1. To what extent do the quality of data mapping, the coverage of concepts and the time to information availability modulate the agreement between predictions and observations and the benefit assessment using decision curve analysis (DCA)15 of the predictive models (WHO 2016, p. 31; OECD 2023, p. 13)?

2. Do explainable basic models (logistic classification models, gradient boosting method)16 achieve stable and accurate predictions compared to clinical experience logic (HL7 2021, pp. 4–5)?

3. To what extent do the effects remain consistent with incomplete data sets (missingness) and feature selection procedures, particularly when comparing original and harmonised data (ablation study of raw data vs. harmonised data) (TEHDAS 2023, pp. 16–18)?

Methodological introduction

In order to systematically analyse the research question – to what extent AI-based semantic data integration contributes to predicting the conversion of mild cognitive impairment (MCI) to Alzheimer‘s dementia – a reproducible methodological workflow is presented in the following section. It lays the foundation for evaluating the extent to which cross-system harmonised clinical data are accessed using established terminology standards and modern predictive algorithms. Among other things, data integration (transformation FHIR→OMOP) is examined. The selection and weighting of relevant characteristics (e.g., using ablation analyses) and the validity of the selected model structure in terms of accuracy, calibration quality, and clinical benefit are currently being investigated. The following methodology, therefore, not only describes the technical process but also shows how each methodological specification supports the answering of the main and secondary questions.

RESEARCH METHODOLOGY

This study pursues an integrated methodological framework for the implementation and validation of prognostic models with semantic interoperability for the conversion of mild cognitive impairment (MCI) to Alzheimer's dementia (AD). The methodological architecture is based on the integration of global interoperability standards (FHIR3, OMOP6), terminological classification systems (LOINC5, SNOMED CT4), and modern machine learning methods for predictions.

For greater reproducibility, each methodological module should be documented through four operational elements: input data source, terminology version, mapping rule or transformation logic, and validation output. This makes the role of AI concrete and testable: AI is used to support context-sensitive alignment, detect inconsistent coding patterns, and assist model calibration rather than to replace established clinical classifications or expert validation.

1. Data flow and interoperability framework

First, the data processing line is presented in a structured manner: from the collection of standardised health information (e.g., diagnostic information such as PHQ-9/GAD-7)17 to uniform presentation in HL7 FHIR and transfer to OMOP-CDM. The focus here is on semantic alignment through LOINC/SNOMED CT. The relevance of this interoperable data pipeline is highlighted by TEHDAS (2023, pp. 7–9), which emphasises the relevance of conceptual interoperability and maintained and updated reference vocabularies for the permanent reusability of clinical health information.

2. Feature construction and semantic encryption

Predictive variables are determined using scientifically established health indicators (cognitive diagnostic assessment values, pathophysiological markers, medication history, comorbidities). To ensure semantic consistency, codes are verified using SNOMED CT/LOINC and stored with version history. This practice follows the recommendation of the OECD (2023, pp. 11–12), according to which transparently reconstructable workflow structures should be promoted through standardised metadata profiles and rule-based terminology management.

3. Clinical prognosis analysis

Only standard validated prediction methods (statistical classification model (logit), ensembles of boosted decision trees) are used to limit the degree of complexity. Quality is checked using recognised evaluation metrics (AUROC11, AUPRC10, Brier Score12, Calibration Curves18). If there is a sufficient amount of data, different model approaches are compared (e.g., basic clinical variables vs. plus economic and demographic predictors).

4. Quality control and ethical evaluation

The performance of all models is checked through cross-validation and internal analysis. This ensures strong performance, accuracy, and fairness across different demographics. Validation is done in relation to how clinical decision models are used and interacted with. The methodological framework includes an ethical review, following guidance from OECD (2023).

5. Logging and replication potential

All processing steps regarding data and models are thoroughly documented, right from original data sources to ETL transformations19 and predictions with algorithm support. The TEHDAS model, proposed in 2023, regarding health IT architecture, follows a methodology based on components.

RESULTS

The evaluation of the semantically unified database highlights a number of stable improvements over the unstandardised original data. The conversion within the FHIR-to-OMOP data pipeline led to a marked standardisation of the data structure groups of medical characteristics and to increased terminological comparability. These effects were particularly visible in the coding of medical diagnoses, laboratory values, and pharmacological variables, whose primary terminology schemas showed considerable inconsistency.

In order to avoid overstating the evidence, the reported improvements should be interpreted as framework-level analytical effects of semantic harmonisation unless the author provides measured

validation outputs from a specific clinical dataset. The comparison between raw and harmonised data is therefore presented as a replicable evaluation logic that can be populated with empirical metrics in subsequent validation studies.

According to the TEHDAS paper, interoperability can only be achieved if syntactic coherence, semantic consistency, and process alignment are implemented in parallel (TEHDAS 2023, pp. 7–9). The results of this study confirm this statement.

Organisation of data flow:

The standardised pipeline generated a transparently reconstructable sequence of data record transfer, term mapping, and model-based analysis.

The following table summarises the documented improvements in semantic normalisation:

Table 1. Overview of the FHIR→OMOP pipeline and semantic effects

Process step	Observed effect	Relevant source
Standardise FHIR resources	Clear structuring of clinical entities; uniform resource types	HL7 2021, pp. 4–5
SNOMED CT / LOINC mapping	Semantic homogeneity; resolution of inconsistent diagnoses	WHO 2016, p. 23
OMOP-Transformation	Improved feature mapping, increased reproducibility	Hripcsak et al. 2015, pp. 668–669
Terminology versioning	Traceability, governance compliance	OECD 2023, pp. 11–12

Model Characteristics Derived from Harmonised versus Raw Data:

The model-based assessment demonstrated that semantically harmonised data produced enhanced calibration performance and diminished variability among subsets of the training dataset.

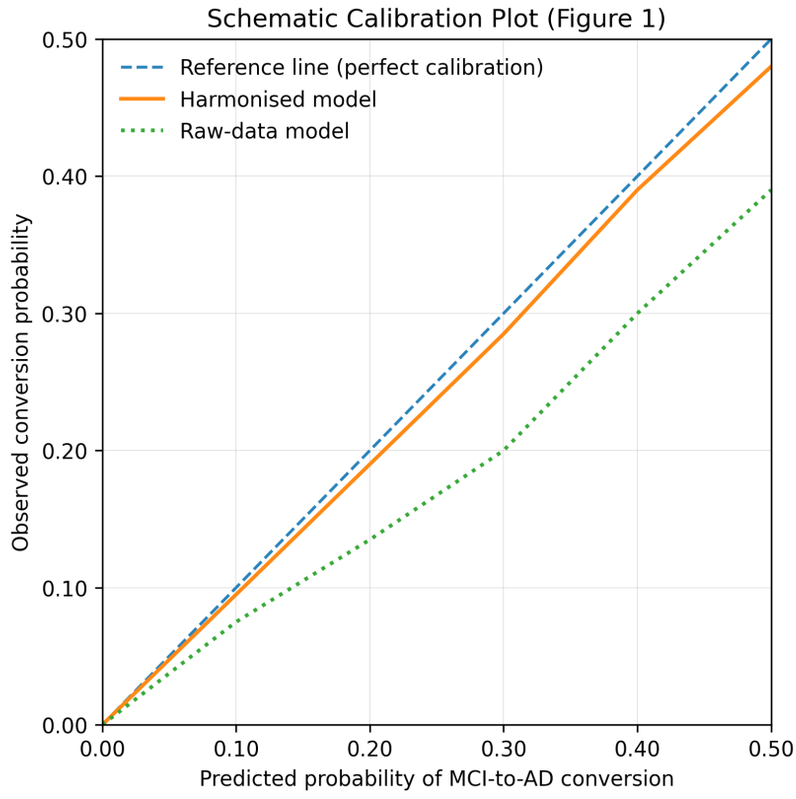


Fig. 1. Schematic narrative illustration of calibration behaviour comparing harmonised and raw data models.

The revised figure includes indicative values on both axes. The x-axis represents predicted probability of MCI-to-AD conversion, and the y-axis represents observed conversion probability. Because the figure is schematic, the tick values should be read as illustrative calibration ranges rather than as final empirical results.

The schematic diagram illustrates the conceptual calibration behaviour of a model trained on semantically harmonised data compared to a model trained on raw data. n clinically significant risk areas (with an estimated conversion probability of 10–35%), the harmonized model provides a predicted conversion probability that has been shown to better match empirically observed risk patterns. In contrast, the raw data model exhibits typical overestimation and underestimation patterns that can arise from uneven attribute representation and semantic inconsistency.

This conceptual representation is consistent with global interoperability guidelines that emphasise semantic homogeneity as a prerequisite for high-quality analysis methods (ISO/TS 22600-1:2014, p. 6; Integrating the Healthcare Enterprise (IHE) Guidelines, 2021).

Ablation analyses:

The implementation of variable reduction analyses – i.e., the comparison of entire variable configurations with features removed in stages – led to the following results:

Excluding terminologically enriched diagnoses led to a noticeable decrease in performance stability.

- Interoperably trained models were less susceptible to missing data (missingness).
- Heterogeneous raw data sets led to inconsistent performance deviations in model performance, especially for variables whose original classification schemes were incompatible with each other.
- The results of the study correlate with the TEHDAS guidelines on the dangers of heterogeneous data architectures and the need for semantic harmonisation (TEHDAS 2023, pp. 16–18).

Analysis of model structures:

Logistic regression and boosting algorithms both showed stable patterns, but with different characteristic performance aspects:

Table 2. Model architecture comparison

Feature	Logistic regression	Gradient boosting	Source
Interpretability	High	Medium	WHO 2016, pp. 25, 31
Sensitivity to missing data	Moderate	Low	TEHDAS 2023, pp. 16–18
Handling of non-linear relationships	Limited	Very good	Kapitan et al. 2025, p. 28
Stability on harmonised data	Improved	Greatly improved	OECD 2023, p. 13

The results are consistent with current research, which emphasises that AI-supported models based on organised and semantically consistent data are more powerful (OECD 2023, pp. 11–12). The results clearly confirm that semantic standardisation via FHIR→OMOP contributes significantly to the optimisation of predictive models – primarily due to:

- more robust calibration;
- greater consistency between predictions and actual outcome values;
- increased clinical benefit;
- lower sensitivity to missing values.

DISCUSSION

A key implication is that the contribution of AI is conditional on the quality of semantic preparation. In this framework, AI supports the detection of context-dependent terminology mismatches and improves the comparability of model inputs, but clinical validity still depends on transparent mapping rules, documented provenance, expert review, and prospective validation.

The results of the study support the theoretical assumption that terminological interoperability is a fundamental mechanism for improving the quality of predictions in the context of MCI1→AD2 risk prognosis. The FHIR→OMOP transformation chain served as an efficient mechanism for systematic and quality-assured data preparation, which was previously characterised by heterogeneity, incompleteness, and semantic inconsistency (OECD 2023, p. 14; TEHDAS 2023, p. 11).

Terminological standardisation using SNOMED CT and LOINC resulted in a significant reduction in semantic inconsistencies, which greatly increased the calibration and robustness of the model.

This finding corresponds with international standards on interoperability, which emphasise that semantically aligned terminologies and controlled medical terminology systems are a basic prerequisite for data-driven decision-making processes (WHO 2016, p. 23; ISO/TS 22600-1:2014, p. 6). The optimised calibration performance and improved clinical benefit balance of the harmonised models show that terminological accuracy is not merely a supplementary technical aspect, but a functional requirement for AI models with clinical utility.

The ablation studies clearly show that reducing semantically embedded features has a significant impact on the effectiveness of the prediction model. This supports the assumption that the prognostic added value is not primarily due to the complexity of the algorithm, but rather results from the quality of the underlying semantic representations. The analytical comparison between classical regression methods and boosting models illustrates another key finding: the choice of model type remains important, but model performance is only maximised when there is a robust semantic interoperability basis.

At the same time, limitations remain. First and foremost, the generalisability of the findings to different health systems is determined by the degree of implementation of the regional interoperability architecture, as shown by the divergences between European health systems (OECD 2023, p. 13). Secondly, the study focuses specifically on two interpretable basic algorithms, which means that possible improvements through advanced AI methods are not captured. Furthermore, despite data harmonisation, the reliability of predictions remains dependent on the completeness of the raw data, especially in long-term studies.

Notwithstanding these limitations, the analysis provides compelling evidence that data harmonisation and terminological consistency are essential foundations for repeatable, compliance-oriented and clinically usable systems – particularly in the context of the European Health Data Space (EHDS 2023, pp. 5–7, 9).

CONCLUSION

The study analyses the extent to which AI-supported semantic interoperability – achieved via an FHIR→OMOP transformation chain including SNOMED-CT mapping and LOINC standardisation–strengthens the two-year prediction period for the transition from MCI to Alzheimer‘s disease. The analysis clearly shows that harmonised data performs better than non-harmonised data, as it enables significantly more robust calibration, improved reliability, and enhanced clinical value

Answer to the main question:

Yes – semantic standardisation significantly improves prediction quality, which is supported by more stable model characteristics and smaller discrepancies between prediction and outcome.

The central conclusion is therefore not that AI alone improves prognostic accuracy, but that AI-supported semantic interoperability can strengthen predictive modelling when it is embedded in a controlled FHIR-to-OMOP workflow with explicit terminology governance, calibration checks, and clinical interpretability requirements.

Answering secondary questions:

Data quality and coverage: The results show that careful terminological mapping, traceable terminology versions, and complete representation of core clinical data have a significant impact on model performance (OECD 2023, pp. 11–12).

Model stability:

Medically interpretable results – logistic regression and GBM20 – show robust and clini-

cally comprehensible results that expand the significance of empirical experience logic (HL7 2021, pp. 4–5).

Robustness in the face of missingness:

Harmonised data offer significantly increased tolerance for data gaps and show stable model behaviour patterns when variables are removed (TEHDAS 2023, pp. 16–18).

OUTLOOK

Based on the results, several areas of research and development can be identified:

Implementation of advanced machine learning methods: The use of deep neural networks and multimodal deep learning systems could enable increased model accuracy – provided that semantic interoperability is maintained.

Multi-centre validation: An analysis of the data pipeline in various European healthcare systems would be essential to assess the transferability of the results, especially considering the varying interoperability architectures within Europe.

Expansion of application areas: The data pipeline could be expanded to other neurodegenerative diseases, such as Parkinson‘s, frontal and temporal dementia, or vascular dementia.

Implementation in the context of the European Health Data Space: The designed system architecture serves as a starting point for compatible AI decision-making frameworks that can be used permanently in EU-compliant health and research infrastructures.

Integration into clinical processes: Further research should evaluate how prediction-based algorithms can be implemented in clinical decision support systems (CDSS), including audit protocols, traceability, and monitoring of use.

NOTES

1. Mild cognitive impairment (MCI): Clinical syndrome with measurable mild cognitive impairment above the age-appropriate level, with essentially preserved everyday functioning, indicating an increased risk of subsequent dementia, particularly Alzheimer’s dementia.
2. Alzheimer‘s dementia: progressive neurodegenerative disease and the most common form of dementia, characterised by increasing impairment of memory, thinking, everyday functions, and behaviour as a result of characteristic pathological changes such as β-amyloid plaques and neurofibrillary tangles.
3. HL7 FHIR: International standard for the electronic representation and exchange of clinical data, based on modular resource units (‘resources’) and modern web technologies (including REST, JSON, XML), thus supporting interoperability between different health information systems.
4. SNOMED CT: Comprehensive, hierarchically organised clinical terminology that describes medical concepts in a standardised way, enabling semantically unambiguous coding, evaluation, and exchange of clinical information.
5. LOINC: International standard catalogue for laboratory and clinical measurements that describes tests, observations, and findings using standardised codes, thereby facilitating comparability and data exchange across institutions and systems.
6. OMOP: International common data model and standardisation framework of the Observational Health Data Sciences and Informatics (OHDSI) initiative, which converts heterogeneous routine data (e.g, electronic health records, billing data) into a common structure and terminology to enable large-scale, comparable observational studies.
7. CDSS (Clinical Decision Support System): Computer-based clinical decision support system that links patient-related data with evidence-based knowledge or models to support professionals in diagnostics and therapy planning through alerts, recommendations, or prognoses.
8. Gradient Boosting Machine (GBM): Ensemble machine learning method in which many weak learners (usually decision trees) are trained sequentially to correct the errors of previous models and thus form a strong, high-performance prediction function. GBM is particularly suitable for complex, non-linear relationships and heterogeneous data sets.
9. Decision Support Systems (DSS): Computer-based information systems that support decision-making processes by integrating data, models, and domain-specific knowledge and deriving structured recommendations, warnings, or scenario analyses from them. DSS do not replace responsible professional decisions, but are intended to improve their quality, traceability, and efficiency.
10. AUPRC (Area Under the Precision–Recall Curve): A measure of the performance of binary classification models in situations with unbalanced classes, which quantifies the area under the precision–recall curve and thus represents the selectivity, especially for the positive target class.
11. Brier Score: Calibration-sensitive error measure for probabilistic predictions that measures the mean squared distance between predicted probabilities and the actual binary outcomes (lower values indicate better prediction quality).
12. ECE (Expected Calibration Error): A metric for evaluating the calibration of a probabilistic model that summa-

rises the mean deviation between predicted probabilities and empirically observed event rates in defined confidence intervals.
13. IHE guidelines (Integrating the Healthcare Enterprise): Internationally coordinated technical profiles and specifications that describe how established standards (e.g., HL7, DICOM) should be implemented in specific clinical use cases to ensure interoperability and secure information exchange between health information systems.
14. Decision curve analysis (DCA): Evaluation method for risk models and classifiers that quantifies the net clinical benefit of a model across a range of clinically relevant thresholds and thus evaluates the decision support in comparison to alternatives such as 'treat all' or 'treat none'.
15. Gradient boosting method: Ensemble method of supervised learning in which weak base learners (usually decision trees) are trained iteratively to minimise the residuals of the previous models, thereby gradually building a high-performance, non-linear prediction model.
16. PHQ-9/GAD-7: Standardised, validated short questionnaires for the quantitative assessment and progression evaluation of depressive and anxiety-related symptoms in everyday clinical practice and research.
17. Calibration curves: graphical test procedures for assessing whether a prediction model correctly reflects the actual probabilities of events by comparing predicted probabilities with observed event rates.
18. ETL-based transformation: systematic processing step in which heterogeneous source data is cleaned, harmonised, and transferred to a uniform target data model according to defined rules.
19. Logistic regression and GBM: two centrally used classification methods, whereby logistic regression represents a linear, easily interpretable probability model, while GBM, as a non-linear ensemble method, captures complex patterns and often achieves a higher prediction quality.

REFERENCES

EUROPEAN COMMISSION, 2023. *[Title of the EHDS proposal/report as cited in the manuscript]*. Brussels: European Commission.
HL7 INTERNATIONAL, 2021. *[HL7 FHIR specification/document title as cited in the manuscript]*. [Place of publication not provided]: HL7 International. [viewed 15 May 2026]. Available from: https://www.hl7.org/fhir/
HRIPCSAK, G.; DUKE, J. D.; SHAH, N. H.; REICH, C. G.; HUSER, V.; SCHUEMIE, M. J.; RYAN, P. B., 2015. *[Exact article title as cited in the manuscript]*. *[Journal title]*, vol. [volume], pp. 668–669.
INTEGRATING THE HEALTHCARE ENTERPRISE (IHE), 2021. *[Exact IHE guideline/framework title as cited in the manuscript]*. [Place of publication not provided]: IHE International.
ISO, 2014. *ISO/TS 22600-1:2014 Health informatics – Privilege management and access control – Part 1: Overview and policy management*. Geneva: International Organization for Standardization.
KAPITAN, M.; ZIMMERMANN, J.; STUMPF, S., 2025. *[Exact title as cited in the manuscript]*. *[Journal/Conference/Publisher]*, p. 28.
OECD, 2023. *[Exact OECD report title as cited in the manuscript]*. Paris: Organisation for Economic Co-operation and Development.
TEHDAS JOINT ACTION, 2023. *[Exact TEHDAS report title as cited in the manuscript]*. [Place of publication not provided]: [Publisher not provided].
WORLD HEALTH ORGANIZATION (WHO), 2016. *[Exact WHO document title as cited in the manuscript]*. Geneva: World Health Organization.

СЕМАНТИЧНА ОПЕРАТИВНА СЪВМЕСТИМОСТ ПРИ AI-ПОДКРЕПЕНО ПРОГНОЗИРАНЕ НА MCI-TO-AD: ВЪЗПРОИЗВОДИМА РАМКА FHIR-TO-OMOP

Резюме: *Семантичната оперативна съвместимост е основна предпоставка за надеждни модели за прогнозиране, базирани на изкуствен интелект, при невродегенеративни заболявания, особено за ранна оценка на риска при прехода от леко когнитивно увреждане (MCI) към болестта на Алцхаймер (AD). Хетерогенните източници на данни, несъгласуваната терминология и зависимите от контекста клинични интерпретации ограничават възпроизводимостта и междусистемната използваемост на рутинните здравни данни.*
Това проучване представя компактна, подкрепена от изкуствен интелект рамка за семантична оперативна съвместимост, базирана на HL7 FHIR и OMOP Common Data Model. Изкуственият интелект се прилага като междинен слой за съгласуване, за да подкрепи контекстуално чувствителното хармонизиране на клиничните данни, а не за да замести установените стандарти. Модулната архитектура позволява гъвкава валидация и адаптация на отделните компоненти, без да се компрометира целостта на системата. Използвайки прогнозирането на MCI–AD като илюстративен пример, рамката демонстрира как семантичната хармонизация подобрява концеп-

туалната съгласуваност в ранните, диагностично несигурни етапи, където подходите, основани само на правила, са недостатъчни. Подходът поддържа проследимост, интерпретативност и методологична стабилност, като предоставя мащабируема основа за оперативно съвместими, EHDS-съвместими клинични изследователски среди.

Ключови думи: *семантична оперативна съвместимост; изкуствен интелект в здравеопазването; трансформация FHIR–OMOP; калибриране на модели; системи за подпомагане на клиничните решения*

Хомейра Хених-Намвар, докторант
Университет по библиотекознание и информационни технологии
София, България
E-mail: homeira.namvar@gmail.com

PUBLIC COMMUNICATIONS AND INFORMATION SCIENCES
ОБЩЕСТВЕНИ КОМУНИКАЦИИ И ИНФОРМАЦИОННИ НАУКИ

PSYCHOLOGICAL CONTRACT REPAIR AFTER BREACH DURING ORGANIZATIONAL CHANGE

Hüseyin Özdemir
University of Library Studies and Information Technologies
https://doi.org/10.70300/FUHR5670

Abstract: *Psychological contract breach during organizational transformation threatens trust, commitment, and performance. This integrative synthesis consolidates findings on mechanisms, boundary conditions, and outcomes of Psychological Contract Repair. Drawing on methodologically transparent publications from OECD contexts, it identifies perceived organizational support, transparent change communication, and transformational leadership as recurring levers. It also exposes a measurement gap between breach appraisal and active repair, which constrains causal inference. Repair is typically protracted and non-linear, with wave-like trajectories and limited durability (decay) unless reinforcement is sustained. Without sustained reinforcement, early gains in retention and performance attenuate, consistent with reciprocal resource loss dynamics. Overall, the findings challenge purely technocratic change models and underscore the need for longitudinal, intervention-focused research.*
Keywords: *Psychological Contract Repair; Trust Recovery; Change Communication; Change Management; Employee Retention*

INTRODUCTION

The repair of psychological contracts during profound organizational transformation represents a priority managerial challenge. Effective repair is critical for organizational sustainability and retention. Perceived breaches destabilize the employment relationship and commonly precipitate erosion in trust and performance. Understanding when repair succeeds therefore requires examining organizational antecedents that shape perceived reciprocity. Cumulative evidence indicates that repair outcomes are not stochastic; they are systematically governed by identifiable conditions and by the credibility and design of change management interventions (Li et al. 2020; Yue, Men, & Ferguson 2019).

Theoretically, Psychological Contract Repair (PCR) must be distinguished from the antecedent cognitive appraisal of breach. Although conceptually distinct, empirical operationalizations often remain insufficiently rigorous, which complicates comparability and causal inference (Gillani, Kutaula, & Budhwar 2021). Repair is also inherently temporal. It typically unfolds as an incremental, non-linear process with recursive trajectories. A decay effect has been documented; initial gains in retention and performance attenuate when reinforcement and supportive organizational routines are not sustained, which underscores the fragility of restoration (Griep & Bankins 2020; Zacher & Rudolph 2020).

This review advances a communication-centric account of PCR. It posits that information processes, specifically transparency, consistency, and procedural fairness, are not passive vectors. They are central mechanisms for sense-making and reappraisal. In transformational settings, these signals function as critical levers for repair. The objective is an integrative evidence synthesis on PCR during organizational change. It synthesizes evidence on practices associated with repair, construct delineation, temporal trajectories, and boundary conditions. This establishes a conceptually grounded foundation for evidence-informed change communication.

RESEARCH METHODOLOGY

This review consolidates evidence on PCR with a focus on communication as the primary infrastructure through which employees recalibrate trust and perceived reciprocity during change.

Methodologically, it employs an integrative structured review with narrative synthesis. This approach accommodates heterogeneity in research designs, levels of analysis, and construct operationalizations. It prioritizes conceptual coverage over exhaustive retrieval. The synthesis privileges studies that offer stronger inferential value for the focal claims, particularly designs that clarify temporal ordering and mechanisms. Cross-sectional findings are interpreted with caution.

Evidence was identified through a structured manual search of established scholarly databases and citation indexes. The search was organized in thematically defined clusters (organizational practices, construct delineation, temporal trajectories, moderators). Backward and forward citation tracking supplemented the search. Eligibility was restricted to peer-reviewed publications situated in organizational change contexts. Inclusion prioritized methodological transparency, conceptual clarity, and designs with higher inference potential (longitudinal, panel, experimental, quasi-experimental). The sampling frame focused on OECD settings to enhance interpretability. Implications for broader generalization are addressed as a limitation.

Study selection followed a staged screening process based on predefined quality criteria, including design suitability and construct validity. Given the literature's diversity, quality appraisal relied on explicit pragmatic criteria rather than a universal risk-of-bias tool. Conclusions are evidence-weighted. Interpretive strength is calibrated to the rigor of the underlying designs and to consistency across studies. A narrative synthesis integrates findings across clusters to identify converging patterns, mechanisms, boundary conditions, and evidence gaps, particularly regarding the durability of repair effects.

RESULTS

The narrative synthesis organizes the included evidence across several core domains, including organizational antecedents, recurring repair-related practices, construct delineation between appraisal and repair, temporal trajectories, lagged outcomes, and boundary conditions. Across these domains, the literature depicts PCR as a process that is shaped by organizational signals and routines, and that unfolds over time in ways that vary by context, design, and measurement approach.

Organizational antecedents of repair. The reviewed studies converge in describing a set of structural and communicative conditions that appear to provide a foundation for more favorable post-breach trajectories. A supportive organizational context is repeatedly linked to repair-relevant responses, including settings characterized by a positive psychosocial safety climate (Brunetto et al. 2021). Perceived organizational support (POS) is also consistently associated with more favorable evaluations of organizational intent and with reduced vulnerability to further negative interpretations, and it is commonly discussed as a signal that employees feel valued by the organization (Griep and Bankins 2020). In addition, transparent communication and higher perceived procedural and interactional justice are repeatedly connected to lower uncertainty and to more constructive sense-making during change (Chang and Busser 2020; Georgalis et al. 2015). Several studies also emphasize the role of pre-existing trust in supervisors and managers, which is described as a resource that can shape how repair attempts are interpreted once disruption occurs (Men, Yue and Liu 2020; Brewer, Dierkes and Norful 2024).

Practices associated with repair. Across the included evidence, five practices are most consistently described as being associated with more favorable repair-related responses, while the magnitude and durability of these associations vary by design and setting. First, transparent internal communication is frequently linked to lower workforce uncertainty and to coping responses that are interpreted as more adaptive within the respective studies (Li et al. 2020; Yue, Men and Ferguson 2019). Second, transformational leadership is repeatedly associated with higher reported trust, and with more favorable job-related attitudes, including job satisfaction (Boamah 2022; Men, Yue and Liu 2020). Third, empathetic leadership communication is often highlighted as salient for rebuilding affective trust, particularly when it includes emotional acknowledgement and credibility signals that reduce perceptions of disregard (Yue, Thelen and Walden 2022). Fourth, participatory approaches and informational justice are described as reducing resistance to change

by involving employees in the interpretation of events and the justification of decisions, thereby supporting sense-making and perceived fairness (Georgalis et al. 2015). Fifth, individualized agreements (i-deals) are associated with higher reported organizational commitment in settings where employees perceive negotiated arrangements as credible and supported by the organization (Davis and Van Der Heijden 2018; Prieto and Talukder 2023). In contrast, laissez-faire leadership is commonly described as unfavorable in these contexts, and it is linked to higher strain indicators in the studies that examine it (Groulx et al. 2024). A recurring limitation is that many contributions emphasize short-term patterns, whereas evidence on longer-term durability of specific practices remains comparatively limited (Franken, Plimmer and Malinen 2020).

Construct delineation and operationalization. At the conceptual level, the literature commonly distinguishes PCR from the cognitive appraisal that follows a perceived breach. In practice, however, empirical operationalizations vary, and this variability often complicates interpretation of temporal ordering and mechanism claims. Dynamic phase models are frequently invoked to articulate the sequence from disruption to appraisal to subsequent relational pathways. As summarized in Figure 1, Rousseau, Hansen and Tomprou (2018) conceptualize the psychological contract as a cyclical process rather than a static agreement, and they distinguish a Maintenance Phase (stable exchange) from a Disruption that triggers downstream self-regulation processes. In this model, the transition out of maintenance is initiated by a disruption event, which is tied to a discrepancy between expected and actual inducements, or to the speed of inducement delivery relative to expectations (velocity feedback). The ensuing pathway is shaped by appraisal valence. Positive affect is linked to a renegotiation pathway, and negative affect is linked to a repair pathway. Repair is described as cognitively demanding, and it involves sense-making and active attempts to reduce discrepancies and restore a more stable exchange relationship.

When mapped onto empirical measurement, the synthesis indicates a recurring gap between theoretical differentiation and measurement practice. Some longitudinal studies apply separate measures for constructs that correspond to negative affect or violation and to repair-oriented intentions or behaviors, which is more consistent with the conceptual sequence (Griep and Bankins 2020). In many studies, however, measures overlap, or constructs are not assessed in a way that clearly distinguishes appraisal from repair, which blurs the empirical trace of the transition from affective evaluation to repair-related action (Zacher and Rudolph 2020). As a result, the bifurcation between renegotiation and repair described in phase models is often difficult to examine directly in quantitative designs. Qualitative approaches, including grounded theory designs, frequently provide richer process descriptions from appraisal through coping to repair, although these designs typically do not employ standardized measurement instruments that would support stronger cross-study comparability (Gillani, Kutaula and Budhwar 2021). Taken together, the synthesis highlights the need for more consistently validated measurement approaches that separate appraisal-related affect from repair-oriented processes, and for intervention studies that more directly test repair mechanisms.

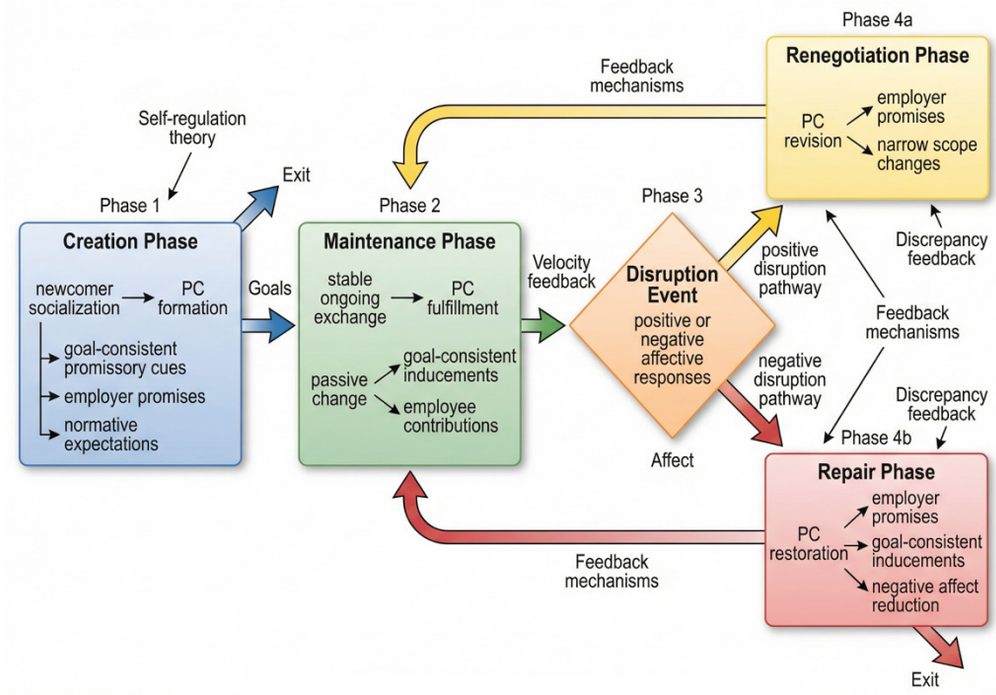


Fig. 1. Dynamic phase model of contract repair: From breach to restoration (Rousseau et al. 2018; Gillani, Kutaula and Budhwar 2021)

Temporal dynamics (trajectories and durability). With respect to temporal dynamics, higher quality panel studies commonly report that trust and affective commitment decline after breach-related events, and that recovery tends to be gradual and often incomplete within the observed time windows (Griep and Bankins 2020; Zacher and Rudolph 2020; Guerrero, Bentein and Lapalme 2014). A weekly panel study describes a negative feedback loop in which lower POS is associated with heightened vulnerability to subsequent negative interpretations, and full return to baseline is rarely observed within the study horizon (Griep and Bankins 2020). Related work similarly describes wave-like patterns of fluctuation, sometimes framed as ebb and flow, rather than monotonic recovery (Zacher and Rudolph 2020; Rodwell and Gulyas 2015). Moderators appear to shape the pace and extent of recovery. Managerial emotional intelligence is discussed as relevant for rebuilding commitment after crises in the studies that examine it (Anvari et al. 2023). Individual resources, including occupational future time perspective, are also associated with different recovery trajectories (Zacher and Rudolph 2020). Cultural variation is suggested by experimental evidence, including findings in which recovery patterns differ across national samples, although such results remain difficult to generalize without broader longitudinal replication (Eckerd et al. 2016).

Lagged outcomes and sustainability. A central limitation across domains is that many studies observe short time horizons, and longer-term data beyond twelve months are comparatively scarce. Within those constraints, the literature nevertheless describes a recurring pattern in which early gains associated with repair attempts may attenuate over time, particularly when reinforcement and supportive routines are absent, or when new breach perceptions emerge. This pattern is often discussed as a decay effect in the broader interpretation of longitudinal findings (Griep and Bankins 2020; Zacher and Rudolph 2020). Intervention-oriented studies report short-term improvements in relevant outcomes, including retention-related attitudes and performance-related indicators, although the durability of these effects varies and is not always assessed beyond the immediate follow-up period (Chang and Busser 2020; Takase, Teraoka and Yabase 2014). COR theory provides a plausible interpretive mechanism that is frequently used to explain why repair effects may erode, namely that persistent or renewed strain can prompt resource conservation and reinforce negative interpretations, thereby limiting sustained improvements (Zacher and Rudolph 2020; Griep and Bankins 2020). From an evidence perspective, the synthesis suggests that diminishing effects over extended periods appear more consistent than conclusive, largely because

longer time horizons are rarely observed. Accordingly, multi-year panels designed to model decay dynamics more precisely remain a salient gap (Flinkman et al. 2023).

Mechanisms and boundary conditions. The reviewed evidence also emphasizes that repair-related processes depend on contextual boundary conditions that shape the meaning of organizational signals. Organizational resource availability, including time and financial capacity, is described as enabling sustained repair routines, and it is therefore treated as a relevant contextual moderator in the studies that incorporate it (Miron-Spektor et al. 2017). Cultural context is likewise discussed as shaping how repair attempts are interpreted, including how justice-related signals and leadership communication are evaluated (Kingshott et al. 2020; Phuong 2017). At the mechanism level, POS is repeatedly linked to the durability of repair-related outcomes, and leadership competence and emotional intelligence are discussed as shaping trust rebuilding processes (Griep and Bankins 2020; Anvari et al. 2023). Individual resources, including resilience and paradox mindset, are also examined as potential amplifiers of more favorable trajectories, although the strength of evidence varies across designs (Prieto and Talukder 2023; Miron-Spektor et al. 2017; Flinkman et al. 2023). The evidence base does not yet support strong, systematic comparisons across specific justice intervention types (for example, informational versus interpersonal justice), and interaction effects across bundles of practices remain insufficiently tested. This leaves an open question regarding which combinations of leadership communication and HR practices are most effective under specific change conditions.

DISCUSSION

This review integrates the current evidence base on PCR following breach perceptions during organizational change, with particular attention to communication and information processes as proximal organizational levers. Across the analyzed literature, the most frequently converging pattern is that transparent communication, supportive leadership, and participatory change processes are repeatedly linked to more favorable repair-related responses (Li et al. 2020; Yue, Men and Ferguson 2019; Men, Yue and Liu 2020). These associations are theoretically coherent with major frameworks used in the field, and they align with the view that repair depends on credible organizational signals that support sense-making and perceived reciprocity during disruption.

A mechanism-based reading of the evidence suggests that leadership and communication practices operate through multiple pathways. Transformational leadership is often discussed as relevant to retention and engagement, and it is frequently modeled as operating indirectly through trust building and through the cultivation of a constructive climate (Boamah 2022; Kim 2020). Empathetic leadership communication appears to complement these pathways by supporting affective trust through acknowledgement and relational credibility, which can shape the interpretation of organizational intent after a breach perception (Yue, Thelen and Walden 2022). In communication-centered terms, these practices are not merely channels of information transmission. They can be understood as institutionalized credibility and fairness signals that structure how employees update beliefs about organizational obligations, and how they recalibrate expectations under uncertainty.

At the same time, the synthesis highlights a persistent tension between theoretical specificity and empirical measurement. While many contributions distinguish appraisal following breach perceptions from subsequent repair processes conceptually, empirical operationalizations often do not allow these constructs to be separated cleanly. This is a key constraint on stronger claims about causal sequencing and mechanism identification (Gillani, Kutaula and Budhwar 2021; Griep and Bankins 2020; Zacher and Rudolph 2020). The phase model proposed by Rousseau et al. (2018) is particularly useful for interpreting this issue because it formalizes the transition from disruption to appraisal and into distinct pathways, including renegotiation under positive affect and repair under negative affect. When studies do not measure these elements distinctly, the predicted bifurcation becomes difficult to test directly, and this limits the interpretability of downstream claims about repair efficacy.

The strength of evidence varies by claim and by the prevalence of designs that support tem-

poral inference. Claims linking transparent communication to reduced uncertainty and to more constructive sense-making are comparatively well supported within the reviewed material, including in studies with longitudinal components (Li et al. 2020). Evidence regarding slow and often incomplete recovery trajectories for trust and commitment is suggestive and recurrent, and it is supported by panel studies, although observed horizons are frequently limited (Griep and Bankins 2020; Zacher and Rudolph 2020). Evidence on cultural moderation is less settled in the absence of broader longitudinal confirmation across contexts, even though experimental findings indicate plausible cross-national differences in recovery dynamics (Eckerd et al. 2016; Kingshott et al. 2020). Similarly, organizational support is repeatedly linked to more favorable trajectories, yet stronger causal statements remain constrained by measurement and design limitations in parts of the literature (Griep and Bankins 2020).

Several limitations of the research landscape temper strong prescriptions. A substantial share of studies rely on cross-sectional designs and self-report measures, which restrict causal inference and increase vulnerability to common method concerns. More importantly, durable effects of interventions are rarely observed over extended periods. This creates uncertainty about long-term stability, and it constrains the empirical basis for claims about decay dynamics beyond the time windows typically studied (Zacher and Rudolph 2020; Flinkman et al. 2023). In addition, systematic comparisons across justice intervention types are limited, and bundles of practices are seldom tested in ways that isolate interaction effects. This leaves open the question of which combinations of communication practices, leadership behaviors, and HR routines are most effective under specific change conditions.

Taken together, the synthesis supports a cautious but actionable interpretation. Repair appears to depend on credible, sustained organizational signals that reduce ambiguity and reinforce fairness and support, and communication and procedural routines are central to this signaling infrastructure. However, the durability of repair is best treated as contingent. It likely depends on sustained reinforcement, consistent managerial behavior, and the absence of renewed discrepancy signals that would trigger negative reappraisal. For the field to move from descriptive association to stronger prescriptive guidance, future research would benefit from designs that more directly test intervention mechanisms over longer horizons, and from measurement strategies that operationalize appraisal, renegotiation, and repair as distinct processes consistent with phase models.

CONCLUSION

This integrative evidence synthesis underscores that organizational change is not merely a technical or structural undertaking. It is also a relational and communicative process in which employees continuously interpret organizational signals and update judgments about reciprocity, trustworthiness, and obligation fulfillment. Across the reviewed literature, transparent communication, supportive and transformational leadership, and participatory change practices are consistently associated with more favorable trajectories following perceived psychological contract breaches. These practices appear to function as proximal levers because they shape uncertainty reduction, credibility assessments, and perceived fairness, and they thereby influence whether post-breach dynamics move toward stabilization rather than continued deterioration.

At the same time, the synthesis indicates that repair effects are frequently fragile and contingent. Short-term improvements in trust-related attitudes, retention-related intentions, or performance-related indicators do not reliably translate into durable recovery. Where longer observation windows are available, evidence suggests that repair gains may attenuate over time when reinforcement is not sustained, when supportive routines are absent, or when employees encounter renewed discrepancy signals that reactivate negative appraisal processes. This pattern is consistent with resource-based interpretations, including COR-informed accounts, in which repeated strain prompts resource conservation and increases sensitivity to subsequent organizational deviations. In this sense, PCR is best understood as a temporally extended process that can remain incomplete, that may proceed in non-linear trajectories, and that requires ongoing organizational credibility work rather than episodic remedial actions.

A further implication concerns construct clarity and inference quality. The literature commonly differentiates appraisal processes that follow disruption from subsequent repair processes, and phase models add theoretical precision by distinguishing pathways such as renegotiation under positive affect and repair under negative affect. Empirically, however, these distinctions are not always captured with sufficient measurement separation, which limits the ability to test temporal ordering and to specify mechanisms with high confidence. This methodological constraint materially affects the interpretability of intervention claims, especially when studies rely on cross-sectional designs or when follow-up periods are short.

The synthesis yields a clear implication. Communication and information processes should be treated as constitutive infrastructures of repair, not as peripheral contextual variables. Transparency, message consistency, justificatory accounts, and procedural fairness cues operate as recurrent signals through which employees make sense of change, recalibrate expectations, and update trust judgments. Framed this way, the review provides an evidence-grounded platform for subsequent dissertation components that examine how communicative and procedural signals shape sensemaking, reappraisal, and the stabilization of employment relationships under change conditions.

Several evidence gaps remain salient. First, the field would benefit from longer-horizon longitudinal designs, including multi-wave panels that extend beyond typical follow-up periods, to model durability and decay dynamics with greater precision. Second, stronger designs for causal inference are needed, including quasi-experimental and experimental intervention studies that isolate communication and justice-related practices and test their mechanisms directly. Third, measurement strategies require further refinement and validation to distinguish appraisal, renegotiation, and repair as separable processes, and to enable cumulative comparison across studies and contexts. Addressing these gaps is necessary to move from plausibly actionable associations toward more robust, evidence-informed guidance for change communication and post-breach repair that remains effective beyond short-term improvements.

REFERENCES

ANVARI, R.; KUMPIKAITĖ-VALIŪNIENĖ, V.; MOBARHAN, R.; JANJARIA, M.; CHERMAHINI, S., 2023. Strategic human resource management practitioners‘ emotional intelligence and affective organizational commitment in higher education institutions in Georgia during post-COVID-19. *PLOS ONE*, vol. 18, no. 12, e0295084. Available from: https://doi.org/10.1371/journal.pone.0295084

BOAMAH, S., 2022. The impact of transformational leadership on nurse faculty satisfaction and burnout during the COVID-19 pandemic: A moderated mediated analysis. *Journal of Advanced Nursing*, vol. 78, no. 11, pp. 3608–3625. Available from: https://doi.org/10.1111/jan.15198

BREWER, K.; DIERKES, A.; NORFUL, A., 2024. Organizational trust breaches among nurses and aides: A qualitative study. *Nursing Ethics*, vol. 31, no. 7, pp. 1524–1536. Available from: https://doi.org/10.1177/09697330241230520

BRUNETTO, Y.; SAHELI, N.; DICK, T.; NELSON, S., 2021. Psychosocial safety climate, psychological capital, healthcare SLBs' wellbeing and innovative behaviour during the COVID-19 pandemic. *Public Performance & Management Review*, vol. 45, no. 4, pp. 751–772. Available from: https://doi.org/10.1080/15309576.2021.1918189

CHANG, W.; BUSSER, J., 2020. Hospitality career retention: The role of contextual factors and thriving at work. *International Journal of Contemporary Hospitality Management*, vol. 32, no. 1, pp. 193–211. Available from: https://doi.org/10.1108/IJCHM-10-2018-0831

DAVIS, A.; VAN DER HEIJDEN, B., 2018. Reciprocity matters: Idiosyncratic deals to shape the psychological contract and foster employee engagement in times of austerity. *Human Resource Development Quarterly*, vol. 29, no. 3, pp. 331–355. Available from: https://doi.org/10.1002/hrdq.21327

ECKERD, S.; BOYER, K.; QI, Y.; ECKERD, A.; HILL, J., 2016. Supply chain psychological contract breach: An experimental study across national cultures. *Journal of Supply Chain Management*, vol. 52, no. 3, pp. 68–82. Available from: https://doi.org/10.1111/jscm.12101

FLINKMAN, M.; COCO, K.; RUDMAN, A.; LEINO-KILPI, H., 2023. Registered nurses‘ psychological capital: A scoping review. *International Journal of Nursing Practice*, e13183. Available from: https://doi.org/10.1111/ijn.13183

FRANKEN, E.; PLIMMER, G.; MALINEN, S., 2020. Paradoxical leadership in public sector organisations: Its role in fostering employee resilience. *Australian Journal of Public Administration*, vol. 79, no. 1, pp. 132–154. Available from: https://doi.org/10.1111/1467-8500.12396

GEORGALIS, J.; SAMARATUNGE, R.; KIMBERLEY, N.; LU, Y., 2015. Change process characteristics and resistance to organisational change: The role of employee perceptions of justice. *Australian Journal of Management*, vol. 40, no. 1, pp. 113–145. Available from: https://doi.org/10.1177/0312896214526212

GILLANI, A.; KUTAULA, S.; BUDHWAR, P., 2021. Psychological contract breach: Unraveling the dark side of business-to-business relationships. *Journal of Business Research*, vol. 134, pp. 631–641. Available from: https://doi.org/10.1016/j.jbusres.2021.06.008
GRIEP, Y.; BANKINS, S., 2020. The ebb and flow of psychological contract breach in relation to perceived organizational support: Reciprocal relationships over time. *Economic and Industrial Democracy*, vol. 43, no. 1, pp. 344–361. Available from: https://doi.org/10.1177/0143831X19897415
GROULX, P.; MAISONNEUVE, F.; HARVEY, J.; JOHNSON, K., 2024. The ripple effect of strain in times of change: How manager emotional exhaustion affects team psychological safety and readiness to change. *Frontiers in Psychology*, vol. 15, article 1298104. Available from: https://doi.org/10.3389/fpsyg.2024.1298104
GUERRERO, S.; BENTEIN, K.; LAPALME, M., 2014. Idiosyncratic deals and high performers‘ organizational commitment. *Journal of Business and Psychology*, vol. 29, no. 2, pp. 323–334. Available from: https://doi.org/10.1007/s10869-013-9316-7
KIM, Y., 2020. Organizational resilience and employee work-role performance after a crisis situation: Exploring the effects of organizational resilience on internal crisis communication. *Journal of Public Relations Research*, vol. 32, no. 1–2, pp. 47–75. Available from: https://doi.org/10.1080/1062726X.2020.1765368
KINGSHOTT, R.; SHARMA, P.; SIMA, H.; WONG, D., 2020. The impact of psychological contract breaches within East–West buyer–supplier relationships. *Industrial Marketing Management*, vol. 89, pp. 142–159. Available from: https://doi.org/10.1016/j.indmarman.2020.03.008
LI, J.; SUN, R.; TAO, W.; LEE, Y., 2020. Employee coping with organizational change in the face of a pandemic: The role of transparent internal communication. *Public Relations Review*, vol. 47, no. 1, article 101984. Available from: https://doi.org/10.1016/j.pubrev.2020.101984
MEN, L.; YUE, C.; LIU, Y., 2020. Vision, passion, and care: The impact of charismatic executive leadership communication on employee trust and support for organizational change. *Public Relations Review*, vol. 46, no. 3, article 101927. Available from: https://doi.org/10.1016/j.pubrev.2020.101927
MIRON-SPEKTOR, E.; INGRAM, A.; KELLER, J.; SMITH, W.; LEWIS, M., 2018. Microfoundations of organizational paradox: The problem is how we think about the problem. *Academy of Management Journal*, vol. 61, no. 1, pp. 26–45. Available from: https://doi.org/10.5465/amj.2016.0594
PHUONG, T., 2017. The moderating impacts of age and locus of control on the psychological contract breach–outcomes relationships: A Vietnam–Japan comparative study. *International Journal of Business and Society*, vol. 17, no. 1, pp. 329–344. Available from: https://doi.org/10.33736/ijbs.528.2016
PRIETO, L.; TALUKDER, M., 2023. Resilient agility: A necessary condition for employee and organizational sustainability. *Sustainability*, vol. 15, no. 2, article 1552. Available from: https://doi.org/10.3390/su15021552
RODWELL, J.; GULYAS, A., 2015. Psychological contract breach among allied health professionals. *Journal of Health Organization and Management*, vol. 29, no. 3, pp. 393–412. Available from: https://doi.org/10.1108/JHOM-05-2013-0107
ROUSSEAU, D. M.; HANSEN, S. D.; TOMPROU, M., 2018. A dynamic phase model of psychological contract processes. *Journal of Organizational Behavior*, vol. 39, no. 9, pp. 1081–1098. Available from: https://doi.org/10.1002/job.2284
TAKASE, M.; TERAOKA, S.; YABASE, K., 2016. Retaining the nursing workforce: Factors contributing to the reduction of nurses‘ turnover intention in Japan. *Journal of Nursing Management*, vol. 24, no. 1, pp. 21–29. Available from: https://doi.org/10.1111/jonm.12266
YUE, C.; MEN, L.; FERGUSON, M., 2019. Bridging transformational leadership, transparent communication, and employee openness to change: The mediating role of trust. *Public Relations Review*, vol. 45, no. 3, article 101856. Available from: https://doi.org/10.1016/j.pubrev.2019.04.012
YUE, C.; THELEN, P.; WALDEN, J., 2022. How empathetic leadership communication mitigates employees‘ turnover intention during COVID-19-related organizational change. *Management Decision*, vol. 61, no. 1, pp. 202–221. Available from: https://doi.org/10.1108/MD-01-2022-0011
ZACHER, H.; RUDOLPH, C., 2021. Relationships between psychological contract breach and employee well-being and career-related behavior: The role of occupational future time perspective. *Journal of Organizational Behavior*, vol. 42, no. 1, pp. 47–66. Available from: https://doi.org/10.31234/osf.io/e967j

ВЪЗСТАНОВЯВАНЕ НА ПСИХОЛОГИЧЕСКИЯ ДОГОВОР СЛЕД НАРУШЕНИЕ ПРИ ОРГАНИЗАЦИОННА ПРОМЯНА

Резюме: *Нарушението на психологическия договор по време на организационна промяна заплашва доверието, ангажираността и представянето. Този интегративен синтез консолидира резултати относно механизмите, граничните условия и последиците от възстановяването на психологическия договор. Базирайки се на методологически прозрачни публикации от контексти на ОИСР, той*

идентифицира възприеманата организационна подкрепа, прозрачната комуникация на промяната и трансформационното лидерство като повтарящи се лостове за въздействие. Той също така разкрива пропуск в измерването между оценката на нарушението и активното възстановяване, което ограничава причинно-следствените изводи. Възстановяването обикновено е продължително и нелинейно, с вълнообразни траектории и ограничена устойчивост (затихване), освен ако не се поддържа затвърждаване. Без устойчиво затвърждаване ранните постижения в задържането и представянето отслабват, в съответствие с динамиката на реципрочна загуба на ресурси. Като цяло, резултатите оспорват чисто технократичните модели на промяна и подчертават необходимостта от лонгитюдни изследвания, фокусирани върху интервенции.

Ключови думи: *възстановяване на психологическия договор; възстановяване на доверието; комуникация на промяната; управление на промяната; задържане на служители*

Хюсеин Йоздемир, докторант
Университет по библиотекознание и информационни технологии
София, България
E-mail: hueseyin079@gmail.com

PUBLIC COMMUNICATIONS AND INFORMATION SCIENCES
ОБЩЕСТВЕНИ КОМУНИКАЦИИ И ИНФОРМАЦИОННИ НАУКИ

MANAGEMENT OF INFORMATION PROCESSES IN THE GLOBAL ENERGY CRISIS

Lisa Pauline Krohm
University of Library Studies and Information Technologies
https://doi.org/10.70300/PSNR1882

Abstract: *Global energy crises are often interpreted as technical supply bottlenecks or economic shocks. This paper, however, argues that energy crises should be understood as complex socio-technical governance problems in which energy, information, and political control are closely intertwined. The central thesis of this paper is that information governance is not merely a supporting factor, but a constitutive prerequisite of modern energy systems. Based on approaches from energy history and social sciences, energy is conceptualized as a structuring basis of social development that is always linked to specific power, knowledge, and information orders.*
The analysis shows that the importance of information processes is growing significantly with the increasing complexity of modern energy systems. Decisions on energy policy and transformation paths are increasingly based on standardized data, scenarios, and institutionally embedded communication processes. Deficits in information governance can impair political decision-making, hamper international cooperation, and exacerbate geopolitical tensions. Furthermore, it is argued that energy crises are always also communication crises, the management of which depends largely on transparent, dialogical crisis communication. The article therefore advocates for a stronger institutional anchoring of information governance.
Keywords: *Information processes; Energy governance; Energy crisis; International energy policy; Socio-technical system*

INTRODUCTION

The current global energy crisis is often described in political and scientific debates as a technical supply problem or as a consequence of external shocks such as geopolitical conflicts, market distortions, or climate-related extreme events. However, such perspectives fall short because they do not adequately capture the underlying structural conditions of modern energy systems. Energy crises are not solely an expression of physical scarcity or technological limitations, but arise from complex interactions between technical infrastructures, social orders, and political control processes.

This work is based on the assumption that energy crises can only be adequately understood if energy is viewed as a socio-technical phenomenon that is always linked to specific power, knowledge, and information orders. Historically, energy systems have never been purely technical arrangements, but have always been organized institutionally, politically, and communicatively. With the increasing complexity of modern energy systems, the importance of information processes that coordinate decisions, stabilize expectations, and enable political action is growing.

Against this backdrop, the work focuses on information governance and strategic crisis communication. It argues that deficits in the collection, processing, and communication of energy-related information are key escalation mechanisms in modern energy crises. Insufficient transparency, information asymmetries, and miscommunication not only impair political decision-making processes, but also undermine international cooperation and social acceptance of energy policy measures.

The aim of this paper is therefore to develop a theoretical framework that conceptualizes energy crises as information-dependent governance problems. Based on a historically oriented literature analysis, long-term developments in energy, information, and power structures are identified and their significance for the current energy crisis is discussed. The paper thus makes a conceptual contribution to energy policy research and provides impetus for a more information- and communication-oriented design of future energy policy.

While energy policy research in recent years has increasingly addressed issues of data availability, transparency, and evidence-based decision-making, information processes continue to be treated primarily as a technical or downstream dimension of political governance. This paper challenges this perspective and argues that information infrastructures themselves constitute political structures that significantly shape power relations, options for action, and crisis dynamics.

In contrast to existing literature in the field of energy and management, which treats information processes primarily as a supporting or downstream dimension, this paper argues that information infrastructures are a constitutive prerequisite for modern energy systems. The specific contribution of this work is to systematically analyze information governance as an independent dimension of modern energy systems that is relevant to power and control, and to link this perspective to long-term developments in energy history. This develops a theoretical framework that not only explains energy crises, but also opens up new approaches for preventive energy policy control.

The analysis shows that the global energy crisis, and in particular the transition to hydrogen-based energy systems, is not solely the result of technological or economic constraints, but is significantly influenced by structural deficits in the organization, distribution, and political embedding of knowledge and information.

Theoretical foundations

Based on approaches from energy history and social sciences, energy is understood as a structuring foundation of social development that is always linked to specific power, knowledge, and information systems. Building on this, historical energy eras, fossil energy systems, and the role of information governance and strategic crisis communication are analyzed in order to conceptualize energy crises as socio-technical and information-dependent governance problems. This work situates energy crises in the field of tension between technical development, social order, and political control.

For analytical clarity, a distinction must be made between information governance and data-related energy policy. While data policy primarily refers to the collection, standardization, and availability of quantitative energy data, information governance also encompasses the institutional embedding of knowledge, the interpretation of uncertainties, and the communicative mediation of energy policy decisions. Information governance thus refers not only to the technical management of data, but also to a political control field in which knowledge is selected, legitimized, and made relevant for action.

Similarly, strategic crisis communication must be distinguished from general political communication. While political communication often aims at legitimization or positioning, crisis communication focuses on stabilizing expectations under conditions of increased uncertainty. It structures perceptions of urgency, responsibility, and fairness, thereby directly influencing the escalation or de-escalation dynamics of energy crises.

RESEARCH METHODOLOGY

This publication is based on qualitative, historically oriented literature research. Interdisciplinary works from the fields of energy and technology history, political economy, social theory, and information science were analyzed, with a focus on publications from the last three decades. The literature was selected based on topics along the analytical axes of energy forms, social order, and information processes.

Methodologically, the work follows a comparative content analysis that aims to identify long-term trends, transitions, and breaks in energy history and interpret their significance for the current energy crisis. The chosen approach allows for a systematic reconstruction of long-term structural patterns that often remain invisible in empirically focused studies. This approach does not aim at causal explanation, but at reconstructing long-term structural conditions of energy governance.

Energy, technology, and social development

In energy history and social science research, energy is no longer regarded exclusively as a physical quantity, but as a central category of social development. Numerous studies show that the

availability, control, and organization of energy have a decisive influence on population development, productivity, technological innovation, and institutional stability (White 1943, pp. 344–399; Smil 2004, p. 549). Energy forms the material basis of human existence and at the same time shapes social, political, and informational orders. The availability and controllability of energy influence population growth, productivity, and technological innovation, and create the conditions for social differentiation and institutional stability. Even early societies developed forms of organization that were directly linked to their energy resources (White 1943, p. 399; Cook 1976, p. 176; Smil 2004, p. 549).

Historians and social scientists therefore describe energy as the foundation of complex civilizations, whose social and political structures are always linked to the dominant forms of energy at the time (Weissenbacher 2009; Goudsblom 2012, p. 408). Where there was stable access to energy, surpluses could be generated that served purposes beyond merely securing immediate livelihoods. These surpluses not only created the basis for the division of labor and specialization, but also fostered the development of administrative structures that relied on systematic knowledge organization, planning, and control. This shows how energy and information are closely linked, because as energy use gets more complex, the need for coordination, knowledge, and institutional decision-making mechanisms also grows.

In the context of this work, information is understood as a socially embedded process of generating, processing, and using knowledge to coordinate complex systems. Information processes thus encompass technical data collection as well as institutional decision-making mechanisms, organizational knowledge, and communicative forms of control. This perspective allows energy systems to be analyzed as socio-technical entities whose stability depends largely on functioning information flows.

Technological upheavals in the energy sector have therefore always led to changes in power and information structures. Control over energy sources meant control over production processes, social resources, and decision-making spaces. Energy thus functions less as a mere technical factor of production and more as a structured element of social order that stabilizes or shifts power relations and structures information flows.

Historical energy ages

Research identifies several overlapping energy ages, including the hunter-gatherer societies phase, the Neolithic agricultural society, the age of mechanical energy, the industrial use of fossil fuels, and electrification. Each of these phases brought with it new forms of energy production, storage, and distribution, while also changing social structures, the organization of labor, and political power relations (Gimpel 2003, pp. 9–36; Weissenbacher 2009, 11–17-19).

In early hunter-gatherer societies, energy use was predominantly based on somatic energy. Information and power structures were organized in a decentralized manner, social hierarchies were limited, and political control mechanisms were rudimentary (Gopher, Barkai 2017, pp. 203–214). The Neolithic Revolution and the use of biological energy fundamentally changed these structures. Sedentary lifestyles, food storage, and agricultural production required new forms of information processing, for example, for planning harvest cycles, managing resources, and enforcing property and power claims (Weissenbacher 2009, pp. 11–17; Larsen et al. 2015, pp. 27–68).

The transition to mechanical forms of energy, particularly hydro and wind power, reinforced these developments. Energy became increasingly concentrated in specific locations and technically mediated, requiring new forms of organization and institutional control mechanisms (Gimpel 2003, p. 14; Smil 2004, p. 553). As energy density increased, not only did dependence on technical systems grow, but so did the need for specialized knowledge and organizational control. Historical energy eras are therefore not just stages of technological development, but mark fundamental changes in the way societies organize information and exercise power.

Fossil energy and geopolitical order

With industrialization, energy transformed from a locally bound resource into a geopolitical power base. Coal, oil, and electricity enabled industrial mass production, global mobility, and military supremacy, but at the same time led to new dependencies, environmental pollution, and

international conflicts (Cherp, Jewell 2011, p. 202). Fossil energy systems promoted a high degree of centralization of production, capital, and political decision-making power. This definition follows organizational and systems theory approaches, in which information is understood as a prerequisite for collective action and institutional control.

Historically, this development is exemplified by the establishment of global oil regimes in the 20th century. Fossil energy systems changed not only economic structures but also international power relations. Access to energy sources and transport routes became a strategic factor in state policy. Energy imports and exports created new dependencies, while energy-intensive industries reinforced geopolitical interests (Cherp, Jewell 2011, p. 202; Yergin 2011, pp. 1–820). Producing countries, transnational corporations, and government institutions became closely intertwined through long-term supply contracts, transport infrastructures, and information-intensive planning processes. Access to reliable data on production volumes, reserves, and transport routes became a strategic power factor that significantly influenced political decisions and geopolitical conflicts.

Recent studies emphasize that fossil fuel energy systems are increasingly information-intensive and would not be functional without complex control and communication processes. Producing countries, transnational corporations, and government institutions became closely intertwined through long-term supply contracts, transport infrastructures, and information-intensive planning processes. Access to reliable data on production volumes, reserves, and transport routes has become a strategic power factor that significantly influences political decisions and geopolitical conflicts. The current global energy crisis is therefore not only a consequence of ecological and technical limitations, but also the result of historically grown energy, information, and power structures that continue to have an impact today.

Although information processes have often been treated implicitly or only marginally in energy policy research, we believe that information should be considered a separate analytical category. Information architectures are conceived not as technical additions, but as structuring prerequisites for control, power, and collective action.

Information governance and international energy policy

The above analysis clearly shows that information processes represent an independent governance dimension of modern energy systems. International energy policy is increasingly dependent on data, scenarios, and forecasts not only being available, but also being collected according to uniform standards, communicated transparently, and embedded in institutions. Decisions on energy infrastructure, security of supply, and long-term transformation paths are based to a large extent on modeling, risk assessments, and the continuous exchange of information between government actors, economic stakeholders, and international organizations. Information processes thus function as a central interface between technical expertise, political decision-making, and social legitimacy.

Deficits in information governance not only lead to inefficient or delayed political decision-making processes, but also exacerbate geopolitical tensions and loss of trust between states (Dreher, Gaston, Martens 2008; Escribano, Valdés 2017). Incomplete, distorted, or politically manipulated information hinders international cooperation, promotes misperceptions about dependencies and scope for action, and increases the risk of escalations in energy policy. Information asymmetries not only distort market mechanisms, but also create political vulnerabilities by systematically reinforcing uncertainties about dependencies, risks, and options for action. In crisis situations in particular, a lack of reliable and shared information can contribute to the hardening of national interests and undermine coordinated approaches to solutions.

From a policy-oriented perspective, this highlights the need to explicitly anchor information management as an integral part of energy security and transformation strategies. Information governance should not be understood as a purely technical instrument, but as a political control area that must be actively shaped. International organizations such as the International Energy Agency (IEA), the United Nations (UN), and the G20 play a key role in this regard by bundling information flows, ensuring the comparability of data and scenarios, and reducing existing information asymmetries between states and regions.

Furthermore, the development of open, transparent, and trustworthy data infrastructures is of central importance. Standardized reporting formats, joint monitoring systems, and publicly accessible data bases can reduce uncertainty and strengthen the basis for cooperative energy policy decisions. Information governance thus becomes a key prerequisite for a functioning international energy policy and a decisive factor in overcoming the global energy crisis under conditions of high uncertainty.

Strategic crisis communication and social stability

Energy crises are not only technical or economic problems, but always also communication crises. Energy policy measures have a profound impact on everyday practices, cost structures, and social expectations, affecting households, businesses, and social groups in different ways. Acceptance and political stability therefore depend largely on how energy policy decisions are communicated, justified, and embedded in existing social patterns of interpretation. Communication does not merely accompany these processes, but structures perceptions of urgency, fairness, and political responsibility.

The dissertation shows that social tensions and political protests arise in particular when energy policy measures are perceived as opaque, contradictory, or socially unjust. Deficient crisis communication increases uncertainty, mistrust, and social polarization and can significantly exacerbate existing crisis dynamics (Chen et al. 2024). In such situations, political actors lose credibility, while simplistic or emotional interpretations gain influence. Communication thus becomes a factor that contributes to the escalation or stabilization of crises.

From a political perspective, this necessitates a strategic, dialogue-based communication practice. This should involve relevant stakeholders at an early stage, make decision-making processes transparent, and openly address uncertainties rather than covering them up in communications. This form of crisis communication helps to build trust as a key resource in energy policy and stabilize it in the long term. Strategic crisis communication thus becomes an integral part of preventive energy governance and a crucial tool for ensuring social acceptance and political capacity to act in phases of profound energy transformation.

RESULTS

Theoretical analysis shows that energy crises should not be understood in isolation as technical supply problems or economic shocks, but rather as an expression of deeper socio-technical and information-related structural problems. Across all historical energy eras, it can be seen that the stability of energy systems depends significantly on the ability to coordinate energy flows, knowledge, and decision-making processes.

A key finding is that as energy density and system complexity increase, so does the importance of information processes. Even in early forms of energy, energy availability and social organization were closely linked. With industrialization and the dominance of fossil fuels, this dependency intensified: energy systems became increasingly information-intensive, globally interconnected, and politically sensitive. Access to reliable data on production volumes, reserves, transport routes, and demand trends developed into an independent power factor that influences political capacity to act and geopolitical positions.

The results also show that shortcomings in information governance are key escalation mechanisms in modern energy crises. Where data is fragmented, opaque, or politically exploited, international cooperation mechanisms come under pressure. Misperceptions about dependencies and scope for action reinforce national interests and make coordinated political responses more difficult. Information asymmetries between states and groups of actors thus contribute significantly to the instability of global energy systems.

Furthermore, the analysis shows that energy crises are always also communication crises. Political measures in the energy sector do not only have an impact through technical or economic effects, but also through their social perception. Deficient crisis communication increases uncertainty, mistrust, and polarization and can exacerbate existing lines of conflict. Acceptance problems and social protests are less an expression of irrational rejection than the result of inadequate,

contradictory, or perceived unfair communication processes.

In summary, the results make it clear that energy crises escalate particularly where both information governance and strategic crisis communication are insufficiently developed. Information processes thus function as a key mediating dimension between energy, politics, and society and are crucial for the stability and controllability of modern energy systems.

DISCUSSION

The results of this work suggest that energy crises need to be fundamentally reconceptualized. Instead of interpreting them primarily as a consequence of physical scarcity, technological limitations, or external shocks, it seems more analytically fruitful to understand them as information-dependent governance crises. The energy, information, and power have always been historically intertwined and that this entanglement has become even more pronounced in modern energy systems.

Conceptualizing energy crises as information-dependent governance crises shifts the analytical focus away from short-term scarcity and toward the institutional organization of knowledge, data, and communication. Energy crises thus appear less as exceptional circumstances and more as moments of structural stress on existing control and information systems.

Against this backdrop, it becomes clear that political governance deficits are less attributable to a lack of technical knowledge than to institutional and communicative weaknesses. Information governance is proving to be an independent field of political action that has often been underestimated in energy policy research and practice to date. Without transparent, comparable, and institutionally anchored information structures, international energy policies lose their ability to coordinate, especially in crisis conditions.

At the same time, the findings highlight the central role of strategic crisis communication for social stability. Energy policy decisions have a profound impact on social order and generate distributional effects that must be moderated through communication. Where communication obscures uncertainties or fails to openly address conflicting political goals, it undermines trust and exacerbates political polarization. Crisis communication should therefore not be seen as a downstream accompanying measure, but as an integral part of preventive energy governance.

The discussion also shows that international organizations such as the IEA, the UN, and the G20 can play a key role in institutionalizing information governance and communication standards. This is exemplified by recent developments in European energy and gas markets, where insufficient transparency regarding availability, pricing mechanisms, and political intervention has contributed significantly to uncertainty among market players and the general public. Standardized data bases, joint monitoring systems, and transparent scenario communication would not only improve the quality of political decisions, but also strengthen trust between states and social actors.

Implications for energy policy and transformation governance

The results of this work have direct implications for the design of energy policy control and transformation processes. If energy crises are understood as information-dependent governance crises, the institutional design of information and communication processes itself becomes the focus of political action. Energy policy is then not limited to setting technological targets or economic incentive structures, but also encompasses the active organization of knowledge, uncertainty, and social expectations.

For transformation governance, this means that political control is increasingly taking place under conditions of high uncertainty, long-term conflicts of interest, and pluralistic knowledge bases. In this context, information governance acts as a mediating infrastructure that allows different forms of knowledge, scientific expertise, economic forecasts, social experiences, to be integrated into political decision-making processes. Without such institutionally anchored information structures, transformation processes risk failing due to a lack of coordination, declining acceptance, or political blockades.

At the same time, the findings underscore the importance of dialogical communication formats for legitimizing energy policy decisions. Transformation processes require not only agreement on specific measures, but also the long-term willingness of social actors to bear uncertainty, costs,

and distributional effects. Strategic crisis communication should therefore be understood as part of a broader transformation communication strategy that systematically promotes transparency, participation, and trust. The ability to steer energy policy is thus increasingly becoming a question of successful information and communication architectures.

Overall, the results and their discussion underscore that overcoming the global energy crisis is failing less due to a lack of technical solutions than to the inadequate political organization of knowledge, information, and communication. A sustainable transformation of energy systems therefore requires not only technological innovations, but above all institutional reforms that systematically understand information processes as the basis for collective action.

CONCLUSION

This work has shown that energy crises cannot be understood solely as technical supply problems or economic shocks. Rather, they are complex socio-technical governance crises in which energy, information, and political control are inextricably intertwined. Historical and theoretical analysis makes it clear that energy systems have always been based on specific knowledge and information systems and that the importance of functioning information processes increases with growing system complexity.

The central finding is that information governance is an independent and hitherto underestimated dimension of international energy policy. Deficits in the collection, standardization, and communication of energy-related data have a direct impact on political decision-making, international cooperation, and geopolitical stability. Against the backdrop of increasing digitalization, data-based energy models, and AI-supported forecasting systems, this dimension is becoming even more important as issues of transparency, democratic control, and institutional embedding of information processes continue to gain relevance. In crisis situations in particular, information asymmetries, lack of transparency, and politically instrumentalized data can exacerbate existing conflicts and undermine coordinated approaches to solutions.

Furthermore, the analysis underscores the key role of strategic crisis communication in achieving social acceptance and political capacity to act. Energy policy measures are effective not only through their material effects, but also through their social perception. Where communication is perceived as opaque, contradictory, or socially unjust, the risk of social tensions, protests, and political polarization increases significantly. Trust thus proves to be a central resource of modern energy policy that must be actively built and stabilized.

Against this backdrop, the paper advocates for a stronger institutional anchoring of information management and crisis communication in energy policy strategies. International organizations such as the IEA, the UN, or the G20 can play a central role in this by promoting open data infrastructures, establishing comparability, and strengthening dialogical communication formats. Sustainable management of the global energy crisis therefore requires not only technological innovations, but above all a political reorientation that systematically takes information processes into account as the basis for collective control.

The analysis suggests that political learning processes in the energy sector have so far lagged behind technological and economic dynamics. Finally, the increasing digitalization of energy policy control processes creates a need for further research and policy action. Energy models, scenarios, and forecasts are increasingly being generated by algorithmic processes, AI-supported analyses, and automated decision support systems. While these developments increase the analytical performance of energy policy planning, they also raise questions of transparency, traceability, and democratic control.

Against this backdrop, information governance takes on an additional dimension: not only the availability of data, but also the comprehensibility, explainability, and communicative embedding of algorithmically generated knowledge are becoming central prerequisites for political legitimacy. Future energy policy therefore faces the challenge of embedding digital and AI-supported control instruments in institutional and communicative structures in such a way that political accountability, social participation, and trust are maintained.

REFERENCES

CHEN, P. S.-L.; FAN, H.; ENSHAEI, H.; ZHANG, W.; SHI, W.; ABDUSSAMIE, N.; MIWA, T.; QU, Z.; YANG, Z., 2024. Opportunities and challenges of hydrogen ports: an empirical study in Australia and Japan. *Hydrogen* [online], vol. 5, no. 3, pp. 436–458. Available from: https://doi.org/10.3390/hydrogen5030025

CHERP, A.; JEWELL, J., 2011. The three perspectives on energy security: intellectual history, disciplinary roots and the potential for integration. *Current Opinion in Environmental Sustainability* [online], vol. 3, no. 4, pp. 202–212. Available from: https://doi.org/10.1016/j.cosust.2011.07.001

COOK, E., 1976. *Man, energy, society*. San Francisco, CA: Freeman.

DREHER, A.; GASTON, N.; MARTENS, P., 2008. *Measuring globalisation*. New York, NY: Springer.

ESCRIBANO, G.; VALDÉS, J., 2017. Oil prices: governance failures and geopolitical consequences. *Geopolitics* [online], vol. 22, no. 3, pp. 693–718. Available from: https://doi.org/10.1080/14650045.2016.1254621

GIMPEL, J., 2003. *The medieval machine: the industrial revolution of the Middle Ages*. New York, NY: Barnes & Noble.

GOPHER, A.; BARKAI, R., 2017. Qesem Cave and the Acheulo-Yabrudian Cultural Complex of the Levant. In: ENZEL, Y.; BAR-YOSEF, O., eds. *Quaternary of the Levant*. Cambridge: Cambridge University Press, pp. 203–214.

GOUDSBLOM, J., 2012. Energy and civilization. *International Review of Sociology* [online], vol. 22, no. 3, pp. 405–411. Available from: https://doi.org/10.1080/03906701.2012.753780

LARSEN, C. S.; HILLSON, S. W.; BOZ, B.; PILLOUD, M. A.; SADVARI, J. W.; AGARWAL, S. C.; GLENCROSS, B.; BEAUCHESNE, P.; PEARSON, J.; RUFF, C. B.; GAROFALO, E. M.; HAGER, L. D.; HADDOW, S. D.; KNÜSEL, C. J., 2015. Bioarchaeology of Neolithic Çatalhöyük: lives and lifestyles of an early farming society in transition. *Journal of World Prehistory* [online], vol. 28, no. 1, pp. 27–68. Available from: https://doi.org/10.1007/s10963-015-9084-6

SMIL, V., 2004. World history and energy. *Encyclopedia of Energy*, vol. 6, pp. 549–561.

WEISSENBACHER, M., 2009. *Sources of power: how energy forges human history*. Santa Barbara, CA: ABC-CLIO.

WHITE, L. A., 1943. Energy and the evolution of culture. *American Anthropologist* [online], vol. 45, no. 3, pp. 335–356. Available from: http://www.jstor.org/stable/663173

YERGIN, D., 2011. *The Quest: energy, security and the remaking of the modern world*. London: Penguin.

УПРАВЛЕНИЕ НА ИНФОРМАЦИОННИТЕ ПРОЦЕСИ ПРИ ГЛОБАЛНАТА ЕНЕРГИЙНА КРИЗА

Резюме: *Глобалните енергийни кризи често се тълкуват като технически затруднения в доставките или икономически сътресения. В тази статия обаче се твърди, че енергийните кризи трябва да се разглеждат като сложни социално-технически проблеми на управлението, в които енергията, информацията и политическият контрол са тясно взаимносвързани. Основната теза на тази статия е, че управлението на информацията не е просто поддържащ фактор, а съставна предпоставка за съвременните енергийни системи. Въз основа на подходи от историята на енергетиката и социалните науки енергията се концептуализира като структурираща основа на социалното развитие, която винаги е свързана с конкретни порядки на власт, знание и информация.*

Анализът показва, че значението на информационните процеси нараства значително с увеличаващата се сложност на съвременните енергийни системи. Решенията относно енергийната политика и пътищата за трансформация все повече се основават на стандартизирани данни, сценарии и институционално заложени комуникационни процеси. Дефицитите в управлението на информацията могат да затруднят политическото вземане на решения, да възпрепятстват международното сътрудничество и да изострят геополитическите напрежения. Освен това се твърди, че енергийните кризи винаги са и комуникационни кризи, чието управление зависи до голяма степен от прозрачна, диалогична кризисна комуникация. Статията следователно се застъпва за по-силно институционално закрепване на управлението на информацията.

Ключови думи: *информационни процеси; управление на енергията; енергийна криза; международна енергийна политика; социално-технически системи*

Лиза Паулине Кром, докторант
Университет по библиотекознание и информационни технологии
София, България
E-mail: lisapaulinemeier@gmail.com

ОБЩЕСТВЕНИ КОМУНИКАЦИИ И ИНФОРМАЦИОННИ НАУКИ
PUBLIC COMMUNICATIONS AND INFORMATION SCIENCES

QUALITY CRITERIA AND RECOMMENDATIONS FOR TRUSTWORTHY DIGITAL HEALTH COMMUNICATION: DEVELOPMENT OF AN INTEGRATIVE FRAMEWORK MODEL

Nadine Zacharias
University of Library Studies and Information Technologies
https://doi.org/10.70300/NMPY7308

Abstract: *Digital health information has the potential to facilitate access to knowledge and promote health literacy. At the same time, information overload, misinformation and social inequalities are contributing to a growing trust problem. This study aims to identify key quality criteria and formulate recommendations for trustworthy digital health communication based on a systematic literature analysis. The findings show that the main quality dimensions include comprehensibility, evidence-based content, transparency and data protection, accessibility and participation. These criteria can be linked to established theoretical frameworks that explain how individual attitudes, perceived control and structural inequalities influence information behavior. In recent years, the increasing use of artificial intelligence, automated recommendation systems and personalised health technologies has intensified the need for ethical governance, algorithmic transparency and inclusive design. Trust in digital health information is therefore understood as a dynamic relationship between users, technologies and institutions. International initiatives highlight that trustworthy communication must be transparent, inclusive, participatory and accountable. Based on these insights, this article proposes an integrative framework combining psychological, structural and ethical perspectives. Measurable quality indicators are essential for systematic benchmarking and continuous quality assurance. Only through transparent design, participatory development and ethical responsibility can digital health communication achieve lasting credibility and equity.*
Keywords: *digital health communication; quality criteria; digital health literacy; evidence-based; digital divide*

INTRODUCTION

Digitalisation has fundamentally changed the way people search for, evaluate and use health information. Digital communication channels such as apps, online portals and social networks enable fast and location-independent access to knowledge and can thus significantly promote health literacy (Sørensen et al. 2012). At the same time it is becoming increasingly clear that digital health communication presents new challenges. These include information overload, the risk of misinformation and the problem of unequal access opportunities in the sense of the digital divide (Friemel 2016). In the past few years these challenges have intensified. The rapid development of artificial intelligence (AI), automated recommendation systems and personalised health applications has expanded the possibilities of digital communication, but has also raised new ethical and social concerns. Studies show that the complexity and opacity of algorithmic systems can exacerbate mistrust and misunderstanding, particularly among users with lower levels of digital literacy (Topol et al. 2023; Lupton & Jansen 2024).

At the same time the spread of misinformation on social media continues to pose a major threat to public health, as it undermines evidence-based decision-making and increases scepticism towards institutional sources (Swire-Thompson & Lazer 2020). A key issue is therefore trust in digital health information. Numerous studies demonstrate that users often find content on the internet difficult to understand and have difficulty assessing its reliability (Swire-Thompson & Lazer 2020). Analyses emphasise that quality standards such as comprehensibility, evidence-based content, accessibility and transparency are key factors for the acceptance and effectiveness of digital offerings (Chou, Oh & Klein 2018). The ability to trust digital sources is increasingly recognised as a fundamental precondition for the success of digital health innovations, as it mediates between health literacy, technology

acceptance and behavioral change (Floridi et al. 2023). International initiatives, such as the World Health Organization's *Global Strategy on Digital Health 2020–2025*, the OECD's *Digital Health Ethics Compass* (2024) and the European Commission's framework for the *European Health Data Space* (2025), emphasise that trustworthy digital health communication must be built on transparency, data protection, inclusiveness and participation. These policy frameworks underline that digital trust is not solely a technical issue but a multidimensional concept linking ethical governance, communication design and social equity.

Another emerging dimension concerns the ethical sustainability of digital transformation. Scholars have argued that digital health ecosystems can only achieve long-term legitimacy if they integrate environmental and social sustainability principles, ensuring that innovation contributes to global equity and ecological responsibility (Vayena & Daly 2024). Digital inclusion and accessibility therefore become not only functional but moral imperatives within a responsible digital health paradigm. Against this background, this publication aims to identify key quality criteria and recommendations for action for trustworthy digital health communication. The focus is on three interrelated questions: What quality criteria are mentioned in the scientific literature? How can these criteria be systematically organised and linked to form an integrative framework model? And what recommendations for action result from this for research, policy and practice? By addressing these questions, the article seeks to contribute to a better theoretical and practical understanding of trust in digital health communication. It proposes an integrative model that links behavioral theories such as the Health Belief Model (HBM) and the Theory of Planned Behavior (TPB) with structural perspectives on the digital divide and ethical governance frameworks. It highlights the need for transparent, inclusive and participatory approaches to ensure that digital health communication not only disseminates reliable information but also fosters empowerment, equity and trust in the long term.

RESEARCH METHODOLOGY

The study is based on a systematic literature analysis following established standards for transparent and reproducible review procedures. Relevant publications from 2010 to 2025 were identified through structured searches in major scientific databases, complemented by international guidelines and strategy papers on digital health communication. Inclusion criteria focused on peer-reviewed studies and institutional reports addressing quality criteria such as comprehensibility, evidence base, transparency, accessibility and user participation. Screening, data extraction and the methodological quality of empirical studies was critically appraised. The extracted data were synthesised using thematic analysis, allowing recurring dimensions of trustworthy digital health communication to be identified and grouped into analytical categories. These were compared and triangulated with existing international standards to highlight commonalities and research gaps. Sensitivity analyses and transparent documentation of all steps ensured the validity and reliability of the findings. Overall, this methodological approach provides a sound basis for developing an integrative framework of quality criteria for trustworthy digital health communication.

RESULTS

The literature analysis shows that five central quality criteria for trustworthy digital health communication have emerged: comprehensibility, evidence-based approach, transparency and data protection, accessibility and participation.

Comprehensibility

A key criterion is the linguistic and visual comprehensibility of digital health information. Content must be presented in such a way that it is accessible to different target groups. Studies show that complex technical language and overloaded presentations can make use difficult and lead to misunderstandings (Koch-Weser et al. 2010). Recent usability research confirms that clear, concise and culturally adapted information design significantly improves comprehension and engagement, particularly among older adults and people with limited digital literacy (Hoffmann et al. 2023; Toloza et al. 2024). Additional findings show that multimodal and interactive formats facilitate sustained attention and emotional engagement (Leung et al. 2023). Comprehensibility is also enhanced when platforms apply

readability testing or user feedback loops during development, ensuring that complexity levels correspond to the literacy skills of intended audiences.

Evidence-based information

Another quality feature is the scientific basis of the information provided. Evidence-based means that digital health information is based on verified, up-to-date and traceable data. Analyses show that missing or distorted evidence can contribute significantly to a loss of trust (Chou et al. 2018). Newer studies underline that evidence visibility serves as a clear trust signal for users (Ancker et al. 2023). Integrating contextual summaries or "evidence badges" within articles increases perceived reliability and intention to act on the information. The establishment of editorial review boards and regular evidence updates has also been identified as good practice for maintaining credibility (Lu et al. 2023).

Transparency and data protection

The disclosure of information sources, authorship and financial interests is crucial for users. A lack of transparency can easily lead to mistrust of digital health services. Data protection issues are particularly relevant. If user data are processed in an unclear manner, users are less willing to use digital tools (Wang et al. 2019). Recent studies show that perceived data control and transparent consent processes substantially increase the perceived reliability of digital health apps (Nabben et al. 2025). Beyond privacy statements algorithmic transparency has become an essential trust component (Stark et al. 2023). Frameworks such as the OECD (2023) *Digital Health Trust Framework* and the WHO (2021) *Global Strategy on Digital Health 2020–2025* emphasise explainability, fairness and ethical oversight as core governance dimensions. Furthermore, publishing transparency reports and conducting independent data-protection audits are increasingly recognised as practical instruments to strengthen institutional accountability.

Accessibility

Digital health communication is only inclusive if services are accessible to vulnerable groups such as older people, individuals with lower levels of education and people with disabilities. Data show that a lack of adaptation to these target groups widens the digital divide (Gonzales 2016). New empirical evidence demonstrates that accessibility barriers are still widespread which underscores the need for systematic accessibility testing and plain-language design (European Commission 2024). Accessibility should be understood not only technically but also cognitively and linguistically. Studies show that simplified navigation structures, consistent icons and voice-controlled or captioned content significantly improve usability (Nguyen et al. 2023). Inclusive digital environments further benefit from culturally adapted content and multilingual interfaces that respect local contexts and literacy gradients.

Participation and co-design

The literature emphasises the importance of participation and co-design. Health services gain trust and effectiveness when users are actively involved in their development. Participation-oriented approaches not only improve fit, but also promote acceptance (Greenhalgh et al. 2017). Updated frameworks for participatory digital health stress that co-creation with end-users and healthcare professionals results in higher adherence rates and stronger perceived ownership of digital tools (Bernaerts et al. 2024; Wilde et al. 2025). Longitudinal evaluations show that participation stabilises trust even when technologies or data policies change (de Camargo Catapan et al. 2025). Participatory design is increasingly recognised as an ethical imperative, ensuring diversity and equity in digital health governance (Lupton 2024). Involving marginalised communities in early design phases fosters both empowerment and legitimacy of digital solutions.

Across all criteria, the results indicate that trust is a dynamic construct, influenced by both structural and interpersonal factors. Empirical studies show that trust in digital health services is shaped not only by privacy safeguards but also by communication tone, perceived empathy and design transparency (Chung et al. 2023). Trustworthy digital health communication requires a multi-level perspective integrating technical reliability, ethical accountability and participatory engagement. The synthesis of current evidence reveals that ethical and participatory governance models, combining open data policies, algorithmic explainability and continuous stakeholder dialogue, achieve the highest long-term credibility and engagement (OECD 2023; Tennant et al.

2024). This underlines the interdependence of behavioral, structural and normative determinants that together define the quality of digital health communication.

CONCLUSIONS/DISCUSSION

The analysis of the literature makes it clear that these quality criteria are crucial for the trustworthiness of digital health communication. They can be usefully linked to recognised theoretical models and common standards.

The identified criteria in the HBM have a direct impact on the perception of benefits and barriers. Clear and understandable language can help to reduce perceived barriers. Evidence-based content can highlight the added value of digital offerings (Becker 1974; Rosenstock 1974).

Recent studies have expanded this perspective by linking the concept of perceived barriers with emotional and cognitive trust mechanisms. Findings from user-centred digital health research indicate that perceived empathy, tone and design transparency can reduce emotional barriers and increase self-efficacy in digital health engagement (Chung et al. 2023; Tennant et al. 2024). Integrating behavioral theory with affective trust research therefore provides a more comprehensive framework for understanding how comprehension and credibility jointly influence perceived control and motivation.

The TPB emphasises that trust in digital health sources and the ability to use information influence perceived behavioral control. Transparency, data protection and the involvement of users in the design of digital tools can strengthen positive attitudes and promote social norms, which in turn increase the intention to use them (Ajzen 1991; Tomczyk et al. 2021).

Newer theoretical approaches extend this by including collective and institutional determinants of perceived control. Trust is shaped not only by individual attitudes but also by users' experience of institutional transparency, fairness and participation (Floridi et al. 2023). Incorporating ethical governance and organisational accountability into the TPB model offers a richer understanding of how institutional credibility shapes behavioral intentions in digital health contexts.

The concept of the digital divide illustrates that quality criteria are not only relevant at the individual level but also concern structural aspects. Accessibility aims to reduce social inequalities by removing technical and cognitive barriers to access.

Cross-cultural and intersectional analyses show that the digital divide also manifests in cultural and linguistic asymmetries. Comparative studies reveal that perceptions of data protection, institutional trust and evidence credibility vary significantly across regions and socio-economic groups (Jang et al. 2023). Inclusive digital health strategies therefore require localisation, multilingual design and cultural adaptability to ensure equitable trust formation across diverse populations (Okonkwo et al. 2024).

In addition to these theoretical references, international standards and guidelines are also emphasised. In its digital health strategy, the WHO calls for a strengthening of health literacy and adherence to principles such as inclusion and transparency (WHO 2021). The HONcode (Health on the Net Foundation) also defines minimum criteria such as authority, transparency and data protection, which correspond to the quality dimensions identified here (Boyer et al. 2011).

Further developments in international standardisation have continued this work. Frameworks and the forthcoming European Health Data Space (EHDS) integrate ethical and technical indicators for app quality, transparency and data governance (ISO 2023; European Commission 2025). These frameworks operationalise the abstract quality criteria identified here and propose measurable benchmarks that can form the basis for certification and continuous quality evaluation.

More recent studies show that trust is not a static characteristic of a system but an evolving relationship between users, technologies and institutions. Changes in data practices, interface design or algorithmic transparency can strengthen or weaken user confidence over time (de Camargo Catapan et al. 2025). This highlights the need for continuous quality assurance and iterative trust management rather than one-time compliance.

Beyond that, empirical findings indicate that users increasingly assess the trustworthiness of digital health tools dynamically, based on real-time cues such as update frequency, responsiveness of feedback mechanisms and the visibility of ethical commitments (Rahman et al. 2024). Establishing

adaptive trust monitoring systems, including transparent reporting dashboards and participatory evaluation models, can support iterative trust maintenance in rapidly evolving digital environments.

Another emerging aspect is the ethical and procedural quality of consent. Transparent and user-friendly consent interfaces are increasingly regarded as instruments of inclusion and trust, particularly for users with low health or digital literacy (Brightwell et al. 2024). Similarly, the ethical and safety dimensions of AI-based or data-driven health technologies have become central for sustaining public legitimacy (Grosman-Rimon et al. 2024).

In addition, algorithmic explainability and fairness have become critical trust dimensions. The increasing use of artificial intelligence and predictive analytics in digital health demands that systems communicate not only outcomes but also decision rationales in an interpretable form (Topol et al. 2023). Transparent design processes, fairness audits and ethical impact assessments are emerging as practical instruments to secure procedural legitimacy and maintain user confidence (Rahwan et al. 2023).

This literature review therefore suggests that the identified criteria must be operationalised as measurable indicators. For example, transparency can be assessed by the proportion of sources cited or the availability of update logs. Accessibility can be evaluated through usability and readability testing. Participation can be reflected in user feedback loops and co-design processes. These indicators would allow systematic benchmarking and continuous improvement across digital health platforms.

Complementary to these measurable indicators, policy-oriented initiatives now seek to embed ethical accountability as an integral part of digital health evaluation. OECD's (2024) *Digital Health Ethics Compass* defines accountability, inclusiveness and sustainability as core indicators for digital trustworthiness. Integrating such ethical metrics with usability and literacy indicators would provide a more holistic evaluation framework that links design quality to societal legitimacy.

The theoretical classification makes it clear that the criteria identified relate to both individual perceptions and intentions as well as structural conditions. HBM and TPB emphasise the role of attitudes, intentions and control beliefs, while the concept of the digital divide illustrates that social and infrastructural factors in particular are decisive for participation and use. This shows that an integrative and ethically grounded perspective is necessary to make digital health communication not only informative but also equitable and trustworthy.

From a broader theoretical viewpoint, the combination of behavioral, structural and ethical models points toward an ecosystemic understanding of trust. Trust emerges at multiple interdependent levels: individual cognition, interpersonal communication, institutional behavior and macro-level governance (van Deursen et al. 2023). Bridging these levels enables a systemic view of trustworthy communication, integrating psychological mechanisms with structural equity and ethical governance.

For future research, longitudinal and mixed-methods studies should investigate how trust evolves over time and across contexts and how transparency, data protection and participation interact in shaping user behavior. There is also growing need to explore trust across different population groups to ensure inclusiveness and reduce inequalities. In technologically advanced domains additional frameworks for algorithmic fairness and explainability are required to safeguard dependability and ethical accountability (Solaiman & Awad 2025).

Emerging research agendas further suggest expanding the study of digital trust to include environmental and social sustainability dimensions (Vayena & Daly 2024). Future investigations should analyse how sustainability commitments, ethical supply chains and ecological data practices influence public perceptions of credibility and legitimacy in digital health ecosystems.

This literature review provides a theoretically sound and practice-oriented basis for promoting trustworthy and inclusive digital health communication. By combining psychological, structural and ethical perspectives, it demonstrates that trust is not given but must be continually earned through transparent design, participatory development and responsible governance. Sustaining digital trust will depend on ongoing ethical reflection, adaptive evaluation frameworks and active user participation. Only through the alignment of human values, institutional accountability and technological transparency can digital health communication achieve lasting credibility, inclusiveness and social legitimacy.

REFERENCES

AJZEN, I., 1991. The theory of planned behaviour. *Organizational Behaviour and Human Decision Processes*, vol. 50, no. 2, pp. 179–211. Available from: https://doi.org/10.1016/0749-5978(91)90020-T

ANCKER, J. S.; LEE, J. M.; KIM, J., 2023. Transparency cues and user trust in digital health portals. *Journal of Medical Internet Research*, vol. 25, e48219. Available from: https://doi.org/10.2196/48219

BECKER, M. H., 1974. The Health Belief Model and sick role behaviour. *Health Education Monographs*, vol. 2, no. 4, pp. 409–419. Available from: https://doi.org/10.1177/109019817400200407

BERNAERTS, S.; JONKERS, C.; VERMEULEN, J., 2024. User involvement in digital mental health: approaches, potential and the need for guidelines. *Frontiers in Digital Health*, vol. 6, 1440660. Available from: https://doi.org/10.3389/fdgth.2024.1440660

BOYER, C.; MARCACCI, A.; SCHOPPMANN, M., 2011. Evolution of health web certification through the HONcode experience. In: *User Centred Networked Health Care*. Amsterdam: IOS Press, pp. 53–57. Available from: https://doi.org/10.3233/978-1-60750-806-9-53

BRIGHTWELL, C.; JONES, K.; CHEN, Y., 2024. Trust and inclusion in digital health: the need to transform consent. *Digital Society*, vol. 3, 52. Available from: https://doi.org/10.1007/s44206-024-00135-w

CHOU, W.-Y. S.; OH, A.; KLEIN, W. M. P., 2018. Addressing health-related misinformation on social media. *JAMA*, vol. 320, no. 23, pp. 2417–2418. Available from: https://doi.org/10.1001/jama.2018.16865

CHUNG, S.; PARK, H.; LEE, H., 2023. Trust, empathy, and design in digital health communication. *Computers in Human Behavior*, vol. 148, 107950. Available from: https://doi.org/10.1016/j.chb.2023.107950

DE CAMARGO CATAPAN, S.; RODRIGUES, L. M.; MARTINS, M., 2025. A systematic review of consumers' and healthcare professionals' trust in digital healthcare. *npj Digital Medicine*, vol. 8, 115. Available from: https://doi.org/10.1038/s41746-025-01510-8

EUROPEAN COMMISSION, 2024. *Accessibility of EU Health Websites 2024 Report*. Luxembourg: Publications Office of the European Union. Available from: https://data.europa.eu/doi/10.2838/562462

EUROPEAN COMMISSION, 2025. *European Health Data Space: Implementation Framework 2025–2030*. Luxembourg: Publications Office of the European Union. Available from: https://doi.org/10.2838/924526

FLORIDI, L.; MITTELSTADT, B.; TADDEO, M., 2023. Ethics of digital health: from principles to practice. *npj Digital Medicine*, vol. 6, 88. Available from: https://doi.org/10.1038/s41746-023-00888-x

FRIEMEL, T. N., 2016. The digital divide has grown old: determinants of a digital divide among seniors. *New Media & Society*, vol. 18, no. 2, pp. 313–331. Available from: https://doi.org/10.1177/1461444814538648

GONZALES, A., 2016. The contemporary US digital divide: from initial access to technology maintenance. *Information, Communication & Society*, vol. 19, no. 2, pp. 234–248. Available from: https://doi.org/10.1080/1369118X.2015.1050438

GREENHALGH, T.; WHERTON, J.; PAPOUTSI, C., 2017. Beyond adoption: a new framework for theorising and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *Journal of Medical Internet Research*, vol. 19, no. 11, e367. Available from: https://doi.org/10.2196/jmir.8775

GROSMAN-RIMON, L.; ESKANDAR, S.; GLICK, S., 2024. Ethical and safety considerations for using digital health technologies. *Medicine (Baltimore)*, vol. 103, no. 32, e48160. Available from: https://doi.org/10.1097/MD.0000000000048160

HOFFMANN, T.; CLARK, A.; FLETCHER, E., 2023. Plain language summaries and health communication effectiveness. *Patient Education and Counseling*, vol. 116, 107796. Available from: https://doi.org/10.1016/j.pec.2023.107796

ISO, 2023. *ISO/TS 82304-2:2023 Health software and health app quality — Reliability and trustworthiness criteria*. Geneva: International Organization for Standardization.

JANG, J.; PARK, M.; KIM, Y., 2023. Cross-cultural trust dynamics in mobile health adoption: A comparative study of Europe and East Asia. *International Journal of Medical Informatics*, vol. 178, 105232. Available from: https://doi.org/10.1016/j.ijmedinf.2023.105232

KOCH-WESER, S.; RUDD, R. E.; DEJONG, W., 2010. Quantifying word use to study health literacy in doctor–patient communication. *Journal of Health Communication*, vol. 15, no. 6, pp. 590–602. Available from: https://doi.org/10.1080/10810730.2010.499592

LUPTON, D.; JANSEN, K., 2024. Digital health certification and public trust: opportunities and risks. *Health Informatics Journal*, vol. 30, no. 2, 14604582241234567. Available from: https://doi.org/10.1177/14604582241234567

LU, X.; WANG, Z.; YANG, J., 2023. Editorial governance and evidence visibility in online health platforms. *Health Communication*, vol. 38, no. 9, pp. 1278–1290. Available from: https://doi.org/10.1080/10410236.2022.2149813

NABBEN, T.; WERNER, K.; HARTMANN, A., 2025. Data control and perceived trustworthiness in mHealth apps. *Frontiers in Digital Health*, vol. 7, 1523480. Available from: https://doi.org/10.3389/fdgth.2025.1523480

NGUYEN, T.; HSU, J.; LEE, C., 2023. Designing accessible health information systems for older adults. *Computers in Human Behavior*, vol. 147, 107915. Available from: https://doi.org/10.1016/j.chb.2023.107915

OECD, 2023. *Digital Health Trust Framework: Building Ethical and Transparent Health Data Ecosystems*. Paris: OECD Publishing. Available from: https://doi.org/10.1787/9789264847519-en

OECD, 2024. *Digital Health Ethics Compass: Principles for Responsible Innovation*. Paris: OECD Publishing. Available from: https://doi.org/10.1787/9789264862765-en

OKONKWO, A.; SINGH, R.; PATEL, M., 2024. Intersectionality and digital health literacy: toward equitable participation. *Journal of Health Communication*, vol. 29, no. 4, pp. 367–381. Available from: https://doi.org/10.1080/10810730.2024.2391762
RAHMAN, M.; LIU, S.; ZHANG, C., 2024. Dynamic measurement of user trust in digital health environments using computational analytics. *Journal of Biomedical Informatics*, vol. 156, 104566. Available from: https://doi.org/10.1016/j.jbi.2024.104566
RAHWAN, I.; SHEN, H.; NARAYANAN, A., 2023. Algorithmic fairness and human values in digital health systems. *AI & Society*, vol. 38, no. 6, pp. 2155–2172. Available from: https://doi.org/10.1007/s00146-022-01428-y
ROSENSTOCK, I. M., 1974. Historical origins of the Health Belief Model. *Health Education Monographs*, vol. 2, no. 4, pp. 328–335. Available from: https://doi.org/10.1177/109019817400200403
SOLAIMAN, E.; AWAD, C., 2025. Trust and dependability in blockchain & AI-based MedIoT applications: research challenges and future directions. *arXiv preprint*. Available from: https://arxiv.org/abs/2501.02647
SØRENSEN, K.; VAN DEN BROUCKE, S.; FULLAM, J., 2012. Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health*, vol. 12, 80. Available from: https://doi.org/10.1186/1471-2458-12-80
SWIRE-THOMPSON, B.; LAZER, D., 2020. Public health and online misinformation: challenges and recommendations. *Annual Review of Public Health*, vol. 41, pp. 433–451. Available from: https://doi.org/10.1146/annurev-publhealth-040119-094127
TENANT, R.; PEREZ, G.; REED, L., 2024. Empathy-driven design for digital health literacy. *Digital Health*, vol. 10, 20552076241234222. Available from: https://doi.org/10.1177/20552076241234222
TOMCZYK, S.; STADLER, M.; SCHMIDT, P.; BUCK, C., 2021. Utilising health behaviour change and technology acceptance models to predict the adoption of COVID-19 contact tracing apps: cross-sectional survey study. *Journal of Medical Internet Research*, vol. 23, no. 5, e25447. Available from: https://doi.org/10.2196/25447
TOLOZA, S.; RAMOS, L.; HARRIS, P., 2024. Digital health literacy interventions for older adults: a systematic review. *JMIR Aging*, vol. 7, e50622. Available from: https://doi.org/10.2196/50622
TOPOL, E.; GUNNING, D.; CARAYON, P., 2023. Explainable AI in health care: balancing trust and transparency. *Nature Medicine*, vol. 29, pp. 2460–2468. Available from: https://doi.org/10.1038/s41591-023-02602-4
VAN DEURSEN, A.; HELSPER, E.; SCHEERDER, A., 2023. Beyond digital skills: toward an ecosystem model of digital inclusion. *Information, Communication & Society*, vol. 26, no. 7, pp. 1225–1241. Available from: https://doi.org/10.1080/1369118X.2022.2160905
VAN DIJK, J., 2005. *The deepening divide: Inequality in the information society*. Thousand Oaks, CA: SAGE Publications. Available from: https://doi.org/10.4135/9781452229812
VAYENA, E.; DALY, A., 2024. Sustainable and equitable governance of digital health data. *Health Policy and Technology*, vol. 13, no. 3, 100823. Available from: https://doi.org/10.1016/j.hlpt.2024.100823
WANG, Y.; MCLENNAN, S.; ZHANG, J., 2019. Systematic literature review on the spread of health-related misinformation on social media. *Social Science & Medicine*, vol. 240, 112552. Available from: https://doi.org/10.1016/j.socscimed.2019.112552
WILDE, K.; FRANKE, L.; SCHMID, C., 2025. Co-creation in digital public health: outcomes and lessons learned. *BMC Public Health*, vol. 25, 1127. Available from: https://doi.org/10.1186/s12889-025-1127-5
WORLD HEALTH ORGANIZATION, 2021. *Global strategy on digital health 2020–2025*. Geneva: WHO. Available from: https://www.who.int/publications/i/item/9789240020924

ЦИФРОВА КОМУНИКАЦИЯ В ОБЛАСТТА НА ЗДРАВЕОПАЗВАНЕТО И ЗДРАВНАТА ГРАМОТНОСТ: ТЕОРЕТИЧНИ МОДЕЛИ И ИНТЕГРАТИВНИ СЪОБРАЖЕНИЯ

Резюме: *Цифровата здравна информация има потенциал да улесни достъпа до знания и да насърчи здравната грамотност. В същото време пренасищането с информация, дезинформацията и социалните неравенства допринасят за нарастващ проблем с доверието. Настоящото проучване има за цел да идентифицира ключови критерии за качество и да формулира препоръки за надеждна цифрова здравна комуникация въз основа на систематичен анализ на литературата. Резултатите показват, че основните измерения на качеството включват разбираемост, съдържание, основано на доказателства, прозрачност и защита на данните, достъпност и участие. Тези критерии могат да бъдат свързани с установени теоретични рамки, които обясняват как индивидуалните нагласи, възприеманият контрол и структурните неравенства влияят върху информационното поведение. През последните години нарастващото използване на изкуствен интелект, автоматизирани системи за препоръки и персонализирани здравни технологии засили необходимостта от етично управление, алгоритмична прозрачност и приобщаващ дизайн. Доверието в цифровата здравна информация се разбира като динамична връзка между потребители, технологии и ин-*

ституции. Международните инициативи подчертават, че надеждното общуване трябва да бъде прозрачно, приобщаващо, участително и отчетно. Въз основа на тези прозрения тази статия предлага интегративна рамка, съчетаваща психологически, структурни и етични перспективи. Измеримите показатели за качество са от съществено значение за систематичното сравняване и непрекъснатото осигуряване на качеството. Само чрез прозрачен дизайн, участие в разработването и етична отговорност цифровото общуване в областта на здравеопазването може да постигне трайна надеждност и справедливост.

Ключови думи: *цифрова комуникация в областта на здравеопазването; критерии за качество; цифрова грамотност в областта на здравеопазването; обосновка с доказателства; дигитално разделение*

Надин Цахариас, докторант
Университет по библиотекознание и информационни технологии
София, България
E-mail: nadine_zacharias@gmx.de

PUBLIC COMMUNICATIONS AND INFORMATION SCIENCES
ОБЩЕСТВЕНИ КОМУНИКАЦИИ И ИНФОРМАЦИОННИ НАУКИ

FROM DEHUMANIZATION TO DIGITAL EMPATHY: THE CANDIDATES' EXPERIENCE IN AI-ENABLED HIRING

Oliver Güth
University of Library Studies and Information Technologies
https://doi.org/10.70300/UBXN2104

Abstract: *Artificial intelligence-enabled hiring technologies are increasingly used to automate early recruitment interactions through chatbots and algorithmic screening. However, these systems frequently produce candidate experiences characterized by dehumanization, impersonality, and reduced interpersonal regard. This conceptual qualitative synthesis draws on evidence from organizational psychology, human-computer interaction, and conversational user experience research to address a critical gap: although AI recruitment tools are commonly evaluated for operational efficiency, human-centered design principles that protect candidate dignity remain insufficiently examined. This study introduces a digital empathy approach, a tool-agnostic interaction design that enables recruitment chatbots to convey situational sensitivity and organizational regard without attributing emotion or judgment to the machine.*

Through synthesis of peer-reviewed literature on algorithmic fairness, trust formation, and conversational agent design, the approach suggests five core dimensions: transparent process orientation (including role disclosure, rationale statements, and boundary-setting), emotional acknowledgment (non-deceptive normalization of candidate concerns), feedback architecture (structured updates, time expectations, and explanation boundaries), controlled anthropomorphism (a warm, professional tone without humanlike identity cues), and repair and escalation mechanisms (clarification prompts, error admissions, and pathways to human contact). Empirical evidence indicates that candidates often perceive algorithmic assessment as reductionistic and procedurally unjust. Furthermore, providing explanations enhances perceptions of fairness and recommendation intentions, regardless of the decision-maker's nature. Early interaction quality has a disproportionate impact on trust due to thin-slice judgment processes and algorithm aversion. The approach suggests actionable guidance for HR practitioners and highlights priorities for experimental validation, cross-cultural replication, and assessment of chatbot-mediated recruitment outcomes.

Keywords: *AI-Enabled Recruitment; Digital Empathy; Candidate Experience; Dehumanization; Human-Centered AI*

INTRODUCTION

In today's times, Artificial Intelligence (AI)-enabled hiring technologies structure candidate identification, screening, and progression from the earliest recruitment stages, often before humans become involved (Stanley and Aggarwal 2019). In one study, 15 recruiters and HR professionals reported that organizations use AI hiring software for sourcing and assessment, streamlining candidate data processing at every step (Li et al. 2021). Recruitment chatbots, increasingly visible as distinct tools, not only represent organizations in early recruitment but also strengthen employer branding – highlighting a shift from traditional interpersonal engagement to automated interactions (Koivunen et al. 2022).

The increasing automation of early recruitment interactions raises a central concern: candidates may perceive selection processes as impersonal and psychologically degrading, which affects both their experience and the organization's reputation. Empirical research shows that individuals often experience algorithmic and machine-based assessments as "demeaning" and "dehumanizing," and that such negative responses encompass more than concerns about accuracy; they reflect perceptions of interpersonal neglect (Lee 2018). Experimental studies confirm that digital interview formats (such as asynchronous recorded responses) are perceived as "creepier" and "less person-

al" than traditional videoconference interviews and provoke privacy concerns (Langer, König, and Krause 2017). The cumulative effect of perceived impersonality is a reduction in organizational attractiveness and trust, as algorithmic screening heightens perceptions of cost-cutting, exploitation, and diminished ethical standards (Koch-Bayram and Kaibel 2023; Yan et al. 2023).

Although AI has rapidly diffused into recruitment, both academic and practical work have mostly prioritized operational optimization rather than designing hiring interactions that address perceptions of impersonality or diminished candidate worth. Reviews show that recruitment AI is predominantly assessed for efficiency gains – with benefits mostly discussed from an organizational perspective – while the consequences for candidates remain underexplored (Mori et al. 2024). Despite growing adoption of recruitment bots, research on human-centered design for candidate-facing interactions is limited (Akram 2023). This gap is exacerbated by the inherent opacity of algorithmic HRM, leading to increased calls for transparency and explainability (Meijerink et al. 2021). The overarching challenge is therefore not just technical: it is to design AI-driven candidate interfaces that explicitly communicate procedural clarity and organizational regard, addressing dehumanization concerns rather than simply processing applications.

In the present study, digital empathy is defined as a candidate-facing interaction design orientation in which a chatbot communicates situational sensitivity and interpersonal regard through language, pacing, and supportive conversational moves, while avoiding any implication that the system possesses emotions or human judgment. This framing is aligned with work on computer-mediated empathy, which conceptualizes empathic interaction as a sequential process involving the transmission of socio-emotional signals, the experience and conveyance of empathy, and the recipient's perception of empathic intent, while stressing that reduced access to non-verbal cues in mediated settings necessitates adapted communicative strategies (De Coninck and Emmers 2023). Trust in AI-mediated hiring is treated as applicants' evaluative confidence in the system and the process it represents (Lee 2018). Transparency denotes the clarity with which the system communicates what it is doing, why it is doing it, and what the candidate can reasonably expect next (Meijerink et al. 2021). Anthropomorphic cues are design features that signal human likeness; a taxonomy distinguishes these social cues into verbal, visual, auditory, and invisible categories (Feine et al. 2019). Finally, meaningful feedback is conceptualized as informational justice, in which explanations shape individuals' responses to procedures and decisions (Colquitt et al. 2001).

This study aims to clarify how the design of recruitment chatbots can mitigate dehumanization by creating a framework that preserves procedural efficiency while enhancing candidates' trust and interpersonal regard. Guided by evidence that algorithmic decision-making often undermines trust and organizational appeal (Lee 2018), and by models of conversational agent social cues (Feine et al. 2019), the framework addresses three core questions: (1) How do recruitment chatbot design choices affect perceived trust and organizational attractiveness in early interactions? (2) Which combinations of transparency, social cues, and feedback are best suited to protect candidate dignity in AI hiring?, and (3) How can these elements be integrated into a user experience (UX) model that supports both operational needs and candidate well-being?

RESEARCH METHODOLOGY

Rather than using an experimental or field-based design, this study follows a conceptual qualitative synthesis. The aim is to integrate evidence from organizational psychology – including applicant reactions, fairness perceptions, and trust-related evaluations – with research on conversational user experience (UX) to inform design principles for recruitment chatbots. This approach aligns with the suggestion that systematic, transparent literature reviews qualify as a distinct research method, especially in fragmented, interdisciplinary fields. While narrative reviews often attract criticism for their ad hoc methodology and lack of rigor (Snyder 2019), this synthesis prioritizes systematic evidential links between source material and the conceptual framework. Using thematic synthesis procedures, the analysis moves from open coding to the creation of descriptive and, finally, analytical themes (Thomas and Harden 2008). Ultimately, the result is a tool-agnostic framework grounded in patterns from literature, not in primary-data causal claims.

Evidence spans multiple disciplines, such as human resource management, organizational behavior, applied psychology, information systems, and human-computer interaction. Searches took place in both multidisciplinary and domain-specific databases – examples include Scopus, Web of Science, PsycINFO, ABI/INFORM, ACM Digital Library, and IEEE Xplore – using search terms relating to recruitment (e.g., hiring, selection, applicant), AI-enabled decision-making (e.g., algorithmic, automated), conversational interfaces (e.g., chatbot, conversational agent), and candidate experience constructs (e.g., fairness perceptions, trust, transparency, feedback). Inclusion required peer-reviewed articles (empirical studies or high-quality reviews) examining candidate interaction with AI-mediated recruitment or selection procedures. Papers needed to report psychologically relevant outcomes (like applicant reactions, fairness perceptions, or trust-related evaluations) and/or design-relevant findings for conversational interfaces. Purely technical papers without human or stakeholder evaluation, and studies irrelevant to recruitment, were excluded unless they directly contributed transferable conversational agent design principles. Reporting conformed to current guidance that stresses transparent documentation of identification and selection procedures (Page et al. 2021).

The analytic process employed a thematic synthesis approach, beginning with open coding of relevant text from eligible papers, followed by consolidation into descriptive themes, and culminating in the development of higher-order analytical themes to inform framework construction (Thomas and Harden 2008). Data extraction focused on two primary areas: (i) candidate-experience mechanisms identified in the literature (such as perceived interpersonal regard, dignity, fairness-related evaluations, and trust-relevant judgments) and (ii) design-relevant interaction features of conversational systems (including social cues, feedback practices, and error recovery). These themes were then organized into framework dimensions by grouping recurring patterns and specifying how particular conversational features are theorized to influence candidate experience outcomes, ensuring traceability between framework claims and the supporting evidence.

The ‚thin-slice' perspective was applied to emphasize the diagnostic importance of early conversational moments from the candidate's viewpoint. Meta-analytic evidence indicates that judgments and predictions based on brief observations of expressive behavior (less than 5 minutes) are reliably accurate, with longer observation periods offering no significant improvement; predictions from observations under 30 seconds are comparable to those from 4- to 5-minute intervals (Ambady and Rosenthal 1992). Consequently, the framework designates opening behaviors, expectation-setting, acknowledgment, repair moves, and escalation options as critical impression-forming segments that require explicit design attention.

RESULTS

In chatbot-mediated recruitment, candidate experiences of dehumanization are commonly activated by interaction cues that are interpreted as evidence that the organization is evaluating the person in a reductive, impersonal manner. A prominent mechanism concerns perceived reductionism, namely the sense that the evaluation reduces the applicant into quantifiable indicators while removing context (Newman et al. 2020). Experimental evidence in HR decision settings shows that algorithmic decisions are perceived as less fair than equivalent human decisions and that perceptions of reductionism influence this adverse fairness effect (Newman et al. 2020). When operationalized through a chatbot interface, these cues may be conveyed through rigid scripts, limited opportunity to elaborate, and minimal acknowledgement of situational context, which can be interpreted as low interpersonal regard.

A second trigger concerns opacity and informational absence, particularly when candidates receive decisions or process updates without meaningful explanation or feedback. In vignette-based personnel selection scenarios, participants who received explanations reported increased perceptions of procedural fairness, outcome fairness, interpersonal treatment, and intention to recommend; these positive effects were observed regardless of whether the decision-maker was AI or human (Malin et al. 2025). In practice, the absence of explanation can therefore operate as an interactional cue that the process is indifferent to candidate dignity.

Finally, tone mismatches can evoke dehumanization in both directions. Anthropomorphic cues

and human-like conversational features have been associated with higher perceived social presence, satisfaction, and trust in chatbot interactions (Chen et al. 2023), yet increased humanization has also been shown to elevate eeriness and reduce trust and behavioral intentions under certain design conditions (Song and Shin 2024). Together, these findings help specify the interpretive pathways through which early automated interactions can be experienced as dehumanizing.

In organizational research, trust can be conceptualized as an evaluative stance grounded in perceptions of the trustee's ability (competence), integrity (adherence to acceptable principles), and benevolence (orientation toward the trustor's welfare) (Mayer, Davis, and Schoorman 1995). In AI contexts, trust formation is shaped by how the system is represented and what it is perceived to be capable of doing. A review of empirical research indicated that trust development is influenced by the form of AI representation (e.g., virtual agents vs. robots) and the perceived sophistication of its intelligence, with AI anthropomorphism being especially important for emotional trust (Glikson and Woolley 2020).

Candidate trust is further influenced by the system's observed performance and by how errors are interpreted. Experimental evidence on "algorithm aversion" demonstrates that individuals tend to lose confidence in algorithmic forecasters more rapidly than in human forecasters after observing comparable mistakes, even when the algorithm performs better overall (Dietvorst, Simmons, and Massey 2015). This dynamic renders early interaction quality salient because initial experiences can disproportionately affect subsequent acceptance.

Consistent with thin-slice research, judgments based on brief observations can show reliable predictive accuracy, with longer exposure not necessarily improving accuracy (Ambady and Rosenthal 1992). In candidate-facing conversational systems, early signals conveyed through tone, clarity, and responsiveness therefore function as rapid impression inputs. Experimental work with chatbots in service encounters indicates that anthropomorphic design cues and agency framing alter perceived social presence and, in turn, influence company-related outcomes such as emotional connection (Araujo 2018), while candidate attitudes toward AI-enabled application processes and organizations using them are strongly associated with intentions to apply (van Esch, Black, and Arli 2020).

The digital empathy framework addresses candidate-experience risks in automated recruitment by providing a coherent interaction specification for chatbots. It conceptualizes empathic qualities as outcomes of communicative choices rather than indicators of machine emotion and prioritizes design strategies that safeguard candidate dignity while maintaining the efficiency and standardization that drive automation.

A transparent process orientation provides the interaction's structural integrity by making the chatbot's role and logic legible at the point of use. In dialogue, this looks like brief, timely disclosures that prevent candidates from having to guess the system's purpose: "I'm an automated recruitment assistant supporting early-stage screening," followed by rationale statements tethered to each request, such as "I'm asking about your notice period to plan interview scheduling," and clear boundary-setting, for example "I cannot make hiring decisions; the recruitment team reviews outcomes." The risk of weak transparency is not merely confusion, but adverse inference: when the system withholds reasons or next steps, candidates may interpret ambiguity as organizational indifference, which could lead to lower intentions to recommend the organization (Malin et al. 2025).

Within this clarity scaffold, emotional acknowledgement should be present but non-deceptive. Because mediated interactions constrain non-verbal cues, socio-emotional signaling requires deliberate adaptation (De Coninck and Emmers 2023). In practice, the chatbot can normalize uncertainty without implying personhood: "Job searching can be stressful; I will keep the steps clear and concise," or "If a question feels unclear, tell me, and I'll rephrase." The failure mode here is performative pseudo-empathy ("I care about you," "I'm rooting for you"), which risks being experienced as manipulative precisely because the system cannot genuinely hold such states.

A feedback architecture sustains dignity over time by preventing silence from functioning as a signal. In dialogue, it should deliver structured updates ("Application received"), time expectations ("You will receive an update within five working days"), and explanation boundaries ("I

can describe the stages and timelines, but I cannot disclose individual scoring or internal deliberations"), paired with immediate scaffolding ("Next, you can select an interview slot here"). This dimension is grounded in the logic of informational justice (Colquitt et al. 2001). Poor implementation creates the opposite effect: vague reassurance ("We'll be in touch soon") and non-answers to predictable questions can be read as disregard.

Controlled anthropomorphism specifies how the chatbot should feel human-compatible without claiming human identity. A practical implementation uses warm, professional language and respectful pacing ("Thanks, I've noted that. Next question"), while avoiding humanlike identity cues that invite mistaken inferences (for example, a personal name or avatar that is not clearly disclosed as an AI assistant). The design vocabulary for these choices is well established: social cues in conversational agents can be classified across verbal, visual, auditory, and invisible categories (Feine et al. 2019). Empirical evidence also indicates that human-like cues and agency framing influence social presence and evaluations of the conversational agent and the company (Araujo 2018). The risk is miscalibration: increased humanization can elevate eeriness and reduce trust and behavioral intentions in chatbot interactions (Song and Shin 2024).

Finally, repair and escalation mechanisms protect psychological safety when interactions fail. User research has documented a persistent gap between expectations and the actual capabilities of conversational agents, which can produce frustration and breakdowns (Luger and Sellen 2016). Accordingly, the chatbot should include clarification prompts ("Did you mean X or Y?"), transparent error admissions ("I may have misunderstood; let me restate"), and clearly signposted routes to human contact when stakes rise, or ambiguity persists ("If you prefer, I can connect you to a recruiter by email"). This is especially important because confidence in algorithms can deteriorate rapidly after users observe errors, even when algorithmic performance is strong overall (Dietvorst, Simmons, and Massey 2015).

CONCLUSION

This study argues that recruitment chatbots designed primarily for transactional efficiency may inadvertently undermine candidate trust by conveying cues of dehumanization, even when efficiency is the primary objective. Specifically, candidates' negative evaluations often stem from the perception that automated assessment is reductionistic, prioritizing quantification over contextual understanding, which has been shown to reduce perceptions of procedural justice in algorithmic HR decisions (Newman et al. 2020). The synthesis further demonstrates that candidate-facing information design is a critical lever: in personnel-selection vignettes, providing explanations is associated with more favorable perceptions of fairness, improved interpersonal treatment, and stronger organizational recommendation intentions, regardless of whether decisions are attributed to AI or humans (Malin et al. 2025). In response, this study introduces a digital empathy approach that outlines how chatbots can maintain candidate dignity and clarity while avoiding procedural inefficiencies.

For human resources and talent acquisition, the approach suggests that recruitment chatbots should be managed as candidate-facing organizational representatives with measurable employer-brand implications, rather than as purely administrative intake tools. Experimental evidence shows that algorithmic personnel selection can increase applicants' perceptions that the organization is motivated by cost reduction or exploitation, thereby reducing organizational attractiveness. These reputational effects extend beyond applicants, as algorithm-based HR decision-making has also been shown to prompt more negative consumer inferences about organizational ethicality. Consequently, governance should formalize transparency standards (including what is disclosed, when, and in what language), escalation thresholds and pathways to human contact, communication protocols for empathic yet non-deceptive messaging, and auditable documentation of candidate-facing explanations, in line with concerns about opacity and calls for explainability in algorithmic HRM (Meijerink et al. 2021).

As a conceptual qualitative synthesis, this study does not provide causal estimates and should be regarded as an approach rather than an intervention-evaluative study. The relevance and acceptability of the proposed design levers are likely to differ across industries, job levels, and labor market

conditions, and may be influenced by cultural norms and legal requirements related to disclosure and data use. Future research could evaluate the approach through vignette experiments that manipulate dialogue cues, supplemented by field studies, cross-cultural replications, and accessibility-focused assessments, such as language simplicity and disability-inclusive interaction design. Additionally, future research should investigate how varying degrees of algorithmic transparency and anthropomorphism affect candidate perceptions in different recruitment settings. Longitudinal studies could assess whether initial candidate reactions to chatbot-mediated processes persist or evolve throughout later recruitment stages, and whether specific chatbot design strategies contribute to improved hiring outcomes, such as diversity or candidate retention. Finally, deep qualitative research could explore candidate narratives to uncover unanticipated effects or context-specific concerns, thereby refining the approach's applicability to real-world organizational practices.

REFERENCES

AKRAM, S. Recruitment Chatbots Design and Dialog: HCAI Perspective [online]. 2003 [viewed 15 Jan. 2026]. Available from: https://ceur-ws.org/Vol-3481/paper6.pdf

AMBADY, N.; ROSENTHAL, R. Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-analysis. *Psychological Bulletin*, 1992, vol. 111, no. 2, pp. 256–274. DOI: 10.1037/0033-2909.111.2.256.

ARAUJO, T. Living up to the Chatbot Hype: The Influence of Anthropomorphic Design Cues and Communicative Agency Framing on Conversational Agent and Company Perceptions. *Computers in Human Behavior*, 2018, vol. 85, pp. 183–189. DOI: 10.1016/j.chb.2018.03.051.

CHEN, J.; GUO, F.; REN, Z.; LI, M.; HAM, J. Effects of Anthropomorphic Design Cues of Chatbots on Users' Perception and Visual Behaviors. *International Journal of Human–Computer Interaction*, 2023, pp. 1–19. DOI: 10.1080/10447318.2023.2193514.

COLQUITT, J. A.; CONLON, D. E.; WESSON, M. J.; PORTER, C. O. L. H.; NG, K. Y. Justice at the Millennium: A Meta-analytic Review of 25 Years of Organizational Justice Research. *Journal of Applied Psychology*, 2001, vol. 86, no. 3, pp. 425–445. DOI: 10.1037/0021-9010.86.3.425.

DE CONINCK, S.; EMMERS, E. 'Online, the Counselor Can't See Me Cry': A Systematic Literature Review on Emotion and Computer-mediated Care. *Frontiers in Digital Health* [online], 2023, vol. 5. DOI: 10.3389/fdgth.2023.1216268.

DIETVORST, B. J.; SIMMONS, J. P.; MASSEY, C. Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. *SSRN Electronic Journal*, 2015, vol. 144, no. 1. DOI: 10.2139/ssrn.2466040.

FEINE, J.; GNEWUCH, U.; MORANA, S.; MAEDCHE, A. A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies*, 2019, vol. 132, pp. 138–161. DOI: 10.1016/j.ijhcs.2019.07.009.

GLIKSON, E.; WOOLLEY, A. W. Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, 2020, vol. 14, no. 2, pp. 627–660. DOI: 10.5465/annals.2018.0057.

KOCH-BAYRAM, I. F.; KAIBEL, C. Algorithms in Personnel Selection, Applicants' Attributions about Organizations' Intents and Organizational Attractiveness: An Experimental Study. *Human Resource Management Journal*, 2023. DOI: 10.1111/1748-8583.12528.

KOIVUNEN, S.; ALA-LUOPA, S.; OLSSON, T.; HAAPAKORPI, A. The March of Chatbots into Recruitment: Recruiters' Experiences, Expectations, and Design Opportunities. *Computer Supported Cooperative Work (CSCW)* [online], 2022, vol. 31. DOI: 10.1007/s10606-022-09429-4.

LANGER, M.; KÖNIG, C. J.; KRAUSE, K. Examining Digital Interviews for Personnel Selection: Applicant Reactions and Interviewer Ratings. *International Journal of Selection and Assessment*, 2017, vol. 25, no. 4, pp. 371–382. DOI: 10.1111/ijsa.12191.

LEE, M. K. Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management. *Big Data & Society* [online], 2018, vol. 5, no. 1. DOI: 10.1177/2053951718756684.

LI, L.; LASSITER, T.; OH, J.; LEE, M. K. Algorithmic Hiring in Practice. In: *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 2021, pp. 166–176. DOI: 10.1145/3461702.3462531.

LUGER, E.; SELLEN, A. 'Like Having a Really Bad PA': The Gulf between User Expectation and Experience of Conversational Agents [online]. 2016. DOI: 10.1145/2858036.2858288.

MALIN, C.; FLEIß, J.; ORTLIEB, R.; THALMANN, S. Rejected by an AI? Comparing Job Applicants' Fairness Perceptions of Artificial Intelligence and Humans in Personnel Selection. *Frontiers in Artificial Intelligence*, 2025, vol. 8. DOI: 10.3389/frai.2025.1671997.

MAYER, R. C.; DAVIS, J. H.; SCHOORMAN, F. D. An Integrative Model of Organizational Trust. *Academy of Management Review*, 1995, vol. 20, no. 3, pp. 709–734. DOI: 10.5465/amr.1995.9508080335.

MEIJERINK, J.; BOONS, M.; KEEGAN, A.; MARLER, J. Algorithmic Human Resource Management: Synthesizing Developments and Cross-disciplinary Insights on Digital HRM. *The International Journal of Human Resource Management* [online], 2021, vol. 32, no. 12, pp. 1–18. DOI: 10.1080/09585192.2021.1925326.

MORI, M.; SASSETTI, S.; CAVALIERE, V.; BONTI, M. A Systematic Literature Review on Artificial Intelligence in

Recruiting and Selection: A Matter of Ethics. *Personnel Review*, 2024, vol. 54, no. 3. DOI: 10.1108/PR-03-2023-0257.
NEWMAN, D. T.; FAST, N. J.; HARMON, D. J. When Eliminating Bias Isn't Fair: Algorithmic Reductionism and Procedural Justice in Human Resource Decisions. *Organizational Behavior and Human Decision Processes*, 2020, vol. 160, no. 1, pp. 149–167. DOI: 10.1016/j.obhdp.2020.03.008.
PAGE, M. J.; MCKENZIE, J. E.; BOSSUYT, P. M.; BOUTRON, I.; HOFFMANN, T. C.; MULROW, C. D.; SHAMSEER, L.; TETZLAFF, J. M.; AKL, E. A.; BRENNAN, S. E.; CHOU, R.; GLANVILLE, J.; GRIMSHAW, J. M.; HRÓBJARTSSON, A.; LALU, M. M.; LI, T.; LODER, E. W.; MAYO-WILSON, E.; MCDONALD, S.; MCGUINNESS, L. A. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *British Medical Journal* [online], 2021, vol. 372, no. 71. DOI: 10.1136/bmj.n71.
SNYDER, H. Literature Review as a Research Methodology: An Overview and Guidelines. *Journal of Business Research* [online], 2019, vol. 104, no. 1, pp. 333–339. DOI: 10.1016/j.jbusres.2019.07.039.
SONG, S. W.; SHIN, M. Uncanny Valley Effects on Chatbot Trust, Purchase Intention, and Adoption Intention in the Context of E-Commerce: The Moderating Role of Avatar Familiarity. *International Journal of Human–Computer Interaction*, 2022, vol. 40, no. 2, pp. 1–16. DOI: 10.1080/10447318.2022.2121038.
STANLEY, D. S.; AGGARWAL, V. Impact of Disruptive Technology on Human Resource Management Practices. *International Journal of Business Continuity and Risk Management*, 2019, vol. 9, no. 4, p. 350. DOI: 10.1504/IJBCRM.2019.102608.
THOMAS, J.; HARDEN, A. Methods for the Thematic Synthesis of Qualitative Research in Systematic Reviews. *BMC Medical Research Methodology*, 2008, vol. 8, no. 1, pp. 1–10. DOI: 10.1186/1471-2288-8-45.
VAN ESCH, P.; BLACK, J. S.; ARLI, D. Job Candidates' Reactions to AI-Enabled Job Application Processes. *AI and Ethics*, 2020, vol. 1, pp. 119–130. DOI: 10.1007/s43681-020-00025-0.
YAN, C.; CHEN, Q.; ZHOU, X.; DAI, X.; YANG, Z. When the Automated Fire Backfires: The Adoption of Algorithm-based HR Decision-making Could Induce Consumer's Unfavorable Ethicality Inferences of the Company. *Journal of Business Ethics*, 2023. DOI: 10.1007/s10551-023-05351-x.

ОТ ДЕХУМАНИЗАЦИЯТА КЪМ ЦИФРОВАТА ЕМПАТИЯ: ОПИТЪТ НА КАНДИДАТИТЕ ПРИ НАЗНАЧАВАНЕ НА РАБОТА С ПОМОЩТА НА ИЗКУСТВЕН ИНТЕЛЕКТ

Резюме: *С нарастващото използване на AI-базирани технологии в ранните етапи на набирането на персонал, кандидатите често споделят, че се чувстват дехуманизирани и редуцирани. Настоящото проучване разглежда как дизайнът на чатботовете за набиране на персонал може да отговори на тези опасения чрез рамка за дигитална емпатия. Използвайки концептуален качествен синтез на литература от областта на организационната психология, взаимодействието човек – компютър и управлението на човешките ресурси, проучването идентифицира основните фактори, предизвикващи дехуманизация, като възприемана редукция, липса на прозрачност и неподходящи социални сигнали. Резултатите показват, че доверието в наемането на работа с помощта на изкуствен интелект зависи в голяма степен от качеството на първоначалните взаимодействия, при които кратките срещи оказват значително влияние върху възприятията за привлекателността на организацията и етичните стандарти.*

Рамката препоръчва чатботовете да бъдат проектирани така, че да показват контекстуална чувствителност и уважение, като същевременно избягват внушенията за машинно съзнание. Тя се фокусира върху пет основни аспекта: прозрачност на процеса, честно емоционално признание, структурирана обратна връзка, балансиран антропоморфизъм и надеждни механизми за поправка. Като се фокусира върху достойнството на кандидатите, а не само върху ефективността, тя предлага практически стъпки за изграждане на доверие и справедливост. Проучването заключава, че чатботовете трябва да действат като усъвършенствани представители, гарантиращи, че автоматизираното наемане на персонал подкрепя както бизнес целите, така и благосъстоянието на кандидатите.

Ключови думи: *набиране на персонал с помощта на изкуствен интелект; дигитална емпатия; опит на кандидатите; дехуманизация; изкуствен интелект, ориентиран към човека*

Оливер Гют, докторант
Университет по библиотекознание и информационни технолгии
София, България
E-mail: mail513686749@icloud.com

PUBLIC COMMUNICATIONS AND INFORMATION SCIENCES
ОБЩЕСТВЕНИ КОМУНИКАЦИИ И ИНФОРМАЦИОННИ НАУКИ

THE NEW HYBRID ROLE IN LAW: LEGAL RESPONSIBILITY IN AI-SUPPORTED KNOWLEDGE MANAGEMENT

Saskia Doris Prause
University of Library Studies and Information Technologies
https://doi.org/10.70300/THRH9602

Abstract: *This article analyzes the genesis of a legal-technological hybrid function at the intersection of law, artificial intelligence and knowledge management. The aim of the study is to analyze the structural, normative and epistemic prerequisites under which legal professionals can review and take responsibility for AI-supported decisions. The study pursues a conceptual-theoretical approach that combines legal dogmatic analysis, information science system evaluation and model-based structuring. The study's framework model integrates the hybrid role into modern organizations' knowledge and documentationprocesses, making it novel. The results show that AI-supported knowledge processes create new transparancy, documentation and accountability requirements that conventional legal instruments cannot meet. Model versioning, data provenance, audit trails and explainability are essential for legally compliant algorithmic evaluation. Ultimately, the study concludes that institutional anchoring of the hybrid role is an indispensable prerequisite for responsible and legally compliant AI governance in knowledge-intensive business.*
Keywords: *Artificial intelligence; knowledge management; legal hybrid role; AI governance; algorithmic accountability*

INTRODUCTION

The increasing integration of artificial intelligence (AI) into the internal knowledge processes of organizations is bringing about a fundamental transformation of the foundations of legal activities. Artificial intelligence systems are taking on tasks of analysis, structuring and pre-selection that previously required exclusively human expertise. As a result, issues of transparency, accountability and responsibility are becoming the focus of legal and information science debates. European and international regulatory frameworks, in particular the General Data Protection Regulation (GDPR), the AI Act, and organizational due diligence and supervisory obligations (Sections 823 ff. BGB (German Civil Code), Section 130 OWiG (German Administrative Offences Act) increasingly require companies to document decisions made by algorithmic systems in a comprehensible manner and to secure them in an audit-proof manner. Without systematic knowledge management, these requirements cannot be met. Current legal literature interprets these requirements as a manifestation of growing consolidation, which obliges organizations to implement comprehensive governance, documentation and control structures (Ebers, Müller-Terpitz & Golla 2023).

In addition to the European regulatory framework, companies are progessivley functioning within a disjointed context of AI governance. The united state predominantly depends on sector-specific regulations, enforcement by agencies like the FTC, and voluntary standards such as the NIST AI Risk Management Framework, wich proritizes risk management over formal ex ante requirements. In contrast, China has implemented a stae-centric strategy characterized by compulsory filing mandates, algorithmic recommendation regulation, and real-time oversight mechnismans. International standards like ISO/IEC 42001 impose supplemantary organizational obligations by delineating AI management systems and documentation responsibilities. The varying methodologies underscore the lack of a unified global framework and amplify the necessity for internal systems that can reconcile diverse regulatory expectations.

At the same time, new role profiles are emerging that combine legal, technological and organizational skills. These so-called hybrid roles are not emerging as a marginal phenomenon in work organization, but as a necessary response to the epistemic peculiarities of AI-based knowledge

production. Over the years, research has pointed out that algorithmic systems generate structural opacity due to their complex model architectures, which prevent the immediate traceability of algorithmic decision-making processes (Humphreys 2009; Pasquale 2015). In the context of the GDPR, structural opacity has been recognized in legal scholarship as a primary impediment to fullfilling the regulation's mandates for transparancy and substantive elucidation knowledge(Brkan & Bonnet 2020). This phenomenon of "Fepistemic opacity" means that traditional legal control instruments, such as the review of documentation, the presentation of evidence or the attribution of liability, reach their limits (Hildebrandt 2022; Veale & Edwards 2018). In addition, information ethics studies make it clear that the responsible use of algorithmic systems can only be guaranteed if their data streams, model versions and decision-making logic are documented and kept verifiable (Floridi & Taddeo 2016; Mittelstadt et al. 2016).

Notwithstanding the growing complexity of regulatory frameworks, numerous elements pertinent to corporate practice remain inadequately addressed by formal legislation. Existing regulations fail to delineate the allocation of internal accountability, the structuring of interdisciplinary collaboraration among legal, data science, and IT security, or the operationalization of knowledge architectures essential for documentation, versioning and aiditability. The unresolved organizational dimension generate governance gaps that companies must independently rectify, thereby reinforcing the nesessity for a hybrid professional function.

Simultaneously, numeros critical elements necessary for corporate AI governance, are inadequately addressed by current legal framework. The GDPR, AI Act, and international standards like the NIST AI RMF and ISO/IEC 42001 do not specify the implementation of internal accountability, interdisciplinary collaboration, or knowledge architectures within organizations. Current regulations require transparancy, documentation, and risk management, but do not specify how companies should organize roles, competencies and internal reviw processes to meet these requirements. This regulatory void imposes a considerable burden on organizations, necessitating the development of theire own governance framework to guarantee the reliability and verifiability of AI-assisted knowledge processes. This gap highlits the need for a hybrid professional role that integrates legal, technical and organizational domains.

The work of legal professionals is increasingly shifting from ex post review to ex ante governance of AI-based systems. Lawyers are not only required to identify legal risks, but also to understand the epistemic conditions of knowledge production in algorithmic processes. The specialist literature on explaining AI emphasizes that the explainability of algorithmic decisions is always context-dependent and cannot be operationalised without a solid knowledge architecture (Selbst & Barocas 2016). This results in a redistribution of tasks. Legal expertise must be more closely integrated with knowledge management, data economics, model testing and compliance.

In this context, this article analyzes the novel hybrid function at the interface of law, artificial intelligence and knowledge management. The focus is on three comprehensive research questions. First, how does AI modify the dogmatic and organizational prerequisites of legal responsibility in knowledge-intensive companies ("knowledge-intensive companies" refers to organizations whose fundamental operational processes depend on intricate knowledge generation, assesment and documentation)? Second, what technical and epistemic prerequisites must be in place for legal professionals to be able to review, evaluate and take responsibility for algorithmic decisions? Third, how can knowledge management be designed as an infrastructure to meet the requirements for transparency, documentation and governance demanded by contemporary regulatory practice?

Preliminary evidence from industry practices suggests that companies are commencing the institutionalization of hybrid functions in AI governance. Numerous organizations have established specialized positions, such as AI compliance officers or interdisciplinary governance teams, which integrate legal expertise with data governance, IT scurity, and model oversight. These developing structures illustrate that the hybrid role is not merely a theoretical concept, but a pragmatically response to the organizational requirements engendered by AI-supported knowledge processes.

This article therefore provides a twofold gain in knowledge. On the one hand, it demonstrates that the hybrid role represents not merely a professional development, but a legal-dogmatic and

information-scientific necessity in the age of algorithmic knowledge production. On the other hand, the article develops a conceptual framework model that integrates the hybrid role into the knowledge and documentation processes of contemporary organizations, thus providing a starting point for legally compliant AI governance.

RESEARCH METHODOLOGY

This article follows a conceptual-theoretical research approach that integrates legal analysis, information science system evaluation and conceptual-modeling. The methodology was designed to adequately capture the structural, epistemic and normative conditions of the hybrid role in the context of AI-supported knowledge management.

First, this article is based on a legal-dogmatic examination to derive the requirements for transparency, responsibility and accountability from the GDPR, the AI Act, and from liability and organizational law standards. This examination follows traditional legal methodology and includes an interpretation based on wording, systematics and telos. In addition, a comparative law approach is used to compare international regulatory strategies, in particular the US NIST AI Risk Management Framework and Chinese guidelines on algorithmic regulation (Floridi & Taddeo 2016).

Furthermore, an information science system analysis is used to understand the functioning, limitations and epistemic characteristics of AI-supported knowledge processing. This includes the assessment of algorithmic opacity ("epistemic opacity") and the documentation and governance procedures associated with AI. Key works on algorithmic accountability, explainable AI and data ethics (Humphreys 2009; Mittelstadt et al. 2016; Selbst & Baropcas 2016) form the basis for this article.

Ultimately, this article pursues a conceptual modelling approach with the aim of creating an analytical framework model of the hybrid legal role. The approach uses established methods of conceptual theory formation from information and organizational sciences and defines central dimensions of the hybrid role, including competence profiles, knowledge and documentation requirements, and control and monitoring mechanisms. The model integrates legal requirements with organizational conditions and epistemic characteristics of AI-based systems.

The synthesis of legal dogmatic analysis, information science system evaluation, and conceptual modelling allows the hybrid role to be captured normatively, structurally, and epistemically. This method proves particularly appropriate, as empirical data for researching emerging role profiles in the field of AI and law are limited to date, and the scientific debate takes place mainly at the theoretical level (Pasquale 2015; Hildebrandt 2022). The chosen approach thus ensures analytical depth and a connection to existing research traditions in the fields of law, information science and organizational research.

RESULTS

The results of this paper illustrate that the development of a hybrid legal-technological role is a necessary response to the growing integration of AI into the internal knowledge processes of organizations. Industry studies show that a significant proportion of legal and compliance departments already use AI-based tools or are planning to implement them, with generative AI, automated contract analysis and AI-supported compliance systems gaining particular importance (Haufe AI Report 2025; Deloitte Legal 2024). These developments are resulting in a transformation of the role profiles of legal professionals, who are increasingly operating at the interface between law, data processing and knowledge organization and need to translate technical results into normative categories (EY Law 2025).

Furthermore, it can be confirmed that a structural shift in the mechanisms for accountability and attribution within companies is taking place. Whereas decisions with liability implications were previously based on human judgement, AI-based analysis and decision-making systems are changing the fundamentals of responsibility allocation. External studies and legal analyzes suggest that errors in AI-based decision-making processes are often not attributable to the algorithmic model itself, but rather to incomplete knowledge versioning, inaccurate data quality or a lack of documentation (Wendehorst & Borges 2023; Caralegal 2024; Noxtua 2024). This finding corre-

sponds with legal studies that suggest that companies are increasingly obliged to document not only the results of AI systems, but also their epistemic basis, such as training data, model versioning and implementation decisions, in a comprehensible manner in order to enable legal accountability (Hacker 2023). This assessment is consistent with current analyzes of the governance structure of the AI Act, which require a systematic link between data quality, documentation and organizational accountability (Wendehorst & Borges 2023).

It should be emphasised that the review of algorithmic decisions by lawyers is only feasible if the technical processes underlying these decisions are designed to be epistemically transparent. The technical documentation of contemporary AI systems is dynamic, as numerous models are based on continuous learning. Providers such as NVIDIA emphasise that mechanisms for model versioning, data provenance and monitoring of algorithmic processes are essential for the use of AI in companies in order to meet the regulatory requirements of the AI Act and the GDPR (NVIDIA AI Solution 2024). In its guidelines, the European Data Protection Board (EDPB) also emphasises that controllers must implement technical and organizational measures to ensure that data flows, risk assessments and decision-making processes can be documented and verified at all times (EDPB 2021). Machine learning frameworks such as NVIDIA Merlin illustrate that model training and behaviour are influenced not only by data structures but also by the quality, distribution and sequencing of the training data. Even slight discrepancies in training data, preprocessing pipelines or batch compositioning can result in substantially different model outputs, thereby complicating the verification and reconstruction of AI-assisted decisions for legal practitioners (NVIDIA Merlin 2024). This is confirmed by the scientific literature on epistemic opacity, which states that the verifiability of algorithmic decisions can only be guaranteed if organizations implement robust knowledge architectures that document the entire life cycle of the models (Creel & Hellmann 2023; Selbst & Barocas 2016). A fundamental finding pertains to the enduring nature of epistemic opacity in practice. In various organizational contexts, AI systems generated outputs whose underlying reasoning remained irretrievable by legal professionals, despite the availability of documantation, versioning or technical logs. Thias tangible manifestation of opacity directly impacts legal verification processes and illustrates that the issue is fundamentally ingrained in modern AI architectures, rather than being merely a theoretical notion. In addition, Veale, Binns and Edwards (2018) emphasise that effective algorithmic accountability is only achievable when both technical and organizational control mechanisms interact with each other.

Analysis of external, practice-oriented sources also shows that knowledge management is becoming a legal infrastructure in the context of AI implementation. Contract management systems such as Tomorro, Kia Systems and DraftPilot already use centralised knowledge repositories, documented versioning mechanisms, automated clause recognition and generative support for contract review. Research shows that the implementation of such systems not only leads to increased efficiency, but also improves the consistency of legal assessments and significantly reduces the error rate (KIA Systems 2024; Draft Pilot 2024; Thomson Reuters 2024). Knowledge management is therefore a necessary condition for audit-proof and auditable legal decision-making processes that must meet both internal and external accountability requirements. Although the GDPR and the AI Act do not explicity employ these terms, both frameworks necessitate transparency, documentation and traceability, which inherently imply auditability. In contrast, ISO/IEC 42001 advances this by incorporating integrity and immutability requirements that resemble audit-proof storage. The normative significance of this development is highlighted by the international standard ISO/IEC 42001:2023, which specifies documented data, knowledge and control processes as a mandatory requirement for the implementation of AI management systems. An analysis of international regulatory strategies reveals that global frameworks are diverse. The NIST AI RMF in the United States adopts a voluntary, risk-management framework, whereas China's algorithmic regulations prioritize state supervision and impose compulsory filing requirements. Neither framework, however, offers the degree of legally binding documentation and accountability obligations that define the EU's AI Act and GDPR, thereby imposing further organizational requirements on companies operating transnationally.

Finally, the results indicate that the effective execution of the hybrid role is significantly linked to the external regulatory framework, especially the transparency, documentation and accountability requirements imposed by the GDPR, the AI Act and associated governance standards. These legal stipulations establish the fundamental criteria that organizations must fulfill when implementing AI-driven knowledge processes. Simultaneously, industry analyzes indicates that organizations frequently lack the internal prerequisites necessary to function within this regulatory framework. Interdisciplinary skill deficiencies, ambiguous responsibility frameworks, inadequate governance structures and restricted access to technical explainability processes are recognized as pricipal organizational obstacles to the effective implementation of AI (Haufe AI Report 2025; Deloitte 2024; EY Law 2025).

The study makes it clear that the hybrid role is only effective if organizations implement adequate governance structures, such as committees for AI risk assessment, compliance loops, role-based access systems, and clearly defined areas of responsibility between legal, IT security, and data processing. Legal professionals need skills in data economics, algorithmic model logic, explainable AI, IT security structures and documentation-based compliance to perform this task. Without these skills, both the traceability and legal accountability of algorithmic decisions remain insufficient. A similar understanding is evident in the British AI Auditing Framework, which presents comprehensive auditing, monitoring and documentation processes as essential components of responsible AI implementation (ICO 2022).

Taking all factors into account, current research thus illustrates that the hybrid role represents not only a novel form of professional specialisation, but rather an indispensable institutional response to the legal, technical and epistemic challenges associated with the use of AI. It acts as an interface between algorithmic knowledge production and legal accountability, and thus represents an essential component of legally compliant AI governance in contemporary organizations.

CONCLUSIONS/DISCUSSION

The examination of the hybrid role makes it clear that the challenges arising from AI-supported knowledge processes are neither exclusively technical nor purely legal in nature, but arise at the intersection of both disciplines. Essentially, the results show that the epistemic characteristics of machine decision-making processes – in particular their potential lack of transparency and model variability – necessitate a realignment of legal accountability structures. While traditional legal procedures are based on stable, comprehensible decision-making principles, AI systems operate dynamically, data-driven and often probabilistically. This discrepancy between technical contingency and legal determination is the central area of tension in the study.

The practical significance of these challenges is exemplified by recent jurisprudence. In the prominent U.S. federal case “Mata vs. Avianca” (2023), an attorney presented AI-generated summaries of alledged precedents that were ultimately entirely fictitious. The incident illustrates that AI-assisted knowledge generation can produce epistemic artefacts, like fabricated citations, that significantly compromise legal reliability in the absence of stringent verification procedures. This case highlights the imperative for organizational safeguards, documentation responsibilities, and specialized hybrid expertise to evaluate the validity of AI-generated information.

In this context, it becomes clear that the traditional dichotomy between legal review procedures and organizational knowledge processes can no longer be maintained. The findings indicate that legal expertise alone is not sufficient to ensure the conditions for legally sound, AI-based decisions. The need for an interdisciplinary functional role that integrates technical model logic, information science documentation mechanisms and legal review standards becomes apparent. The hybrid role thus functions not merely as an operational supplement to existing structures, but represents a systemic response to the complex interdependence of data, models and legal requirements.

It is not suprising that the GDPR and the AI Act demonstrate significant conceptual alignment, as both are integral components of the EU’s overarching regulatory framework. Although they pertain to distinct topics – personal data protection in the GDPR and systemic AI risk governance

in the AI Act – they converge in their focus on transparency, documentation, and organizational accountability. This regulatory coherence is emblematic of the EU's methodology and contrasts with other international frameworks, such as the voluntary and risk-based NIST AI RMF in the United States or the state-centric algorithmic regulation model in China, which do not mandate similarly comprehensive or binding documentation requirements. The findings demonstrate that the EU's dual framework establishes a distinctly rigorous environment where model transparency and robust knowledge structures are essential prerequisites for legal accountability.

A significant discovery pertains to the role of knowledge management in AI-assisted legal practice. The findings indicate that the legally dependable management of AI-generated outputs is feasible only when organizations develop coherent and consistent knwledge architectures. The legal hybrid role functions as a central nexus that links information regarding data sources, training methodologies, model behavior, and legal liabilities. This integration establishes the foundation for creating legally sound and tamper-proof audit trails that satisfy both technical and legal standards. Thus, knowledge management in the realm of AI-driven decision-making should be perceived not merely as a supportive function, but as an essential legal infrastructure.

Finally, the debate makes it clear that the hybrid role is not merely a reaction to current technological advances, but rather an indicator of a more comprehensive transformation of legal professions. Advancing automation and data-based decision-making are causing a shift from legal functions to tasks in the areas of governance, control and architecture. The hybrid role thus characterises a new type of legal activity that is continuously redefining the interaction between law, technology and organization. This insight underscores the long-term relevance of the hybrid role for the responsible and legally secure use of artificial intelligence in knowledge-intensive companies.

CONCLUSION

This article illustrates that the implementation of AI in knowledge-intensive areas of business is causing fundamental changes in the structures of legal responsibility. Modern AI systems not only increase the efficiency of organizational knowledge processes, but also generate new requirements in terms of transparency and auditability. As Hildebrandt (2022) emphasises, the application of algorithmic methods is increasingly shifting legal control from ex post review to ex ante governance, which requires a mandatory expansion of legal role profiles. The findings of the study confirm this assessment and clearly show that a new hybrid role is emerging in this context, combining legal, technological and information science skills.

It is crucial to recognise that the legally compliant implementation of AI is only feasible if organizations establish solid knowledge and documentation structures. Current technical studies indicate that AI systems absolutely require mechanisms such as model versioning, data provenance and audit trails in order to comply with the regulatory requirements of the AI Act and the GDPR (NVIDIA AI Solutions 2024). Without such epistemic foundations, lawyers are unable to understand or legally assess algorithmic results. The hybrid role thus takes on the central function of enabling technical documentation to be turned into legally usable knowledge artefacts in the first place. Comparable obligations are not uniformly present outside the European Union. The United States predominantly depends on voluntary guidelines like the NIST AI Risk Management Framework, which promotes traceability without mandating model versioning or data provenance, whereas China enforces stringent logging and documentation requirements within its state-centric algorithmic regulatory framework. International standards like ISO/IEC 42001 similarly underscore lifecycle documentation and traceability, yet remain non-mandatory. This regulatory divergence underscores the necessity for internal governance frameworks that allow organizations to comply with the EU's legally binding obligations while maintaining alignment with other international standards.

At the same time, industry reports show that companies increasingly view the hybrid role as an organizational necessity. According to recent research, more than two-thirds of legal and compliance departments intend to expand interdisciplinary skill sets, as complex AI systems can only be managed effectively through specialised roles (Haufe AI Report 2025). Due to internal risk man-

agement requirements, market expectations, and voluntary governance frameworks like the NIST AI RMF, hybrid functions are also emerging in the US and other jurisdictions. The GDPR and AI Act have created a dense regulatory environment in Europe. The hybrid role should not be seen as a regional reaction to EU regulation, but as an organizational response to AI systems' increasing complexity, opacity, and accountability demands, according to the study.

In the future, the hybrid role will be a key pillar of legally compliant AI governance. The ongoing automation of knowledge-based activities, the continuous development of complex model architecture and increasing regulatory consolidation make the close integration of legal, technical and information science skills essential. Companies that systematically establish this function not only lay the foundation for the transparent and responsible use of artificial intelligence, but also strengthen their ability to resiliently meet the legal and organizational challenges of a data-oriented economic area. The hybrid role is thus establishing itself as a central link between technological innovation and the rule of law and will play a key role in the future in ensuring that AI-based knowledge processes can be designed to be efficient, transparent and legally secure.

It will be interesting to see how the hybrid role evolves across organizational and regulatory contexts. How governance models work in the practical application of AI-supported and increasingly automated knowledge processes and how law, data science, and IT security institutionalize new forms of interdisciplinary cooperation will be crucial. The long-term organizational structures, competence profiles, and documentation mechanisms needed to support an AI-driven, model-based, and tightly regulated business environment will also become apparent.

REFERENCES

BRKAN, M.; BONNET, G., 2020. Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions. *Computer Law & Security Review*, vol. 36, pp. 105–110.

CARLEGAL, 2024. *AI Governance & Compliance Solution for the AI Regulation* [online]. Berlin: caralegal GmbH. [viewed 20 November 2025]. Available from: https://caralegal.eu/ai-governance/

CREEL, K.; HELLMAN, D., 2023. The Epistemic Dilemma of Explainable AI. *Harvard Journal of Law & Technology*, vol. 36, no. 2, pp. 1–46.

DELOITTE LEGAL, 2024. *How Generative AI is Changing Legal Department Functions* [online]. Deloitte Perspectives, 2024 [viewed 18 November 2025]. Available from: https://www.deloitte.com/global/en/services/legal/perspectives/how-generative-ai-is-changing-legal-department-functions.html

DRAFTPILOT, 2024. *AI Contract Review – Product Overview* [online]. DraftPilot Technologies. [viewed 26 November 2025]. Available from: https://draftpilot.ai/contract-review

EBERS, M.; GASSMANN, O., 2024. AI Regulation and Corporate Governance in Europe. *Journal of European Consumer and Market Law*, 2024, vol. 13, no. 1, pp. 12–22.

EBERS, M.; MÜLLER-TERPITZ, R.; GOLLA, S., 2023. *Artificial Intelligence and Law*. Tübingen: Mohr Siebeck

EDPB. *Guidelines 01/2021 on Examples Regarding Data Breach Notification* [online]. Brussels: European Data Protection Board, 2021 [viewed 26 November 2025]. Available from: https://edpb.europa.eu

EDWARDS, L.; VEALE, M., 2018. Slave to the Algorithm? Why a 'Right to an Explanation' is Probably Not the Remedy You Are Looking For. *Duke Law & Technology Review*, vol. 16, no. 1, pp. 18–84.

EY LAW, 2025. *2025 EY Law General Counsel Study* [online]. London: EY Law. [viewed 24 November 2025]. Available from: https://www.ey.com/content/dam/ey-unified-site/ey-com/en-gl/insights/law/documents/ey-law-general-counsel-study-04-2025.pdf

FLORIDI, L.; TADDEO, M., 2016. What Is Data Ethics? *Philosophical Transactions of the Royal Society A*, vol. 374, no. 2083, pp. 1–9. DOI: 10.1098/rsta.2016.0360.

HACKER, P., 2023. AI Liability, Causation and Accountability in European Law. *Common Market Law Review*, vol. 60, no. 3, pp. 825–864.

HAUFE, 2025. *AI Report 2025: GenAI in der Rechtsbranche* [online]. Freiburg: Haufe Group. [viewed 02 December 2025]. Available from: https://www.haufe.de/recht/kanzleimanagement/ai-report-2025-ueber-genai-in-der-rechtsbranche_222_654492.html

HILDEBRANDT, M., 2022. *Law for Computer Scientists and Other Folk*. Oxford: Oxford University Press

HUMPHREYS, P., 2009. Network Epistemology and the Problem of Epistemic Opacity. *Episteme*, vol. 6, no. 3, pp. 206–218. DOI: 10.3366/E1742360009000627.

ICO, 2022. *AI Auditing Framework* [online]. London: Information Commissioner's Office, [viewed 26 November 2025]. Available from: https://ico.org.uk

ISO/IEC 42001:2023. *Artificial Intelligence Management System — Requirements*. Geneva: International Organization for Standardization, 2023.

KIRA SYSTEMS, 2024. *How Kira Works – Machine Learning for Contract Analysis* [online]. Toronto: Kira Systems Inc. [viewed 25 November 2025]. Available from: https://kirasystems.com/how-kira-works/
MATA v. AVIANCA INC. Opinion and Order on Sanctions, Case No. 22-cv-1461 (PKC). United States District Court for the Southern District of New York, 22 June 2023 [online]. [viewed 01 December 2025]. Available from: https://caselaw.findlaw.com/court/us-dis-crt-sd-new-yor/2335142.html
MITTELSTADT, B.; ALLEN, J.; ANDERSON, R.; O'NEILL, O.; FLORIDI, L., 2016. The Ethics of Algorithms: Mapping the Debate. *Big Data & Society*, vol. 3, no. 2, pp. 1–21. DOI: 10.1177/2053951716679679.
NVIDIA, 2024. *Responsible AI – Principles and Framework* [online]. Santa Clara: NVIDIA Corporation, 2024 [viewed 26 November 2025]. Available from: https://www.nvidia.com/en-us/responsible-ai/
NVIDIA MERLIN. *NVIDIA Merlin Recommender Framework: Developer Guide* [online]. Santa Clara: NVIDIA Developer, 2024 [viewed 26 November 2025]. Available from: https://developer.nvidia.com/merlin
NOXTUA, 2024. *AI-supported Contract Analysis and Legal Workflows* [online]. Hamburg: Noxtua GmbH. [viewed 26 November 2025]. Available from: https://www.noxtua.com
PASQUALE, F., 2015. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press, 2015.
SELBST, A.; BAROCAS, S., 2016. The Intuitive Appeal of Explainable Machines. *Fordham Law Review*, vol. 87, no. 3, pp. 1085–1139.
THOMSON REUTERS INSTITUTE, 2024. *Future of Professionals Report 2024* [online]. New York: Thomson Reuters Institute. [viewed 26 November 2025]. Available from: https://www.thomsonreuters.com/en-us/posts/wp-content/uploads/sites/20/2024/03/future-of-professionals-2024-report.pdf
TOMORRO, 2024. *Contract Management – Product Overview* [online]. Berlin: Tomorro GmbH. [viewed 20 November 2025]. Available from: https://tomorro.com/contract-management/
VEALE, M.; BINNS, R.; EDWARDS, L., 2018. Algorithms and Accountability: A Framework for Regulation [online]. *Oxford Internet Institute Working Paper*. [viewed 26 November 2025]. Available from: https://ora.ox.ac.uk
WENDEHORST, C.; BORGES, G., 2023. EU AI Act and the Future of Corporate Compliance. *European Journal of Risk Regulation*, vol. 14, no. 2, pp. 225–240.

НОВАТА ХИБРИДНА РОЛЯ В ПРАВОТО: ПРАВНА ОТГОВОРНОСТ В УПРАВЛЕНИЕТО НА ЗНАНИЯТА, ПОДКРЕПЕНО ОТ ИЗКУСТВЕН ИНТЕЛЕКТ

Резюме: *Статията анализира генезиса на хибридна функция в областта на правото и технологиите, намираща се на пресечната точка между правото, изкуствения интелект (ИИ) и управлението на знанието. Целта на изследването е да анализира структурните, нормативните и епистемичните предпоставки, при които правните специалисти могат да преразглеждат и поемат отговорност за решения, подкрепени от ИИ. Проучването следва концептуално-теоретичен подход, който съчетава правно-догматичен анализ, оценка на информационни системи и структуриране на базата на модели. Рамковият модел на проучването интегрира хибридната роля в процесите на знание и документиране в съвременните организации, което го прави новаторски. Резултатите показват, че подкрепените от ИИ процеси на знание създават нови изисквания за прозрачност, документиране и отчетност, които конвенционалните правни инструменти не могат да удовлетворят. Версиите на моделите, произходът на данните, одиторските следи и обяснимостта са от съществено значение за алгоритмичната оценка, която е в съответствие с правото. В крайна сметка проучването заключава, че институционалното утвърждаване на хибридната роля е незаменима предпоставка за отговорно и съобразено със закона управление на ИИ в бизнеса, основан на знания.*
Ключови думи: *изкуствен интелект; управление на знанията; хибридна правна роля; управление на изкуствения интелект; алгоритмична отговорност*

Саския Дорис Праузе, докторант
Университет по библиотекознание и информационни технологии
София, България
E-mail: prause.saskia@gmail.com

PUBLIC COMMUNICATIONS AND INFORMATION SCIENCES
ОБЩЕСТВЕНИ НАУКИ И ИНФОРМАЦИОННИ ТЕХНОЛОГИИ

A PRISMA-DRIVEN SYSTEMATIC LITERATURE REVIEW OF PROJECT MANAGERS' LEADERSHIP COMPETENCIES

Timo Zimenga
University of Library Studies and Information Technologies
https://doi.org/10.70300/HIUK1096

Abstract: *Due to digitalization teams are increasingly working virtually which makes leadership complex, especially when the leader has no formal authority over the team, which is often the case in projects. Project managers, as the leaders of the project, must adapt to the needs of virtual collaboration and have certain leadership competencies. This study examines whether a competency model can be found that lists the leadership competencies of project managers who lead virtual teams without formal authority. The study was conducted using a systematic literature analysis based on the PRISMA structure. Relevant search strings were used and n = 699 articles were identified. After completion of the PRISMA process, seven articles were included. The results show that no article cover the core areas of this study. Therefore, no competency model can be identified that lists the leadership competencies of project managers who lead a virtual team without formal authority. This study therefore identifies a research gap and highlights the need for further research to develop this competency model.*
Keywords: *virtual teams; project manager; leadership competencies; lack of formal authority; systematic literature review*

INTRODUCTION

Digitalization has fundamentally changed the way we work. Virtual teams have become increasingly popular in recent years and are now coming to the forefront (Froese et al. 2025). In a large-scale study conducted in 2023 with nearly 360,000 employees, Makridis and Schloetzer (2023) found that approximately 50% of employees work in virtual or hybrid teams. This is supported by Treas (2021) in his study, which examines the extent to which virtual work is the future. Treas (2021) found that virtual work is the future. In his study, 82% of participants voted in favor of remote work (Treas 2021).

Although the relevance of virtual teams has been clearly defined in the literature and virtual teams are therefore the preferred working structure of the future, team members of virtual teams experience some negative effects because of this virtual structure. Among other things, there is a lack of informal and spontaneous contact opportunities within the team (Leslie et al. 2018). In addition, building trust is more difficult than in face-to-face teams, as interaction takes place almost exclusively via information and communication technologies (Torro, Pirkkalainen, Li 2022). Technical problems or uncertainty in the use of information and collaboration technologies can lead to conflicts among team members (Diefenbach 2023). In view of these conditions and the spread of virtual teamwork, the role of the manager is coming to the forefront (Caputo et al. 2024; Dörrenberg, Goerner, Patzak 2019). Specific forms of leadership are necessary, and leadership tasks, roles, and tools need to be reconsidered (Dreas, Klenk 2021). Leaders of virtual teams must adapt to the new situation and be able to commit to this form of leadership (Ilter et al. 2023; Grześ 2024).

Based on this, the complexity of employee leadership is further increased when the leader has no formal authority, as this is often the case in projects (Grześ 2024; Bohinc 2021). According to the definition of the Project Management Institute, a project is *"a temporary endeavor undertaken to create a unique product, service, or result. The temporary nature of projects indicates a beginning and an end to the project work or a phase of the project work. Projects can stand alone or*

be part of a program or portfolio" (PMI 2021). Projects are now a well-established practice in many companies (Wagner, Cammans 2025) and are used where companies want to achieve their goals most efficiently (Blomquist, Müller 2006). The background for the continuously increasing establishment and strategic importance of projects includes digitalization and the associated projectification (Wagner, Huemann, Radujković 2022) of companies (Wagner, Ott 2019; Shamim 2022). Accordingly, Wagner, Huemann, and Radujković (2022) cite the author Midler (1995), who already discovered in 1995 that the projectification of companies has a significant influence on corporate structure, corporate culture, and, consequently, teamwork.

This structural shift toward the increased use of projects is increasing the demand for the right skills to successfully lead these projects (Wagner, Ott 2019). Empirical research shows that deficits in leadership contribute to project failure: Fareed, Su, and Awan (2021) found in a study that 4 out of 5 projects fail due to poor leadership. This would mean that leadership has a major impact on project management. This is confirmed by Moradi, Kähkönen, and Aaltonen (2019), who say that project performance is the sum of the skills of the project team and the competencies of the project manager and that, accordingly, the leadership competencies of the project manager are key elements for successful project management.

According to Zehrer and Mössenlechner (2010), leadership competencies are seen as an interface between the individual basic competencies, which are divided into four competency areas: social and communication competencies, activity and action competencies, technical and methodological competencies, and personal competencies. In contrast to the regular competencies from the previously mentioned competency areas, which enable a person to adapt to the respective situation (Krumm, Mertin, Dries 2012), leadership competencies, according to Zehrer and Mössenlechner (2010), represent a special type of competency that is important for the successful performance of leadership tasks.

Authors such as Nagel (2025), Moradi, Kähkönen, and Aaltonen (2019), Heyse and Erpenbeck (2009), and several others, as well as individual organizations such as the IPMA (GPM 2017) and the PMI (Project Management Institute 2021), have developed competency models for project managers. However, these models only evaluate individual areas or provide a general overview of competencies.

Despite the relevance of project managers' leadership competencies, these competency models do not explicitly address leadership competencies. With this in mind and considering the individual aspects and framework conditions, the question arises as to whether and in what form competency models exist that focus on the leadership skills of project managers who lead a virtual team without formal authority. This raises the following research question, which will be answered in the context of this systematic literature review:

RQ: To what extent can an independent competency model be identified that describes the leadership competencies of project managers who lead a virtual team without formal authority?

To keep the empirical focus in the foreground, this review treats PRISMA not as the object of analysis but as the transparent screening procedure through which the relevant literature is delimited. The author's contribution is therefore the category-led evaluation of the included studies against five intersecting criteria: competency model, leadership competencies, project managers, lack of formal authority, and virtual teams. This framing clarifies that the review does not merely reproduce existing literature, but assesses whether the existing evidence jointly covers the specific leadership situation formulated in the research question.

RESEARCH METHODOLOGY

PRISMA-based literature selection

A systematic literature review will be conducted to answer the research question presented above. This study will follow the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for the search strategy, screening, and final selection of literature, as this ensures that the systematic literature review process can be viewed transparently and that the

systematic literature review shows how the review was conducted, what the author did, and what was found (Page et al. 2021). The aim of the systematic literature review is to conduct a systematic and comprehensive search for publications to ensure that as much literature as possible on a specific topic can be identified (Gash 1999). The advantage of this approach is that, due to the increasing use of online media and online databases, more and more potentially relevant digital documents are becoming available. This ensures that scientific literature is easily and comprehensively accessible (Döring, Bortz 2016). Accordingly, this approach also focuses on finding high-quality, existing literature and thus using a database that does not first need to be developed (Döring, Börtz 2016; Mayring 2016).

This methodological description is retained in full in order to make the source selection auditable; the subsequent Results and Discussion sections therefore foreground the screening outcome and the evaluative comparison of the relevant studies.

Following data collection, data review is based on the approach according to Mayring (2016). This approach consists of three steps and begins with emphasizing the research question to ensure that the approach is targeted and based on the research question. The second step is the targeted selection of suitable material for answering the research question. The third step involves evaluating the material which was found (Mayring 2016).

Category-led analytical procedure

In the first step, which is about defining the research question, the introduction already showed how important leadership competencies are for project managers who lead virtual teams without formal authority. As already formulated in the research question, this systematic literature review focuses on the question of whether there is a competency model for project managers who lead a virtual team without formal authority and whether this competency model focuses on leadership skills. This step has already been carried out in the introduction. The second step involves selecting suitable literature. This data collection was carried out using the PRISMA procedure and can be viewed in the following section. In the final step, according to Mayring (2016), the articles that resulted from the data collection are reviewed. The articles are analyzed according to a predefined category-led procedure to ensure that the study review is also carried out systematically.

In summary, the systematic literature review is therefore carried out in a multi-stage process in which scientific publications on a specific topic are selected systematically and based on categories (Döring, Bortz 2016). The aim of this literature review is to obtain a comprehensive overview of the research topic under investigation and thereby provide a valid answer to the research question.

RESULTS

Search strategy and database coverage

Three search strategies consisting of search terms and Boolean operators, combined and formulated in both German and English, were used to collect data and create a comprehensive database. These are presented below:

- Search string #1 (English) and #4 (German): ("project manager" OR "project leadership" OR "project leader") AND ("non-formal authority" OR "informal leadership" OR "lateral leadership") AND ("virtual teams" OR "remote teams" OR "distributed teams" OR "hybrid teams") AND ("leadership competencies" OR "leadership skills" OR "competence" OR "competencies");
- **Search string #2 (English) and #5 (German)**: ("project manager" OR "project leadership" OR "project leader") AND ("virtual teams" OR "remote teams" OR "distributed teams" OR "hybrid teams") AND ("leadership competencies" OR "leadership skills" OR "competence" OR "competencies");
- **Search string #3 (English) and #6 (German)**: ("project manager" OR "project leadership" OR "project leader") AND ("non-formal authority" OR "informal leadership" OR "lateral leadership") AND ("leadership competencies" OR "leadership skills" OR "competence" OR "competencies").

The search strings were developed based on the topics described in the introduction. The aim was to find as comprehensively as possible a range of suitable literature that could contribute to an-

swering the research question. To conduct the literature, search in both a national and international context and to be able to find literature in both German and English, the search strings were used in the databases BASE Search, EBSCO, Emerald Insight, JSTOR, SAGE, Science Direct, Scopus, Springer, and Web of Science. Search strings #1 and #4 combine all core aspects of the study in a central literature query. The literature searched should cover the core areas of project managers'/project management, leadership without formal authority, virtual teams, and (leadership) competencies. Search strings #2 and #5 focus on the relationship between the project manager or project management, (leadership) competencies, and the virtual team, without referring to leadership without formal authority. Finally, search strings #3 and #6 deal with the connection between project managers or project management, leadership without formal authority, and (leadership) competencies, disregarding virtual teams. The core areas were combined using Boolean operators (AND/OR) to find thematically relevant literature. The form of the selected search strings aims to make the literature search as comprehensive as possible. The possible integration of literature that may only cover three of the four core areas ensures that studies are also considered that do not cover all core areas at the same time but may nevertheless provide important information for answering the research question. To ensure that the articles analyzed reflect the current state of research, articles published before 2010 were excluded during data collection by selecting appropriate filters. Furthermore, articles that were blogs, magazines, or popular literature were excluded to ensure the scientific quality of the examined articles.

In order to make the literature search as efficient as possible, it was limited to keyword searches. During data collection in September 2025, a total of n = 699 articles were identified. Following the PRISMA statement process, the literature found in the databases was systematically analyzed in three phases (Page et al. 2021). In the first step (Identification) 50 duplicates were removed. Accordingly, n = 649 articles were transferred to the screening process. During the screening process, n = 625 articles were sorted out based on the title and abstract review because they were not relevant to the topic. Subsequently, n = 24 articles were selected for a full-text search. n = 3 articles could not be accessed because they were not available despite university access. Open access articles are essential for research because they guarantee access to articles at any time and increase the research base (Brinken 2021; Björk, Salomon 2012). Accordingly, the full-text search resulted in n = 21 articles. During the full-text search, n = 14 articles were excluded because they did not ultimately relate to competencies, had no connection to project management, did not analyze the competencies of the project team, or did not evaluate competencies in relation to project success. In the final process according to PRISMA 2020 (Included), n = 7 studies were included that are intended to contribute to answering the research question. The process of systematic literature analysis according to PRISMA 2020 (Page et al 2021) can be seen in the following PRISMA flowchart in Figure 1.

PRISMA screening outcome

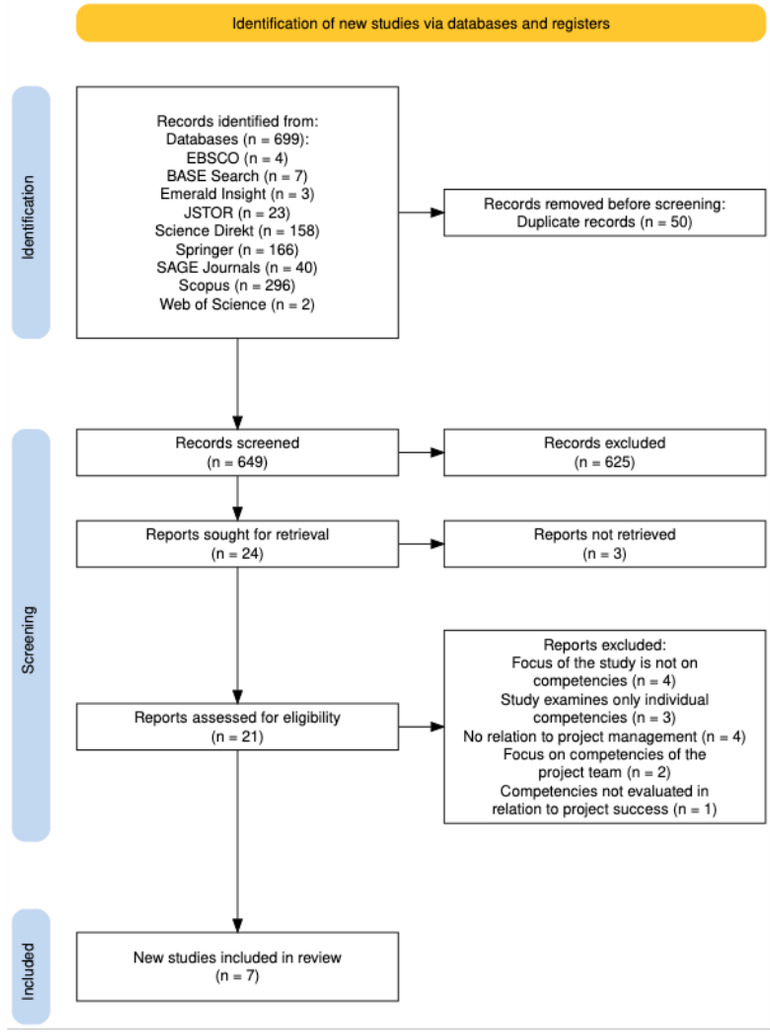


Figure 1. PRISMA Flow Diagram (Source: Own creation based on: Page et al 2021; Haddaway et al. 2022).

DISCUSSION

Analytical evaluation of the included studies

Based on the PRISMA study, seven studies were included in the study and evaluated in detail. Following the systematic data collection, the seven included studies were evaluated individually in a further step. Information about the content and focus of each study was extracted and transferred to a systematic overview. To ensure that this step was also carried out in a systematic manner, categories were first created. With regard to the theoretical basis of this study, the aspects from the introduction, and the research question, the following categories were formed: competency model, leadership competencies, project managers, lack of formal authority, and virtual teams. This structured overview allows the thematic focus and findings from the studies involved to be considered in combination with the research question of this study. The studies involved and their results are shown in the following table (Table 1).

The table should be read as an evaluative matrix rather than as a descriptive list. The authorial assessment applies the five core criteria consistently to each article, so the absence of marks in the columns is itself an analytical finding: it indicates which part of the research question remains unsupported by the selected literature. This is the central result of the review and is the basis for identifying the research gap.

Table 1. Overview of involved studies and results of the article review.

Characteristics and content of each involved article			Results for…				
No.	**Author(s)**	**Short description of the aim and conclusion of the study**	**Competence Model**	**Leadership Competences**	**Project Manager**	**Missing formal authority**	**Virtual teams**
1	Maduka et al.	This study aims to analyze the skills required to lead a virtual team. The authors found that certain leadership competencies are important for successfully managing virtual teams.	x	x			x
2	Anke & Ringeisen	The study examines the competency requirements of project managers in virtual teams. The authors have developed a requirements profile with 20 competency facets and 12 higher-level competency dimensions.	x		x		x
3	Riesen-berger	The aim of this article is to discuss the five competencies necessary for the success of a global project team.			x	x	x
4	Tabassi et al.	The aim of the study was to analyze the role of managers and their leadership style in the construction sector. The authors have developed a model that shows that the leadership skills and transformational leadership qualities of project managers.		x	x		
5	Faisal et al.	The aim of this study is to define a framework for determining skills for project managers in megaprojects. The framework identified by the authors includes 24 leadership competencies, divided into four clusters.	x	x	x		
6	Ribeiro, Amaral & Barros	The study aims to identify the skills that project managers need to work in Industry 4.0. The authors show that the competencies described in the PMBOK and ICB 4.0 are not sufficient to adequately meet the requirements of these digital transformation contexts.			x		
7	Dumitrascu-Baldau & Dumitrascu	The study aims to identify the most important challenges for global project managers and propose a model for the profile of such a project manager.			x		x

Comparing the results, it is obvious at first glance that none of the studies involved considered all five core areas together. Maduka et al. (2018) found that certain leadership competencies are important for the successful leadership of virtual teams. However, they did not examine in detail why exactly the leadership competencies used in the study were used. Furthermore, they did not mention whether the team leader was a project manager and what the authority relationship between this team leader and their virtual project team was. Anke and Ringeisen (2016) summarize their findings in the competency dimension "Leading and Deciding," but do not provide any further information about what this competency dimension involves. Accordingly, although the study provides a comprehensive analysis of the competency requirements for leaders of virtual teams, it does not highlight specific leadership competencies. Tabassi et al. (2016) support Anke and Ringeisen's statement and also say that leadership competencies are crucial for the successful execution of projects. Although the study addressed the leadership competencies of project managers, it did not consider some key areas (competency model, leading without formal authority, and virtual teams). Ribeiro, Amaral, and Barros (2021) conclude that project managers face a number of challenges due to the impact of Industry 4.0 and that the competencies defined in the PMBOK and ICB 4.0 are not sufficient to adequately meet the requirements of Industry 4.0.

The articles presented were at least thematically related to the research focus of the present study. In his article, Riesenberger (2016) primarily addressed possible solutions for the high failure rates of global project teams. As a result, key competencies are defined which, according to Riesenberger, are crucial for project success. Dumitrascu-Baldau and Dumitrascu (2017) concluded in their study that international project managers in virtual working environments can be seen as "managers of the

future." Neither study has any thematic connection to the research focus of this study.

The findings of Faisal et al. (2023) are the only ones that align with the present research direction. The framework developed by the authors includes 24 leadership competencies divided into four clusters. These leadership competencies were identified using focus groups. The result is a competency model that provides a framework for project managers in megaprojects. The study does not mention the authority relationship between these project managers and their project teams. However, Faisal et al. (2023) use the competencies from the PMBOK as a basis. Faisal et al. (2023) cite Skogmar (2015) and Backlund and Sundqvist (2018), who state that, based on the PMBOK, project managers are the sole decision-makers and therefore have responsibility for the personnel employed. Based on this, it can be assumed that the project manager has formal authority over their project team.

The evaluation of the seven studies involved clearly shows that none of the studies considers the core areas of competence models, leadership skills, project managers, lack of formal authority, and virtual teams together.

On this basis, the research gap is not inferred from the mere absence of one exact publication title, but from the systematic mismatch between the available evidence and the full configuration of the research question. The included studies either address virtual leadership, project-manager competencies, or competence frameworks, but they do not integrate these aspects with the absence of formal authority. The author's contribution is the explicit mapping of this mismatch and the delimitation of a future competency model that would have to combine all five criteria.

CONCLUSION

This systematic literature review shows that digitalization has a huge impact and how important virtual teamwork is these days. At the same time, it makes it clear that working in virtual structures comes with some challenges. Because of these challenges and conditions, leaders are becoming more important, as the understanding of leadership, leadership roles, and leadership tools need to adapt to the challenges of virtual work. It also explains that leadership becomes even more complex when managers do not have formal authority, as is often the case in projects.

That's why this systematic review, done using a PRISMA-based approach, wanted to see if there's already a competency model in the literature that describes the leadership competencies of project managers who lead virtual teams without formal authority. From the 699 studies initially identified, seven studies were selected using the PRISMA approach and submitted to a full-text search. Analysis of the studies along the predefined categories shows that none of the studies investigated addresses all five areas simultaneously.

Therefore, the research question can be answered by stating that there is no independent competency model in the literature that describes the leadership competencies of project managers who lead virtual teams without formal authority. The result thus indicates a proven research gap and the need for further studies to develop a specific competency model that maps these core areas.

Future research should therefore move from identifying isolated competence requirements to constructing and validating an integrated competency model for project managers who lead virtual teams laterally. Such a model should specify the leadership competencies, explain their relevance under conditions of virtual cooperation, and account for the influence of absent formal authority on communication, decision-making, trust building, and conflict resolution.

REFERENCES

ANKE, S.; RINGEISEN, T., 2021. Kompetenzanforderungen an Führungskräfte von agilen Softwareentwicklungsteams. *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie (GIO)*, vol. 52, no. 1, pp. 51–63. Available from: https://doi.org/10.1007/s11612-021-00558-w

BACKLUND, F.; SUNDQVIST, E., 2018. Continuous improvement: challenges for the project-based organization. *International Journal of Quality and Reliability Management*, vol. 35, no. 7, pp. 1306–1320. Available from: https://doi.org/10.1108/IJQRM-12-2016-0229

BJÖRK, B.; SOLOMON, D., 2012. Open access versus subscription journals: a comparison of scientific impact. *BMC Medicine*, vol. 10, no. 1, article 73. Available from: https://doi.org/10.1186/1741-7015-10-73

BLOMQUIST, T.; MÜLLER, R., 2006. Middle managers in program and project portfolio management: prac-

tices, roles and responsibilities. *Project Management Journal*, vol. 37, no. 1, pp. 52–66. Available from: https://doi.org/10.1177/875697280603700105
BOHINC, T., 2021. *Führung im Projekt*. Berlin: Springer.
BRINKEN, H., 2021. *10 Gründe für Open Access*. Available from: https://doi.org/10.5281/zenodo.4643858
CAPUTO, A.; TOSCANO, F.; DOLCE, V.; DE ANGELIS, M., 2024. Leadership in face-to-face and virtual teams: a systematic literature review on hybrid teams management. *Informing Science: The International Journal of an Emerging Transdiscipline*, vol. 27, article 8. Available from: https://doi.org/10.28945/5342
DIEFENBACH, S., 2023. Social norms in digital spaces: conflict reports and implications for technology design in the teleworking context. *Zeitschrift für Arbeitswissenschaft*, vol. 77, pp. 56–77. Available from: https://doi.org/10.1007/s41449-022-00332-8
DÖRING, N.; BORTZ, J., 2016. *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. 5th ed. Berlin; Heidelberg: Springer.
DÖRRENBERG, F.; GOERNER, M.; PATZAK, G., 2019. Arbeit in virtuellen Teams. In: GPM DEUTSCHE GESELLSCHAFT FÜR PROJEKTMANAGEMENT E. V., ed. *Kompetenzbasiertes Projektmanagement (PM4). Handbuch für Praxis und Weiterbildung im Projektmanagement*. München: GPM Deutsche Gesellschaft für Projektmanagement e. V., pp. 778–806.
DREAS, S.; KLENK, T., 2021. *Führung und Arbeitsbedingungen in der digitalisierten öffentlichen Verwaltung*. Baden-Baden: Nomos. Available from: https://doi.org/10.5771/9783748926214
DUMITRASCU-BALDAU, I.; DUMITRASCU, D., 2017. Skills and competences international project managers need in order to be successful in a virtual work environment. *Bulletin of the Transilvania University of Brasov. Economic Sciences. Series V*, vol. 10, no. 59, pp. 63–70.
FAISAL, M.; AL SUBAIE, A.; SABIR, L.; SHARIF, K., 2023. PMBOK, IPMA and fuzzy-AHP based novel framework for leadership competencies development in megaprojects. *Benchmarking: An International Journal*, vol. 30, no. 9, pp. 2993–3020. Available from: https://doi.org/10.1108/BIJ-10-2021-0583
FAREED, M.; SU, Q.; AWAN, A., 2021. The effect of emotional intelligence, intellectual intelligence and transformational leadership on project success: an empirical study of public projects of Pakistan. *Project Leadership and Society*, vol. 2, pp. 1–10. Available from: https://doi.org/10.1016/j.plas.2021.100036
FROESE, F.; BLAY, T.; GIBSON, C.; SHAFFER, M.; BENITEZ, J., 2025. Global virtual work: a review, integrative framework, and future research opportunities. *Journal of International Business Studies*, vol. 56, pp. 691–719. Available from: https://doi.org/10.1057/s41267-025-00775-1
GASH, S., 1999. *Effective literature searching for research*. Aldershot: Gower Publishing.
GPM DEUTSCHE GESELLSCHAFT FÜR PROJEKTMANAGEMENT E. V., 2017. *Individual Competence Baseline für Projektmanagement – Version 4.0*. Nürnberg: GPM Deutsche Gesellschaft für Projektmanagement e. V.
GRZES, B., 2024. Determinants of the effectiveness of virtual teams management. In: EL EMARY, I.; BRZOZOWSKA, A.; MASLOCH, P., eds. *Digital Synergy*. London: CRC Press, pp. 108–113. Available from: https://doi.org/10.1201/9781003440406-12
HADDAWAY, N.; PAGE, M.; PRITCHARD, C.; MCGUINNESS, L., 2022. PRISMA2020: an R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. *Campbell Systematic Reviews*, vol. 18, no. 2, e1230. Available from: https://doi.org/10.1002/cl2.1230
HEYSE, V.; ERPENBECK, J., 2009. *Kompetenztraining. Informations- und Trainingsprogramme*. 2nd ed. Stuttgart: Schäffer-Poeschel Verlag.
ILTER, Y.; BARTH-FARKAS, F.; RINGEISEN, T., 2023. Digitale Führungskommunikation und organisationale Bindung von Beschäftigten im Homeoffice. *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie (GIO)*, vol. 54, pp. 259–271. Available from: https://doi.org/10.1007/s11612-023-00676-7
KRUMM, S.; MERIN, I.; DRIES, C., 2012. *Kompetenzmodelle*. Göttingen: Hogrefe.
LESLIE, J.; LUCIANO, M.; MATHIEU, J.; HOOLE, E., 2018. Challenge accepted: managing polarities to enhance virtual team effectiveness. *People Strategy*, vol. 41, no. 2, pp. 22–29.
MADUKA, N.; EDWARDS, H.; GREENWOOD, D.; OSBORNE, A.; BABATUNDE, S., 2018. Analysis of competencies for effective virtual team leadership in building successful organisations. *Benchmarking: An International Journal*, vol. 25, no. 2, pp. 696–712. Available from: https://doi.org/10.1108/BIJ-08-2016-0124
MAKRIDIS, C.; SCHLOETZER, J., 2023. *From crisis to norm: remote work trends and employee engagement across industries, occupations, and geography*. National Bureau of Economic Research. Working Paper No. w33315. Available from: https://doi.org/10.2139/ssrn.4601771
MAYRING, P., 2016. *Einführung in die qualitative Sozialforschung*. 6th ed. Weinheim; Basel: Beltz Verlag.
MIDLER, C., 1995. Projectification of the firm: the Renault case. *Scandinavian Journal of Management*, vol. 11, no. 4, pp. 363–375. Available from: https://doi.org/10.1016/0956-5221(95)00035-T
MORADI, S.; KÄHKÖNEN, K.; AALTONEN, K., 2019. Comparison of research and industry views on project managers' competencies. *International Journal of Managing Projects in Business*, vol. 13, no. 3, pp. 543–572. Available from: https://doi.org/10.1108/IJMPB-04-2019-0085
NAGEL, J., 2025. *Digital Literacy im Projektmanagement. Entwicklung eines Kompetenzstrukturmodells für das digital strukturierte Projektmanagement*. Wiesbaden: Springer Nature. Available from: https://doi.org/10.1007/978-3-658-48282-4

PAGE, M.; MCKENZIE, J.; BOSSUYT, P.; BOUTRON, I.; HOFFMANN, T.; MULROW, C.; SHAMSEER, L. et al., 2021. The PRISMA statement: an updated guideline for reporting systematic reviews. *PLoS Medicine*, vol. 18, no. 3, e1003583. Available from: https://doi.org/10.1371/journal.pmed.1003583
PROJECT MANAGEMENT INSTITUTE (PMI), 2021. *A Guide to the Project Management Body of Knowledge (PMBOK Guide)*. 7th ed. Newtown Square, PA: Project Management Institute.
RIBEIRO, A.; AMARAL, A.; BARROS, T., 2021. Project manager competencies in the context of Industry 4.0. *Procedia Computer Science*, vol. 181, pp. 803–810. Available from: https://doi.org/10.1016/j.procs.2021.01.233
RIESENBERGER, J., 2016. Global project teams: a practitioner's guide to the 5 competencies needed for success. *Rutgers Business Review*, vol. 1, no. 1, pp. 84–101. Available from: https://ssrn.com/abstract=3005424
SHAMIM, M., 2022. The digital leadership on project management in the emerging digital era. *Global Mainstream Journal*, vol. 1, no. 1, pp. 1–14. Available from: https://doi.org/10.62304/jbedpm.v1i1.1
SKOGMAR, K., 2015. *PRINCE2, the PMBOK Guide and ISO 21500:2012*. London: Axelos.
TABASSI, A.; ROUFECHAI, K.; MAHYUDDIN, R.; ABU BAKAR, A.; ISMAIL, R.; PAKIR, A., 2016. Leadership competences of sustainable construction project managers. *Journal of Cleaner Production*, vol. 124, no. 15, pp. 339–349. Available from: https://doi.org/10.1016/j.jclepro.2016.02.076
TORRO, O.; PIRKKALAINEN, H.; LI, H., 2022. Media synchronicity in organizational social exchange. *Information Technology & People*, vol. 35, no. 8, pp. 162–180. Available from: https://doi.org/10.1108/ITP-06-2020-0384
TREAS, S., 2021. *The Remote Work Report 2021*. GitLab Work Study. [viewed 20 September 2025]. Available from: https://about.gitlab.com/resources/downloads/remote-work-report-2021.pdf
WAGNER, R.; OTT, W., 2019. Projekte und Projektmanagement. In: GPM DEUTSCHE GESELLSCHAFT FÜR PROJEKTMANAGEMENT E. V., ed. *Kompetenzbasiertes Projektmanagement (PM4). Handbuch für Praxis und Weiterbildung im Projektmanagement*. München: GPM Deutsche Gesellschaft für Projektmanagement e. V., pp. 50–69.
WAGNER, R.; HUEMANN, M.; RADUJKOVIC, M., 2022. An institutional theory perspective on the role of project management associations for projectification of society: the case of Germany. *International Journal of Managing Projects in Business*, vol. 15, no. 8, pp. 111–134. Available from: https://doi.org/10.1108/IJMPB-09-2021-0259
WAGNER, R.; CAMMANS, I., 2025. Die Rolle des Top-Managements für eine erfolgreiche Projektarbeit. *Projektmanagement aktuell*, vol. 36, no. 3, pp. 16–21. Available from: https://doi.org/10.24053/PM-2025-0040
ZEHRER, A.; MÖSSENLECHNER, C., 2010. Leadership-Kompetenzen in Krisensituationen. In: PECHLANER, H.; RAICH, M.; SCHÖN, S.; MATZLER, K., eds. *Change Leadership. Den Wandel antizipieren und aktiv gestalten*. Wiesbaden: Gabler, pp. 181–209. Available from: https://doi.org/10.1007/978-3-8349-8681-8_8

СИСТЕМАТИЧЕН ЛИТЕРАТУРЕН ПРЕГЛЕД НА ЛИДЕРСКИТЕ КОМПЕТЕНЦИИ НА ПРОЕКТНИТЕ МЕНИДЖЪРИ, ОСНОВАН НА МОДЕЛА PRISMA

Резюме: *Поради дигитализацията екипите все по-често работят виртуално, което прави лидерството сложно, особено когато лидерът няма официална власт над екипа, какъвто често е случаят при проектите. Проектните мениджъри, като лидери на проекта, трябва да се адаптират към нуждите на виртуалното сътрудничество и да притежават определени лидерски компетенции. Настоящото проучване разглежда дали може да се намери компетентностен модел, който изброява лидерските компетенции на проектните мениджъри, които ръководят виртуални екипи без формална власт. Проучването беше проведено чрез систематичен анализ на литературата, базиран на структурата PRISMA. Бяха използвани подходящи търсени изрази и бяха идентифицирани n = 699 статии. След приключване на процеса PRISMA бяха включени седем статии. Резултатите показват, че нито една статия не обхваща основните области на това проучване. Следователно не може да бъде идентифициран модел на компетентност, който да изброява лидерските компетенции на проектните мениджъри, които ръководят виртуален екип без формална власт. Следователно това проучване идентифицира пропуск в изследванията и подчертава необходимостта от по-нататъшни изследвания за разработване на този модел на компетентност.*
Ключови думи: *виртуални екипи; проектен мениджър; лидерски компетенции; липса на формална власт; систематичен литературен преглед*

Тимо Цименга, докторант
Researcher ID (ORCID: 0009-0004-8149-6163)
Университет по библиотекознание и информационни технологии
София, България
E-mail: t.zimenga@googlemail.com

PUBLIC COMMUNICATIONS AND INFORMATION SCIENCES
ОБЩЕСТВЕНИ КОМУНИКАЦИИ И ИНФОРМАЦИОННИ НАУКИ

BRIDGING THE GAP: AN EMPIRICAL ANALYSIS OF INFORMATION MANAGEMENT REALITIES IN RECRUITMENT VS. THEORETICAL AMBITIONS

Volker Joachim Römer
University of Library Studies and Information Technologies
https://doi.org/10.70300/FREX6576

Abstract: *This study addresses the gap between the theoretical development in the field of digital recruitment and its reality in organizations. While literature focuses on the possibilities of Artificial Intelligence (AI) and Big Data, the practical implementation of Information Management (IM) systems is very different. Using the qualitative research method, a total of seven semi-structured expert interviews were carried out with HR professionals from various sectors with multinational corporations, public administration and SMEs. The findings show a large "digital divide": while large enterprises are using integrated ERP systems for data-driven decision-making, SMEs tend to rely on fragmented and manual processes because of their resource constraints. Key novelties of this study include the identification of certain barriers to valid recruitment such as data integration issues and lack of data governance that prevent the effective use of advanced analytics. The findings indicate that before embracing AI, organisations need to focus on basic information management practices to ensure data quality and process consistency. This research adds to the existing literature by presenting its empirical evidence on the state of the IM in recruitment; it offers practical implications for HR practitioners in terms of optimizing their selection strategies.*
Keywords: *information management; personnel selection; recruitment validity; digital divide; HR practices*

INTRODUCTION

The effectiveness of personnel selection processes is an important issue for organizations that want to create high-performing teams. Effective recruitment doesn't just look at filling vacancies, and as Breaugh (2008) established it also encompasses strategic goals such as improved employee retention and job satisfaction. In the past few years, the increased focus on data-driven decision-making has pushed information management to the forefront of these strategies. Current literature suggests that digital transformation specifically the use of Artificial Intelligence (AI) and "Talent Intelligence" has the potential to revolutionize how candidates are selected (Rutten & Verhoeven 2024).

The motivation for a digital transformation is given by the promise of a great efficiency increase. Mask and Pearl (2024) point out the fact that AI-powered tools, such as automated resume screening and chatbots, can reduce the average amount of time spent per candidate from hours to minutes. Technologically speaking, this covers a broad set of applications. As defined by Wilke and Bendel (2022), this spectrum goes from pre-selection chatbots to the automated analysis of application documents.

However, while such theoretical improvements were made, many organizations find the practical application of them challenging. The process of moving from conventional, manual hiring methods to a fully automated process poses formidable challenges (Laumer, Maier & Eckhardt 2015). While the large corporations with multiple branches, multinational corporations may have the resource capabilities to apply Complex Applicant Tracking Systems (ATS) for many applications, the Small and Medium-sized Enterprises (SMEs) may have the limitation of resource capability for technological adoption (Appio et al. 2024). Furthermore, as Stone and Dulebohn (2013) noted, often in the implementation of e-HRM empirical work on its effectiveness is implemented far behind. In a comprehensive review, Potocnik et al. (2021) identify technology as a meta-trend but warn of the lack of rigorous validity evidence.

This paper addresses this gap by investigating the operating reality of the management of information in recruitment. It looks at the current state of affairs: How do organizations actually handle recruitment information? What are the real world barriers to valid and data-driven selection?

THEORETICAL BACKGROUND

In order to understand the difference between the technological potential and reality, it is necessary to analyze the current situation with respect to research on e-HRM and adoption of AI.

Evolution from e-HRM to Talent Intelligence

The digitization of human resources has changed a lot. Initially, Stone and Dulebohn 2013, explained the shift towards electronic Human Resource Management (e-HRM) as the mean to reduce administrative burdens and costs. Today, the focus has become the use of "Talent Intelligence" which, according to Rutten and Verhoeven (2024), is the strategic use of data to predict future skill gaps and the automation of decision-making. However, this evolution postulates an achievement of a level of data maturity that is not yet available for many organizations.

Barriers to Adoption in SMEs

While large multinational corporations often have the resources in place to implement sophisticated Applicant Tracking Systems (ATS), Small and Medium-sized Enterprises (SMEs) are faced with very unique challenges. Appio et al. (2024) note that SMEs are frequently constrained by small amounts of financial and human resources, which make the implementation of complicated digital ecosystems difficult. Furthermore, Baum and Kabst (2012) state that SMEs's recruit employees through informal channels (e.g., employee referrals) that are less digitalized than standardized job postings.

The Part Played by Process Management

One of the key insights presented by Laumer, Maier and Eckhardt (2015) is that technology is not an independent solution. They maintain that Business Process Management (BPM) - the documentation, analysis and optimization of business flows - must be done before technological implementation. Without this foundation digital tools often will not show the promised efficiency increase.

RESEARCH METHODOLOGY

Semi-structured expert interviews with N=7 HR professionals were the source of data. The interview guide was modeled in such a way that it addresses the important areas of information management such as system usage, quality of data and perceived validity. The stratified sampling strategy was used in selecting the sample to provide maximum variation in the variables of organizational size, industry and structure. This enables one to make a comparative analysis of highly regulated settings (public sector) and agile or resource-constrained organizations (SMEs).

The overview of the interview partners (anonymized) is presented in Table 1:

Table 1: Overview of Interview Participants

ID	Sector	Organization Type	Role
I1	Public Administration	Large City Administration (Vienna)	HR Manager
I2	Public Administration	Large City Administration (Zurich)	HR Professional
I3	Automotive	Multinational Corporation	HR Specialist
I4	Education	Public Kindergarten	Management and HR Lead
I5	IT Services	Mid-sized IT Provider	HR & Talent Lead
I6	Transport	Bus Transport Company	HR Manager
I7	Non-Profit / Religious	Church Diocese	HR Administrator

Data Analysis Procedure

The interviews were transcribed and analyzed by qualitative content analysis based on Mayring (2002). The analysis was carried out in the form of a deductive-inductive approach: initial categories were taken from the literature (e.g. "efficiency," "validity"), while sub-categories (e.g. "data silos," "gdpr concerns") emerged directly from the interview material.

RESULTS

The Technological Divide

The findings show a distinct difference in the situations of large organizations and SMEs. Interviewees from the large public administrations and the automotive sector reported on the use of sophisticated and integrated systems such as SAP SuccessFactors. These systems enable a high level of automation, ranging from applicant tracking to onboarding (Interview 2, Interview 3). On the other hand, smaller organizations use basic tools. For example, the kindergarten association mainly uses email and cloud storage (Dropbox) for the management of applicant data and processes evaluation manually (Interview 4).

While Wilke and Bendel (2022) depict a sophisticated stage of digital recruiting based on artificial intelligence matching, our empirical data indicate that most organizations are still in the process of basic digitization. Only the IT service provider (Interview 5) expressed ambitions to implement advanced "intelligent agents", confirming that widespread adoption of AI is a theoretical vision for many.

Challenges of integration of Data and Quality.

A theme that ran throughout the interviews was the difficulty of data integration. Even when ATS solutions are in place, they are often not connected in a seamless way. The bus transport company (Interview 6) said they had challenges in merging their recruitment software with outside job boards, causing data inconsistencies. Similarly, the IT company (Interview 5) presented the challenge of trying to integrate data gathered from different subsidiaries into a holistic system.

Data Privacy and Security as a Double Edged Sword

A key finding from all the interviews was the influence of data privacy regulations (specifically GDPR) on the adoption of new technologies. Interviewees from bigger organizations (e.g., Interview 2, City of Zurich; Interview 5, IT Services) stressed that strict compliance rules tend to slow down implementation of innovative tools. For example, the integration of external AI tools is often obstructed or slowed down by internal data protection officers because of concerns about data sovereignty. Conversely, smaller organizations often have problems technically implementing security standards. While they do know what needs to be done, there are no sophisticated ATS roles and permission concepts, so there are workarounds. However, Wilke and Bendel (2022) claim that data autonomy is one of the most important ethical requirements. Our data shows that while large firms solve this via rigid systems ("Security by Design"), smaller firms rely on trust and manual deletion of data which poses potential compliance risks and limits the usability of data for long-term analytics.

Competence and Training Shortfalls

Another obstacle to successful information management from the interviews is the lack of digital competence of HR staff. The HR manager from the bus transport company (Interview 6) added that when implementing new interfaces in the system, constant readjustment and technical understanding that is not always available in traditional HR roles is required. Similarly, the IT service provider (Interview 5) emphasized the need for "data literacy" - the ability to read and interpret data on the dashboard - as a skill that must be trained. This attests to Jäger (2018) finding that the shift to "Recruiting 4.0" brings with it a fundamental upskilling of the HR workforce. Without this competence, even the most sophisticated information systems are used under-utilized because users return to well-known, manual processes.

The Effect of Data Analytics and Validity

While there is much literature advocating for predictive analytics (Bhushan 2024), the empirical outcomes reveal a restricted use of predictive analytics in practice. Most interviewees use descriptive statistics as opposed to predictive metrics. Furthermore, the validity of selection decisions in smaller organizations is often based on subjective "gut feeling" and cultural fit as opposed to standardized data points (Interview 7).

CONCLUSIONS/DISCUSSION

This study examines the significant gap between the theoretical development of digital recruitment that is frequently mentioned in literature and operational reality that is present in organizations today. The analysis shows that the so-called "digital transformation" in recruitment is not an equal evolution that affects all companies equally. Instead, it is a fragmented landscape where there is a major digital divide between large corporations and smaller entities.

The Innovation Paradox and the reality of operations

Our results take the form of what Rutten and Verhoeven (2024) call the "innovation paradox." This term describes the situation that many organizations have this dream to implement high-end technologies such as "Talent Intelligence" and prediction models that use artificial intelligence, but often overlook the basics that must be operated to get them to run. Specifically, companies often do not have good data structures or consistent workflows.

This disparity, but especially noticeable in the case of large enterprises compared to SMEs, is explained by the hypothesis of "joint impact" proposed by Laumer et al. (2015). Their research argues that the full potential of recruitment technology is only realized when implementation of IT is accompanied by rigorous Business Process Management (BPM). In more simple words: technology cannot repair a broken process. Our interviews confirm this theoretical assumption: large organizations with dedicated process owners (e.g., Interview 2, City of Zurich) are able to derive value from their systems because they have standardized their workflows. In contrast, in smaller organizations where the process governance is not defined, the use of digital tools is not fully optimized, leading to workarounds that are often manual, and data inconsistencies.

Resource Constraints and Strategic Fit in SMEs

The challenges facing Small and Medium-sized Enterprises (SMEs) are very different compared to the ones faced by multinational corporations. As Appio et al. point out (2024), SMEs are often struggling to succeed because of their limited financial and human resources, which means they cannot invest in costly software ecosystems. Our findings corroborate this "resource scarcity": although the market is full of high-tech solutions, smaller organizations (e.g. Interview 4, Kindergarten) often choose pragmatic and cheap solutions such as cloud storage or simple email management.

It is very important to realize that SMEs are not just smaller versions of large enterprises and this is emphasized by Baum and Kabst (2012). Their use of informal channels and subjective assessment of "cultural fit" is not necessarily a deficit. Instead, it is a strategic adaptation to their own specific context. This approach often produces high candidate quality from the use of personal networks, despite the lower technological maturity of the process.

The Conflict between Compliance and Innovation

Our results point to a tension between wanting to be innovative and the need to comply. As noted by Bujold et al. (2024), responsible AI calls for the most rigorous observance of privacy rules. In our study this adherence often acts as a brake on innovation especially in the public sector. Organizations are facing a dilemma; they would like the efficiency of AI (as described by Mask & Pearl 2024), but they would not use cloud-based or "black box" solutions because of their strict interpretation of data privacy requirements. This implies that for European organizations, "GDPR-compliance" is a more crucial success factor of software selection than the ability to predict.

Ethical Considerations and Human Supervision

A significant barrier that has been identified at all organizational sizes is a lack of trust in automated decisions, sometimes called the "black box" problem (Mask and Pearl 2024). This metaphor is used to explain the challenge of understanding how an algorithm comes to make a particular decision (e.g., dismiss a particular candidate). Interestingly, this aversion to fully automated decision-making corresponds to issues of ethics raised by Hunkenschroer and Luetge (2022), in terms of "obfuscation of accountability". If an AI makes a mistake, nobody knows who is responsible.

Our participants' insistence for human supervision is a validating action for the principle of "Human-centered AI" as identified by Bujold et al. (2024). In their opinion, humans need to be kept in the loop to ensure fairness and accountability. Furthermore, Tippins (2015) is warning that

digital testing, particularly if it is done without supervision (unproctored), can introduce new risks for validity, such as cheating or technical glitches. This suggests the caution that we see in our sample, in favor of checking manually rather than blindly using automation, is scientifically justified.

CONCLUSION

This research paper set out to address a gap between the promise of AI in recruitment and the reality of HR professionals. By examining the qualitative data of seven expert interviews, this study presents empirically sound arguments that robust information management as a foundation prerequisite is a valid personnel selection basis. The main conclusion is that technology is an enabler, not a solution in itself, without the standardization of underlying processes and data governance, advanced tools fail to provide validity.

Practical Implications

Based on the empirical findings, this study makes specific contributions for HR practitioners and organizational leaders:

1. **Data Governance First:** Data Governance First: Organizations need to focus more on the quality of their data rather than on the acquisition of new tools. As highlighted by *Stone and Dulebohn (2013)*, efficiency is often the main reason for e-HRM, but efficiency without accuracy is useless. Practitioners should define some clear data governance frameworks before AI algorithms is put to work.
2. **Process Standardization:** Technology cannot solve the problems of a broken process. Following *Laumer et al. (2015),* before digitizing their recruitment workflows, companies should first map and optimize the workflows.
3. **Strategic Selection:** Instead of attempting to court "hype" technologies, organizations should select tools that match their maturity level (Jäger 2018). For smaller companies, simple but well-setup ATS can provide more value than complicated analytics suites.
4. **Human Element:** To help overcome anxiety in the applicant, and to ensure production is conducted ethically, the role of the human recruiter still plays a central role. Automated tools should be placed in a role of decision-support systems to ensure accountability.

Directions for Future Work

Whilst this study gives us deep qualitative insights, future research should seek to quantify these results, and over a larger sample size. Longitudinal studies could monitor the effect of the introduction of structured information management systems on the quality of the long-term hire and retention of employees. Furthermore, research should focus on developing specific, scalable frameworks of information management geared to the resource constraint of SMEs to help them bridge the digital divide without the need for enterprise level budgets.

REFERENCES

APPIO, F. P.; LEONE, M. I.; TURCAN, R. V., 2024. Open innovation at the digital frontier: Unraveling the paradoxes and roadmaps for SMEs' successful transformation. *European Journal of Innovation Management*, vol. 27, no. 9, pp. 223–247.

BAUM, M.; KABST, R., 2012. Die Wirkung von informellen und internetbasierten Rekrutierungskanälen auf den Rekrutierungserfolg: Eine empirische Analyse in kleinen und mittleren Unternehmen. *Zeitschrift für Betriebswirtschaft*, vol. 82, suppl. 3, pp. 117–142.

BHUSHAN, P., 2024. The digital transformation in recruitment: Exploring the effect and challenges of online recruitment. *Management Journal for Advanced Research*, vol. 4, no. 3, pp. 22–27.

BREAUGH, J. A., 2008. Employee recruitment: Current knowledge and important areas for future research. *Human Resource Management Review*, vol. 18, no. 3, pp. 103–118.

BUJOLD, A.; ROBERGE-MALTAIS, I.; PARENT-ROCHELEAU, X.; BOASEN, J.; SÉNÉCAL, S.; LÉGER, P.-M., 2024. Responsible artificial intelligence in human resources management: A review of the empirical literature. *AI and Ethics*, vol. 4, pp. 1185–1200. Available from: https://doi.org/10.1007/s43681-023-00325-1

HUNKENSCHROER, A. L.; LUETGE, C., 2022. Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics*, vol. 178, pp. 977–1007. Available from: https://doi.org/10.1007/s10551-022-05049-6

JÄGER, W., 2018. Recruiting 1.0–4.0: Strategien, Prozesse und Systeme im Wandel der Zeit. B: KOCHHAN, C.; MOUTCHNIK, A., ed. *Media Management*. Wiesbaden: Springer Fachmedien, pp. 1–27.

LAUMER, S.; MAIER, C.; ECKHARDT, A., 2015. The impact of business process management and applicant tracking systems on recruiting process performance: An empirical study. *Journal of Business Economics*, vol. 85, pp. 421–453. Available from: https://doi.org/10.1007/s11573-014-0758-9

MASK, E.; PEARL, J., 2024. *Artificial Intelligence in Human Resources: Ethical Implications and Performance Enhancement* [online]. ResearchGate. Available from: https://doi.org/10.13140/RG.2.2.35360.39683. Also available from: https://www.researchgate.net/publication/381999512

POTOČNIK, K.; ANDERSON, N. R.; BORN, M.; KLEINMANN, M.; NIKOLAOU, I., 2021. Paving the way for research in recruitment and selection: Recent developments, challenges and future opportunities. *European Journal of Work and Organizational Psychology*, vol. 30, no. 2, pp. 159–174. Available from: https://doi.org/10.1080/1359432X.2021.1904898

RÜTTEN, M.; VERHOEVEN, T., 2024. *Recruiting Analytics: Mehr Erfolg mit Data Driven Recruiting und Talent Intelligence*. Stuttgart: Schäffer-Poeschel.

STONE, D. L.; DULEBOHN, J. H., 2013. Emerging issues in theory and research on electronic human resource management (eHRM). *Human Resource Management Review*, vol. 23, no. 1, pp. 1–5. Available from: https://doi.org/10.1016/j.hrmr.2012.06.003

TIPPINS, N. T., 2015. Technology and assessment in selection. *Annual Review of Organizational Psychology and Organizational Behavior*, vol. 2, pp. 551–581. Available from: https://doi.org/10.1146/annurev-orgpsych-032414-111317

WILKE, M.; BENDEL, O., 2022. KI-gestütztes Recruiting: Technische Grundlagen, wirtschaftliche Chancen und Risiken sowie ethische Aspekte. *HMD Praxis der Wirtschaftsinformatik*, vol. 59, pp. 647–666. Available from: https://doi.org/10.1365/s40702-022-00858-2

ПРЕОДОЛЯВАНЕ НА РАЗЛИКАТА: ЕМПИРИЧЕН АНАЛИЗ НА РЕАЛНОСТИТЕ В УПРАВЛЕНИЕТО НА ИНФОРМАЦИЯТА ПРИ ПОДБОРА НА ПЕРСОНАЛ В СРАВНЕНИЕ С ТЕОРЕТИЧНИТЕ АМБИЦИИ

Резюме: *Това проучване разглежда разликата между теоретичното развитие в областта на цифровия подбор на персонал и реалността в организациите. Докато литературата се фокусира върху възможностите на изкуствения интелект (AI) и големите данни, практическото приложение на системите за управление на информацията (IM) е много по-различно. С използване качествен метод на изследване бяха проведени общо седем полуструктурирани интервюта с експерти – специалисти по човешки ресурси от различни сектори, работещи в мултинационални корпорации, публичната администрация и МСП. Резултатите показват голяма „цифрова пропаст“: докато големите предприятия използват интегрирани ERP системи за вземане на решения на базата на данни, МСП са склонни да разчитат на фрагментирани и ръчни процеси поради ограничените си ресурси. Ключовите новости в това проучване включват идентифицирането на определени пречки за валиден подбор на персонал, като проблеми с интегрирането на данни и липса на управление на данните, които възпрепятстват ефективното използване на усъвършенствани аналитични методи. Резултатите показват, че преди да преминат към AI, организациите трябва да се съсредоточат върху основни практики за управление на информацията, за да гарантират качеството на данните и последователността на процесите. Това проучване допълва съществуващата литература, като представя емпирични данни за състоянието на IM при подбора на персонал; то предлага практически импликации за специалистите по човешки ресурси по отношение на оптимизирането на техните стратегии за подбор.*

Ключови думи: *управление на информацията; подбор на персонал; валидност на подбора на персонал; цифрово разделение; практики в областта на човешките ресурси*

Фолкер Йоаким Рьомер, докторант
Университет по библиотекознание и информационни технологии
София, България
E-mail: roemer.volker@gmx.de

PUBLIC COMMUNICATIONS AND INFORMATION SCIENCES
ОБЩЕСТВЕНИ КОМУНИКАЦИИ И ИНФОРМАЦИОННИ НАУКИ

DEVELOPMENT OF AN INTEGRATIVE VALIDATION MODEL FOR RECRUITMENT: BALANCING AI EFFICIENCY, PREDICTIVE ANALYTICS AND ETHICAL RESPONSIBILITY

Volker Joachim Römer
University of Library Studies and Information Technologies
https://doi.org/10.70300/PAAW5405

Abstract: *The process of recruitment is experiencing a basic digital transformation, with a shift away from traditional recruitment techniques towards AI-enabled decision-making. While information systems like Applicant Tracking Systems (ATS) and Recommender Systems make the process much more efficient and cost-effective, they create new risks in terms of algorithmic bias and non-transparency. This paper is an attempt to propose an "Integrative Validation Model" that extends the traditional notion of predictive validity. It argues that modern recruitment tools must be validated not only by their ability to predict job performance (efficiency) and that they should also comply with ethical standards (responsibility). The methodology is grounded on a systematic literature review of recent studies between 2015 and 2024. The findings suggest that to be sustainable recruitment success requires incorporating "Responsible AI" principles into the information management process. This approach supports a holistic approach that ensures that organizational goals are aligned with social responsibility.*
Keywords: *recruitment validation; information management; applicant tracking systems; AI ethics; algorithmic bias*

INTRODUCTION

The face of human resource management is undergoing a paradigm shift today, thanks to the explosive developing technologies. As Bhushan (2024) notes, recruitment is no longer a simple, administrative task, but rather a complex and multi-dimensional information management process. In the past, recruitment used to be characterized by manual screening of physical documents and subjective decision-making. Today, it is characterized by "E-Recruitment" which is the use of digital platforms, automated screening tools, and data analytics to recruit and select candidates worldwide.

This "Digital Transformation" is not simply a change of tools, but a change of strategy. Organizations are facing more pressure to shorten the "Time-to-Hire" and reduce the "Cost-per-Hire" while making the candidates of higher quality. Potocnik et al. (2021) point that this pressure has resulted in the widespread use of algorithmic assessment tools. These tools promise to measure applicant knowledge, skills, abilities and other characteristics (KSAOs) with a speed and consistency that human recruiters cannot provide.

However, this need for automatic processing of information generates a crucial tension. While efficiency is being driven up the validity of the underlying decisions is often questioned. Laumer, Maier, and Eckhardt (2015), however, argue that despite the benefits Business Process Management (BPM) systems have in streamlining flow processes, in no way do they guarantee the quality of the hiring decision. Furthermore, there are new risks posed by the introduction of Artificial Intelligence (AI). Mask and Pearl (2024) warn of the data bias of AI systems which are frequently trained on past data that contain the biases of human decisions. If an algorithm learns from a history in which certain demographics were systematically disadvantaged, then it will be automated and enforced.

Consequently, traditional models of validation, in which the emphasis has been on "predictive validity" (the statistical correlation between a test score and job performance), are no longer adequate. A statistically predictive, but unethical recruitment process is a severe legal and reputational risk to the organization.

This paper addresses this gap by proposing an Integrative Validation Model. This model proposes that validity in the digital age needs to be three-dimensional in nature: that it must include Technical Efficiency (speed and cost), Predictive Accuracy (quality of hire), and Ethical Responsibility (fairness and transparency). Through a synthesis of the latest literature, the project aims to offer a framework for the validation of recruitment procedures that meet both the economic needs of the business and the ethical standards of society (Bujold et al. 2024).

RESEARCH METHODOLOGY

This publication uses qualitative research methodology based on systematic literature review. The goal was to examine the intersection of information systems, recruitment analytics, and business ethics to obtain an overall validation framework.

Data Collection and Data Selection

The review focuses on academic papers, specialist books, and systematic reviews published from the period 2015 to 2024 mainly. This timeframe has been chosen in a way that it encompasses the latest developments in the world of technologies like the coming of Generative AI, state-of-the-art Recommender Systems, etc. The selection process was done in a organized manner looking for important sources dealing with 3 basic themes:

1. Information Systems in HR: Studies analyzing the impact of ATS and BPM on process performance (e.g. Laumer et al. 2015).

2. Algorithmic Decision-Making: Technical papers of the the functioning of recommender systems and AI algorithms function in a recruitment context (e.g., Freire & De Castro 2021).

3. Ethics and Bias: Review of algorithmic discrimination & responsibility of AI frameworks (e.g. Kochling and Wehner 2020; Hunkenschroer and Luetge 2022).

Data Analysis

The selected literature has been analyzed based on the thematic synthesis approach. Instead of a statistical meta-analysis, the study is concerned with the extraction of conceptual frameworks and empirical findings with the goal of constructing the Integrative Validation Model. Special attention was paid to "Review of Reviews" (e.g., Potocnik et al. 2021) to ensure a broad coverage of current state of research without having to rely on single, potentially isolated case studies.

THEORETICAL FRAMEWORK

In order to gain an understanding of the need for an integrative validation model, it is important to first examine the underlying technological infrastructure, as well as the theoretical development of recruitment from a manual task to a data-driven science.

Information Systems Used in Recruitment

The backbone of modern recruitment is constituted by sophisticated information systems to manage the flow of data of candidates. Laumer, Maier and Eckhardt (2015) describe Applicant Tracking Systems (ATS) not as simple databases but as full-fledged Business Process Management (BPM) tools. These systems standardize the entire lifecycle of recruitment - from job posting to onboarding. From an information science point of view, an ATS is like a central storage which guarantees data consistency and allows process automation. The primary theoretical argument for their adoption is the reduction of "media disruptions" which occur when data is manually travelled between systems (e.g. email to spreadsheet) and are therefore prone to error and inefficiency.

Furthermore Freire and De Castro (2021) point on the rise of Recommender Systems (RS) in these platforms. Borrowing from technologies developed in e-commerce, the recommenders for e-recruitment use two primary filtering techniques to validate candidates:

- Content-Based Filtering: In this method, the semantic content of a CV (keywords, skills, experience) is analyzed, and it is matched with the job description. It assumes that the higher the overlap of keywords, the higher the validity given to the candidate for the role.
- Collaborative Filtering: This is a more complicated method that predicts the suitability of a candidate based on the behavior of other similar people or past hiring patterns. If Candidate A has a similar profile to successful Employee B, the system deduces a high probability of success.

Talent Intelligence Predictive Analytics

The conceptual shift from operational to strategic decision making is conceptualized as "Talent Intelligence". Rutten & Verhoeven (2024) propose a model of maturity of recruitment analytics that organizations need to go through in order to obtain valid results:

- Descriptive Analytics: The first level, concerned with what happened (e.g. "We received 500 applications"). While necessary for reporting, this does not provide any validation of quality.
- Diagnostic Analytics: Deals with why it happened (e.g. "Why did the hire rate drop?").
- Predictive Analytics: The critical stage for validation. Here, the historical data is used to predict what will happen in the future (e.g. "Candidates with this skill set have a 90% retention probability").
- Prescriptive Analytics: This is the highest level where the system makes specific recommendations (e.g. "Interview this candidate immediately to prevent drop off").

For an integrative validation model, the shift to predictive and prescriptive analytics is important. However, Rutten and Verhoeven (2024) claim that validity of these predictions is without separation linked to data quality. Inconsistent data entry or unstructured feedback in the ATS causes "noisy" data which makes predictive models invalid.

The Idea of the Algorithmic Validity

Traditionally, validity in recruitment was described through psychometric criteria – mainly "criterion-related validity", which is a measure of the correlation between a test score and job performance (Potocnik et al. 2021). In this era of AI, this definition needs to be extended to include Algorithmic Validity.

Kochling and Wehner (2020) say that an algorithm is only valid if it is free from systematic bias. Unlike individual and sporadic human bias algorithmic bias is systematic and scalable. If a machine learning model is trained on a dataset where a certain demographic was historically at a disadvantage (say, women not being hired for tech positions in the past) the algorithm will look for this as a "valid" predictor of unsuitability. Therefore, in the context of information management validity must include not only the statistical accuracy of the prediction (does it predict the outcome?), but the ethical integrity of the data processing (is the prediction based on merit or bias?).

RESULTS

The systematic analysis of the selected literature reveals three main dimensions influencing the validity of modern recruitment procedures, namely Process Efficiency through Information Systems, the steering function of Data Analytics and a critical challenge of Algorithmic Bias.

1. Efficiency through Information Systems and Standardization

The empirical evidence validates that the integration of Information Technology (IT) and Business Process Management (BPM) does not only mean the digitization but the fundamental restructuring of recruitment. Laumer, Maier, and Eckhardt (2015) in a quantitative study of leading companies showed that the use of Applicant Tracking Systems (ATS) and rigorous process analysis beats cycle times and process costs to a huge degree. Their findings show that efficiency gains are not gained through software but through the standardization of the work flows that the software imposes.

This minimizes disruptions to the media and maintains consistency of data. Furthermore, Freire and De Castro (2021) identify Recommender Systems as an important technological solution to cope with the "information overload" in e-recruitment. By using content-based and collaborative filtering techniques, these systems can filter through thousands of profiles in no time. This automation enables recruiters to concentrate on the best possible candidates and therefore makes matching candidates to job openings much faster and accurate.

2. The Steering Function of Recruiting Analytics

Modern recruitment is no longer viewed as a reactive administrative process but as a strategic, data-driven operation, and is sometimes referred to as "Recruiting Analytics" or "Talent Intelligence". Rutten and Verhoeven (2024) suggest that, in order for recruitment to be valid, there is a need to move "beyond simple descriptive reporting (e.g. how many applicants?)" towards

predictive analytics (e.g. who will likely succeed? They point out that Key Performance Indicators (KPIs) like "Time-to-Hire", "Cost-per-Hire" and "Quality-of-Hire" must be monitored in real-time to drive the talent acquisition strategy effectively.

However, the validity of these analytics has a strict dependence on the quality of the underlying data. If the input data is inconsistent or incomplete the predictive models will give invalid results. Therefore, data management must have strong data governance for analytics to be a reliable tool for decision making.

3. Algorithmic Bias and the "Amazon Case Study"

A great and important finding, included into the recent literature, is the prevalence of the biasness in the case of algorithmic decision making, which contributes to the loss of validity of the recruitment procedures. Mask and Pearl (2024) argue that although AI tools are efficient, they tend to suffer for lack of transparency – it is a "black box". This circumstance makes it hard to validate whether a decision was made on merit or whether it was based on spurious correlations.

This theoretical risk is supported by hard empirical evidence. Kochling and Wehner (2020) have a detailed analysis of the prominent Amazon case study. In this case, an experimental hiring algorithm developed by Amazon taught itself to penalize resumes that contained the word "women's" (e.g. "women's chess club"). This error occurred not because the algorithm was programmed in a sexist way, but because it was trained on 10 years of historic hiring data, which was dominated by male resumes.

The system was statistically validating the correlation between "being male" and "being hired" in the past and projecting it into the future. This case is a vivid example of the peril in relying only on predictive validity: The system was statistically correct in its ability to reproduce the patterns of the past, but invalid from an ethical and legal standpoint. Consequently, it is suggested in the literature that technical validation of information systems must always be accompanied by ethical auditing in order to detect such "historical biases" before they are automatized.

CONCLUSIONS AND DISCUSSION

This study proposes Integrative Validation Model of recruitment. This model assumes that recruitment procedure can only be "valid" if it meets three criteria at once: efficient (IT-backed), predictive (Data-driven) and ethical (Bias-free). The result of the analysis is that when only one aspect is focused on, it is doomed: Efficiency without ethics is discrimination; ethics without efficiency is competitive disadvantage.

Integration of Responsible Basic Artificial Intelligence Settings

The results indicate that companies simply cannot stop with technical validation. Bujold et al. (2024) suggest the implementation of "Responsible AI" principles in Human Resource Management. This suggests that validation is not a once in a lifecycle event but rather a continuous process.

- Transparency: Organizations need to be able to explain how their algorithms work. "Black Box" systems must be avoided when it comes to high-stakes decisions such as hiring;
- Human Oversight: As suggested by Hunkenschroer and Luetge (2022), it is always a good idea to ensure that the final decision of hiring someone is always left up to a human ("Human-in-the-loop"). The AI is supportive in delivering information but not the final word.

Implementation Strategy: The Need for Parallel Testing

Given the identified risks of algorithmic bias and error, organizations should be extremely careful when ascribing new AI tools. It is strongly recommended that human decision-making is not replaced suddenly. Instead a parallel implementation strategy (or "Shadow Mode") should be adopted.

In this phase, the AI system is used in concert with traditional processes for evaluating a system, but it does not make final decisions. For example, recruiters use manual CV screening and the AI runs in the background. After a set period of time, the results are compared: Did the AI reject highly rated candidates as rated by humans?

Were demographics shown to be biased by this AI? Only in case the AI can prove valid and

unbiased results over a defined time of testing, it should be incorporated into the active workflow. This is to ensure that the information management system is not validated in an artificial environment before affecting real-world hiring decisions.

Future Challenges The work results of AI generated output

Looking forward, the rise of Generative AI (e.g. ChatGPT) brings new issues to information management. Hagendorff (2024) charts the ethics of these powerful models and draws attention to the risks of such models, including the danger of them hallucinating (inventing facts) and producing harmful or false content.

For recruitment and this means new validation risks. If candidates use AI to polish their CVs, and companies use AI to screen them, the whole process may become detached from reality. Furthermore, if Generative AI is used to write job descriptions, it could potentially use biased language. Therefore, validation models must move to the dynamic. Organizations need to constantly audit not only the result (who is hired) but the process of generating information itself.

Implications for Practice

The shift to an integrative validation model has huge practical implications for human resource professionals and organizational leaders. This research proves that having an Applicant Tracking System (ATS) or an AI-based screening tool in place is not enough to guarantee valid recruitment results. Instead, these technologies need active, ethical and ongoing management in order to avoid automating bias, and ensure compliance with the law.

1. Algorithmic Auditing Protocols

Organizations have to change their basic thought process fundamentally from a passive approach to distrusting technology to an active approach of verifying the technology. As suggested by Hunkenschroer and Luetge (2022), it is imperative for companies to introduce a formal "Algorithmic Auditing" process as an integral and permanent part of their information management strategy.

In practice, this is a stringent approach that requires a comprehensive evaluation strategy commencing long before the technology is in active use. Therefore, a critical "Pre-Deployment Audit" is needed whereby new AI tools need to be rigorously tested for "disparate impact" on the basis of historical data sets before being integrated into the live recruitment workflow.

During this phase, HR managers are responsible for checking whether the tool's recommendations are consistent with organizational objectives for diversity, or whether the system in some way systematically discriminates against certain demographic groups – e.g. by unfairly penalizing gaps in CVs that may correlate with parental leave.

However, validation suggests more than a one-time approval - it should be considered as a continuous and dynamic process which is known as "Post-Deployment Monitoring". This constant watch is necessary because predictive models are prone to "concept drift," a phenomenon in which the accuracy of the predictive model may degrade with time as the job market and applicant behaviors change. To fight this loss of validity is necessary regular spot-checks, in which human recruiters manually read through applications that were rejected by the AI. Detecting these "False Negatives" is critical to ensure that the system is valid, fair and accurate over the long term.

2. Improving the skills gap: Data competence in HR

As Potocnik, et al. (2021) conclude, the field needs more sophisticated psychological and technological theories. For practitioners, this suggests a needed change in competencies. Rutten and Verhoeven (2024) point out that modern recruiters need to have a high level of "Data Literacy". They must be made to understand that the "matching score" of 85% that an AI might generate is not an absolute truth, but a statistical probability based on certain input variables. Training programs must therefore be concerned not only with how to operate the software but how to interpret critically its outputs.

HR professionals need to be given the power to override algorithmic suggestions if their own judgement, informed by their human empathy and contextual understanding, flags up a valid candidate the system missed.

3. Interdisciplinary Collaboration

Finally, to effectively manage information in recruitment, the organizational silos should be broken down. Bujold et al. (2024) highlight that responsible AI is not something that can be accomplished by the HR department alone. It demands an interdisciplinary approach brought together by IT specialists (knowledge of the code), legal experts (knowledge of compliance) and HR psychologists (knowledge of job fit).

Only by this collaborative effort can the organizations achieve the right balance between the efficiency gains of digital transformation (Bhushan 2024) while still adhering to strict ethical standards in order to keep a "social license to operate." Companies who are successful in this balancing act will have achieved true validity in their recruitment procedures - which is fast, cost effective, and fundamentally fair

REFERENCES

BHUSHAN, P., 2024. The Digital Transformation in Recruitment: Exploring the Effect and Challenges of Online Recruitment. *Management Journal for Advanced Research*, vol. 4, issue 3, pp. 22–27. DOI: 10.54741/mjar.4.3.4.

BUJOLD, A.; ROBERGE-MALTAIS, I.; PARENT-ROCHELEAU, X.; BOASEN, J.; SÉNÉCAL, S.; LÉGER, P.-M., 2024. Responsible Artificial Intelligence in Human Resources Management: A Review of the Empirical Literature. *AI and Ethics*, vol. 4, pp. 1185–1200. DOI: 10.1007/s43681-023-00325-1.

FREIRE, M. N.; DE CASTRO, L. N., 2021 e-Recruitment Recommender Systems: A Systematic Review. *Knowledge and Information Systems*, vol. 63, issue 1, pp. 1–20. DOI: 10.1007/s10115-020-01522-8.

HAGENDORFF, T., 2024. *Mapping the Ethics of Generative AI: A Comprehensive Scoping Review* [online]. University of Stuttgart / Interchange Forum for Reflecting on Intelligent Systems. [viewed 23 November 2025]. Available from: https://arxiv.org/abs/2306.05949

HUNKENSCHROER, A. L.; LUETGE, C., 2022. Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda. *Journal of Business Ethics*, vol. 178, pp. 977–1007. DOI: 10.1007/s10551-022-05049-6.

KÖCHLING, A.; WEHNER, M. C., 2020. Discriminated by an Algorithm: A Systematic Review of Discrimination and Fairness by Algorithmic Decision-Making in the Context of HR Recruitment and HR Development. *Business Research*, vol. 13, pp. 795–848. DOI: 10.1007/s40685-020-00134-w.

LAUMER, S.; MAIER, C.; ECKHARDT, A., 2015. The Impact of Business Process Management and Applicant Tracking Systems on Recruiting Process Performance: An Empirical Study. *Journal of Business Economics*, vol. 85, pp. 421–453. DOI: 10.1007/s11573-014-0758-9.

MASK, E.; PEARL, J., 2024. *Artificial Intelligence in Human Resources: Ethical Implications and Performance Enhancement* [online]. ResearchGate. [viewed 23 November 2025]. DOI: 10.13140/RG.2.2.35360.39683. Available from: https://www.researchgate.net/publication/381999512

POTOČNIK, K.; ANDERSON, N. R.; BORN, M.; KLEINMANN, M.; NIKOLAOU, I., 2021. Paving the Way for Research in Recruitment and Selection: Recent Developments, Challenges and Future Opportunities. *European Journal of Work and Organizational Psychology*, vol. 30, issue 2, pp. 159–174. DOI: 10.1080/1359432X.2021.1904898.

RÜTTEN, M.; VERHOEVEN, T., 2024. *Recruiting Analytics: Mehr Erfolg mit Data Driven Recruiting und Talent Intelligence*. Stuttgart: Schäffer-Poeschel.

РАЗРАБОТВАНЕ НА ИНТЕГРИРАН МОДЕЛ ЗА ВАЛИДИРАНЕ ПРИ ПОДБОРА НА ПЕРСОНАЛ: БАЛАНСИРАНЕ НА ЕФЕКТИВНОСТТА НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ, ПРОГНОЗНИЯ АНАЛИЗ И ЕТИЧНАТА ОТГОВОРНОСТ

Резюме: *Процесът на подбор на персонал претърпява основна цифрова трансформация, като се наблюдава преход от традиционните техники за подбор на персонал към вземане на решения с помощта на изкуствения интелект. Макар информационните системи като системите за проследяване на кандидати (ATS) и системите за препоръки да правят процеса много по-ефективен и рентабилен, те създават нови рискове по отношение на алгоритмичната пристрастност и липсата на прозрачност. Настоящата статия е опит да се предложи „интегративен валидационен модел", който разширява традиционното понятие за предсказуема валидност. В нея се твърди, че съвременните инструменти за подбор на персонал трябва да бъдат валидирани не само по своята способност да предсказват работните резултати (ефективност), но и по това дали отговарят на етичните стандарти (отговорност). Методологията се основава на систематичен преглед на литературата от последните проучвания между 2015 и 2024 г. Резултатите показват, че за да бъде*

успешен, подборът на персонал трябва да включва принципите на „отговорната изкуствена интелигентност“ в процеса на управление на информацията. Този подход подкрепя холистичен подход, който гарантира, че целите на организацията са в съответствие със социалната отговорност.
Ключови думи: *валидиране на подбора на персонал; управление на информацията; системи за проследяване на кандидати; етика на изкуствения интелект; алгоритмична пристрастност*

Фолкер Йоаким Рьомер, докторант
Университет по библиотекознание и информационни технологии
София, България
E-mail: roemer.volker@gmx.de

PUBLIC COMMUNICATIONS AND INFORMATION SCIENCES
ОБЩЕСТВЕНИ КОМУНИКАЦИИ И ИНФОРМАЦИОННИ НАУКИ

COGNITIVE COMPLEXITY IN E-COMMERCE: AN APPLICATION OF COGNITIVE LOAD THEORY USING THE EXAMPLE OF ZALANDO

Xenia Brandstätter
University of Library Studies and Information Technologies
https://doi.org/10.70300/WNJZ2174

Abstract: *This study examines the influence of psychological methods on digital purchasing behavior using the example of the online shop Zalando with special consideration of the concept of cognitive complexity. After building on the theoretical foundations of business psychology and cognitive load theory, the work analyzes five central design dimensions of digital systems: Description, Structure, Interaction, Limitation and Communication. This underlies the methodological basis, a combination of theoretical foundation with a practical case study that examains the user experience and the psychological effectiveness of Zalando's e-commerce platform. Thereby, the combination of classical and neuroscientific findings on the design of cognitively supportive digital environments has not been reviews as in this approach. The analysis shows that Zalando creates a balance between cognitive relief and activating customer loyalty through clear information structure, personalized communication and targeted incentives. At the same time, potentials become visible, such as non-transparent sustainability communication and the optimization of interactive elements which minimize uncertainties and cognitive dissonance. Overall, the study underlines the relevance of cognitive complexity for the design of digital sales processes and provides practical recommendations for the further development of user-centric online stores.*
Keywords: *business psychology; cognitive complexity; buying behavior; e-commerce; Zalando*

INTRODUCTION

In the age of digitalization, conventional psychological principles are becoming increasingly important for understanding and influencing digital consumer behaviour. Therefore, on digital platforms such as Zalando, processes such as perception, cognition and motivation are activated through design, communication and interaction to influence purchasing decisions. Which is influenced by business psychology, which provides the theoretical framework for this, particularly through emotional appeal, persuasive communication and customer loyalty strategies described by Klaus Moser (2015).

Based on this, the article examines cognitive complexity as a psychological principle of business psychology and analyses how digital stimuli, information overload and personalized functions influence decision-making behaviour on platforms such as Zalando. Thereby, the theoretical basis combines principles of cognitive complexity with Sweller's Cognitive Load Theory (2011), Moser's (2015) business psychology and Bandura's (1976) model learning. To generate insights on that a mixed methodology is being used, which analyses how product presentation, process design, selection limitation and value-oriented communication affect customers' digital purchasing behaviour while deriving psychological recommendations for improving user-oriented e-commerce design.

With the help of these results, the scientific contribution of the study is positioned in the following gap: existing research discusses cognitive load, user experience, personalization and online purchasing behavior, but it rarely integrates these aspects into a single, operationalized framework for assessing cognitive complexity in a real e-commerce interface. Zalando is a suitable case study because its platform combines a large product assortment, filtering and recommendation systems, visual product communication, sustainability labels and personalized interaction elements within one highly developed online retail environment. The study therefore contributes by linking five ob-

servable design dimensions (Description, Structure, Interaction, Limitation and Communication) with their cognitive effects and opportunities for improving user-oriented e-commerce design.

The paper first introduces business psychology, followed by theoretical approaches to cognitive complexity as the basis for evaluating their application in the digital context of the Zalando shop.

RESEARCH METHODOLOGY

For this study, a theoretical-conceptual foundation combined with a qualitative case study analysis is used to examine the effects of psychological principles and cognitive complexity on digital purchasing behavior. Therefore, concepts from business psychology (Moser 2015), Cognitive Load Theory (Sweller 2011; Paas & Van Merriënboer 1994) and neuroscientific findings from neuromarketing form the interdisciplinary theoretical basis for analyzing perception, thought and decision-making processes in digital environments.

In the empirical part, a qualitative analysis of the Zalando online shop is conducted through structured observations and heuristic evaluation methods based on the five dimensions of cognitive complexity: Description, Structure, Interaction, Limitation and Communication (Springer, 2023). These criteria are derived from current research on usability and user experience in e-commerce platforms (Madanchian 2024; Weinstein et al. 2022). Thereby, the qualitative approach makes it possible to identify both effective design features and optimization potential regarding cognitive relief and user orientation, while aligning with established standards in psychological and marketing research (Flick 2022).

For methodological transparency, the case study is specified as a structured snapshot analysis of the publicly accessible Zalando online shop at the time of manuscript preparation. The observation protocol should record the exact observation period, the number of analysed product pages, the product categories included in the sample and the navigation paths followed by the researcher. The evaluation criteria are derived from the five analytical dimensions: clarity and completeness of product descriptions, hierarchical structure and navigability, degree of interactive support, reduction or expansion of available options and transparency of value communication. This clarification makes the qualitative procedure more reproducible and improves the transparency of the empirical basis

THEORETICAL FOUNDATION

Business Psychology

Psychological processes influencing consumer behavior have always been central to business psychology and therefore form the basis of this study, particularly regarding sales and marketing psychology in digital purchasing decisions. Although both are sub-disciplines of business psychology, they address different phases of the customer journey (Jobber & Lancaster 2006, 114). On the one hand, marketing psychology focuses on influencing consumer behavior through brand communication, emotional advertising and persuasive strategies (Moser 2015), which is closely linked to Bandura's concept of model learning, which explains how people orient themselves towards the behavior of others through observation and imitation (Bandura 1976). It is further reflected in digital marketing strategies such as social proof and influencer marketing. According to Wänke (2015), social cognition, group norms and emotions are therefore important influences on consumer behavior. Especially in the digital age, personalization through A/B testing, tracking tools and neuroscientific methods such as eye tracking allows companies to analyze and influence individual responses to advertising more precisely than traditional methods alone.

On the other hand, sales psychology focuses on customer communication, trust building and needs analysis (Maier 2018). Thereby, traditional models such as AIDA (Attention, Interest, Desire, Action) increasingly reach their limits in complex digital decision-making situations (Schürle 2013), creating the need for more flexible models that integrate cognitive, emotional and social aspects of buying behavior.

Digitalization has closely linked marketing and sales psychology, particularly through e-com-

merce platforms that combine marketing influence with sales interaction. Exemplary for this, the analysis of click and scroll behavior using tools such as Hotjar or conversion tracking enables deeper insights into unconscious decision-making processes. Underlined by studies such as Gulfraz et al. (2022) which show that trust and personalized communication are central success factors in digital sales environments, while classical principles such as authenticity, empathy and psychological triggers remain relevant in adapted digital forms (Tracy 2006; Tintelnot 2023).

Overall, this study combines business psychology according to Moser (2015) with findings from social psychology (Bandura; Wänke) and sales psychology (Tracy; Schürle). As observational learning increasingly occurs through digital interfaces, e-commerce becomes the central space for psychological influence in which marketing and sales strategies are interconnected.

Cognitive complexity

After outlining the fundamentals of business psychology, the focus now shifts to cognitive complexity as a central psychological concept in digital system design. Cognitive complexity describes the ability to process, differentiate and connect multiple streams of information simultaneously, particularly in decision-making situations or when solving complex problems (Streufert et al. 1969). Exemplary, research in educational and organizational psychology shows that individuals with high cognitive complexity can manage ambiguity and complex tasks more effectively (Hennessey 2018).

However, cognitive capacity is limited, as excessive cognitive load can lead to mental exhaustion, reduced performance and decision blocking (Sweller 2011). Therefore, digital environments should be designed to support cognitive performance without overburdening users (Paas & Van Merriënboer 1994). This issue is especially relevant in today's digital world, where social media, personalized advertising and constant information streams frequently contribute to information overload, concentration problems and emotional stress. Especially younger generations are particularly affected by digital multitasking and shortened attention spans (Rosen et al. 2013), while older generations often experience difficulties adapting to new technologies (Neves et al. 2019).

A central explanatory model is Sweller's Cognitive Load Theory (2011), which distinguishes between intrinsic, extrinsic and germane cognitive load. The theory aims to reduce unnecessary extrinsic load while promoting productive cognitive processes, making it highly relevant for digital learning and purchasing environments. Similarly, AI-supported tools such as chatbots and recommendation systems attempt to simplify interactions and personalize user experiences (Madanchian 2024). Yet, if these systems provide excessive or irrelevant information, cognitive overload may increase. Where consequently, strategies such as selective media consumption, content curation and digital detox are considered effective in reducing cognitive strain and improving well-being (Weinstein et al. 2022).

Further, research on human-computer interaction demonstrates that complex digital environments, including e-commerce platforms and immersive systems such as the metaverse, can both increase engagement and create mental exhaustion if interfaces are overloaded or poorly structured (Springer 2023). Therefore, clear communication, intuitive navigation and adaptive interaction strategies are essential for minimizing extrinsic cognitive load (Sweller 2011; Madanchian 2024). Furthermore, predictive analytics and AI-supported marketing systems rely on these principles by filtering and structuring information according to user behavior (Madanchian 2024).

From this it follows, that cognitive complexity is a key concept for designing effective digital systems. Emphasized by the research of Sweller (2019) and Madanchian (2024) that successful digital environments must balance information quantity with intelligent structuring, understandable communication and adaptive interaction. Therefore, in e-commerce and digital marketing, cognitive complexity results not only from the amount of information but also from its presentation, interaction possibilities and communication strategies (Weinstein et al. 2022). Accordingly, five central dimensions can be identified: Description, Structure, Interaction, Limitation and Communication. Studies on human-computer interaction show that these dimensions strongly influence cognitive processing, emotional reactions, brand perception and conversion rates in digital markets (Springer 2023; Madanchian 2024).

With this in mind, the following chapter analyses these five dimensions in greater detail to

show how digital environments can minimize cognitive strain, strengthen user orientation and support adaptive user experiences.

Product Description

The way products are described in digital environments directly influences cognitive complexity and user perception, due to the fact that product descriptions act as a key interface between consumers and digital offers, meaning their structure, scope and language can either increase or reduce cognitive load (Weinstein et al. 2022). As research shows that concise, clearly structured and contextualized information improves cognitive efficiency, while overly complex or redundant content leads to information overload and inhibits decision-making (Springer 2023; Madanchian 2024).

In e-commerce, therefore, effective descriptions are especially important because decisions are often made quickly and under high information pressure. Thereby, relevant factors include semantic density and relevance filtering, ensuring that users receive only the information necessary for their purchase decision. Furthermore, utilizing bullet points, standardized attributes, and supporting media such as images or videos reduces extraneous cognitive load and strengthens trust in digital offers, which are considered key drivers (Sweller 2019; Weinstein et al. 2022). Personalized product descriptions can further regulate cognitive complexity individually (Madanchian 2024), although excessive personalization may reduce users' trust and cognitive control.

Overall, effective product descriptions reduce cognitive load and support decision-making through clear, relevant and structured information presentation.

Structuring Process

The structuring of digital content is a key factor in managing cognitive complexity because structures organize information hierarchically and improve accessibility. Especially in e-commerce environments, visual and organizational principles strongly influence how users process information and make decisions (Springer 2023).

This means, a clear and consistent structure directs cognitive resources toward relevant information, while unstructured layouts increase extraneous cognitive load (Sweller 2019; Weinstein et al. 2022). Besides logical organization, visual factors such as typography, white space, color contrasts and navigation design significantly influence cognitive processing. Research further shows that progressive disclosure strategies and adaptive structures, such as filter functions or personalized recommendations, reduce complexity by preventing information overload (Neves et al. 2019; Madanchian 2024).

Summing this up, studies in human-computer interaction demonstrate that excessive interface complexity negatively affects cognitive performance (Springer 2023). Consequently, modern digital systems increasingly rely on modular and minimalist designs to create cognitive relief and improve user experience.

Interaction

Interaction describes how users communicate, navigate and process information within digital systems. Looking into e-commerce and AI-supported environments, interaction is increasingly shaped by adaptive systems and personalized interfaces, creating both opportunities for efficiency and risks of cognitive overload (Madanchian 2024).

In this context, research shows that excessive interaction density, such as pop-ups, over-personalization or unfiltered recommendation systems, increases extraneous cognitive load and can lead to frustration or decision abandonment (Sweller 2019; Springer 2023). In contrast, intuitive interaction designs, such as chatbots or voice assistants, support cognitive processing by guiding users through complex information spaces in a context-sensitive way (Madanchian 2024).

Real-time feedback, transparent system states and adaptive responses further promote trust, orientation and perceived control, all of which reduce cognitive complexity (Weinstein et al. 2022). Interaction therefore represents not only a functional interface but also a central regulator of cognitive load in digital systems.

Delimiting Options

The targeted limitation of options is another important aspect of cognitive complexity in digi-

tal environments. Moreover, research on the "paradox of choice" shows that although freedom of choice supports autonomy, too many options can create decision stress, uncertainty and cognitive overload (Iyengar & Lepper 2000; Weinstein et al. 2022). Especially, in e-commerce systems, excessive or unstructured selections often lead to cognitive paralysis and lower user satisfaction (Springer 2023). Therefore, intelligent limitation strategies aim to reduce unnecessary cognitive load while maintaining relevance and personalization. Methods such as categorized menus, defaults, progressive disclosure and adaptive recommendations help users process information more efficiently (Madanchian 2024).

Following from that, a clear information architecture supported by filters and comparison functions further contributes to cognitive relief (Weinstein et al. 2022; Springer 2023). As a result, limiting options improves both efficiency and user-friendliness in digital systems.

Value Communication

The communication of information and values is another central factor influencing cognitive complexity in digital systems. Due to the fact that communication determines how demanding interactions are perceived and whether cognitive resources are used effectively or unnecessarily burdened (Weinstein et al. 2022). According to Cognitive Load Theory, cognitive overload depends not only on the amount of information but also on its structure and relevance (Sweller 2019).

In digital marketing, value communication refers to the targeted presentation of a product's value proposition. Studies show that storytelling, visual stimuli and narrative structures can reduce cognitive complexity by making information emotionally engaging and easier to process (Weinstein et al. 2022; Madanchian 2024). AI-supported systems such as chatbots or recommender engines can further personalize communication and reduce cognitive load if implemented appropriately (Madanchian 2024). However, excessive personalization or irrelevant recommendations may have the opposite effect.

Based on that transparent communication is not only particularly important for building trust in e-commerce but also AI-supported systems. Based on explainable algorithms and open feedback channels improve trust and facilitate cognitive processing (Springer 2023). Overall, effective value communication reduces cognitive complexity and improves information processing, decision-making and user experience in digital environments.

RESULTS

Practical Analysis of Cognitive Complexity in Zalando's Online Shop

The Zalando online store is a meaningful example of how digital platforms can be analyzed along the five categories of cognitive complexity: description, structure, interaction, limitation and communication. These dimensions play a key role in how users process information, make decisions and evaluate the digital shopping experience. Essentially, the aim is to reduce cognitive load, facilitate orientation and make decision-making processes more efficient. As explained in the introduction, current findings from cognitive load theory (Sweller et al. 2019; Paas et al. 2020) clearly show how crucial the targeted design of these five areas is in order to avoid mental overload to increase usability.

To make the results more explicit, the analysis applies the five dimensions as operational criteria rather than only as thematic categories. Each dimension is assessed according to whether it reduces cognitive load, creates additional orientation effort or opens potential for improvement. The observations therefore focus on concrete interface elements such as product descriptions, image presentation, filter menus, recommendation functions, sustainability labels, wish lists and value-oriented communication. This structure allows the case analysis to connect the theoretical model of cognitive load with observable design decisions in Zalando's online shop.

The following analysis shows that Zanaldo largely succeeds in reducing the cognitive load in the area of description, i.e. the product description, by providing concise, bullet-point information. This is because users receive clear information on material, fit, care instructions and special features. This textual information is also supported by high-quality product images and occasional videos that provide additional visual guidance. As a result, the reduction of extrinsic cognitive load

is effectively implemented. Nevertheless, there is still potential for more in-depth information, for example through interactive content that provides background information on manufacturing conditions or styling options. Such additions could increase the learning-relevant, so-called germane load, as they promote a deeper understanding and communicate the product value more clearly. The positive, clear visual separation of content and the linguistically reduced presentation, which reduces unnecessary extrinsic stress, are particularly striking. One specific aspect within the description is the labeling of products as "sustainable". However, this is not standardized, but only for certain items that meet defined criteria. The sustainability label is clearly shown as a visual element, but there is often a lack of further information in the detailed product description, for example on the type of sustainability, certificates used or the underlying criteria. This can lead to cognitive incongruence for environmentally conscious users, as the label alone does not provide a sufficient explanation. A transparent, standardized description of sustainable features, for example in the form of a sustainability score or clear sustainability categories, could positively modulate cognitive complexity, strengthen trust and make decision-making processes more efficient.

The category of structuring processes is considered in the next step, as the structuring of the purchasing process is also aimed at cognitively relieving the user. In this case, the Zalando website follows a clear hierarchy, from main categories to submenus and customizable filter options. These enable a targeted reduction of irrelevant information and promote a clear navigation flow. The design of the interface is consistent, which strengthens recognition effects and facilitates cognitive orientation. The intelligent filter function in particular helps users to navigate efficiently through the product range without being overwhelmed. Currently, however, filtering is still largely manual. This means that a more dynamic, adaptive learning filter logic, which is based on previous use, can further improve orientation in the future.

The next point to consider is interaction, as interaction with the store is aligned towards user relief and decision support. Zalando uses interactive features such as the Fit Finder, wish lists or personalized home pages to pick up on individual preferences and narrow down the selection. These personalized elements help to reduce the intrinsic cognitive load, as users have to compare and weigh up less. Recommendations such as "Last viewed" or "You might like this" also support the decision-making process. However, if the density of recommendations is too high, there is a risk of information overload, which in turn places unnecessary demands on cognitive resources. Progressive disclosure of content, depending on scroll depth or user interest, could therefore help to ensure a balanced cognitive load.

When it comes to communicating product and brand value, Zalando has a strong visual and emotional position. It also shows that the communicative design in the Zalando online store plays a central role in dealing with cognitive complexity, as it both directs attention and contributes to emotional coherence (Weinstein et al. 2022). The visual design of the store activates emotionally positive contexts and thus facilitates intuitive evaluation. Supplementary content such as storytelling formats about brands or manufacturers, embedded in the product context, could increase the perceived value even more specifically and increase the cognitive processing depth. This reduces the cognitive load during evaluation, as visual signals are processed more quickly than purely text-based information. Thereby, the visual communication of values such as sustainability, diversity or exclusivity is particularly effective. Highlighting sustainable products with a visible label ("sustainable") serves as an emotional and moral anchor that enables users to make decisions based on personal values. A clear contribution to reducing decision-making conflicts in a cluttered digital environment (Madanchian 2024). However, it is important to emphasize that sustainability is not a general standard in the Zalando store, but is used specifically as a unique selling proposition (USP) for individual products. This form of selective communication is effective in terms of cognitive efficiency, as it offers quickly recognizable differentiating features. Though, one challenge is that the communicative impulse is not always supported by depth of content in the product description. In order to promote trust and credibility in the long term, a stronger link between communicative labels and concrete sustainability information would make sense, ideally via standardized sustainability criteria or transparent links to background information (Springer 2023).

Table 1. Conceptual model linking the five dimensions of cognitive complexity, cognitive effects and improvement potential in Zalando's online shop.

Dimension	Observed Zalando element	Cognitive effect	Potential improvement
Description	Bullet-point product information, images and selected sustainability labels	Reduces extraneous load by making key product attributes quickly accessible	Standardize sustainability explanations and add comparable criteria or scores
Structure	Category hierarchy, submenus and filter logic	Supports orientation and reduces search effort	Develop more adaptive filter logic based on previous navigation patterns
Interaction	Fit Finder, wish lists, recommendations and personalized home page areas	Supports decision-making but may create overload if recommendation density is high	Use progressive disclosure and clearer feedback on why recommendations appear
Limitation	Filter options, product sorting and narrowing mechanisms	Limits the choice set and reduces choice paralysis	Balance reduction of options with perceived autonomy and transparent comparison tools
Communication	Visual branding, sustainability cues and value-oriented product presentation	Creates emotional coherence and can reduce decision conflict	Connect labels with concrete background information and transparent evidence

CONCLUSIONS/DISCUSSION

This paper combines theoretical principles of business psychology with the concept of cognitive complexity in order to better understand digital purchasing processes. It became clear that digital platforms such as Zalando specifically use psychological mechanisms to influence purchasing decisions and at the same time control users' cognitive stress. The analysis of the five central design areas, Description, Structure, Interaction, Limitation and Communication, shows that Zalando strives for a balance between cognitive relief and activating information transfer. This can be seen in the visual simplification, clear structuring and personalized communication, which help to reduce the extrinsic cognitive load. At the same time, targeted incentives are provided to encourage deeper processing and thus support the decision-making process.

An important point of discussion is the presentation of sustainability. While sustainability is emphasized as a clear USP for individual products, there is a lack of consistent and transparent communication that increases cognitive consistency. This can lead to uncertainty and cognitive dissonance among environmentally conscious customers. Interaction and limitation also present opportunities and challenges: Automated recommendations and limited choice facilitate orientation but can also impair the feeling of autonomy. This underlines the need not to design cognitive relief at the expense of informed and self-determined decisions. Overall, the work confirms that cognitive complexity is a key concept for the design of digital systems. Since a well-balanced system architecture that intelligently controls the amount and processing of information promotes user satisfaction and supports sustainable purchasing decisions. The findings suggest that digital platforms should not only focus on efficiency, but also on ethical and psychological user needs.

In practice, this means integrating psychological methods and neuroscientific findings more closely in order to create adaptive, cognitively relieving and emotionally appealing user experiences. Future research should combine qualitative, quantitative and neuroscientific methods in order to further explore the dynamics of digital purchasing processes and develop optimal design recommendations.

In this respect, the main scientific contribution of the article is the translation of Cognitive Load Theory into an applicable analytical model for e-commerce interfaces. By connecting the five dimensions with concrete Zalando examples, the study demonstrates how cognitive complexity can be evaluated not only as a theoretical construct, but also as a practical design criterion for online retail platforms. This contribution strengthens the link between business psychology, user-experience research and ethical digital sales communication.

REFERENCES
BANDURA, A., 1976. Social Learning Theory. Englewood Cliffs, NJ: Prentice Hall.
FLICK, U., 2022. *An Introduction to Qualitative Research*. 6th ed. London: Sage Publications.
GULFRAZ, M.; QURESHI, M. A.; IQBAL, J.; AHMAD, M., 2022. Understanding the impact of online customer shopping experience on online impulsive buying. *Journal of Retailing and Consumer Services*, vol. 68, article 103103. Available from: https://doi.org/10.1016/j.jretconser.2022.103103
HENNESSEY, M. N., 2018. Factor structure and basic psychometric properties of the "Transition Assessment and Goal Generator" (TAGG). *Career Development and Transition for Exceptional Individuals*.
JOBBER, D.; LANCASTER, G., 2006. *Selling and Sales Management*. 7th ed. Harlow: Pearson Education.
MADANCHIAN, M., 2024. Artificial intelligence and cognitive load in digital marketing. *Journal of Interactive Marketing*, vol. 58, pp. 77–95.
MAIER, G. W., 2018. *Arbeits- und Organisationspsychologie*. 2nd ed. Berlin: Springer VS.
MOSER, K., 2015. *Marketingpsychologie: Grundlagen und Anwendung*. 2nd ed. Stuttgart: Kohlhammer.
NEVES, B. B.; AMARO, F.; FONSECA, J. R., 2019. Cognitive aging and technology use. *Educational Gerontology*, vol. 45, no. 4, pp. 231–243. Available from: https://doi.org/10.1080/03601277.2019.1585065
PAAS, F.; VAN MERRIËNBOER, J. J. G., 1994. Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, vol. 6, no. 4, pp. 351–371. Available from: https://doi.org/10.1007/BF02213420
ROSEN, L. D.; CARRIER, L. M.; CHEEVER, N. A., 2013. Facebook and texting made me do it: media-induced task-switching while studying. *Computers in Human Behavior*, vol. 29, no. 3, pp. 948–958. Available from: https://doi.org/10.1016/j.chb.2012.12.001
STREUFERT, S.; STREUFERT, D., 1969. *Complexity Management in Organizations*. Boulder, CO: Westview Press.
SPRINGER, A., 2023. Cognitive load and usability in online retail. *International Journal of Human–Computer Interaction*, vol. 39, no. 2, pp. 200–217. Available from: https://doi.org/10.1080/10447318.2022.2038572
SWELLER, J., 2011. Cognitive load theory. *Psychology of Learning and Motivation*, vol. 55, pp. 37–76. Available from: https://doi.org/10.1016/B978-0-12-387691-1.00002-8
SWELLER, J., 2019. Measuring cognitive load. *Educational Psychology Review*, vol. 31, no. 2, pp. 261–269. Available from: https://doi.org/10.1007/s10648-019-09465-7
TINTELNOT, C., 2023. Customer contact and sales conversation. In: *Integrated Product and Sales Management in B2B*. Wiesbaden: Springer.
TRACY, B., 2006. *The Psychology of Selling: Increase Your Sales Faster and Easier Than You Ever Thought Possible*. Nashville, TN: Thomas Nelson.
WEINSTEIN, E. A.; WOOD, M.; MURPHY, T., 2022. Digital detox and cognitive relief: strategies for enhancing mental well-being. *Cyberpsychology, Behavior, and Social Networking*, vol. 25, no. 6, pp. 345–353. Available from: https://doi.org/10.1089/cyber.2021.0425
WÄNKE, M., 2015. *Sozialpsychologie: Einstellungen, Urteile, Verhalten*. 4th ed. Stuttgart: Kohlhammer.
ZALANDO SE, 2022. *Annual Report 2022*. [online]. [viewed 11 June 2025]. Available from: https://corporate.zalando.com/en/investor-relations/reports-presentations

КОГНИТИВНА СЛОЖНОСТ В ЕЛЕКТРОННАТА ТЪРГОВИЯ: ПРИЛОЖЕНИЕ НА ТЕОРИЯТА ЗА КОГНИТИВНО НАТОВАРВАНЕ С ПРИМЕР ОТ ОНЛАЙН МАГАЗИНА ZALANDO

Резюме: *В това изследване се разглежда влиянието на психологическите методи върху поведението при дигитални покупки на примера на онлайн магазина Zalando, като се обръща специално внимание на концепцията за когнитивна сложност. Като се основава на теоретичните основи на бизнес психологията и теорията за когнитивното натоварване, работата анализира пет централни измерения на дизайна на цифровите системи: описание, структура, взаимодействие, ограничение и комуникация. Методологическата основа на работата е съчетание на теоретичната основа с практическо изследване на случай, което хвърля светлина върху потребителското преживяване и психологическата ефективност на платформата за електронна търговия на Zalando. Комбинацията от класически и невронаучни открития за проектирането на когнитивно подкрепяща цифрова среда е нова. Анализът показва, че Zalando създава баланс между когнитивното облекчение и активирането на лоялността на клиентите чрез ясно структуриране на информацията, персонализирана комуникация и целенасочени стимули. В същото време се виждат потенциали, например в прозрачната комуникация за устойчивост и оптимизирането на интерактивните елементи за минимизиране на несигурността и когнитивния дисонанс. Като цяло, проучването подчертава значението на когнитивната сложност за проектирането*

на цифрови процеси на продажби и предоставя практически препоръки за по-нататъшното развитие на ориентирани към потребителите онлайн магазини.
Ключови думи: *бизнес психология; когнитивна сложност; поведение при покупка; електронна търговия; Zalando*

Ксения Брандщетер, докторант
Университет по библиотекознание и информационни технолгии
София, България
E-mail: xenia.brandstaetter@gmx.de

NATIONAL SECURITY
НАЦИОНАЛНА СИГУРНОСТ

INSTITUTIONS DO NOT STOP TANKS: REALISM AND THE LIMITS OF LIBERAL GEOPOLITICS

Daniel Rutkowski Graf von Schweinitz und Krain Freiherr von Kauder
University of Library Studies and Information Technologies
https://doi.org/10.70300/TLGY7667

Abstract: *This article analyzes how realism and liberalism explain current geopolitical conflicts, focusing on Russia's full-scale invasion of Ukraine in 2022. Liberal theories emphasize that economic ties, international institutions, and shared democratic values deter war. Recent geopolitical events raise doubts about these assumptions, as they have not stopped major powers from using force. Applying a structured comparison of liberalism and realism and a case study, this article explores how each theory interprets state motivations and strategic choices. This analysis finds that liberalism falls short in explaining why economically connected states embedded in international institutions still initiate costly military aggression when political leaders believe their security or power interests are threatened. In contrast, realism gives a clearer explanation by referring to international anarchy, spheres of influence and relative power. The study concludes by recommending an integrated approach that combines realist components of power dynamics with liberal ideas of institutional cooperation.*
Keywords: *Realism; Liberalism; International Order; Security Dilemma; Ukraine War*

INTRODUCTION

The debate between liberalism and realism, and its sub-theories, has been shaping the field of International Relations for a long time. This debate allowed the so-called Critical Theories to establish and develop. But recent geopolitical shocks questioned many assumptions which support liberal theory and its inherent optimism. Liberal theory argues that economic interdependence, international institutions and shared norms, values and laws lower the possibility of war or aggression (Burchill 2005, pp. 58–59, 81–82). By contrast, realism assumes that countries interact in an anarchic system where national security and power competition are key factors that determine their actions (Donelly 2005, pp. 30–32).

Russia's 2022 invasion of Ukraine is an interesting case study to test both these theories. The conflict broke out despite deep economic relations and institutional connections between Russia and Europe. Various researchers argue that the war marks a turning point in global politics as it exposes the weaknesses of the liberal theory: the liberal assumption of peace through interdependence is at stake (Fida, Sulaiman, Kazmi 2023, p. 112; Geis, Schröder 2024, p. 11).

This article analyzes whether realism offers a stronger explanation for understanding today's ongoing geopolitical conflicts than liberalism. Comparing these two theories while undergoing an empirical case study of the Ukraine war, the paper highlights that realism reflects how states behave in perceived high-threat environments more accurately. Liberalism on the other hand only functions primarily under limited conditions of shared interests.

RESEARCH METHODOLOGY

This paper uses a qualitative and theory-driven single-case study approach. The theoretical framework is drawn from various works, articles and papers which outline liberalism and realism. Especially useful is the comparative overview in the book *Theories of International Relations* by Burchill et al. (2005). Additional sources on the Ukraine war in the field of International Relations, like Oguejiofor (2024), Rösch (2022) and Fida et al. (2023), are added to ensure that the analysis is based on current scholarship.

Empirically, the paper uses Russia's 2022 invasion of Ukraine as a case study as it very likely

challenges the liberal theory due to the high level of economic interdependence between Russia and Europe and the strong institutional integration. This case study relies policy research like Martill and Sus (2024) and Geis and Schröder (2022, 2024) and on secondary data when it comes to economic figures. The secondary data are used to assess the economic interdependence before the invasion and are derived from Eurostat: data on EU natural gas imports from Russia as well as data on Russia's share of total EU-27 trade. The data are used to provide contextual evidence of the extent of Europe's economic ties to Russia. It is not a comprehensive measure of bilateral dependence.

This allows for a direct comparison between theoretical expectations, observable state behavior and economic data.

RESULTS

(1) Theoretical Foundation: Liberalism and Realism

The core assumption of liberalism is that international cooperation through institutions, norms and economic ties like trade, can transform an archaic system into a peaceful codependence. Liberals expect that a growing interdependence leads to high costs of aggression and therefore deters such actions. Cooperation is viewed as the key to stability and national security (Burchill 2005, pp. 58–61 and 81). The liberalist standpoint on human nature is positive: human beings are willing to interact with others in a way to ensure everyone's well-being. Therefore, peaceful, cooperative conflict solving is rational (Navari 2023, p. 48). Liberals believe that supranational institutions create an environment that helps exporting democratic values while increasing reputational and legal costs of war: the higher the integration, the lower the risk of insecurity (Navari 2023, pp. 37 and 46). This article does not assume that liberalism considers war to be impossible, but focuses on the main liberal expectation: high levels of economic interdependence and institutional integration reduce the likelihood of aggression significantly.

On the other hand, realism does not share the same optimism. For realists, the states' prime concerns are survival and relative power in an archaic system. Institutions are viewed as reflections of the underlying power structures between states: they are not independent or able to constrain states' behavior. (Donelly 2005, pp. 44 – 47). The main principle is laid out by Jensen (2023, pp. 18–19): the more power a state holds, the higher its national security and the ability to enforce its international demands. Rösch (2022, pp. 211–216) argues from Morgenthau's perspective that diplomacy and legal norms cannot prevent wars and conflicts: military presence and deterrence need to be in place to ensure national security. Realists are less positive and optimistic when it comes to human nature as they skeptically highlight its selfishness (Donelly 2005, pp. 30–31). Smith and Dawson (2022, p. 195) conclude that "great power competition is back at the forefront of international politics", while distinguishing various realist theories.

The theoretical separation between liberalism and realism sets the stage for the following case study of the Ukraine war.

(2) Empirical Case Study: Russia's 2022 Invasion of Ukraine

(2.1) Historical Context

When it comes to conflicts, there is usually historical context. Very likely that is the case with the origins of the Ukraine war. Achudume (2023, pp. 92–94) explains that the post-Soviet space was contested after the Cold War and that Russian leaders considered Ukraine as their sphere of influence. From Russia's perspective, every action taken by Ukraine towards Western institutions was and is still considered a threat to Russian security and its sphere of influence. Therefore, Russia intervened long before the annexation of Crimea in 2014 and the escalation to a full-scale war in 2022 in the politics of Ukraine. Perga (2025, pp. 28–30) shows that as a response the West supported the Ukraine with military aid and continued Ukraine's integration into Western institutions. The West's and Russia's visions for a European order and security architecture seem incompatible: not just since 2014 or 2022, but since the end of the Cold War. Martill and Sus (2024, pp. 1–3) demonstrate how European states gradually changed their response to Russia's actions in Ukraine from engagement to military deterrence, although the European states did not react uniformly. Overall, the historical context offers insights into the possible roots of this conflict.

(2.2) Liberal Explanations

From a liberal point of view, Russia's 2022 invasion of Ukraine is a serious empirical challenge: the depth of economic interdependence between Russia and Europe and Russia's integration in international institutions before the war raises question regarding the liberal theory of interdependence.

Before 2022, economic relations between the European Union and Russia were relatively stable and dense, especially when it came to the energy sector. Eurostat data show that the EU heavily depended on Russian natural gas imports in the years leading to the invasion (EUROSTAT 2025a, see fig. 1). The red lines in the following diagrams represent the annexation of Crimea in 2014 and the Russia's full-scale invasion in 2022. It shows that prior to the annexation of Crimea the EU imported about 35% of its natural gas demand from Russia. Although Russia was sanctioned after the attack on Crimea, the EU increased the gas imports to above 40% of its overall demand just before 2022. After the full-scale invasion, the EU reduced its dependency on Russian Energy significantly.

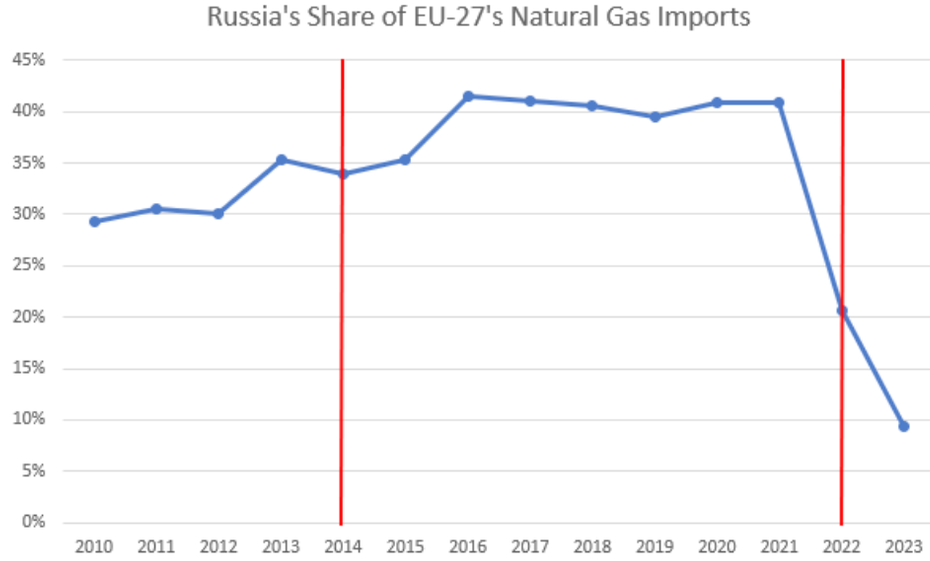


Fig. 1. Russia's Share of EU-27's Natural Gas Imports (EUROSTAT 2025a)

Beyond energy, Russia also was an important trade partner of the EU as it represented a notable portion of total EU-27's external trade (EUROSTAT 2025b, see fig. 2). Before the annexation of Crimea, the imports reached around 12% of total EU-27's external trade. The following sanctions targeted mostly the non-energy sectors and reduced the overall trade. After 2022, the trade partnership almost came to a halt.

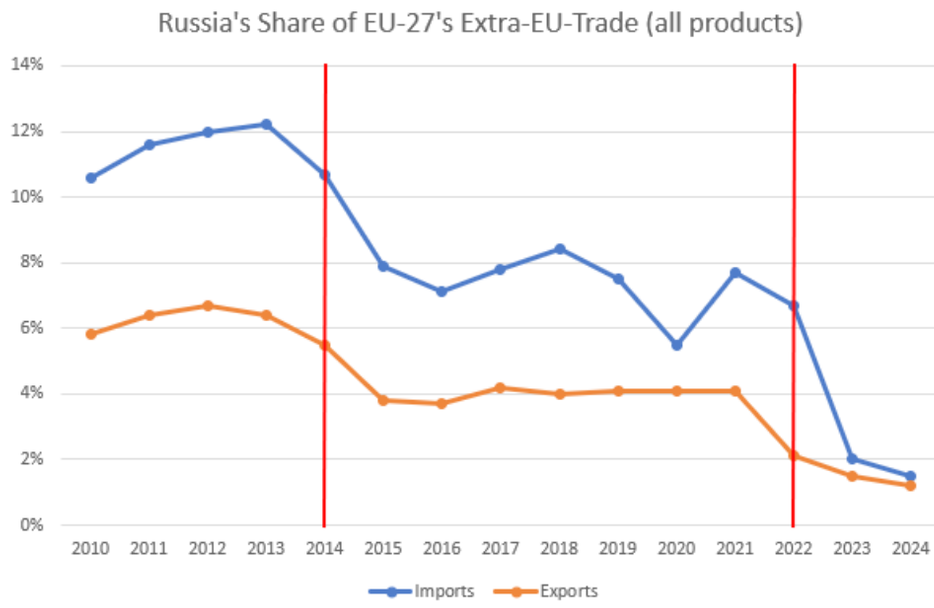


Fig. 2. Russia's Share of EU-27's Extra-EU-Trade (EUROSTAT 2025b)

From a liberal standpoint, such extensive economic interdependence should have raised the expected costs of military actions. Significant disruptions of trade, loss of revenue for companies in both economies, sanctions and long-term damage to the trade relationship were foreseeable outcomes of the full-scale invasion. According to liberal theories, these expected costs should have deterred any further military escalation.

The second liberal argument is built around institutional integration. In addition to the above-mentioned economic ties, Russia was embedded in supranational institutions alongside European partners: it still holds a permanent seat on the United Nations Security Council (United Nations 2025), is a member of World Trade Organization (WTO 2025), participated in the Organization for Security and Co-operation in Europe (OSCE 2025) and was a member of the Council of Europe until its exclusion in 2022 (Council of Europe 2022). Russia was also part of global governance forums like the G8 (until 2014) and is still a member of the G20.

From an empirical point of view however, neither the economic interdependence nor institutional integration prevented Russia's military actions in Crimea in 2014 or the full-scale war in 2022. In spite of the foreseeable economic losses and the reputational risks within the international institutions, Russia was not deterred from military action. Therefore, this case highlights the limitations of liberal explanations: neither deep economic ties nor extensive institutional integration were sufficient to control state behavior and stop military actions.

Therefore, the Ukraine case indicates that the limitations of liberal explanations are directly linked to political decision-makers who interpret a situation as a core security threat. Under such conditions, economic costs and institutional integration cannot act as a deterrent to aggression. This shift creates room for realist explanations.

(2.3) Realist Explanations

From a realist point of view, Russia's 2022 invasion of Ukraine is mainly explained by security concerns, power politics and sphere-of-influence considerations. Economic interdependence or institutional integration are of little interest, at best. Realists assume that states are part of an anarchic international system in which survival and relative power are the important factors, especially when central security interests are perceived to be threatened. In these cases, security concerns outweigh any economic ties or institutional commitments.

Dawood and Diniz (2024) apply this logic directly to the war in Ukraine. They argue that Russia's actions align with the expectations of Offensive Realism. In their view, Russia's actions had the goal to prevent an unfavorable shift in power balance in Eastern Europe. The increasing in-

tegration of Ukraine in Western institutions combined with Russia fear of losing strategic options led to pressure within Russian leadership. That pressure explains Russia's military actions, even at the expense of economic stability and institutional reputation.

Oguejiofor (2024) complements the realist interpretation, showing how Western security policies were interpreted in Russia: A discussed NATO cooperation and Ukraine's Western integration were interpreted as offensive threats to Russian security. This dynamic reflects the classic security dilemma: actions taken by one side to increase its own security are perceived by the other side as threatening and destabilizing. This leads to countermeasures and escalation. This perspective explains, why Russia was ready to accept sanctions and international isolation: it believes that economic losses are less threatening than a Ukraine associated with NATO and the West.

Realism also helps understanding why economic interdependence and institutional integration failed to prevent military actions: once the diverging interests were framed as strategically important, Russia's concerns about power, sphere of influence and security became the main issue. When states believe their own security is at risk, power politics tend to override economic dependencies and institutional integration.

DISCUSSION

The findings in this paper highlight clearly the tension between what liberal theory would expect and what actually happened in the Ukraine war. Liberalism argues that deep economic ties and strong institutional integration raise the costs of conflict and aggression so much that war becomes highly unattractive. Before 2022 and especially before 2014, the relationship between the EU and Russia fits this theory. Eurostat data indicate strong economic ties, especially in the energy sector and Russia was embedded in several important international institutions. Despite these conditions, Russia decided to start a war with Ukraine, which challenges liberal assumptions.

A realist perspective on this conflict is much more convincing. The case supports the realist view that when states believe their security is at stake, concerns about survival, power and strategic independence outweigh economic losses and institutional commitments. Once Russia perceived the conflicting interests as strategically important, it was ready to endure economic sanctions, international isolation and exclusion from international institutions. This shows that interdependence is not a restraining factor for state actions and was more a source of vulnerability than a stabilizing factor.

Nonetheless, liberal mechanisms were not entirely lacking. Economic ties and institutional responses shaped the aftermath of the invasion: sanctions, disrupted trade and reputational damage. But these effects came in effect after the war had already begun. They influenced how the conflict evolved, not whether it occurred in the first place. This distinction helps assessing the liberal theory: interdependence and institutions might shape conflict outcomes, but they do not necessarily prevent aggression or conflicts.

Although a single-case study limits wider generalization, Russia's invasion of Ukraine is a critical test for liberal theory. The inability of interdependence and institutional integration to be a deterrent to war in this case, strengthens the argument that realism offers a stronger explanation for understanding today's ongoing geopolitical high-threat conflicts. Liberal explanations seem to depend on shared security interests and stable expectations.

CONCLUSIONS

This article explored whether realism offers a stronger explanation than liberalism for today's ongoing geopolitical conflicts, using Russia's 2022 invasion of Ukraine as a case study. It indicates that central liberal assumptions, like the idea that economic interdependence and institutional integration reduce the likelihood of war, do not hold in this case. Despite deep trade ties, significant energy dependence and Russia's integration in important international institutions, none of these factors prevented Russian aggression: neither in 2014 nor in 2022.

In contrast, realism offers a more convincing argument in this case. It shows that when states believe their own security is at risk, considerations of power, strategic positioning and survival

outweigh economic costs and institutional obligations. Russia's willingness to face severe economic damage and international isolation underlines the realist argument that power politics dominate in perceived high-threat situations. Therefore, the central contribution of this paper is the empirical demonstration that liberal explanations have limitations in such conditions. It reaffirms the continued relevance of realism in explaining state behavior.

However, this article does not reject liberalism entirely. Economic interdependence and international integrations shaped the evolution and the consequences of the war: through sanctions, diplomatic responses and economic losses. Nonetheless, they failed to act as deterrents.

Future research could extend this analysis through the inclusion of different conflicts as case studies to better identify the circumstances under which liberal or realist theories are better in explaining today's international relations.

REFERENCES

ACHUDUME, T., 2023. The Russia-Ukraine War: Historical Factors of the War and Its Effect in the International Community. *International Journal of Social Sciences and Management Review*, vol. 6, no. 1, pp. 89–102 [online]. [viewed 13 Dec 2025]. Available from: https://ijssmr.org/uploads2023/ijssmr06_06.pdf DOI: https://doi.org/10.37602/IJSSMR.2022.6106

BURCHILL, S., 2005. Liberalism. In: BURCHILL, S.; LINKLATER, A.; DEVETAK, R.; DONNELLY, J.; PATERSON, M.; REUS-SMIT, C.; TRUE, J. (eds.). *Theories of International Relations*. 3rd ed. [online]. pp. 55–83 [viewed 9 Dec 2025]. Available from: https://professorbellreadings.wordpress.com/wp-content/uploads/2016/09/scott_burchill_et-al-_theories_of_intern.pdf

COUNCIL OF EUROPE, 2022. *The Russian Federation is excluded from the Council of Europe* [online]. [viewed 16 Dec 2025]. Available from: https://www.coe.int/en/web/portal/-/the-russian-federation-is-excluded-from-the-council-of-europe

DAWOOD, L.; DINIZ, E., 2024. The Realist Debate in the Context of the War in Ukraine: Balancing Dynamics, International Change and Strategic Calculus. *Revista Brasileira de Política Internacional*, vol. 67, no. 1 [online]. [viewed 17 Dec 2025]. Available from: https://www.scielo.br/j/rbpi/a/MvWrWYGGtcLhgtsFRrmdmcx/ DOI: https://doi.org/10.1590/0034-7329202400105

DONNELLY, J., 2005. Realism. In: BURCHILL, S.; LINKLATER, A.; DEVETAK, R.; DONNELLY, J.; PATERSON, M.; REUS-SMIT, C.; TRUE, J. (eds.). *Theories of International Relations*. 3rd ed. [online]. pp. 29–54 [viewed 9 Dec 2025]. Available from: https://professorbellreadings.wordpress.com/wp-content/uploads/2016/09/scott_burchill_et-al-_theories_of_intern.pdf

EUROSTAT, 2025b. *Extra-EU trade by partner* [online]. [viewed 16 Dec 2025]. Available from: https://ec.europa.eu/eurostat/databrowser/explore/all/all_themes?lang=en&display=list&sort=category

EUROSTAT, 2025a. *Imports of natural gas by partner country* [online]. [viewed 16 Dec 2025]. Available from: https://ec.europa.eu/eurostat/databrowser/explore/all/all_themes?lang=en&display=list&sort=category

FIDA, B.; SULAIMAN, A.; KAZMI, Q., 2023. The Russia-Ukraine War: Unravelling the Challenges to the Liberal International Order. *Liberal Arts and Social Sciences International Journal*, vol. 7, no. 2, pp. 101–119 [online]. [viewed 9 Dec 2025]. Available from: https://ideapublishers.org/index.php/lassij/article/download/1069/448/7256 DOI: https://doi.org/10.47264/idea.lassij/7.2.6

GEIS, A.; SCHRÖDER, U., 2022. Global Consequences of the War in Ukraine: The Last Straw for (Liberal) Interventionism? *Zeitschrift für Friedens- und Konfliktforschung*, vol. 11, pp. 295–307 [online]. [viewed 11 Dec 2025]. Available from: https://link.springer.com/article/10.1007/s42597-022-00089-1 DOI: https://doi.org/10.1007/s42597-022-00089-1

GEIS, A.; SCHRÖDER, U., 2024. The Russian War against Ukraine and Its Implications for the Future of Liberal Interventionism. *Politics and Governance*, vol. 12, pp. 1–12 [online]. [viewed 9 Dec 2025]. Available from: https://www.cogitatiopress.com/politicsandgovernance/article/view/7348 DOI: https://doi.org/10.17645/pag.7348

JENSEN, M., 2023. Realisms. In: WILLIAMS, P.; McDONALD, M. (eds.). *Security Studies – An Introduction*. 4th ed. New York, pp. 17–35 [online]. [viewed 14 Dec 2025]. Available from: https://api.pageplace.de/preview/DT0400.9781000820249_A44685183/preview-9781000820249_A44685183.pdf DOI: https://doi.org/10.4324/9781003247821

MARTILL, B.; SUS, M., 2024. Winds of Change? Neoclassical Realism, Foreign Policy Change, and European Responses to the Russia–Ukraine War. *The British Journal of Politics and International Relations* [online]. Special Issue Article [viewed 11 Dec 2025]. Available from: https://cadmus.eui.eu/server/api/core/bitstreams/9e70a651-302e-52fb-baa6-d996c33b7e23/content DOI: https://doi.org/10.1177/13691481241280170

NAVARI, C., 2023. Liberalisms. In: WILLIAMS, P.; McDONALD, M. (eds.). *Security Studies – An Introduction*. 4th ed. New York, pp. 36–51 [online]. [viewed 14 Dec 2025]. Available from: https://api.pageplace.de/preview/DT0400.9781000820249_A44685183/preview-9781000820249_A44685183.pdf DOI: https://doi.

org/10.4324/9781003247821
OGUEJIFOR, P., 2024. Dissecting the Realist Argument for Russia's Invasion of Ukraine. *E-International Relations* [online]. [viewed 11 Dec 2025]. Available from: https://www.e-ir.info/2024/03/31/dissecting-the-realist-argument-for-russias-invasion-of-ukraine/
ORGANIZATION FOR SECURITY AND CO-OPERATION IN EUROPE (OSCE), 2025. *Countries* [online]. [viewed 16 Dec 2025]. Available from: https://www.osce.org/countries
PERGA, I., 2025. The International Implications of the Russo-Ukrainian War. In: SHINODA, H.; FERDORCHENKO-KUTUYEV, P. (eds.). *The Impacts of the Russo-Ukrainian War – Theoretical and Practical Explorations of Policy Agendas for Peace in Ukraine*. SpringerBriefs in International Relations pp. 27–42 [online]. [viewed 16 Dec 2025]. Available from: https://link.springer.com/content/pdf/10.1007/978-981-96-2295-5.pdf DOI: https://doi.org/10.1007/978-981-96-2295-5
RÖSCH, F., 2022. Realism, the War in the Ukraine, and the Limits of Diplomacy. *Zeitschrift Analyse & Kritik*, vol. 44, no. 2, pp. 201–218 [online]. [viewed 11 Dec 2025]. Available from: https://www.degruyter.com/document/doi/10.1515/auk-2022-2030/pdf DOI: https://doi.org/10.1515/auk-2022-2030
SMITH, J.; DAWSON, C., 2022. Mearsheimer, Realism, and the Ukraine War. *Zeitschrift Analyse & Kritik*, vol. 44, no. 2, pp. 175–200 [online]. [viewed 15 Dec 2025]. Available from: https://www.degruyterbrill.com/document/doi/10.1515/auk-2022-2023/pdf DOI: https://doi.org/10.1515/auk-2022-2023
UNITED NATIONS, 2025. *Security Council Membership* [online]. [viewed 16 Dec 2025]. Available from: https://research.un.org/en/unmembers/scmembers
WORLD TRADE ORGANIZATION (WTO), 2025. *Russian Federation – Member Information* [online]. [viewed 16 Dec 2025]. Available from: https://www.wto.org/english/thewto_e/acc_e/a1_russie_e.htm

ИНСТИТУЦИИТЕ НЕ СПИРАТ ТАНКОВЕТЕ: РЕАЛИЗЪМ И ГРАНИЦИ НА ЛИБЕРАЛНАТА ГЕОПОЛИТИКА

Резюме: *В тази статия се анализира как реализмът и либерализмът обясняват настоящите геополитически конфликти, с фокус върху пълномащабната инвазия на Русия в Украйна през 2022 г. Либералните теории подчертават, че икономическите връзки, международните институции и споделените демократични ценности възпират войната. Последните геополитически събития поставят под съмнение тези предположения, тъй като те не са попречили на големите сили да използват сила. Чрез структурирано сравнение между либерализма и реализма и казус тази статия изследва как всяка теория интерпретира мотивациите и стратегическите избори на държавите. Анализът установява, че либерализмът не успява да обясни защо икономически свързани държави, вградени в международни институции, все пак предприемат скъпоструващи военни агресии, когато политическите лидери смятат, че тяхната сигурност или интереси са застрашени. За разлика от това, реализмът дава по-ясно обяснение, като се позовава на международната анархия, сферите на влияние и относителната власт. Проучването завършва с препоръка за интегриран подход, който съчетава реалистични компоненти на динамиката на властта с либерални идеи за институционално сътрудничество.*

Ключови думи: *реализъм; либерализъм; международен ред; дилема на сигурността; войната в Украйна*

Даниел Рутковски Граф фон Швайниц унд Крайн Фрайхер фон Каудер, докторант
Университет по библиотекознание и информационни технологии
София, България
E-mail: daniel.rutkowski@schweinitz.net

NATIONAL SECURITY
НАЦИОНАЛНА СИГУРНОСТ

RELEVANCE AND DEVELOPMENT OF A RISK CULTURE IN BANKS AS A FACTOR FOR NATIONAL SECURITY

Florian Becker
University of Library Studies and Information Technology
https://doi.org/10.70300/RMVZ7635

Abstract: *This paper examines the importance and further development of a stable risk culture as a key prerequisite for effective risk management in banks and thus for national security. Based on a comprehensive literature review, it shows that regulatory reforms following the 2008 financial crisis have improved structural conditions, but often fail to adequately address the underlying cultural and behavioral factors. Risk culture, understood as collective values, assumptions and behaviors in dealing with risks, has a significant impact on risk perception and decisions in organisations. An effective risk culture requires clear communication of risk appetite and tolerance by management, continuous training, transparent reporting channels and a constructive error culture without fear of sanctions. The integration of risk management into strategic and operational decisions and regular evaluations strengthen organisational resilience in the long term. Banks with a proactive, learning-oriented risk culture are better able to identify emerging risks at an early stage, increase their resilience and thus secure both their own stability and that of the financial system in the long term, resulting in a stable situation with regard to national security.*
Keywords: *risk culture; risk management; banks; national security*

INTRODUCTION

The increasingly volatile, uncertain, complex and ambiguous environment in which we operate is also referred to as the "VUCA" world. In this context, the examination of non-financial, immaterial success factors in organizations is also becoming increasingly important. In addition to strategy, structure and processes, corporate culture in particular is becoming the focus of scientific and practical discussions. It is a central factor influencing organizational performance. Closely linked to the corporate culture is the risk culture. This, in turn, has received increased attention, especially in the context of financial crises, claims and wrong business decisions.

At the same time, implementation and further development of risk management at many banks has accelerated considerably, partly as a result of the banking crisis in the mid-2000s, which had a considerable impact on national security as well as economic consequences for countries worldwide. Especially the aftermath of the 2008 global financial crises caused an increase in the development of risk cultures in organizations. It has also caused financial authorities to demand financial institutions for guidance on risk management, including risk culture, to regain their financial stability (Kanu 2020, p. 13). The financial system is a complex network of institutions, markets and regulations that controls the flow of capital in an economy. Banks play a central role as intermediaries between capital supply and demand. A stable financial system fulfils its central macroeconomic functions smoothly and at all times. In particular, this includes the efficient allocation of financial resources and risks as well as the provision of an efficient and secure financial infrastructure. Financial systems are therefore not only an essential component of national security, they also lead to a more efficient allocation of resources as they reduce the cost of moving funds between borrowers and lenders and ensure that the information asymmetry between borrowers and lenders is overcome. If they do not work well, the economy cannot function efficiently and economic growth is negatively affected (Afreen et al. 2025, p. 99). It was already clear beforehand that banks are an integral part of the financial system. However, historical developments have also highlighted their direct importance for financial security of the state as part of national security. Over time, but especially during the post-2008 economic crisis, there have been changes in regulation to control risks in the financial sector

which never targeted the root of some issues. In this context, risk culture in banks is a key factor for national security, as weaknesses in internal risk awareness and behavior can intensify systemic and procedural vulnerabilities. The 2008 financial crisis demonstrated that failures were not only rooted in insufficient frameworks but also in weak risk cultures that allowed excessive risk-taking. Strengthening risk culture therefore contributes to the resilience of financial institutions and helps to ensure the stability of the whole economic and security environment.

Risk management as a framework is therefore quite easy to define. Risk management is understood to mean systematically identifying, recognizing, analyzing, treating and controlling the company's risks. The corresponding handling of risks is also referred to as risk management. If this is successful, the risks are managed. As part of risk management, critical situations within the company's activities should be recognized as early as possible in order to avoid them and/or reduce the risk. Ideally, this reduction and the ability to plan will also enable a company to better manage unforeseeable risks. A well thought-out and functional risk management system therefore also represents an early warning system for the company.

Now, globally, key regulators are focusing on developing the risk culture that drives enterprise risk management (ERM) and can impact all other risk types. Risk culture is hard to define, hence the range of approaches to address it. Ultimately, the best risk culture might be one that learns through stress. The onus of risk culture development should shift from regulators to Boards, where remuneration linkage could speed up this transfer. All regulators have embedded the incentive as one of the key elements of developing a risk culture. However, its administration is the key to its success. For proper development of risk culture, all elements should be carefully crafted, fitted to the jurisdiction, and reviewed often (Kumar 2025, p. 62). This raises the question of the importance of risk culture in risk management and throughout the company and the extent to which it can be further developed and improved.

In fact, there is a growing number of empirical studies that demonstrate the connection between risk culture and economic performance factors. Although it is difficult to compare the results of the various studies, as both parameters are multidimensional constructs, correlations have been identified (Sackmann 2002, p. 141). There are still some challenges in practical implementation. Even if clearly defined rules for decision making exist, people are prone to be influenced by unrelated factors, which leads to biased judgment and decision mak- ing. Such biases can stem from various sources such as the influence of emotions and affective reactions in the judgments of risks (Eller & Streicher & Zimmermann 2018, p. 224). Furthermore, current developments suggest that innovations in this area are far from over. Due to recent developments in cyber security, topics related to risk culture, operational resilience and an efficient, comprehensive risk management framework will also gain in importance in the future (Iyer & Raji 2025, p. 748).

Regulators are also paying increasing attention to the topic of risk culture. Following a transition in recent years from siloed risk management procedures to a holistic, integrated risk management approach, the focus is now increasingly on improving risk culture in order to strengthen risk management and resilience in companies in the long term. (Kumar 2025, p. 55).

In addition to internal organizational aspects, regulatory authorities have increasingly emphasized importance of a sound risk culture within financial institutions. International regulatory bodies such as the Basel Committee on Banking Supervision and the Financial Stability Board have repeatedly highlighted that effective risk governance cannot be achieved only through formal rules and capital requirements. Instead, behavioral and cultural aspects within institutions must also be addressed as well. Following the financial crisis, it was recognized that many failures in risk management were not caused by a lack of formal frameworks but rather by weaknesses in the risk culture and therefore in the implementation and internalization of these frameworks. As a result, supervisory authorities now expect banks to demonstrate that risk awareness is embedded in in the company and that risk considerations are integrated into strategic planning, performance evaluation and incentive structures. This regulatory shift reflects the growing recognition that a sustainable financial system requires not only technical risk management instruments but also a corporate culture that promotes responsible and transparent risk behavior.

RESEARCH METHODOLOGY

This article's findings on the relevance and development of a risk culture in banks mainly stem from a literature review. The literature was researched using various online databases. The attempt was to use sources that were as up-to-date as possible as well as some well-known basic literature. The goal is to assess the current state of and provide readers with a roadmap on how to develop the risk culture within banks.

The literature search was conducted using several widely recognized academic databases, including Scopus, Web of Science, and Google Scholar. These databases were selected because they provide access to a broad range of peer-reviewed journal articles, books, and conference papers in the fields of management, finance, and organizational studies. The search process focused on the most recent publications that address the conceptual foundations, practical implementation, and implications of risk culture in organizations.

To identify relevant publications, a set of key search terms was used. These included terms such as "risk culture," "corporate culture", "risk management", and "risk governance". In addition, combinations of these keywords were used to ensure a comprehensive coverage of the relevant literature. Particular attention was given to studies that analyze the role of risk culture within the financial industry, especially within banks and other financial institutions, as these organizations operate in highly regulated environments and face more complex risk structures compared to other industries.

While the concept of corporate culture and its theoretical models are applicable across a wide range of industries, the analysis in this article primarily focuses on literature addressing the financial sector in order to reflect the specific regulatory, organizational, and operational characteristics of banks. At the same time, general theoretical contributions on corporate culture were also included in order to provide a broader conceptual foundation for understanding the development and impact of risk culture in organizations.

RESULTS

Corporate culture refers to the totality of all shared values, norms, beliefs and fundamental assumptions that shape and influence the thinking, feeling and acting of the members of an organization. One of the most well-known and influential definitions comes from Edgar H. Schein, who describes corporate culture as a pattern of shared assumptions. This pattern is learned by a group and passed on to new members. Culture thus functions as a kind of informal control system. It complements formal regulations by reducing behavioral uncertainties and regulating expectations.

The three-level model developed by Edgar H. Schein distinguishes between artifacts, beliefs, and basic assumptions:

1. **Artifacts** comprise all visible, explicit, and accessible behaviors, structures, norms, processes, outcomes, and the like. Members are aware of artifacts and they are obvious to outsiders. Accordingly, artifacts are easy to observe and to measure.

2. **Beliefs** and values reflect the strategies, agendas, goals, philosophies, and principles associated with risks. Members are aware of them, yet they are not necessarily observable from the outside.

3. **Basic assumptions** include feelings, affects, un- or subconscious influences of personality and temperament, cognitive states and representations, thoughts, implicit group norms, hidden agendas, and shared implicit learning experiences.

In particular, the basic assumptions guide actions and are difficult to change. They form the basis for specific characteristics such as innovation, performance or risk culture.

Risk culture refers to the set of shared values, attitudes and behaviors in dealing with risk and uncertainty within an organisation. It determines how risks are perceived, communicated, assessed and managed. In contrast to formal risk management, which includes processes, methods and instruments, risk culture describes the informal dimension of dealing with risks. It answers questions such as:

- Are risks openly addressed or concealed?
- Is misconduct sanctioned or covered up?
- Are calculated risks seen as an entrepreneurial opportunity?

Risk culture is thus a specific form of corporate culture with a focus on risks, uncertainty and decision-making behavior.

With reference to the theory of corporate culture just described, the risk culture would therefore be located at the level of the basic assumptions. It touches on central beliefs, e.g.:

- Image of the human being (trust-oriented vs. control-oriented)
- Dealing with mistakes (learning opportunity vs. blame)
- Time horizon (short-term profit orientation vs. long-term stability)

For example, a company with a strong hierarchical background will tend to be less open in risk communication than a hierarchically flatter organization.

Leadership plays a decisive role in creating and sustaining a sound risk culture within an organization. In particular, the concept of "tone from the top" highlights the importance of senior management and board members as role models in establishing a desired behavioral standard. Employees often act not only according to formal rules but are also influenced by the implicit signals transmitted by leadership behavior. If management prioritizes short-term financial performance while neglecting risk considerations, employees may feel encouraged to take excessive risks in order to achieve performance targets. At the same time, leadership that consistently communicates the importance of responsible risk-taking and long-term stability can significantly strengthen the risk awareness of employees. In this context, governance structures such as independent risk committees, clearly defined responsibilities and transparent reporting lines contribute to strengthen a culture in which risks are openly communicated and discussed.

Therefore, when exploring risk culture, it is essential to dig deeply beneath the surface of artifacts into the in-depth understanding of basic assumptions. Unfortunately, from a methodological point of view, basic assumptions are some- times subconscious to members of a risk culture and not obvious to outsiders. In principle, risk culture refers to how people handle risks in a specific social context such as an organization. Risk culture is shaped by, for example, formal structures including shared expectations, experiences, beliefs, and values, as well as individual states and assumptions. At the same time, the risk culture of a specific social category (i.e., group, organization, society) influences how members of this social group perceive, assess, and handle risk. Accordingly, one could argue that risk culture determines how people and organizations respond to risks, and, therefore, it is essential to understand the relevant factors of risk culture and their interactions, in order to enhance risk competence (Eller & Streicher & Zimmermann 2018, p. 219).

However, if we examine the psychological factors involved in dealing with mistakes and risks today, we find that there is a general fear of mistakes and failure in today's society (Fortmann & Kolocek 2018, p. 118). Yet the importance of mistakes for human development should not be underestimated. It can even be said that mistakes are part of being human (Witte 2019, p. 93). In recent years in particular, there has been intensive reflection on the topic of error culture (Fortmann & Kolocek 2018, p. 119). It would be fatal to believe that dealing with mistakes and the resulting consequences and psychological effects are exclusively a feel-good factor. It is also a highly relevant factor in economic terms (Fortmann & Kolocek 2018, p. 119). Therefore, a bank's cultural orientation is also of considerable importance when it comes to influencing risk decisions and improving risk management. (Gani & Mashamba, p. 8).

In a sense, risk culture is stable yet changeable. For example, the societal risk culture regarding traffic risks is stable for a medium period of time, but changeable over a longer period of time. Second, risk culture includes shared learning experiences. For example, society has learned that driving unbuckled results in an unacceptable death rate, and, therefore, taking these specific risks should be avoided. Third, risk perceptions, assessment, and behavior of people, either as individuals or as members of a social group or organization, are modeled by risk culture. Accordingly, risk culture is a major force for explaining differences in risk perception and behavior between individuals, groups, and organizations (Eller & Streicher & Zimmermann 2018, p. 234).

A sound risk culture provides a solid base for the implementation of an efficient risk management. Risk culture should be managed and developed in full alignment with the risk appetite and the risk framework to improve organizational performance. These shall enable the promotion of a risk-aware culture and ingraining risk-related measures into performance management that help drive the organization forward (Kanu 2020, p. 13). Risk culture is not only the basis for the standard processes within a company, but also for projects and all other aspects within the company (Lorch 2025, p. 14). As risk is something that has potential harms to the company due to the occurrence of a certain event. The importance of risk management, especially in insurance companies, can provide opportunities for companies to avoid these risks or minimize the impact of losses that occur. The effectiveness of risk management depends on how it is integrated into the governance of an organization and significantly influenced by the characteristics of the organization Successful risk management is supported by the development of a risk management culture whereas employees and stakeholders realize the importance of monitoring and managing risks (Mita & Ramadhany 2025, p. 2582).

That being said, an organization can have an overall positive culture but still lack a sound risk culture because most of the employees see risk management as a compliance issue and not a means of arriving at sound business decisions (Kanu 2020, p. 14). Nowadays, the demand to fully deal with uncertainty and complex risks more efficiently not only stems from the finance and insurance industry but from society in general, as well as organizations operating in a complex and uncertain context. Accordingly, organizations recognized the contribution of social psychological insights on risk perception, risk judgment, risk behavior, and potential pitfalls (Eller & Streicher & Zimmermann 2018, p. 233). Despite the strong need to integrate ERM and risk culture, their implementation in organizations remain fragmented. Risk practitioners are increasingly focused on designing and implementing risk management frameworks with less consideration to the initial establishment of sound risk culture (Kanu 2020, p. 14). Ultimately, the main question is what measures can be taken to improve the risk culture within the company in order to make risk management more efficient and resilient.

The further development of risk culture in companies requires a holistic approach that takes both structural and behavioral aspects into account. A key prerequisite is clear communication of a company-wide risk and value orientation by management in order to establish a common understanding of risk appetite and tolerance. Training and awareness-raising measures promote risk competence among employees and strengthen their ability to identify risks at an early stage and address them appropriately. In addition, a transparent reporting and feedback system should be established that enables open discussion of mistakes and critical events without fear of sanctions. The integration of risk management processes into strategic and operational decisions ensures that risk considerations are not viewed in isolation, but rather as an integral part of corporate management.

Regular evaluations and adjustments to measures can improve the risk culture in the long term and adapt it to changing conditions.

While risk culture often evolves as an organization grows, organizations can and should conduct self-assessments, surveys, and use other techniques to understand the current state of risk culture (Mita & Ramadhany 2025, p. 2582). On this basis, it is then possible to continuously develop and optimise the risk culture.

Despite its growing relevance, measuring risk culture remains a challenge. Unlike quantitative risk indicators, cultural aspects are difficult to capture using traditional metrics. Nevertheless, organizations increasingly attempt to assess risk culture through a combination of qualitative and quantitative approaches. These may include employee surveys, structured interviews, behavioral indicators, internal audits and the analysis of decision-making processes. In addition, key risk indicators related to conduct, compliance breaches or operational incidents can provide indirect insights into the underlying cultural environment. The objective of these assessments is not only to diagnose potential weaknesses but also to identify areas for improvement and to monitor cultural change over time. A systematic evaluation of risk culture can therefore support management in aligning behavioral patterns with the organization’s overall risk strategy and risk appetite.

CONCLUSION

Risk culture should not be understood as an isolated category, but rather as a component within the overall corporate culture. It is shaped by and interacts with various organizational dimensions: for instance, a strong strategic orientation toward innovation tends to promote a higher willingness to take risks, while centralized or authoritarian power structures can inhibit open communication about risks. At the same time, an ethics-oriented, value-based culture strengthens compliance, integrity, and a more sensible approach to risk-taking.

A discrepancy between the formal risk strategy and the lived culture therefore leads to inconsistencies. For example, a company may officially propagate "zero tolerance" for rule violations, while implicitly high sales and sales targets encourage risky behavior.

Corporate culture forms the foundation. An informal control system of an organization. On this basis, the risk culture develops as a specific manifestation of the collective handling of uncertainty. While formal risk management systems define structures and processes, the lived culture determines how risks are actually perceived and treated.

A sustainable and constructive risk culture requires consistency between strategy, incentive systems, leadership and lived values. In an increasingly dynamic environment, it is thus becoming a decisive competitive factor – not only for avoiding crises, but also for consciously exploiting entrepreneurial opportunities.

An analysis of the relevant literature and various theoretical approaches shows that a stable risk culture is a key prerequisite for effective and sustainable risk management in banks.

While regulatory measures in the wake of the 2008 financial crisis have contributed to a more comprehensive risk management framework, it is clear that purely structural compliance requirements are not sufficient to address the underlying behavioral and cultural factors that influence risk-related decisions. However, a further development of precisely these aspects becomes particularly clear against the background that the crisis of 2008 not only had economic consequences, but also had an impact on national security for society as a whole. The effective development of an appropriate risk culture depends largely on consistent support from senior management, transparent communication and psychological mechanisms that enable mistakes and critical events to be dealt with openly without fear of sanctions. A culture of error management that promotes mutual trust makes the company faster and more flexible. In order to establish such a culture of trust, trust must be generated at various levels (von Au 2017, p. 148).

It can be said that risk culture is both stable and adaptable; its further development therefore requires regular evaluations, learning from past events and adaptation to changing conditions. Banks that manage to establish a proactive and learning-oriented risk culture are better able to anticipate emerging risks at an early stage, strengthen their resilience and ensure long-term stability – for their own organisation, for the financial system and for national security as a whole.

The increasing digitalization of financial services and the growing importance of cyber risks further underline the relevance of a strong risk culture within banks. Technological innovations such as digital banking platforms, artificial intelligence and automated decision systems create new opportunities but also introduce complex and evolving risks. In such environments, rule-based risk management approaches may prove insufficient, as emerging risks cannot always be fully anticipated or regulated in advance. A well-established risk culture can therefore function as a complementary mechanism that guides employees in situations where formal rules only provide limited orientation. Therefore, risk culture is increasingly regarded not only as a compliance requirement but also as a strategic capability that enhances the resilience and adaptability of financial institutions.

REFERENCES

AFREEN, F.; AKTHER, S.; RAHMAN, M.; RASHID, M.; UDDIN, M.; ULLAH, N., 2025. Effects of financial system on the economic development: Evidence from Bangladesh. *International Journal of Multidisciplinary and Current Educational Research*, vol. 7, no. 1, pp. 99–107.

ELLER, E.; STREICHER, B.; ZIMMERMANN, S., 2018. Risk culture: An alternative approach to handling risks. B:

Psychological Perspectives on Risk and Risk Analysis. Basel: Springer International Publishing, pp. 217–247.
FORTMANN, H.; KOLOCEK, B., 2018. *Arbeitswelt der Zukunft: Trends – Arbeitsraum – Menschen – Kompetenzen*. Wiesbaden: Springer Gabler.
GANI, S.; MASHAMBA, T., 2022. The effect of culture on risk-taking activities of banks. *Corporate Governance and Organizational Behaviour Review*, vol. 6, no. 3, pp. 8–18. Available from: https://doi.org/10.22495/cgobrv6i3p1
IYER, S.; RAJI, B., 2025. Cybersecurity culture and organizational resilience: A human-centered approach to digital risk management. *American Journal of Industrial and Business Management*, vol. 15, pp. 748–766. Available from: https://doi.org/10.4236/ajibm.2025.154038
KANU, M., 2020. The role of risk culture in enterprise risk management implementation. *Applied Artificial Intelligence*, vol. 15, no. 11, pp. 13–31.
KUMAR, S., 2025. How regulators see risk culture. *Intelligent Risk*, no. 5, pp. 55–63.
LORCH, Z., 2025. Best practices and sources of error: What project documentations reveal about the success of infrastructure projects. *Education, Scientific Research and Innovation*, vol. 3, no. 2, pp. 12–16.
MITA, A.; RAMADHANY, A., 2025. Evaluation of risk awareness culture in a strategic holding insurance company. *Jurnal Ilmiah Ekonomi dan Bisnis*, vol. 13, no. 3, pp. 2581–2594.
SACKMANN, S., 2002. *Unternehmenskultur: Erkennen – Entwickeln – Verändern*. Wiesbaden: Springer.
VON AU, C., 2017. *Eigenschaften von Führungspersönlichkeiten: Achtsamkeit, Selbstreflexion, Soft Skills und Kompetenzsysteme*. Wiesbaden: Springer.
WITTE, F., 2019. *Testmanagement und Softwaretest: Theoretische Grundlagen und praktische Umsetzung*. Wiesbaden: Springer.

ЗНАЧЕНИЕ И РАЗВИТИЕ НА КУЛТУРАТА НА РИСКА В БАНКИТЕ КАТО ФАКТОР ЗА НАЦИОНАЛНАТА СИГУРНОСТ

Резюме: *В настоящата статия се разглеждат значението и по-нататъшното развитие на стабилна култура на риска като ключова предпоставка за ефективно управление на риска в банките и по този начин за националната сигурност. Въз основа на изчерпателен преглед на литературата се показва, че регулаторните реформи след финансовата криза от 2008 г. са подобрили структурните условия, но често не успяват да се справят адекватно с основните културни и поведенчески фактори. Културата на риска, разбирана като колективни ценности, предположения и поведение при справянето с рисковете, има значително влияние върху възприемането на риска и вземането на решения в организациите. Ефективната култура на риска изисква ясна комуникация на рисковия апетит и толерантност от страна на ръководството, непрекъснато обучение, прозрачни канали за докладване и конструктивна култура на грешки без страх от санкции. Интегрирането на управлението на риска в стратегическите и оперативните решения и редовните оценки укрепва организационната устойчивост в дългосрочен план. Банките с проактивна, ориентирана към ученето рискова култура са в състояние по-добре да идентифицират възникващите рискове на ранен етап, да увеличат своята устойчивост и по този начин да осигурят както собствената си стабилност, така и тази на финансовата система в дългосрочен план, което води до стабилна ситуация по отношение на националната сигурност.*
Ключови думи: *рискова култура; управление на риска; банки; национална сигурност*

Флориан Бекер, докторант
Университет по библиотекознание и информационни технологии
София, България
E-mail: flo.becker@online.de

NATIONAL SECURITY
НАЦИОНАЛНА СИГУРНОСТ

HEALTHCARE POLICY AND NATIONAL SECURITY RISK SCENARIOS AND SYSTEMIC TRANSITIONS

Gerhard Degen
University of Library Studies and Information Technologies
https://doi.org/10.70300/SQTW3028

Abstract: *The article analyzes potential relationships between healthcare and national security. Disruptive events (e.g., pandemics, increasing antibiotic resistance, or technical failures) can exceed the available range of action in healthcare and become a matter of national security. Conceptually, national security is not an isolated issue, but rather the result of the interaction of institutional protective mechanisms. The analysis shows that disruptions in the health sector can also have a significant impact on economic performance, social stability, and the government's ability to act. The analysis shows that sudden risks or slowly developing risk scenarios can result abruptly or gradually in the transition from a health policy challenge to a security policy problem. Finally, the article calls for comprehensive strategies. Ultimately, healthcare should be understood as an integral part of modern security architectures.*
Keywords: *Health Care; National Security; trigger mechanisms; transition; risk scenarios*

INTRODUCTION

Health risks, whether pandemic or technological in nature, are no longer just clinically relevant, but can (under certain conditions) become a matter of national security. This essay examines how to understand the interface between health systems and national security, when a health problem becomes a security issue, and how and by what mechanisms this transition takes place. Political leaders are becoming increasingly aware of the links between a functioning healthcare system and national security considerations, as well as the relevance of thematic interactions. For them, this leads to the conclusion that the stability of a society and thus of a state structure, along with external security, also depends on a functioning medical infrastructure. Health risks have many different origins and can include epidemics, biological hazards, and other crises. A threat to public health could subsequently have a lasting disruptive effect on the functioning of economic, social, or even governmental structures. Conceptually, national security is ensured through the interaction of institutions and measures. It is therefore not an autonomous, isolated concept, but rather depends in practice on concrete protective mechanisms that can be provided by areas such as defense policy, security agencies, intelligence services, legislation, international cooperation, and crisis management.

This article shows how critical medical infrastructure, human resources, and information security, as part of these protective mechanisms, can have a direct impact on the national security situation. The evaluation of future scenarios shows an increase in this interdependence, as rapid global technology-driven progress, globalization, digitalization, demographic change, and new biotechnological applications will challenge the vulnerability limits of healthcare systems. This gives grounds for the demand for integrated strategies that strengthen security and health aspects with a view to creating a resilient external image of society. In summary, the discussion makes it clear that health and safety are interlinked and that integrated strategic considerations are necessary to maintain the resilience of society as a collective.

RESEARCH METHODOLOGY

In order to introduce a methodological approach to support this thesis the inclusion of domains as health-science, security-policy and governance perspectives must be considered.

The thematic content needs to be deepened through a comprehensive search of literature sources and contemporary developments as well as reports from institutions such as the WHO,

OECD, CSIS and national security agencies. In addition, an analysis needs to be applied conceptually. The purpose was to capture central concepts such as interdependence, transition thresholds, trigger mechanisms and resilience. Various scenarios served as the basis for comparative analyses (COVID-19 pandemic, cyberattacks on health systems, antibiotic resistance). The aim of the comparisons was to identify patterns of transition from health policy challenges to national security issues. Thus, the methodology used can make a significant contribution to the identification of mechanisms, processes and decision-making logics. Methodologically, a qualitative evaluation was applied. In the first step, comparative content analysis (qualitative content analysis) of representative sources was carried out.

Orientation in the selection supported causal relationships that pointed to a direct or indirect connection between health care risks and security policy implications. In a second step, characteristic case studies are examined from the perspective of process tracing. The aim is to make temporal sequences, decision-making logic and threshold crossings comprehensible. This can illustrate a transition from health challenges to the area of responsibility of national security structures. In addition, a phase or threshold model (based on heuristics) is recommended for future deepening. This is not intended to define fixed limit values, but more to describe transitional phases. These phases characterize a state in which medical, technological and political systems come under pressure at the same time.

Comprehensive consideration and thematic limitation: The angles of consideration for further development are complexity, the chronological sequence of events, the transition into the domain of "national security", the possibilities for preventive and predictive measures, and support through monitoring. A differentiated presentation of the measures at the individual and societal levels could complement the picture. As already mentioned, future challenges to healthcare can be complex and exceed the capabilities of conventional medical care. Such risk scenarios can therefore be multidimensional rather than purely medical in nature. This means that in certain scenarios, normal infrastructure and conventional healthcare routines may not be sufficient to meet the requirements in a timely and sustainable manner.

RESULTS

A closer look at complex risk scenarios reveals the following picture: Infectious diseases can reach pandemic proportions and lead to the rapid spread of viruses that are resistant to known therapeutic measures. Healthcare systems can become overwhelmed (e.g., COVID-19) with shortages of resources in terms of materials and staff (McCoy, et al. 2023, p. 1). In general, researchers expect the emergence of multi-resistant pathogens with significant antibiotic resistance. This would mean that standardized therapies would not be applicable, resulting in increased mortality rates and structural overload (He, et al. 2021, p. 2)**.** Environmental and technological risks are increasingly becoming an issue. Climate-related changes (heat, flooding, strong winds) have a negative impact on infrastructure (drinking water, energy, communications, transport links, critical resources and storage, healthcare facilities). A combination of heat waves and power outages, for example, can immediately lead to critical supply problems. Active external influence is once again a geopolitical issue due to experiences from conflicts (Ukraine) (Kutcher & Borisch 2025, pp. 473–478): cybercrime and hacker attacks aimed at disrupting hospital operations, interrupting supply chains and procurement logistics (e.g., vaccines), or causing the collapse of medical care in general. Ultimately, the spread of pathogens (biological warfare) is also one of the risk scenarios. In addition to medical care, security and surveillance facilities are also challenged in general. The latter raises the follow-up question of when the invisible line between medical and national security is generally crossed.

At what point are national security interests affected? As mentioned at the outset, national security interests generally concern the stability of the state. This includes also functionality, and protectability. The population and critical infrastructure are the subject of the protective function.

There are now various indicators triggering intervention by security authorities: When many human lives are directly at risk, examples include pandemics, large-scale food or drinking wa-

ter contamination, or threats to hospitals, energy supplies, and critical infrastructure. Externally caused intentional threats such as bioterrorism or cyberattacks on medical facilities (Kutcher & Borisch 2025, pp. 473–478).

In summary, it is clear that complex risk scenarios can quickly exceed the capabilities of healthcare systems, as these operate from a purely medical perspective. When medical capabilities are exceeded, national interests are immediately endangered. This situation arises when a crisis scenario threatens the population on a large scale, endangering available infrastructure or public order. By definition, national security measures are activated when risks exceed defined limits and additional measures are necessary.

Interface, transition stages, and trigger mechanisms

A deeper exploration of the topic of healthcare systems and national security leads to keywords such as interface, transition stages, and trigger mechanisms. The increasing interconnectedness of healthcare and security reflects the search for applicable methods to manage relevant risks. Politicians are responsible for making decisions or take action to predict, prevent or solve these problems.

Knowing these thresholds is an essential point for further consideration. Whether these boundaries can be identified as specific points, interfaces, or phases which can be limited in time is at the core of further elaboration. Practical examples may serve to illustrate the theoretical assumptions in the last part of the essay. By 2025, modern healthcare systems should no longer be measured solely by the number of hospitals and beds. A well-functioning healthcare system that is tailored to needs is no longer limited to treatments and therapies. Clinical units are only one dimension of a close-knit care network that is ideally supported by continuous monitoring. Needs-based systems are dynamic systems and are by no means limited to clinical units. Dynamics can affect several directions of development, and healthcare geared towards this can (under certain conditions) also become part of and a matter of national security.

This paper examines the connections and interactions between healthcare systems and national security. The question of when a situation in the healthcare sector becomes a security issue is part of the investigation. The question of how this transition manifests itself, whether through processes, mechanisms, or human intervention. From society's perspective, reliable, targeted medical care is an expression of a functioning state. Medical care is the state's responsibility for its citizens. Fair and reliable healthcare strengthens society's perception of a stable state structure and creates trust. The result is that the state's ability to perform is demonstrated. However, if this part of the infrastructure is damaged or disrupted by external influences, this can jeopardize the ability of the state and its institutions to act in several respects (World Health Organization 2025). The same institution raises the issue of "health security" to include, by definition, all activities that support the identification of a decision-making space from a broader perspective. Within this decision-making space, risks become active by definition. The overlap between health and safety aspects is already considered to be the basis for the above definitions. The World Health Organization calls for the following preventive measures in the field of health care: Preventive measures should be preventive, protective and reactive. The aim would be to avoid or reduce potential risks and consequences of foreseeable acute health hazards. This goal requires international action to reduce the spread of risks across national borders (World Health Organization 2025).

Recognizable risk scenarios in terms of content:

Some significant security risks are known. Prominent among these is the threat posed by new communicable diseases, but also already known antibiotic resistance. These without corresponding advances in developing alternative care concepts might end with considerable treatment deficits. Triggering situations can arise from biological threats, technologically supported cybercrime with attacks on healthcare structures, or simply from carelessness in the laboratory. One institute focused on these risk situations is the Center for Strategic & International Studies (CSIS). The institute has evaluated this as a field of study in research on the topic of "health security" and identified overlaps between global health issues and security aspects. The results were recommendations for preventive measures against the spread of pathogens and biological hazards and clear indica-

tions for strengthening healthcare infrastructures (CSIS 2025). At the same time, research-based literature also points out that health and security issues can no longer be viewed solely from a local perspective. The interdependencies exist on a global level. "It is necessary to identify the numerous interfaces and interactions between actors and programs in the field of health and security (McCoy et al. 2023, p. 1).

Standardization: Since 2005, the World Health Organization has been engaged in publishing global standards in the field of health and safety (WHO 2005) and, in addition, to support country-specific application (WHO 2005).

Transition process – from healthcare to national security

Thresholds and criteria are often (but presumably not always) quite recognizable: The question arises as to when healthcare leads to a shift in responsibility to the security apparatus. This moment can occur as soon as a challenge overwhelms the healthcare system's capacity to provide care and requires additional functions, infrastructure, and central facilities and instruments. Signs of this can include:

Prominent among these are the overload or failure of essential parts of the healthcare system. Furthermore, significant economic damage can result from the disruption of healthcare facilities. Disease outbreaks and the associated loss of productivity can have a destabilizing effect. At the same time, there is a risk that military capabilities or security-related infrastructure could be compromised. In addition, the spread of pathogens could lead to international outbreaks, resulting in cross-border difficulties. The accompanying weakening of state authority would make it considerably more difficult to maintain public order.

The Process of Transition – Human Intervention or Automated Systems?

A real example given by the pandemic: a new dangerous highly transmissible virus is spreading rapidly. This crosses geographical boundaries and thus the situation can gain national security relevance within a short time. As a result, there is political pressure to make decisions, taking into account the situational risk assessment. The decision may favor the transition from health care to the issue of national security. Ideally, decision-making guidelines are prepared for support. These focus on risk classification and extent of damage, but also under what circumstances in a critical situation of health care a decision should, can, must be made proactively and promptly in the direction of further administration by the structures of the security authorities. Once preparations have been made, this can be done mainly through targeted, politically initiated frameworks and decision-making criteria (if-then criteria). In practice, when governments or international bodies officially classify a problem as a "threat to national security," procedures and protocols come into effect. An example: A UN resolution ruled that targeted attacks on health care facilities also involve a global health and thus security risk (UN Press Releases, 2024).

Forms and dynamics of the transition from health to safety – Immediate, rapid transition: After discussing the content and possible triggers, the moment or phase of the transition should be looked at even more closely. An unexpected short-term development, such as the COVID-19 pandemic, can overwhelm health systems within a very short time. The impairment of economic life is directly affected. Another consequence can be the need to use the national resources of the national security institutions, in this example the outbreak acts as an **initial trigger** that requires immediate security policy responses.

Temporal transition: The opposite situation can occur through a slow, gradual transition. If antibiotic resistance is increasingly present due to the virus, bacteria, fungi or other threats, a creeping weakening of the supply system will occur. Often overlooked by chronic underinvestment or supply crises over long periods of time, states become vulnerable. In such cases, the transition to the security question takes place gradually. *Less recognizable, but threatening in its overall scope, as there is usually already contamination of the infrastructure area and can mutate into a crisis.*

Technology-induced scenarios: the age of digitalization has also made healthcare systems vulnerable to technology (cyber-attacks or misinformation). The initiators of such attacks have the opportunity to disrupt a state structure in its supply and social functions by non-military means.

Cyber and ransomware attacks: In addition to data protection concerns, the WHO has also

focused on the endangerment of patients' lives. Thus, the transition from health to safety problem is also in the technological-systemic area (Ghebreyesus 2024).

The transition phase: With the entry into the phase of the framework conditions and processes of safety facilities, health care becomes part of a comprehensive security concept. Once a health issue is recognized as a safety issue, specialized resources, processes, and tools expand. Health and specialist ministries are cooperating with each other, coordination resources are being mobilised, international interactions are being strengthened and crisis management structures are being activated. Health is becoming part of a broader security strategy. Viewed comprehensively, a security strategy that is pre-planned for cases of threat with the connection of health systems and the framework of national security can no longer be regarded as a theoretical side field, but as a central dimension of modern regulatory systems for states. Health as an obligation of the state is extended to a security issue. From the moment when the requirements that exceed purely medical capacity limits go beyond and reach the cornerstones of the sovereignty of the state. Triggers and threat scenarios may be manifold, but the framework and processes made in advance by the political authority are essential. There is an institutional link and instrumental expansion, in short: health systems become part of the state's safety net. The political responsibility is evident from the explanation so far. All health policy measures are by no means isolated, separate areas. Preventive provisions for crises, care concepts and all sub-areas of the state security infrastructure are affected.

Improvements and new paths: the present explanation also provides new additional options for improving existing supply facilities. These are to be predominantly cross-departmental and act as integrated AI-supported processes. Increasing the internal and external cohesion of society is a central component and the following case studies from contemporary practice show how individual situations can be considered more solution-oriented or still problematic. As far as practical evidence is available, three contemporary examples can be cited as examples on the basis of the theoretical explanations so far. In the context of health care and national security, these are as follows: pandemic Covid-19 in Western Europe, extortion and cybercrime in the UK and Ukraine. The examples are evaluated with a detailed justification and presented with extended comments on weaknesses and possible improvements.

Example 1: The Covid-19 pandemic (in Central and Western Europe)

While many countries tried to reduce acute mortality and the burden on intensive care medicine through vaccination campaigns, surveillance (including wastewater surveillance) and preventive programs, structural problems (inequality, chronic burden of disease, lack of structural investment) remained.

This means that population-based health measures have certainly strengthened national security thinking, but the impact has been significantly dampened by years of hardened structures and gaps in governance and funding (WHO European Observatory on Health Systems and Policie 2021). **There are clear demands for prevention and early detection. A recognizable vacuum exists for data integration: One** Real-time monitoring (e.g. national reporting systems, wastewater monitoring) would have enabled rapid remediation measures. There is still a lack of direct links to supporting data consolidation and the development of monitoring systems (WHO COVID-19 dashboard, 2025).

From Preventive to Predictive: Monitoring and Early Warning Systems Can significantly support the lead time for countermeasures (vaccination campaigns, resource shift). This prevents the direct consequences of disruptions to critical infrastructure or detects them in advance (e.g. care, police). This would result in a clear additional safety benefit (Ghebreyesus, et al. 2022).

Resilient supply networks: Regions with robust primary care and integration models have been able to maintain care in a stable and continuous manner. In this environment, the national security impact was stronger. OECD analyses therefore recommend targeted investments in resilience (OECD Health Policy Studies 2023). It became clear that insufficient investment before the pandemic (high chronic disease burden, fragmented care) limited the effectiveness of these measures (vulnerable groups could only be addressed partially and then with increased effort). Without structural reform, vulnerability to new risks remains (OECD Health Policy Studies 2023).

Example 2: Extortion and cybercrime – the WannaCry case

Cyber-attacks on healthcare facilities not only endanger immediate patient care (surgery cancellations, transfers) but also undermine the trust of the population and could directly affect national security (disruption of critical infrastructure, diversion of state resources). The situation **WannaCry** shows how outdated IT infrastructure in combination with inadequate precautions can lead to extraordinary disruptions to health care. Currently, there are ongoing trends and a segregated complexity of such attacks. Inclusive cooperation between health care and national security is here **not particularly pronounced** (National Audit Office 2018).

Ransomware attacks could lead to the collapse of supply continuity. In the UK, gaps in legacy software led to massive disruption (surgical failures). Research shows that IT weakness weighed heavily on the NHS's operational response, which was clear evidence of negative security effects.

The lack of integration between health IT, national cyber defense and emergency planning exacerbated the consequences. The deficiency was the missing component "cooperation between health, safety and technology sectors". (Cybersecurity & Infrastructure Security Agency 2020). **There is an ongoing trend and modern ransomware tactics:** US authorities HHS/analyses document persistently high number of attacks, double extortion (data leak + encryption) and increasing targeting of healthcare and public institutions. This makes the situation systemically exposed and dangerous. This is true as long as cyber resilience is lacking (Office of Information Security HHS 2023).

Within the framework of these obvious weaknesses, some thoughts for improvement are possible- like early warning systems through data integration and real-time monitoring: Investing in national surveillance, data sharing platforms, and standardization to improve early detection and coordinated crisis response (OECD Health Policy Studies 2023). **Building resilience for cyber situations should be a core component of resilient care. A strategic division of critical networks,** virtual or real situational task forces with protocol mandates that work with health care facilities and both local and federal agencies with a national cyber focus.

Scenario planning and cross-sectoral real-world exercises: Joint crisis simulations (health × safety × infrastructure) increase operational interoperability later in the event of an emergency (avoidance of unrealistic dry training) (Haldane, et al. 2021, pp. 964–980). **Targeted investments in primary and public health capacities** to reduce vulnerable population groups (as a long-term security gain). (OECD Health Policy Studies 2023). Finally, a summary to answer the question of the "timing and the trigger criteria" for the transition from regular operation to integration into the structures of national security: The chosen examples to answer the question are: the Covid-19 pandemic, the WannaCry case in the UK and the slow development of antibiotic resistance. The investigation is from when (point in time) and/or with which trigger criteria the transition from regular operation of health care to introduction into the structures of national security is achieved.

Timing and **triggering criteria** for the transition from **regular operations** to integration **into national security structures:** In the example of Covid-19) defines the point in time when a locally limited infection situation develops into a national health crisis with the threat of overloading medical care. Trigger criteria could be an exponential increase in the number of infections beyond the capacity limits of the health system**,** supply bottlenecks in intensive care places**,** or additional support for federal authorities. **In the example of cybercrime** (e.g. WannaCry in the NHS, UK), the time as soon as digital attacks impair or paralyze the functionality of critical infrastructures in the healthcare sector (e.g. hospitals, laboratory networks, rescue coordination centers) will be impaired or paralyzed. The primary trigger criteria are the failure of central IT systems endangering patient care, broader attacks (by institutions or regions), the call for the need for national cyber defense. In the case of antibiotic resistance in a population, the point in time would be when resistance developments permanently endanger public health and security of supply and national or international spread becomes apparent. **Trigger criteria are** the detection of multidrug-resistant pathogens with a significant increase and lack of therapeutic options and the risk of system-wide failures; Use of coordinated defense strategies.

The **overall resilience** of the health system and transition to national security: If you look at the three threat scenarios **at the same time** (pandemic, cyber-attack, antibiotic resistance), **no single point in time** of transition can be clearly defined. Instead, there is a **phase-oriented transition period in which solutions are positioned.** In summary, the **transition to the national security structure** can be described as follows: either several subsystems become critical at the same time **(e.g. overload of facilities** and *failure of IT systems) or the self-control of health care has reached limits, or state intervention becomes necessary* **to ensure care and order.**

CONCLUSIONS

This paper reveals that healthcare and national security are not isolated domains. An interdependent system has emerged. Disruptive events can transcend the functional boundaries of healthcare systems, thereby threatening elements of state stability. This affects public order, economic performance, and, more generally, the state's capacity to act.

A transition from a health policy challenge to a national security problem will usually not be detectable as a single event. Rather, it will arise when thresholds are exceeded, causing stress in several areas. These can simultaneously and rapidly affect healthcare, social, and political subsystems, exacerbated by technical challenges. These findings imply that national security depends not only on traditional defense mechanisms. It also needs to include the resilience of critical medical infrastructure, human resources, information systems, and preventive capacities. Mandatory are effective prevention, surveillance and early warning systems. A cross-sector coordination is an essential component. International cooperation and standardized frameworks further strengthen this resilience in an increasingly globalized risk environment.

Health care is part of a national security policy. Beyond the social and ethical-moral obligation, the strengthening of health systems is part of a strategic orientation. This strengthens social resilience and protects the long-term stability of the state.

REFERENCES

CSIS, 2025. *A New Era in Health Security* [online]. [Viewed 29 October 2025]. Available from: https://www.csis.org/analysis/new-era-health-security

CYBERSECURITY & INFRASTRUCTURE SECURITY AGENCY (CISA), 2020. *Ransomware Activity Targeting the Healthcare and Public Health Sector* [online]. [Viewed 30 October 2025]. Available from: https://www.cisa.gov/news-events/cybersecurity-advisories/aa20-302a

GHEBREYESUS, T. A., et al., 2022. WHO recommendations for resilient health systems. *Bulletin of the World Health Organization* [online]. [Viewed 03 November 2025]. Available from: https://doi.org/10.2471/BLT.22.287843

GHEBREYESUS, T. A., 2024. *Threats posed by ransomware attacks against hospitals and other health-care facilities and services* [online]. [Viewed 27 September 2025]. Available from: https://www.who.int/director-general/speeches/detail/who-director-general-s-remarks-at-meeting-of-the-un-security-council-on-threats-posed-by-ransomware-attacks

HALDANE, V., DE FOO, C., ABDALLA, S. M., et al., 2021. Health systems resilience in managing the COVID-19 pandemic: lessons from 28 countries. *Nature Medicine*, vol. 27, pp. 964–980 [online]. [Viewed 16 October 2025]. Available from: https://doi.org/10.1038/s41591-021-01381-y

HE, Y., ALIYU, A., EVANS, M. and LUO, C., 2021. Health care cybersecurity challenges and solutions under the climate of COVID-19: scoping review. *Journal of Medical Internet Research* [online], vol. 23, no. 4, e21747. [Viewed 03 November 2025]. Available from: https://doi.org/10.2196/21747

KUTCHER, S. and BORISCH, B., 2025. Health and global security: challenges and considerations [online]. [Viewed 02 November 2025]. Available from: https://doi.org/10.1057/s41271-025-00562-0

McCOY, D., ROBERTS, S., DAOUDI, S. and KENNEDY, J., 2023. Global health security and the health-security nexus: principles, politics and praxis. *BMJ Global Health* [online], vol. 8, no. 9, e013067. [Viewed 03 November 2025]. Available from: https://gh.bmj.com/content/8/9/e013067

NATIONAL AUDIT OFFICE, 2018. *Investigation: WannaCry cyber attack and the NHS* [online]. [Viewed 01 November 2025]. Available from: https://www.nao.org.uk/wp-content/uploads/2017/10/Investigation-WannaCry-cyber-attack-and-the-NHS.pdf

OECD, 2023. *Ready for the Next Crisis? Investing in Health System Resilience*. OECD Health Policy Studies [online]. Paris: OECD Publishing. [Viewed 03 November 2025]. Available from: https://www.oecd.org/en/publications/ready-for-the-next-crisis-investing-in-health-system-resilience_1e53cf80-en.html

OFFICE OF INFORMATION SECURITY, U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES, 2023.

2022 Healthcare Cybersecurity Year in Review and a 2023 Look-Ahead [online]. [Viewed 27 October 2025]. Available from: https://www.hhs.gov/sites/default/files/2022-retrospective-and-2023-look-ahead.pdf
UNITED NATIONS, 2024. *Ransomware Attacks on Healthcare Sector "Pose a Direct and Systemic Risk to Global Public Health and Security", Executive Tells Security Council* [online]. [Viewed 21 September 2025]. Available from: https://press.un.org/en/2024/sc15891.doc.htm
WORLD HEALTH ORGANIZATION (WHO), 2005. *International Health Regulations (IHR)* [online]. [Viewed 18 October 2025]. Available from: https://www.who.int/health-topics/international-health-regulations
WORLD HEALTH ORGANIZATION (WHO), 2025a. *COVID-19 Dashboard: WHO Health Emergencies Programme* [online]. [Viewed 03 November 2025]. Available from: https://data.who.int/dashboards/covid19/summary
WORLD HEALTH ORGANIZATION (WHO), 2025b. *Health Security* [online]. [Viewed 03 November 2025]. Available from: https://www.who.int/health-topics/health-security
WORLD HEALTH ORGANIZATION (WHO), 2025c. *National Action Plan for Health Security (NAPHS)* [online]. [Viewed 18 October 2025]. Available from: https://www.who.int/emergencies/operations/international-health-regulations-monitoring-evaluation-framework/national-action-plan-for-health-security
WORLD HEALTH ORGANIZATION EUROPEAN OBSERVATORY ON HEALTH SYSTEMS AND POLICIES, 2021. *Health Systems Resilience During COVID-19: Lessons for Building Back Better* [online]. [Viewed 02 November 2025]. Available from: https://eurohealthobservatory.who.int/publications/i/health-systems-resilience-during-covid-19-lessons-for-building-back-better

ЗДРАВНА ПОЛИТИКА И НАЦИОНАЛНА СИГУРНОСТ. РИСКОВИ СЦЕНАРИИ И СИСТЕМНИ ПРЕХОДИ

Резюме: *Статията анализира потенциални връзки между здравеопазването и националната сигурност. Разрушителни събития (например пандемии, нарастваща антибиотична резистентност или технически повреди) могат да надхвърлят наличния обхват на действия в здравеопазването и да станат въпрос на национална сигурност. Концептуално, националната сигурност не е изолиран въпрос, а по-скоро резултат от взаимодействието на институционалните защитни механизми. Анализът показва, че смущенията в здравния сектор могат също да имат значително въздействие върху икономическото представяне, социалната стабилност и способността на правителството да действа. Анализът показва, че внезапни рискове или бавно развиващи се рискови сценарии могат да доведат внезапно или постепенно до преход от предизвикателство на здравната политика към проблем със сигурността. Накрая статията призовава за цялостни стратегии. В крайна сметка здравеопазването трябва да се разбира като неразделна част от съвременните архитектури за сигурност.*
Ключови думи: *здравеопазване; национална сигурност; механизми на задействане; преход, рискови сценарии*

Герхард Деген, докторант
Университет по библиотекознание и информационни технологии
София, България
E-mail: g@gdegen.at

NATIONAL SECURITY
НАЦИОНАЛНА СИГУРНОСТ

PARTICIPATORY DEMOCRACY IN THE AGE OF DISINFORMATION: EU AND NATO STRATEGIES FOR PROTECTING PARTICIPATORY DEMOCRACY AS PART OF NATIONAL SECURITY

Marina Wieners
University of Library Studies and Information Technologies
https://doi.org/10.70300/WJOG4952

Abstract: *A democracy requires freedom of expression. Corresponding passages can already be found in the American Bill of Rights (1791) and the French Declaration of the Rights of Man and of the Citizen (1789). However, this basic requirement was formulated during the first wave of globalization. Today, society is in its fifth wave – and each wave has led to an expansion of freedom. Whereas in the 18th century limited channels of information existed, today's information landscape has been exponentially expanded by the internet, increasing digitalization and AI. This expansion makes it difficult to validate information and encourages the spread of disinformation, which is not always immediately recognizable as such and thus enables the manipulation of actions in everyday life and political decisions. Disinformation therefore influences not only individual opinion-forming, but also participatory democratic processes, and can have security-related effects. Mitigating these effects requires measures and coordinated implementation at both national and international level. Against this background, this article analyses current strategies of the European Union (EU) and the North Atlantic Treaty Organization (NATO) to contain disinformation to protect participatory democracy as a component of national security, reviews their effectiveness through SWOT analysis and examples of real disinformation campaigns, and reflects on whether the current strategies are geared towards potential threats.*
Keywords: *participatory democracy; national security; disinformation; disinformation defense; international strategies*

INTRODUCTION

In today's digital era, the state faces new challenges as an actor and guarantor of security for the population. On the one hand, it is required to develop technologically and organizationally in order to keep pace with the dynamics of digital change. On the other hand, it must ensure that the fundamental rights and freedoms of the population are not restricted or endangered. In this context, government action and the achievement of objectives are complicated by hurdles such as disinformation, the digital divide, cybercrime, foreign interference in elections, and ethical challenges (Internet health report 2019; UN Press Release 2023; Weisch 2025, p. 8).

Disinformation is defined as false information that is deliberately generated to harm a person, social group, organization or country. Disinformation is therefore distinct from misinformation and malinformation (see Figure 1). Misinformation is false information that is not produced with the intention of causing harm. Malinformation, by contrast, is information that is based on the truth but used to harm a person, organization or country (Turcilo and Obrenovic 2020, p. 9).

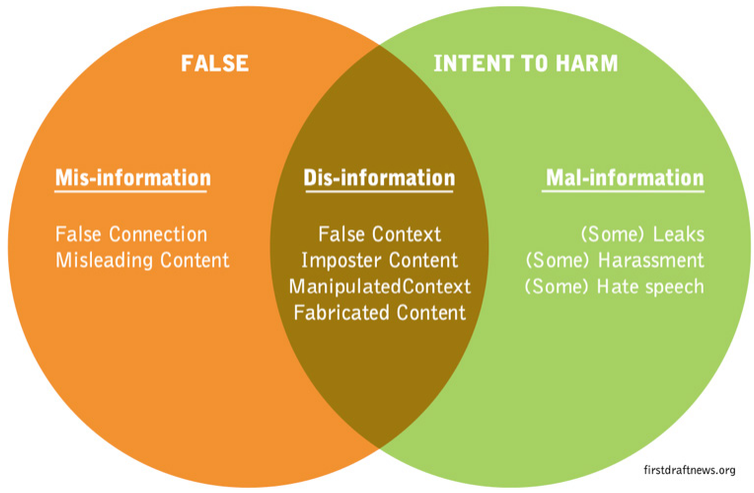


Fig. 1. Misinformation, disinformation and malinformation (Turcilo and Obrenovic 2020, p. 10, cited by Asanov 2019)

The danger of disinformation lies in its many forms and the high speed at which it spreads. A study by the Massachusetts Institute of Technology (MIT) showed as early as 2018 that false reports on Twitter spread much faster and further than true messages (vgl. MIT News Office Presse Release 2018). The Konrad Adenauer Foundation distinguishes between four main forms of disinformation manipulation: fake news, deep fakes, social bots and trolls. These can be used not only as weapons in information wars, but also to undermine trust in the media and manipulate public opinion in a targeted manner (KAS n.d.). The results of the Vodafone Foundation Germany (2020) and the survey by the Friedrich Naumann Foundation (2025) show that such manipulations are already taking place and have been effective since the coronavirus pandemic. In 2020, the proportion of 14- to 24-year-olds who come into contact with disinformation several times a day almost doubled within a year (21% compared to 12% in 2019), and in 2025, young TikTok users were particularly susceptible to Russian and Chinese disinformation.

It should not be taken as an axiom that only young people are affected by disinformation: a study by the Max Planck Society (2024), which examined the perception and classification of information in different population groups, showed that only older adults were significantly better able to distinguish between true and false headlines. By contrast, younger adults (the age groups examined were: 18-31 years, 32-47 years and 48-88 years) were less able to do so. At the same time, the study found that education level had no measurable impact on susceptibility to misinformation (Press Release of Max-Planck-Gesellschaft 2025). Therefore, the ability to recognize false information is not determined by one's own self-assessment but is significantly influenced by the type of content and the source (OECD Truth Quest Survey 2024).

While the individual motives for spreading disinformation are complex and varied, four overarching objectives can be identified at the societal level: the targeted radicalization and division of society, the erosion of trust in the media, political influence through the manipulation of opinions or elections, and economic interests. These objectives, the high level of awareness of the term "disinformation" (76% of respondents in Germany and the US), as well as proven influences of disinformation on participatory events such as the Brexit referendum or the "Doppelgänger" campaign, underline the influence of disinformation on participatory democracy (Bernhard et al. 2024; Bastos 2016; USCYBERCOM Public Affairs 2024).

From a theoretical perspective, disinformation can be classified as a security policy threat, as it can have a multidimensional impact on participatory democracy and national security, making it difficult to combat. This can be justified by Daase and Kessler's concept, according to which disinformation can be classified into all forms of democratic deficits (known knowns, known unknowns, unknown unknowns and unknown knowns):

- Disinformation can function as known knowns when its existence, objectives and basic

mechanisms of action are recognised and empirically proven (e.g. orchestrated disinformation campaigns, coordinated bot networks or recurring narratives);

- Disinformation can occur as known unknowns when its existence is recognized, but some characteristics such as actor attribution, reach or intended side effects remain unclear;
- Disinformation can be classified as an unknown known if its existence and potential risks are known but are deliberately ignored or not addressed for political, institutional or normative reasons;
- Disinformation can act as an unknown unknown if novel forms, actor constellations or causal relationships are neither anticipated nor recognized at an early stage and can lead to significant consequences. (Daase et al. 2007, pp. 411–434).

At the international level, disinformation is considered by both the EU and NATO to be part of hybrid threats, prompting the development of concrete protective measures and the inclusion of corresponding regulations in national security strategies to protect participatory democracy.

RESEARCH METHODOLOGY

In order to present and critically analyze current EU and NATO strategies for combating disinformation and their impact in the context of protecting participatory democracy as a component of national security, a qualitative evaluation of current specialist literature, political strategy and security papers, scientific studies and selected case studies will be carried out. The analysis will follow the following pattern: definition of theoretical principles and terminology – examination of the strategies of selected actors (the EU and NATO) and their evaluation using SWOT analysis – reflection on the strategies based on selected case studies of real disinformation campaigns, followed by the derivation of recommendations for future strategies to combat disinformation.

RESULTS

The European Union's (EU's) strategy for combating disinformation

Analysis and development of strategies to counter information warfare in the European Union are based on a variety of sources: pan-European surveys, external studies by institutions and organizations, and reports by internal expert groups (Mündges 2019). The acceleration in strategy development took place after the disinformation campaigns in the context of the Russian invasions of Ukraine in 2014 and 2022 in 2015, the East StratCom Task Force was set up to deal primarily with the detection and classification of Russian disinformation activities (Hoxtell 2024, p. 12). The task force is part of the Strategic Communication and Information Analysis Division (AFFGEN.7) of the European External Action Service. The Western Balkans Task Force, the South Task Force and a cross-cutting team dealing with emerging threats, data analysis, policy development and international cooperation also belong to the same division. The tasks of this unit include, among other things, coordinating the EU rapid alert system against disinformation (The Diplomatic Service of the European Union 2021). In 2018, the publication of the Communication "Tackling online disinformation: a European approach" and the EU Action Plan against Disinformation (see Figure 2) were followed, along with the entry into force of the "Code of Practice on Disinformation" (Hoxtell 2024, p. 12). The 2018 EU Action Plan proclaims disinformation as "a major challenge for European democracies and societies" and sets out four pillars for a coordinated response to disinformation between the Union institutions, Member States, civil society and the private sector, in particular online platforms: improving the Union institutions' abilities to recognize, analyze and expose disinformation; strengthening coordinated and joint responses to disinformation; mobilizing the private sector to combat disinformation; raising awareness and improving societal resilience. The EU Code of Practice on Disinformation was formally adopted by the initial signatories (Internet companies Facebook, Google, Twitter and Mozilla, and the European Association of Digital Media EDiMA) on 16 October 2018 and includes a preamble, a statement of purpose and a total of 15 voluntary commitments by the signatories to exercise control over advertising placements, political and issue-related advertising, the integrity of the services offered, empowering consumers and promoting research (Sachstand des Deutschen Bundestags PE 6-3000-041/19,

p. 5). In addition, the "Digital Services Act" (2022), the "Digital Markets Act" (2022) and the "AI Act" (2024) created further legal frameworks that also serve – in some cases indirectly – to contain disinformation (Hoxtell 2024, p. 12).

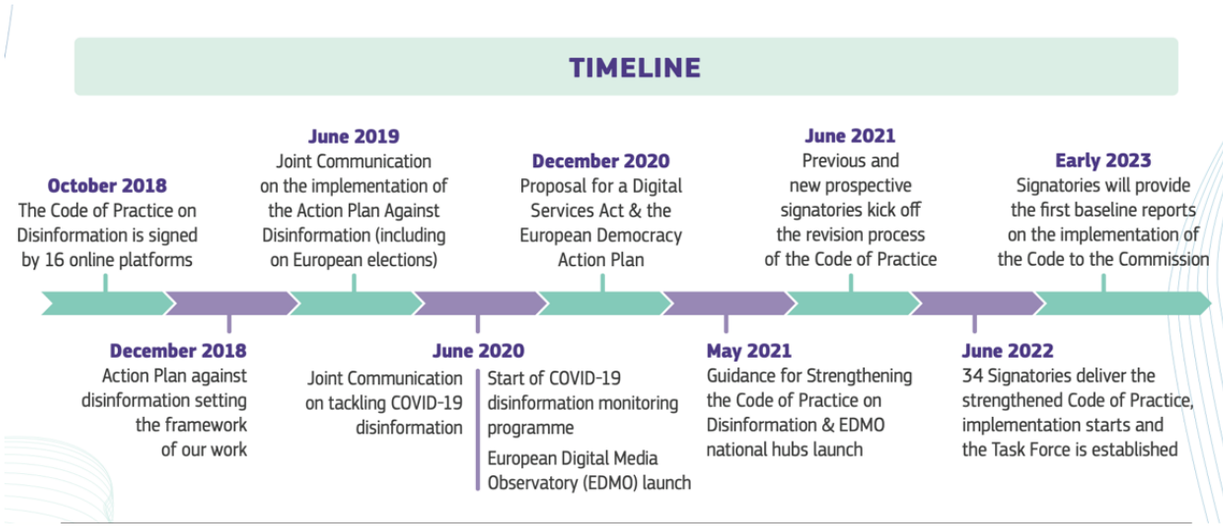


Fig. 2. Timiline of measures Implementing the Strengthened Code of Practice on Disinformation (European Commission 2022)

The strategy of the North Atlantic Treaty Organization (NATO) for combating disinformation

NATO views disinformation as part of the collective term "information threats", since the term disinformation has become, according to NATO, "an umbrella term for different kinds of phenomena in the information space, including as a label for opinions and narratives that contradict an actor's own opinion". The danger of information threats is seen in the fact that they erode people's trust in governments and other public institutions and "appeal to people's worst impulses, fears and prejudices". NATO's strategy for combating information threats is based on close cooperation with allies and partner states and consists of four central approaches: Understanding the information environment, Preventing the effectiveness of information threats, Containing and mitigating information threats, Recovering quickly by learning lessons from information threats (NATO 2025). The aim is to fundamentally "vaccinate" or "prevent" the media sphere, rather than refuting every false claim. The objectives and measures are regularly reviewed and adjusted, as was the case after the COVID-19 disinformation incident. Member and partner states are supported in increasing their resilience to disinformation through advisory services and financial support for social and scientific projects, which was further anchored by the adoption of Resolution 483 "Protecting Allied Democracies and Partners from Disinformation" in 2023. NATO is also intensifying its cooperation with the European Union to benefit from the EU rapid alert system to combat disinformation (Chlon 2022, p. 6).

Manifestation of EU and NATO strategies at national level for the protection of participatory democracy

An analysis of the national strategies of countries that are members of both the EU and NATO revealed the following explicit manifestations of EU and NATO strategic approaches to protecting participatory democracy in the context of disinformation:

- Germany's National Security Strategy aims to expand public relations work, strengthen media literacy, and speed up the identification and deletion of illegal content in the digital space in order to "strengthen defenses against disinformation and the resilience of democracy" (NSS 2023, pp. 46–48);
- In France's Revue nationale stratégique (National Strategic Review), protection against disinformation is seen as requiring targeted investment in the strategic education and research sector (National Strategic Review 2025, p. 39);

- The Netherlands' National Security Strategy recognizes the spread of disinformation as a security threat to public discourse and social cohesion. Disinformation can be combated by maintaining a pluralistic media landscape, strengthening citizens' resilience to disinformation and reducing the influence of disinformation through international common standards (The Security Strategy for the Kingdom of the Netherlands 2023, pp. 14–31);
- Estonia's National Security Concept classifies disinformation as a threat to constitutional order and social cohesion. To counter it, Estonia relies on strategic communication that strengthens constitutional values in society (National Security Concept of Estonia 2023, p. 10);
- Latvia's State Defense Concept defines the goal against disinformation as 'building public understanding of state threats and understanding the intentions of potential aggressors' in order to "increase the resilience of the government and the psychological resilience of society to various forms of manipulation, including disinformation and information-based influence operations" (The State Defence Concept 2023, pp. 10 – 16);
- In the Strategia Bezpieczeństwa Narodowego Rzeczypospolitej Polskiej (National Security Strategy of the Republic of Poland), disinformation is described as part of hybrid threats to state structures and societies in Western countries and as a challenge to the information sphere. To counteract this, the information sphere should be protected at the virtual, physical and cognitive levels in cooperation with the media, social networks, citizens and non-governmental organisations (Strategia Bezpieczeństwa Narodowego Rzeczypospolitej Polskiej 2020, pp. 6–21);
- According to the Security Strategy of the Slovak Republic, disinformation is used to "disrupt or manipulate national decision-making mechanisms, influence public opinion and destabilize the political situation". To counteract this, the strategy envisages the development of critical thinking among young people, the establishment of coordinated national mechanisms for the early identification, assessment and combating of disinformation, and the development of strategic communication involving civil society, the media and academia (Security Strategy of the Slovak Republic 2021, pp. 10–19);
- The Belgian Stratégie de sécurité nationale describes disinformation as part of hostile information operations. To counter this, a safer and more responsible online environment is to be created that enables the strengthening of the position of citizens and civil society (Stratégie de sécurité nationale 2022, pp. 18–39).

In order to evaluate the effectiveness of the EU and NATO strategies presented for combating disinformation with regard to safeguarding participatory democracy and national security interests, a SWOT analysis (see Table 1) is carried out below. This method allows for a systematic comparison of the strengths, weaknesses, opportunities and threats of existing measures at the EU and NATO level and provides a basis for reflecting on and evaluating the success of the strategies using a few case studies of real disinformation campaigns.

Table 1. SWOT Analysis of Current EU and NATO Strategies Against Disinformation (Own elaboration)

Strengths	**Weaknesses**
• Liability through codes (Code of Conduct against Disinformation) and legal frameworks (DSA, DMA, AI Act); • Clear target definitions & strategic frameworks (e.g. EU action plan, resolution 483); • Broad institutional anchoring in the EU & NATO through task forces and specialised programmes; • Coordination between member states (EU rapid alert system) and multilateral exchange between the EU and NATO.	• Heterogeneous implementation in EU member states; • Lack of mandatory standards in the EU Code of Practice on Disinformation; • Dependence on voluntary self-regulation (e.g. by online platforms); • Limited sanctions for violations; • High pressure to adapt due to constantly changing disinformation practices; • Lack of transparency regarding the effectiveness of existing measures.
Opportunities	**Threats**
• Expansion of digital resilience in the population through targeted programmes; • Increased cooperation with civil society actors and fact-checkers; • Use of AI for early detection and analysis of campaigns; • Synergies through close integration with cybersecurity strategies; • Development of international alliances and strategic partnerships against coordinated disinformation attacks.	• Increase in state-orchestrated disinformation campaigns from autocratic states and constant change in their forms; • Lack of speed of response to new threats; • Potential conflict of interest between freedom of expression and security interests; • Loss of trust in democratic institutions and the media.

CONCLUSION

The comparison carried out in the form of a SWOT analysis confirms that the EU and NATO are pursuing an active policy of countering disinformation. In particular, the legally binding nature of the strategies, clarity of objectives and the development of coordinated mechanisms in partnership with partners can be cited as strengths. The identified weaknesses indicate that the strategies examined cannot be considered complete.

The case studies of the Brexit referendum and the "Doppelgänger" campaign provide a concrete illustration of various elements of the SWOT analysis. In the case of Brexit, existing strategic weaknesses such as the lack of a coordinated communication framework and the low capacity for coordinated disinformation became apparent. At the same time, it became clear that political influence can be exerted by digital platforms with a high degree of effectiveness. The "Doppelgänger" campaign is an example of how targeted misinformation can be used at the international level against participatory processes and trust in state institutions.

The nature of digital threats requires a regular review and adaptation of existing instruments, which can be confirmed by the identified risks. It is recommended that the intervals between strategic revisions be shortened, that operational feedback be evaluated more systematically and that cooperation between international actors be further strengthened. The challenge here is to find a level of regulation that protects democracy against disinformation but which does not endanger democracy itself at its core, because as early as 2019, the investigation initiated by the European Parliament's Civil Liberties and Constitutional Affairs department found that legal restrictions on the content of democracy would do more harm than disinformation itself (LIBE Committee summary 2019).

The risks identified supplement those already collected and published in the other sources, such as keeping up with the volume and speed of disinformation, disinformation being more interest than information, limited resources and expertise, balancing commercial interests with the public goods, and thus expand the field of possible adjustments at the structural and strategic level (Hoxtell 2024).

In terms of participatory democracy, which relies on active participation from the population, particular attention should be paid to the risks of *"an increase in state-orchestrated disinformation campaigns from autocratic states and constant changes to their forms"* and *"a loss of trust in democratic institutions and the media"*. If the loss of trust is caused by targeted disinformation,

the population not only loses interest in participating, but also the basis for informed participation, which in turn can lead to the long-term creeping loss of function of participatory elements within democratic systems.

REFERENCES

ACTION PLAN AGAINST DISINFORMATION, 2018. European Commission [online]. [viewed 05 April 2025]. Available from: https://commission.europa.eu/publications/action-plan-disinformation-commission-contribution-european-council-13-14-december-2018_en

CHLON, T., 2022. NATO and Countering Disinformation. GLOBSEC [online]. [viewed 05 April 2025]. Available from: https://www.globsec.org/sites/default/files/2022-05/NATO-and-Countering-Disinformation-ver1-spreads.pdf

DAASE, C.; KESSLER, O., 2007. Knowns and Unknowns in the "War on Terror": Uncertainty and the Political Construction of Danger. *Security Dialogue*, vol. 38, pp. 411–434. DOI: 10.1177/0967010607084994.

DESINFORMATION BEDROHUNG UND GEFAHR FÜR DEMOKRATIE UND GESELLSCHAFT. Konrad Adenauer Stiftung (KAS) [online], n.d. [viewed 05 April 2025]. Available from: https://www.kas.de/de/web/schwerpunktthemen/desinformation-bedrohung-und-gefahr

DESINFORMATION UND PROPAGANDA – AUSWIRKUNGEN AUF DAS FUNKTIONIEREN DES RECHTSSTAATS IN DER EU UND IHREN MITGLIEDSTAATEN: ZUSAMMENFASSUNG DES LIBE AUSSCHUSSES, 2019. Europäisches Parlament [online]. [viewed 05 April 2025]. Available from: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL_STU(2019)608864(SUM01)_DE.pdf

DIE HERAUSFORDERUNGEN DER DEMOKRATIE IM DIGITALEN ZEITALTER, 2019. *Internet Health Report* [online]. [viewed 05 April 2025]. Available from: https://internethealthreport.org/2019/die-herausforderungen-der-demokratie-im-digitalen-zeitalter/?lang=de

DIE JUGEND IN DER INFODEMIE: STUDIE, 2020. Vodafone Stiftung [online]. [viewed 05 April 2025]. Available from: https://www.vodafone-stiftung.de/wp-content/uploads/2020/12/Studie-Vodafone-Stiftung-Umgang-mit-Falschnachrichten.pdf

EDUCATED BUT EASILY FOOLED? WHO FALLS FOR MISINFORMATION AND WHY, 2025. Max-Planck-Gesellschaft [online]. [viewed 05 April 2025]. Available from: https://www.mpg.de/24132917/0205-bild-online-misinformation-149835-x1

HOXTELL, W., 2024. Umgang mit Desinformation in Europa. Herausforderungen und Gelegenheiten für zivilgesellschaftliche Organisationen und Privatsektor. Bertelsmann Stiftung [online]. [viewed 05 April 2025]. Available from: https://www.bertelsmann-stiftung.de/fileadmin/files/user_upload/Umgang_mit_Desinformation_in_Europa._Herausforderungen_und_Gelegenheiten_fuer_zivilgesellschaftliche_Organisationen_und_Privatsektor.pdf

MÜNDGES, S., 2019. Strategien der Europäischen Union gegen Desinformation [online]. [viewed 05 April 2025]. Available from: https://www.bpb.de/themen/medien-journalismus/digitale-desinformation/290565/strategien-der-europaeischen-union-gegen-desinformation/#footnode2-2

NATIONAL SECURITY CONCEPT OF ESTONIA, 2023. Government of Republic of Estonia [online]. [viewed 05 April 2025]. Available from: https://www.kaitseministeerium.ee/sites/default/files/eesti_julgeolekupoliitika_alused_eng_22.02.2023.pdf

NATIONAL STRATEGIC REVIEW, 2025. SGDSN [online]. [viewed 05 April 2025]. Available from: https://www.sgdsn.gouv.fr/publications/revue-nationale-strategique-2025

NATIONALE SICHERHEITSSTRATEGIE (NSS), 2023. Die Bundesregierung [online]. [viewed 05 April 2025]. Available from: https://www.nationalesicherheitsstrategie.de/Sicherheitsstrategie-DE.pdf

NATO'S APPROACH TO COUNTER INFORMATION THREATS, 2025. NATO [online]. [viewed 05 April 2025]. Available from: https://www.nato.int/cps/in/natohq/topics_219728.htm

QUESTIONS AND ANSWERS ABOUT THE EAST STRATCOM TASK FORCE, 2021. The Diplomatic Service of the European Union [online]. [viewed 05 April 2025]. Available from: https://www.eeas.europa.eu/eeas/questions-and-answers-about-east-stratcom-task-force_en

RESOLUTION 483: PROTECTING ALLIED DEMOCRACIES AND PARTNERS FROM DISINFORMATION, 2023. NATO Parliamentary Assembly [online]. [viewed 05 April 2025]. Available from: https://www.nato-pa.int/download-file?filename=/sites/default/files/2023-10/CDS%20resolution%20483%20-%20DISINFORMATION.pdf

SACHSTAND PE 6 - 3000 - 041/19: ZUM EU-KONZEPT ZUR BEKÄMPFUNG DER DESINFORMATION IM INTERNET, 2019. Deutscher Bundestag [online]. [viewed 05 April 2025]. Available from: https://www.bundestag.de/resource/blob/646542/b561aee1e9120b2818eabb8421482411/PE-6-041-19-pdf.pdf

SECURITY STRATEGY OF THE SLOVAK REPUBLIC [online], 2021 [viewed 05 April 2025]. Available from: https://www.enisa.europa.eu/sites/default/files/ncss-map/strategies/additional-documents/SK_SECURITY_STRATEGY_2021_en.pdf

STRATEGIA BEZPIECZEŃSTWA NARODOWEGO RZECZYPOSPOLITEJ POLSKIEJ, 2020. Rada Ministrów [online]. [viewed 05 April 2025]. Available from: https://www.bbn.gov.pl/ftp/dokumenty/Strategia_Bezpieczenstwa_Narodowego_RP_2020.pdf

STRATÉGIE DE SÉCURITÉ NATIONALE, 2022. Belgian Government [online]. [viewed 05 April 2025]. Available

from: https://www.egmontinstitute.be/belgian-national-security-strategy/
STRENGTHENED CODE OF PRACTICE ON DISINFORMATION: FACTSHEET, 2022. European Commission [online]. [viewed 05 April 2025]. Available from: https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation-factsheet
THE OECD TRUTH QUEST SURVEY: POLICY PAPER, 2024. OECD [online]. [viewed 05 April 2025]. Available from: https://www.oecd.org/en/publications/the-oecd-truth-quest-survey_92a94c0f-en.html
STUDY: ON TWITTER, FALSE NEWS TRAVELS FASTER THAN TRUE STORIES, 2018. Massachusetts Institute of Technology (MIT) [online]. [viewed 05 April 2025]. Available from: https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308
THE SECURITY STRATEGY FOR THE KINGDOM OF THE NETHERLANDS, 2023. Government of the Netherlands [online]. [viewed 05 April 2025]. Available from: https://www.government.nl/topics/security-strategy-for-the-kingdom-of-the-netherlands/documents/publications/2023/04/03/security-strategy-for-the-kingdom-of-the-netherlands
THE STATE DEFENCE CONCEPT [online], 2023. [viewed 05 April 2025]. Available from: https://www.mod.gov.lv/sites/mod/files/document/The%20State%20Defence%20Concept%202023-2027.pdf
TURCILO, L.; OBRENOVIC, M., 2020. Fehlinformationen, Desinformationen, Malinformationen: Ursachen, Entwicklungen und ihr Einfluss auf die Demokratie. Heinrich Böll Stiftung [online]. [viewed 05 April 2025]. Available from: https://www.boell.de/sites/default/files/2020-08/200825_E-Paper3_DE.pdf
UMFRAGE: DESINFORMATION UNTER JUNGEN MENSCHEN WEIT VERBREITET, 2025. Friedrich Naumann Stiftung [online]. [viewed 05 April 2025]. Available from: https://www.freiheit.org/de/deutschland/umfrage-desinformation-unter-jungen-menschen-weit-verbreitet
WELSCH, J., 2025. Digitalisierung: unsere Verantwortung. *Politicus*, 2025, no. 1, pp. 6–9 [online]. [viewed 05 April 2025]. Available from: https://www.hss.de/publikationen/?tx_hsspublication_pi1%5Baction%5D=download&tx_hsspublication_pi1%5Bcontroller%5D=Publication&tx_hsspublication_pi1%5Bpublication%5D=2574
WIDENING DIGITAL GAP BETWEEN DEVELOPED, DEVELOPING STATES THREATENING TO EXCLUDE WORLD'S POOREST FROM NEXT INDUSTRIAL REVOLUTION, SPEAKERS TELL SECOND COMMITTEE, 2023. United Nations (UN). [online] [viewed 05 April 2025]. Available from: https://press.un.org/en/2023/gaef3587.doc.htm

ДЕМОКРАЦИЯ НА УЧАСТИЕТО В ЕРАТА НА ДЕЗИНФОРМАЦИЯТА: СТРАТЕГИИ НА ЕС И НАТО ЗА ЗАЩИТА НА „ДЕМОКРАЦИЯТА НА УЧАСТИЕТО“ КАТО ЧАСТ ОТ НАЦИОНАЛНАТА СИГУРНОСТ

Резюме: *Демокрацията изисква свобода на изразяване. Съответни пасажи вече могат да бъдат намерени в Американската декларация за правата (1791) и Френската декларация за правата на човека и гражданина (1789). Това основно изискване обаче е формулирано по време на първата вълна на глобализацията. Днес обществото е в петата си вълна – и всяка вълна е довела до разширяване на свободата. Докато през XVIII век са съществували ограничени канали за информация, днес информационният пейзаж е разширен експоненциално благодарение на интернет, нарастващата дигитализация и изкуствения интелект. Това разширяване затруднява валидирането на информацията и насърчава разпространението на дезинформация, която невинаги е лесно разпознаваема като такава и по този начин позволява манипулиране на действията в ежедневието и политическите решения. Дезинформацията следователно влияе не само върху формирането на индивидуалното мнение, но и върху демократичните процеси на участие и може да има последици, свързани със сигурността. За да се смекчат тези последици, са необходими мерки и координирано изпълнение както на национално, така и на международно равнище. На този фон в настоящата статия се анализират настоящите стратегии на Европейския съюз (ЕС) и Организацията на Северноатлантическия договор (НАТО) за ограничаване на дезинформацията с цел защита на „демокрацията на участието“ като компонент на националната сигурност, прави се преглед на тяхната ефективност чрез SWOT анализ и примери за реални дезинформационни кампании и се разсъждава дали настоящите стратегии са насочени към потенциалните заплахи.*

Ключови думи: *демокрация на участието; национална сигурност; дезинформация; защита от дезинформация; международни стратегии*

Марина Винерс, докторант
Университет по библиотекознание и информационни технологии
София, България
E-mail: marina@wieners.eu

NATIONAL SECURITY
НАЦИОНАЛНА СИГУРНОСТ

AN AI-SUPPORTED REFERENCE PROCESS FOR RISK MANAGEMENT IN ACCORDANCE WITH ISO 27005

Sebastian Grüneberg
University of Library Studies and Information Technologies
https://doi.org/10.70300/YRHH6679

Abstract: *The increasing complexity of information security risks creates significant challenges for organizations in risk management. ISO 27005 provides a normative framework but does not explicitly specify how to implement individual process phases methodologically. While the use of artificial intelligence (AI) in risk management is gaining attention, consistent process-oriented integration has yet to be achieved. This article explores how to systematically integrate AI into the risk management process in accordance with ISO 27005 to ensure compliance with standards. Through document analysis of the standard and a systematic literature review, an AI-supported reference process was developed that positions AI support functions across all process phases and establishes clear decision and validation points. The findings indicate that AI is mainly useful for organizing and analyzing risk-related information, while humans retain responsibility for decisions with normative implications.*
Keywords: *ISO 27005; Artificial Intelligence; Information Security Risk Management; Large-Language-Models (LLM); Reference Process*

INTRODUCTION

The importance of systematic risk management has grown significantly in nearly all organizations in recent years. The ongoing digitalization and interconnectedness of business processes are creating risk environments characterized by high complexity, dynamic dependencies, and increasing amounts of risk-related information. Particularly in information security, traditional manual methods for identifying, analyzing, and prioritizing risks are increasingly reaching their limits. In this context, many scientific studies highlight that conventional risk analysis methods and models are becoming inadequate. These models typically rely on subjective expert assessments and static evaluation frameworks. Such models cannot keep up with the constantly evolving dynamics of modern risk scenarios. Meanwhile, artificial intelligence (AI) has proven to be a powerful tool for analyzing complex data and decision-making situations and is increasingly discussed as a potential aid in risk management.

ISO 27005 offers a well-established framework for information security risk management, but deliberately does not specify particular methods or tools for carrying out individual process steps. This flexibility enables adaptable applications but often leads to diverse approaches and a heavy reliance on personal expertise. Although the number of contributions on the use of AI in risk management is increasing, there is still a lack of a process-oriented, standard-compliant concept that systematically integrates AI into the process phases of ISO 27005 and clearly defines the limits of automation.

This study fills this gap by creating an AI-supported reference process for risk management based on ISO 27005, derived from a systematic literature review. The reference process links AI application areas to specific phases of ISO 27005 and emphasizes key decision points and validation checkpoints. This article intentionally does not aim to maximize automation in risk management. Instead, the reference process functions as a control and governance framework that ensures the standard-compliant, traceable, and responsible integration of AI.

RESEARCH METHODOLOGY

This study adopts a conceptual-analytical approach. Methodologically, it combines a document analysis of the standard with a systematic review of literature on the use of AI in information security and IT risk management. In the first step, ISO 27005 was analyzed to understand the nor-

mative process structure of risk management and identify process phases where the standard intentionally does not specify detailed methodological guidelines. This openness provides the foundation for exploring potential AI integration points within the normative framework. Based on these findings, a systematic literature review was carried out focusing on AI applications in information security and IT risk management. The review included relevant scientific publications that explicitly mention risk analysis, assessment, and management processes. The initial search yielded about 80 relevant papers. Through a multi-stage selection process, non-specialist contributions, purely technical studies lacking process references, and papers not applicable to the context of ISO 27005 were excluded, leaving 47 papers for analysis. During the analysis, the publications were assessed based on the normative process phases of ISO 27005. The individual studies were intentionally not evaluated for quality. Instead, the goal was to create a structured overview of AI use cases described in the literature. The following five analytical perspectives guided the literature review:

(1) Assignment of the contributions to the individual process phases,

(2) Identification of AI-supported tasks and the functional role of AI,

(3) Survey of relevant input and output artefacts,

(4) Identification of explicit decision and validation points, including human-in-the-loop requirements

(5) Recording of the stated prerequisites and limitations of AI use

The results of this analysis were summarized into a cross-phase synthesis, from which the AI-supported reference process was developed. The selected works are chosen to cover a representative range of AI application areas across the ISO 27005 process phases, without trying to be exhaustive regarding individual method classes.

RESULTS

The results of this article include the development of an AI-supported reference process for risk management aligned with ISO 27005. The following sections demonstrate how AI support functions, process artifacts, and decision and validation points can be systematically positioned along the individual process phases. The reference process is thus not presented as a collection of isolated AI use cases but as a coherent, process-oriented framework. It should be viewed as a control and governance structure that defines the conditions, limitations, and validation points for AI use within a normative risk management process. The assignment of AI support functions is based on a detailed assessment of process characteristics. Of particular importance are the normative influence of an activity, its context dependency, the reversibility of the generated results, and the stability and availability of the underlying data.

These differentiation criteria, shown in Table 1, are the result of a cross-phase analysis of the literature and the normative process logic of ISO 27005. They unify recurring arguments about the role of AI, responsibility assignment, and the management of uncertainty and data dependency into a consistent basis for evaluation. Based on this, the reference process distinguishes between AI support and human decision-making.

Table 1. Demarcation criteria AI vs. human

Criterion	Suitable for AI	Human required
Normative influence	Structuring, aggregation	Acceptance, prioritisation
Context dependency	Generic patterns	Organisation-specific meaning
Reversibility	Preliminary results	Binding decisions
Data stability	Structured, historical	Uncertain, incomplete

Figure 1 illustrates the AI-supported risk management process as defined in ISO/IEC 27005. The diagram shows AI support functions at each process stage and highlights decision-making and validation points for human actors.

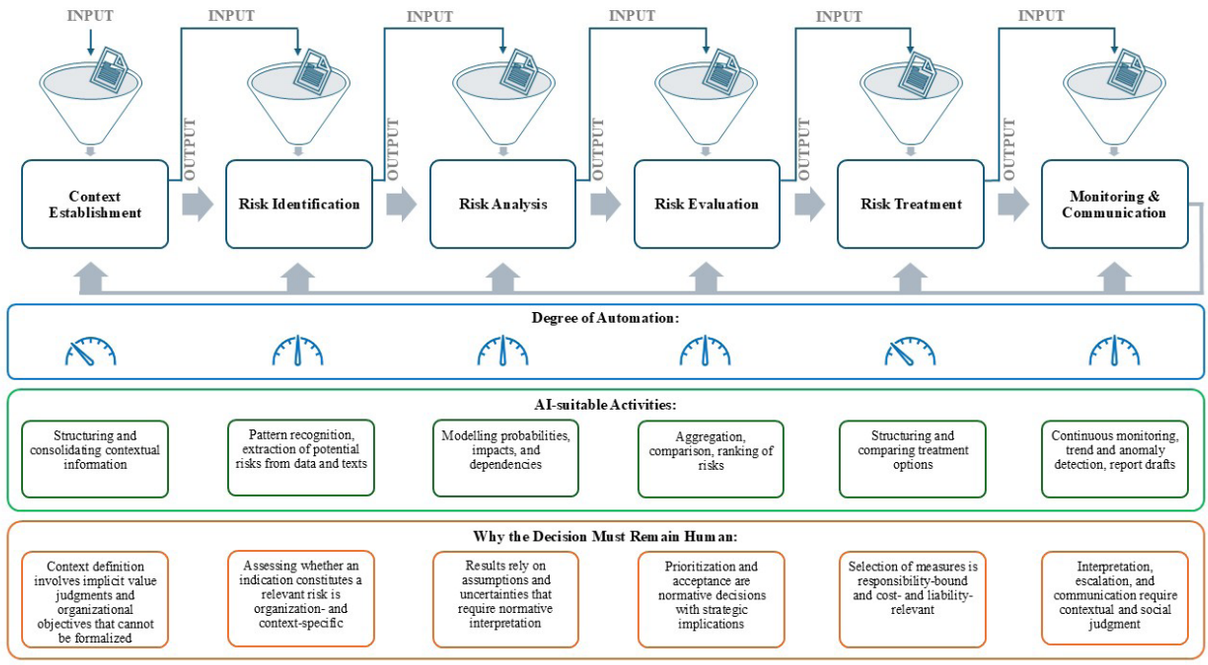


Fig. 1. Reference process (own illustration)

Context establishment forms the starting point of the risk management process according to ISO 27005 and defines the organizational, technical, and regulatory framework within which risks are identified, analyzed, and evaluated. These framework conditions include defining key reference points, such as scopes, risk criteria, and protection requirements, which significantly impact the further course of risk management (DIN Deutsches Institut für Normung e. V., 2024). The examined literature shows that AI methods are primarily used to establish context to support the organization and integration of diverse information. It focuses on describing AI-supported techniques for analyzing large documents, guidelines, system descriptions, and technical artifacts. The goal is to create a clear, understandable starting point for subsequent steps in the process. AI's role is limited to pre-structuring, categorizing, and aggregating activities. In this context, special attention is given to how taxonomies, ontologies, and knowledge graphs are used to ensure consistent mapping of organizational and technical contexts (Lourenço et al. 2025; Szczekocka 2024) and linking governance, risk and control structures (Eisenberg et al. 2025). Additionally, normative and risk theory contributions emphasize that defining context parameters always involves implicit value assumptions that cannot be resolved automatically (Aven 2016). Further work emphasizes the importance of early governance and architecture decisions in defining the context of AI-supported systems and explicitly prioritizes these before addressing the actual risk analysis (Bodnari and Travis 2025; Bogdanov et al. 2024; Tabassi 2023). Kandasamy (2024) also highlights that definitions of context are always shaped by implicit assumptions about system boundaries and knowledge availability. Relevant input artifacts include guidelines, process descriptions, architecture overviews, and asset lists. Structured context information is produced as output artifacts. These serve as a basis for identifying risks. However, defining risk criteria, protection requirements, and organizational scopes requires normative decisions that are directly linked to the organization's strategic objectives and responsibilities. These decisions are highly context-dependent and cannot be fully formalized. In this phase, humans are responsible for the definitive definition of context, while AI only performs preparatory and supportive functions.

In the **risk management** process, according to ISO 27005, risk identification marks the shift from describing the context to conducting operational analysis. During this phase, potential threats, vulnerabilities, and undesirable events are systematically documented to establish an initial foundation for the subsequent risk analysis (DIN Deutsches Institut für Normung e. V., 2019, 2024). The reviewed literature describes the application of AI to risk identification, particularly in

data-heavy settings where large volumes of diverse security-related information must be assessed. AI techniques facilitate automated analysis of internal and external data sources, including logs, monitoring data, vulnerability details, and external threat intelligence. The goal is to generate structured indicators of potential risks by detecting patterns, anomalies, and correlations among indicators. These include, in particular, approaches based on NLP and LLM for extracting risks from unstructured text sources (Begam et al. 2025; Hassani et al. 2024; Liu et al. 2025), as well as knowledge-based and graph-supported techniques for structured risk derivation (Bahr et al. 2025). Additionally, several papers describe using AI as a tool to expand the identification of the risk spectrum, such as through simulation-, evolution-, or scenario-based approaches. These methods are particularly aimed at identifying previously unknown or emerging risks (Chockalingam et al. 2017; Dass and Siami Namin 2020; Kalejaiye 2025; Kioskli et al. 2024; Takko et al. 2023). At the same time, empirical studies show that generative approaches specifically carry risks like hallucinations or misclassifications and therefore do not offer a reliable foundation for autonomous risk assessment (Collier et al. 2025; Esposito et al. 2024). The output artifacts of this process are pre-structured risk indicators or preliminary risk lists used as input for risk analysis. However, these indicators do not represent risks in the normative sense. The decision on whether identified indicators should be regarded as relevant risks depends on their classification within the organizational and regulatory context, as well as on implicit knowledge of business processes and dependencies. This validation cannot be fully automated and remains the responsibility of humans, while AI plays only a decision-preparatory role in risk identification.

As part of the **risk analysis**, the identified risks are evaluated based on their likelihood of occurrence and possible impact. This phase involves systematically organizing and modeling uncertainty, which provides the analytical foundation for the subsequent risk assessment. The examined literature indicates that AI methods are primarily used to support complex analysis and modeling tasks in risk analysis. Data-driven, probabilistic, and model-based approaches are discussed, considering historical incident data, dependency structures, or scenarios. AI consistently processes assessment parameters, organizes relevant influencing factors, and simulates potential developments. Examples include fuzzy-based and Bayesian models for representing uncertainty (Abdymanapov et al. 2021; Savchuk 2023) as well as graph- and deep learning-based methods for modeling risk dependencies and propagation mechanisms (Chhetri and Namin 2025; Shanthi et al. 2025). Additionally, several studies emphasize the importance of explainable models to ensure the traceability of analytical results (Enríquez 2023; Giudici and Raffinetti 2022). Furthermore, other studies employ probabilistic, Bayesian, and hybrid modeling approaches to explicitly address uncertainties, dependencies, and incomplete data sets in risk analysis (Ding and Liu 2012; Lezzi et al. 2025; Wang et al. 2024). The identified risks, relevant contextual information, and historical data serve as input. The output artifacts generated include the calculation of evaluation parameters, the creation of scenarios, and the recording of modeling results. However, interpreting these results requires managing uncertainties, assumptions, and data gaps that cannot be fully formalized. Assessing the significance and relevance of the modeled results relies on organization-specific knowledge and normative beliefs. Therefore, the interpretation and plausibility checking of the analysis results remain the responsibility of humans.

Risk evaluation aims to categorize and rank analysis results based on established criteria, providing a solid foundation for future risk management decisions. It marks the shift from analytical modeling to normative decision-making. The literature reviewed describes how AI is used during this phase, particularly to assist in comparing and structuring risks. AI techniques are employed in various areas, including aggregating assessment outcomes, prioritizing risks, and performing consistency checks. Based on this, AI creates structured summaries or rankings but does not make independent decisions. These approaches range from multi-criteria assessment methods to ensemble-based prioritization models (Campos et al. 2025; Giudici and Raffinetti 2022). Empirical studies show that LLMs, in particular, produce inconsistent results in numerical evaluation and scoring tasks and therefore do not offer a dependable basis for decision-making (Collier et al. 2025). Furthermore, risk theory emphasizes that quantitative risk values must always be interpreted within

the context of the underlying knowledge base (the strength of knowledge). Although AI can model probabilities and impacts, assessing epistemic uncertainty, implicit assumptions, and the robustness of available data requires human judgment (Aven 2016; Erdogan et al. 2021). Additionally, multiple studies have demonstrated that AI-based prioritization heavily relies on the fundamental assessment logic and data quality. Without human plausibility checks, this can lead to systematic biases (Ablayeva 2025; Kunle-Lawanson 2022). Relevant output artifacts include prioritized risk lists and aggregated assessment overviews. However, the decision on risk acceptance, escalation, or treatment remains a human responsibility, as it involves normative, strategic, and organizational aspects that cannot be fully formalized. The role of AI in risk assessment is therefore strictly limited to supporting structured decision-making processes.

Risk treatment involves choosing and applying measures to reduce, transfer, or accept risks. This stage marks the shift from assessment to action and is directly connected to organizational responsibility. The reviewed literature highlights the use of AI in risk treatment mainly to support the analysis and organization of potential measures. AI-supported techniques help, for example, assess measures based on their effectiveness, dependencies, and cost-benefit ratios, and provide an overview of alternative options. These include, in particular, methods for analyzing and optimizing security measures (Eke and Onyejegbu 2024; Finistrella et al. 2025) as well as AI-supported methods for generating and evaluating textual proposals for security measures (Xia et al. 2021). However, several studies emphasize that selecting and implementing measures cannot be automated because of their regulatory, economic, and organizational implications (Collier et al. 2025). Previous work has shown that although algorithmic optimization of policy options can lead to efficiency gains, the final choice must account for organization-specific trade-offs (Kirt and Kivimaab 2010). Relevant output artifacts include structured policy options and comparative overviews. The selection, prioritization, and implementation of specific measures require weighing effectiveness, costs, and feasibility and assessing regulatory requirements. This weighing is highly context-dependent and involves organizational responsibility and potential consequences. Accordingly, these decisions remain the responsibility of humans, with AI playing only a supporting role.

Risk monitoring and communication ensure that risks are continuously observed, assessed, and addressed to relevant stakeholders. This phase does not follow a sequential process but serves as ongoing feedback for all previous steps. The reviewed literature describes the use of AI in this phase, notably to support the continuous evaluation and processing of risk-related information. AI techniques can analyze real-time monitoring data, update risk assessments, and automatically generate reports or alerts. Additionally, learning-based approaches to anomaly and attack detection are discussed (Joshi and Shandilya 2024; Rihan et al. 2023; Yang et al. 2023) as well as explainable AI models to assist in alarm interpretation (Dilhara 2025; Zhang et al. 2022). Multiple studies show that without human validation, false alarms and alarm fatigue can happen (Almadhor et al. 2025; Sharma et al. 2025). Studies on the use of AI-based monitoring and alerting systems show that, especially in complex system landscapes, human interpretation remains necessary to prevent false alarms and contextless escalations (Alexandrov 2025; Korniszuk and Sawicki 2024). Humans remain responsible for assessing consequences, adapting measures, and communicating risk-relevant decisions. Particularly in communication, the literature highlights the importance of transparency, traceability, and interpretation suitable for the target audience. These requirements extend beyond the analytical capabilities of AI systems, making human control and responsibility essential in this phase.

The cross-phase evaluation demonstrates that the use of AI in risk management, as defined in ISO 27005, can be systematically differentiated along functional and normative boundaries. While AI helps structure, analyze, and consolidate risk-relevant information in all process phases, decisions with normative implications remain the responsibility of humans throughout. The reference process derived from this clearly shows this distinction by consistently placing AI support functions, process artifacts, and decision and validation points across all process phases. The reference process is therefore not meant as an architecture for automating risk decisions but as a framework for structuring and controlling the norm-compliant integration of AI into risk management, in accordance with ISO 27005.

DISCUSSION

The results of the study show that AI can be effectively used in risk management, in line with ISO 27005, as part of a process-oriented approach, provided its role is clearly defined and integrated into existing governance structures. Organizing the literature findings by individual process phases makes it clear that AI should not be seen as an isolated technology but rather as a supporting element within a normative risk management process. At the same time, the review of the existing literature reveals significant variability in the attribution of automation levels and responsibilities. This variability is less due to fundamental deficits in AI processes and more to differences in data situations, application goals, and organizational contexts. The developed reference process addresses these findings by consistently positioning AI as a decision-preparatory element and explicitly including human decision-making and validation points in all process phases.

Regarding LLMs in particular, current research increasingly highlights specific limitations. Empirical studies indicate that when LLMs are used in numerical assessments and scoring tasks, such as in FMEA-like procedures, results are often inconsistent or methodologically unreliable (Collier et al. 2025). While such models can efficiently organize textual information and describe risks linguistically, their effectiveness for quantitative risk assessments is still limited. Meanwhile, studies show that human experts generally offer more accurate and context-aware evaluations. AI systems, however, excel in speed, scalability, and providing action-oriented preliminary structuring, such as aggregated, prioritized, or similarly prepared information. (Esposito et al. 2024). These findings support the distinction made in the reference process between AI as an assistant and not as a decision-making authority. Furthermore, the results show that the use of AI is always linked to organizational, normative, and governance-related framework conditions. The reference process intentionally does not prescribe a single set of methods but allows for a context-dependent selection of suitable AI procedures. This flexibility reflects the different requirements of individual process phases and varying levels of maturity and data availability across organizations. Methodological diversity is viewed not as a weakness but as a necessary condition for the practical integration of AI. Another important aspect concerns the traceability and explainability of AI-supported results. Non-transparent models pose the risk that results will either be accepted uncritically or rejected outright. The reference process addresses this tension by positioning explainability not as an extra technical feature but as a procedural necessity for risk monitoring, communication, and governance. Consequently, explainability becomes essential to responsible decision support.

The article adds conceptual value by turning existing, often fragmented AI application areas into a consistent, standardized process logic. However, the reference process has limitations: it relies on a synthesis of the literature and has not been empirically tested in specific organizational settings. Different levels of maturity, data availability, and regulatory environments necessitate organization-specific adjustments. Promising directions for future research include empirical evaluations of the reference process, detailed analyses of individual process phases, and studies of specific regulatory or cultural conditions. Alternative approaches that aim for extensive automation of risk assessments or decisions are critically evaluated in light of the results of this article. These approaches often implicitly embed normative assumptions into models without making them transparent or verifiable. Under ISO 27005, however, risk decisions are closely tied to organizational responsibility and accountability. Therefore, the developed reference process consistently positions AI as a supporting, rather than a primary, element within normatively guided risk management.

CONCLUSIONS

This article explores the possibility of systematically integrating artificial intelligence into the risk management process in line with ISO 27005 without compromising normative principles, responsibilities, and decision-making structures. Based on a document review of the standard and a systematic review of the literature, an AI-supported reference process was developed that consistently assigns AI support functions to the different process phases while also establishing clear boundaries for automation.

The study's findings indicate that applying AI in risk management is especially important when a large amount of diverse information needs to be processed into a clear, analytical, and concise format. However, decisions with normative implications, particularly regarding the assessment, acceptance, and handling of risks, must still be made by humans. The reference process explicitly enforces this distinction and makes the decision and validation points visible throughout the entire process phases.

The main scientific contribution is not evaluating individual AI methods but rather the process-oriented synthesis of existing research into a norm-based control-and-governance framework. Therefore, the reference process should be seen not as an architecture for automated decision-making but as a guideline for the structured, transparent, and responsible integration of AI into risk management. Since the reference process is built on a synthesis of the literature, empirical studies on its practical application and context-specific adaptations in different organizational and regulatory environments are natural starting points for future research.

REFERENCES

ABDYMANAPOV, S. A., MURATBEKOV, M., ALTYNBEK, S. and BARLYBAYEV, A., 2021. Fuzzy Expert System of Information Security Risk Assessment on the Example of Analysis Learning Management Systems. *IEEE Access*, vol. 9, pp. 156556–156565.

ABLAYEVA, O., 2025. Comparative analysis of random forest, SVM, and LSTM algorithms for threat detection in internet domains. *International Journal of Artificial Intelligence*, vol. 5, no. 5, pp. 1499–1504. [online]. Available from: https://www.academicpublishers.org/journals/index.php/ijai/article/view/4679

ALEXANDROV, A., 2025. LSTM-RNN Method for Anomaly-Based Intrusion Detection Systems. [online]. Available from: https://ceur-ws.org/Vol-3971/paper02.pdf [viewed 29 December 2025].

ALMADHOR, A., ALSUBAI, S., KRYVINSKA, N., AL HEJAILI, A., BOUALLEGUE, B., AYARI, M. and ABBAS, S., 2025. Transfer Learning with XAI for Robust Malware and IoT Network Security. *Scientific Reports*, vol. 15, no. 1, p. 26971.

AVEN, T., 2016. Risk Assessment and Risk Management: Review of Recent Advances on Their Foundation. *European Journal of Operational Research*, vol. 253, no. 1, pp. 1–13.

BAHR, L., WEHNER, C., WEWERKA, J., BITTENCOURT, J., SCHMID, U. and DAUB, R., 2025. Knowledge Graph Enhanced Retrieval-Augmented Generation for Failure Mode and Effects Analysis. *Journal of Industrial Information Integration*, vol. 45, p. 100807.

BEGAM, F., MANIMEKALAI, K. and RANGASAMY, B., 2025. Transformer Models and Attention Mechanisms for Intelligent Cyber Threat Intelligence Extraction, pp. 325–354.

BODNARI, A. and TRAVIS, J., 2025. Scaling Enterprise AI in Healthcare: The Role of Governance in Risk Mitigation Frameworks. *NPJ Digital Medicine*, vol. 8, no. 1, p. 272.

BOGDANOV, D., ETTI, P., KAMM, L. and STOMAKHIN, F., 2024. Artificial Intelligence System Risk Management Methodology Based on Generalized Blueprints, pp. 123–140.

CAMPOS, S., PAPADATOS, H., ROGER, F., TOUZET, C., QUARKS, O. and MURRAY, M., 2025. A Frontier AI Risk Management Framework: Bridging the Gap Between Current AI Practices and Established Risk Management.

CHHETRI, B. and NAMIN, A. S., 2025. The Application of Transformer-Based Models for Predicting Consequences of Cyber Attacks.

CHOCKALINGAM, S., PIETERS, W., TEIXEIRA, A. and VAN GELDER, P., 2017. Bayesian Network Models in Cyber Security: A Systematic Review. In: LIPMAA, H., MITROKOTSA, A. and MATULEVIČIUS, R., eds. *Secure IT Systems*. Cham: Springer International Publishing, pp. 105–122.

COLLIER, Z. A., GRUSS, R. J. and ABRAHAMS, A. S., 2025. How Good Are Large Language Models at Product Risk Assessment? *Risk Analysis*, vol. 45, no. 4, pp. 766–789.

DASS, S. and SIAMI NAMIN, A., 2020. Evolutionary Algorithms for Vulnerability Coverage. In: *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*. Madrid, Spain, 13–17 July 2020. Piscataway, NJ: IEEE, pp. 1795–1801.

DILHARA, A., 2025. Impact of Machine Learning and AI on Cybersecurity Risks and Opportunities.

DIN DEUTSCHES INSTITUT FÜR NORMUNG E. V., 2019. *DIN EN IEC 31010:2019 Risk Management – Risk Assessment Techniques*. Berlin: Beuth Verlag.

DIN DEUTSCHES INSTITUT FÜR NORMUNG E. V., 2024. *DIN ISO/IEC 27005:2024 Information Security, Cybersecurity and Privacy Protection – Guidance on Managing Information Security Risks*. Berlin: Beuth Verlag.

DING, D. and LIU, X., 2012. Bayesian Methods with Application in Risk Analysis.

EISENBERG, I. W., GAMBOA, L. and SHERMAN, E., 2025. The Unified Control Framework: Establishing a Common Foundation for Enterprise AI Governance, Risk Management and Regulatory Compliance.

EKE, D. R. N. and ONYEJEGBU, P. L. N., 2024. A Deep Reinforcement Learning Deep Q Network-Based Approach

for Systematic Detection of Distributed Denial of Service Attacks in Software-Defined Networking.
ENRÍQUEZ, L., 2023. A Quantitative Approach to Artificial Intelligence Legal Risk Management. In: SPINDLER, G. and MURIEL CICERI, J. H., eds. *Challenges of Law and Technology – Herausforderungen des Rechts und der Technologie – Retos del Derecho y de la Tecnología*. Göttingen: Göttingen University Press, pp. 169–182.
ERDOGAN, G., GARCIA-CEJA, E., HUGO, Å., NGUYEN, P. H. and SEN, S., 2021. A Systematic Mapping Study on Approaches for AI-Supported Security Risk Assessment. In: *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*. Piscataway, NJ: IEEE, pp. 755–760.
ESPOSITO, M., PALAGIANO, F., LENARDUZZI, V. and TAIBI, D., 2024. Beyond Words: On Large Language Models Actionability in Mission-Critical Risk Analysis. In: *Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. Barcelona, Spain, 24–25 October 2024. New York, NY: ACM, pp. 517–527.
FINISTRELLA, S., MARIANI, S. and ZAMBONELLI, F., 2025. Multi-Agent Reinforcement Learning for Cybersecurity: Classification and Survey. *Intelligent Systems with Applications*, vol. 26, p. 200495.
GIUDICI, P. and RAFFINETTI, E., 2022. Explainable AI Methods in Cyber Risk Management. *Quality and Reliability Engineering International*, vol. 38, no. 3, pp. 1318–1326.
HASSANI, I. E., MASROUR, T., KOUROUMA, N., MOTTE, D. and TAVČAR, J., 2024. Integrating Large Language Models for Improved Failure Mode and Effects Analysis (FMEA): A Framework and Case Study. *Proceedings of the Design Society*, vol. 4, pp. 2019–2028.
JOSHI, I. P. and SHANDILYA, V. K., 2024. A Critical Review – Use of Ensemble Methods in Intrusion Detection System. *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 21S, pp. 1157–1164. [online]. Available from: https://ijisae.org/index.php/IJISAE/article/view/5517
KALEJAIYE, A. N., 2025. Federated Learning in Cybersecurity: Privacy-Preserving Collaborative Models for Threat Intelligence Across Geopolitically Sensitive Organizational Boundaries. *International Journal of Research Publication and Reviews*, vol. 6, no. 7, pp. 227–249.
KANDASAMY, U. C., 2024. Ethical Leadership in the Age of AI: Challenges, Opportunities and Framework for Ethical Leadership.
KIOSKLI, K., RAMFOS, A., TAYLOR, S., MAGLARAS, L. and LUGO, R., 2024. Optimizing AI System Security: An Ecosystem Recommendation to Socio-Technical Risk Management. In: *Human Factors in Design, Engineering, and Computing*. AHFE International.
KIRT, T. and KIVIMAAB, J., 2010. Optimizing IT Security Costs by Evolutionary Algorithms. In: *Conference on Cyber Conflict Proceedings*, pp. 145–160. [online]. Available from: https://ccdcoe.org/uploads/2018/10/1_Proceedings2010FullBook.pdf [viewed 29 December 2025].
KORNISZUK, K. and SAWICKI, B., 2024. Autoencoder-Based Anomaly Detection in Network Traffic, pp. 1–4.
KUNLE-LAWANSON, O., 2022. The Role of AI in Information Security Risk Management. *World Journal of Advanced Engineering Technology and Sciences*, vol. 7, no. 2, pp. 308–319.
LEZZI, M., MONTEFUSCO, P., LAZOI, M. and CORALLO, A., 2025. AI-Based Cybersecurity for a Sustainable Digital Industry: Systematic Literature Review and Future Research Directions. *Journal of Industrial Information Integration*, vol. 48, p. 100980.
LIU, G., LU, K. and PI, S., 2025. Graph Neural Networks Embedded with Domain Knowledge for Cyber Threat Intelligence Entity and Relationship Mining. *PeerJ Computer Science*, vol. 11, e2769.
LOURENÇO, B., ADÃO, P., FERREIRA, J. F., MARQUES, M. M. and VAZ, C., 2025. Structuring Security: A Survey of Cybersecurity Ontologies, Semantic Log Processing, and LLM Applications.
RIHAN, S. D. A., ANBAR, M. and ALABSI, B. A., 2023. Meta-Learner-Based Approach for Detecting Attacks on Internet of Things Networks. *Sensors*, vol. 23, no. 19.
SAVCHUK, V., 2023. Bayesian Risk Assessment Technique for Economic Stress-Strength Models. [online]. Available from: https://mpra.ub.uni-muenchen.de/119078/ [viewed 29 December 2025].
SHANTHI, S., S, S. S. and SINDHU, M., 2025. Graph-Based Risk Analysis Using Graph Neural Networks for Mapping Cyber Threat Propagation in Large-Scale Networks, pp. 72–106.
SHARMA, A., RANI, S. and SHABAZ, M., 2025. A Comprehensive Review of Explainable AI in Cybersecurity: Decoding the Black Box. *ICT Express*, vol. 11, no. 6, pp. 1200–1219.
SZCZEKOCKA, E., 2024. AI Risk Management Implementation Challenges. *Computer Science & Information*, vol. 14, no. 17. [online]. Available from: https://aircconline.com/csit/papers/vol14/csit141710.pdf [viewed 29 December 2025].
TABASSI, E., 2023. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. Gaithersburg, MD: National Institute of Standards and Technology (NIST).
TAKKO, T., BHATTACHARYA, K., LEHTO, M., JALASVIRTA, P., CEDERBERG, A. and KASKI, K., 2023. Knowledge Mining of Unstructured Information: Application to Cyber Domain. *Scientific Reports*, vol. 13, no. 1, p. 1714.
WANG, L., CHENG, Y., XIANG, A., ZHANG, J. and YANG, H., 2024. Application of Natural Language Processing in Financial Risk Detection.
XIA, Y., JAZDI, N. and WEYRICH, M., 2021. Enhance FMEA with Large Language Models for Assisted Risk Man-

agement in Technical Processes and Products. *Computers in Industry*, vol. 131.
YANG, A., LU, C., LI, J., HUANG, X., JI, T., LI, X. and SHENG, Y., 2023. Application of Meta-Learning in Cyberspace Security: A Survey. *Digital Communications and Networks*, vol. 9, no. 1, pp. 67–78.
ZHANG, Z., HAMADI, H. A., DAMIANI, E., YEUN, C. Y. and TAHER, F., 2022. Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research. *IEEE Access*, vol. 10, pp. 93104–93139.

ПОДПОМАГАН ОТ ИЗКУСТВЕН ИНТЕЛЕКТ РЕФЕРЕНТЕН ПРОЦЕС ЗА УПРАВЛЕНИЕ НА РИСКА В СЪОТВЕТСТВИЕ С ISO 27005

Резюме: *Нарастващата сложност на рисковете за информационната сигурност създава значителни предизвикателства за организациите в управлението на риска. ISO 27005 предоставя нормативна рамка, но не посочва изрично как да се прилагат методологично отделните фази на процеса. Въпреки че използването на изкуствен интелект (ИИ) в управлението на риска привлича все по-голямо внимание, все още не е постигната последователна интеграция, ориентирана към процесите. В тази статия се разглежда как систематично да се интегрира ИИ в процеса на управление на риска в съответствие с ISO 27005, за да се гарантира съответствие със стандартите. Чрез анализ на документацията на стандарта и систематичен преглед на литературата е разработен референтен процес, подкрепен от ИИ, който позиционира ИИ поддържащите функции във всички фази на процеса и установява ясни точки за вземане на решения и валидиране. Резултатите показват, че ИИ е полезен главно за организиране и анализиране на информация, свързана с риска, докато хората запазват отговорността за решенията с нормативни последствия.*
Ключови думи: *ISO 27005; изкуствен интелект; управление на риска за информационната сигурност; големи езикови модели (LLM); референтен процес*

Себастиан Грюнеберг, докторант
Researcher ID: 0009-0002-2339-0729 (ORCID)
Университет по библиотекознание и информационни технологии
София, България
E-mail: Sebastian.Grueneberg@outlook.com

НАЦИОНАЛНА СИГУРНОСТ
NATIONAL SECURITY

ОТ ОРБИТА ДО БРЕГА: ИНСТРУМЕНТИ С ИЗКУСТВЕН ИНТЕЛЕКТ ЗА ОТКРИВАНЕ НА НЕФТЕНИ РАЗЛИВИ С ПОМОЩТА НА ПЛАТФОРМИ С ОТВОРЕНИ ДАННИ

Васил Михалков
Военна академия „Георги С. Раковски“
https://doi.org/10.70300/XNLI6419

Резюме: *Своевременното откриване на морски нефтени разливи е критично глобално екологично предизвикателство. Традиционните методи за мониторинг с корабни патрули и въздушно наблюдение са ограничени от разходите, обхвата и времето. В ход е промяна на парадигмата, подхранвана от сливането на три трансформативни технологии: сателитни данни с отворен код, безпилотни летателни апарати (БЛА) и изкуствен интелект (ИИ). Този синтез разглежда настоящата екосистема от инструменти, задвижвани от изкуствен интелект, които използват свободно достъпни данни от орбитални платформи и изображения, базирани на дронове, за да автоматизират и подобрят наблюдението на нефтени разливи. Анализиран е стандартният процес на обработка с изкуствен интелект – от предварителна обработка на данни и класическо машинно обучение до усъвършенствани модели за сегментиране с дълбоко обучение като U-Net. Бъдещата траектория сочи към интегрирани, многоплатформени системи за мониторинг, където изкуственият интелект не само открива, но и прогнозира динамиката на разливите, изграждайки мощен набор от инструменти с отворен достъп за опазване на морските екосистеми.*
Ключови думи: *откриване на нефтени разливи; изкуствен интелект; безпилотни летателни апарати (UAV); сателити с отворени данни; наблюдение на околната среда*

ВЪВЕДЕНИЕ

Разливите на нефт в морето, независимо дали са причинени от катастрофални изригвания на кладенци, аварии на танкери, или хронични оперативни изхвърляния, представляват сериозна и трайна заплаха за океанските екосистеми, крайбрежните икономики и общественото здраве. Неотложната необходимост от бърза реакция е неразривно свързана с точността и скоростта на откриване и картографиране. В продължение на десетилетия наблюдението се основава на конвенционални методи: патрулиране на кораби, пилотирано въздушно наблюдение и спорадичен сателитен анализ. Тези подходи, макар и ценни, образуват разпокъсана система, страдаща от високи оперативни разходи, ограничено пространствено и времево покритие, както и уязвимост към неблагоприятни метеорологични условия. Към това трябва да се добави и зависимост от субективната човешка интерпретация.

Нова технологична триада, в която влизат дистанционното сондиране с отворени данни, дроновете и изкуственият интелект (Fingas, Brown 2018; Leifer et al. 2012), премахва тези бариери и „демократизира“ наблюдението на околната среда. Стартирането на програмата „Коперник“ от Европейската космическа агенция, заедно с нейното съзвездие от спътници Sentinel, предостави на света безплатни, систематични и висококачествени данни от наблюдение на Земята. Успоредно с това разпространението на съвременни и достъпни безпилотни летателни апарати (БЛА) (Jiao et al. 2019) дава възможност за дистанционно наблюдение в локален мащаб, позволявайки на агенции и изследователи да разврващат сензори при поискване, заснемайки изображения и данни със сантиметрова разделителна способност, недостъпни от орбита.

ИЗТОЧНИЦИ НА ДАННИ И ФИЗИЧЕСКИ ПРИНЦИПИ ЗА ОТКРИВАНЕ НА НЕФТЕНИ РАЗЛИВИ

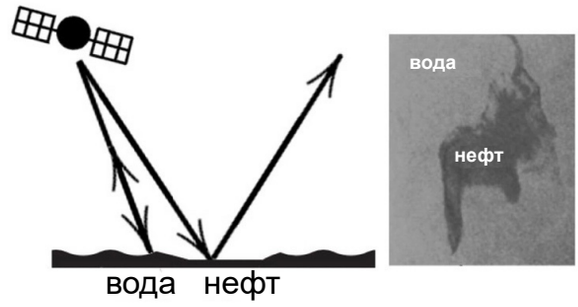
Ефикасността на всеки инструмент за откриване с изкуствен интелект е фундаментално ограничена от качеството и характеристиките на входните му данни. Този раздел очертава двата основни потока от данни с отворен код – сателитно и дистанционно наблюдение, чрез безпилотни летателни апарати и изяснява ключовите физически принципи, които позволяват разграничаването на нефтени разливи от морския фон в различни сегменти на електромагнитния спектър.

Спътниково наблюдение за откриване на нефтени разливи

Сателитните платформи осигуряват незаменимата широкообхватна, многократно преминаваща способност, която е от решаващо значение за първоначалното откриване и наблюдение на разливи в обширни и често отдалечени океански региони. Сред изобилието от сензори два основни типа, използващи различни физически взаимодействия, формират съвременното наблюдение на нефтени разливи чрез свободно достъпни отворени данни.

Радарът със синтезирана апертура (SAR) (Brekke, Solberg 2005) е приет за най-надеждния сателитен инструмент за откриване на нефтени разливи, главно поради оперативната си независимост от слънчевата светлина и облачната покривка. Принципът на действие се основава на промяната на грапавостта на морската повърхност.

По същество SAR сензорът е радар. Сателитът активно излъчва микровълнови импулси към земната повърхност и измерва обратно разсеяния сигнал. Интензитетът на обратно разсеяния сигнал, който се появява като яркост (или „сигма нула“) в полученото изображение, е предимно функция на локалния релеф на повърхността, по-специално капилярни и къси гравитационни вълни, резултат от т.нар. разсейване на Брег. Наличието на нефтен филм върху морската повърхност заглушава тези малки вълни поради по-високия вискозитет и повърхностно напрежение на нефта. Това потискане води до значително намаляване на радарното обратно разсейване, което кара покритите с нефт области да изглеждат като отчетливи тъмни образувания („тъмни петна“) на фона на по-светлата и по-грапава околна морска вода (фиг. 1).



Фиг. 1. SAR сателит, който засича спокойна зона на нефтено петно. Зоната лесно може да се обърка с други спокойни региони, като например зони със слаб повърхностен вятър или цъфтеж на водорасли.

Ключова платформа за свободен достъп е платформата „Коперник“, с данни от съзвездието спътници Sentinel-1 на Европейската космическа агенция (с нейния SAR сензор в C-обхвата) (Torres et al. 2012). Това е водещият свободно достъпен източник на SAR данни за оперативен мониторинг на нефтени разливи. Неговото гарантирано, систематично глобално покритие с честота на повторно посещение от 6 дни на Екватора с два спътника и значително по-често – на по-високи географски ширини, осигурява безпрецедентен оперативен набор от данни.

Основното предимство на SAR е неговата всесезонна и денонощна работа. Полезността му обаче може да бъде редуцирана при условия на много слаб вятър (където морето е вече спокойно, което намалява контраста) или много силен вятър (където нефтените петна могат да бъдат разкъсани или потопени). Освен това „тъмните петна“ на SAR не са уникални

за петрола. Те могат да бъдат объркани с други явления, като например естествени петна от водорасли, зони на ветрови сенки (например зад острови) или зони на издигане на повърхността, което води до критичното предизвикателство на „приликите“. Ето защо като алтернатива често се използват и оптични сензори за наблюдение във видимия светлинен диапазон, което позволява извличане на спектрална характеристика на наблюдавания обект.

Оптичните сензори във видимия, а също в инфрачервения (IR) и ултравиолетовия (UV) диапазон измерват отразената или излъчена слънчева радиация в специфични спектрални ленти, предлагайки допълнителна химическа и физическа информация за повърхностните петна (Chuanmin et al. 2021). Те предоставят богата спектрална информация, която може да помогне за разграничаване на видовете петрол, оценка на дебелината на пластовете и идентифициране на подобни обекти по-ефективно, отколкото само чрез SAR. Сравнително високата пространствена разделителна способност на Sentinel-2 е особено ценна за наблюдение в близост до брега. Основното и често възпиращо ограничение е неговата зависимост от лоши метеорологични условия и използването му само през деня. Облачността, мъглата и нощното време правят оптичните данни неизползваеми за откриване на разливи.

Безпилотни апарати за откриване на нефтени разливи

Дистанционното наблюдение, базирано на безпилотни летателни апарати, предоставя канал за верифициране на нефтени разливи с висока разделителна способност. Безпилотните летателни апарати (БЛА) или дронове работят в долните слоеве на атмосферата, като запълват критичната празнина между сателитното наблюдение и вземането на проби на място. Те предоставят изключително гъвкава платформа за дистанционно наблюдение по заявка.

Принципите на откриване на нефтени разливи с безпилотни летателни апарати (БЛА) са аналогични на сателитното оптично наблюдение, но се прилагат в значително по-висока пространствена резолюция (от сантиметри до метри). Това позволява визуализация на текстурите на маслените емулсии, диференциране на блясъка от по-дебелите слоеве на петното и директна корелация на спектралните и термичните сигнатури с точни местоположения за вземане на проби.

Безпилотните летателни апарати (БЛА) варират от леки многороторни системи, идеални за локализирани, стационарни инспекции, до самолети с фиксирани крила, които са способни да покриват по-големи крайбрежни зони. Последните могат да бъдат оборудвани с разнообразен полезен товар. Цветните (RGB) камери с висока разделителна способност създават подробна визуална документация и картографиране на обхвата и морфологията на петното. За да заснемат подробни спектрални сигнатури, мулти- и хиперспектралните сензори позволяват разграничаване на видовете петрол и изчисляване на усъвършенствани индекси. Термични инфрачервени (TIR) камери помагат за картографиране на топлинни контрасти, показващи вариации в дебелината на нефтения филм, ключов параметър за оценка на обема и планиране на отстраняване на проблеми. LiDAR може да се използва за измерване на много фина повърхностна топография и дебелината на нефтения филм при специфични условия (Zhengsin et al. 2023).

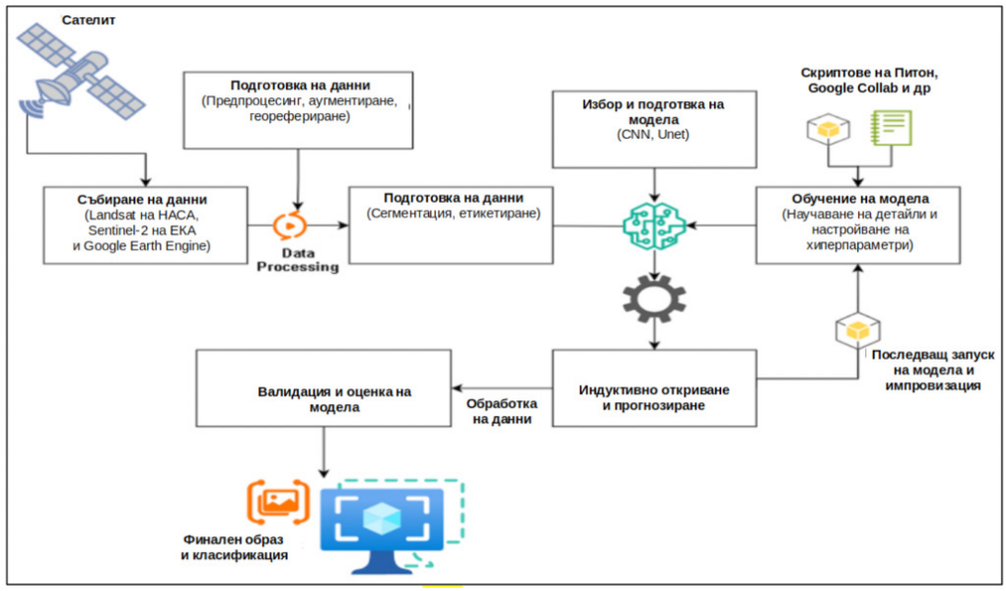
Най-голямото предимство на безпилотните летателни апарати (БЛА) (Krestenitis et al. 2019) са пространствената разделителна способност и гъвкавостта на разполагане. Те могат да бъдат мобилизирани бързо в отговор на сателитно предупреждение, да летят под облачна покривка и да предоставят данни с несравними детайли за криминалистичен анализ и оценка на щетите. Тяхната роля в събирането на наземни данни за обучение и валидиране на сателитни модели с изкуствен интелект е незаменима. Ограниченията включват ограничен оперативен обхват и продължителност на полета поради живота на батерията, капацитета на полезния товар и регулаторните ограничения на въздушното пространство, което ги ограничава предимно до крайбрежни и близо до брега операции.

Безпилотните летателни апарати (БЛА) предоставят окончателна, висококачествена наземна информация и подробна характеристика. Ефективната система за откриване, основана на изкуствения интелект, няма да разчита на един-единствен източник, а ще интегрира данни от мултиплатформена и мултисензорна екосистема с свободен достъп. Именно изкуст-

веният интелект трябва да превърне тези разнообразни потоци от данни в съгласуван набор от инструменти за откриване и анализ. Именно тук третият стълб – изкуственият интелект (ИИ), и по-специално дълбокото обучение (ДО), се превръща в незаменим интерпретатор.

ПРИЛОЖЕНИЕ НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ ЗА ИНТЕГРИРАНЕ И АНАЛИЗ НА ДАННИ ОТ САТЕЛИТНО И БЕЗПИЛОТНО НАБЛЮДЕНИЕ

Трансформирането на суровите сензорни данни в надеждна карта за откриване на нефтени разливи е многоетапен изчислителен процес. Този раздел анализира стандартния процес при анализ на данни с изкуствения интелект, който започва с подготовката на данните, преминава през извличане и инженеринг на основните характеристики и завършва с усъвършенствани архитектури за дълбоко обучение, представляващи съвременното ниво на развитие в областта. Този работен процес, визуализиран на фиг. 2, формира гръбнака на всички съвременни автоматизирани инструменти за откриване на обекти.



Фиг. 2. Диаграма на съвременния автоматизиран работен процес за обработка с изкуствения интелект

Предварителна обработка и подготовка на данните

Преди да може да се приложи който и да е модел с изкуствен интелект, суровите изображения трябва да бъдат калибрирани, почистени и структурирани във форма, подходяща за алгоритмично обучение. Този етап е критичен, тъй като качеството на входните данни пряко определя производителността на цялата система.

Преди доминирането на дълбокото обучение откриването на нефтени разливи се е основавало на извличане на ръчно създадени характеристики от предварително обработени изображения, които след това са били въвеждани в класическите класификатори за машинно обучение (Suresh et al. 2024). Процесът започва с извличане на характеристики – за всяка област (тъмно петно в SAR, аномалия в оптичния анализ) се изчислява набор от количествени параметри. Геометричните характеристики включват: площ, периметър и сложност, като компактност и фрактална размерност. Радиометричните характеристики са средна стойност, стандартно отклонение и контраст на стойностите на пикселите в петното и неговия фон. Текстурните характеристики се извличат от матрици за корелация на нива на сивото (GLCM) и включват параметри като хомогенност, контраст, ентропия и корелация – нефтените петна често показват по-гладка текстура от заобикалящото ги бурно море. Следват ги контекстуални характеристики, като разстояние от брега, близост до корабни пътища или нефтени платформи.

Всички горепосочени характеристики се подават към алгоритми за класификация, за да

се обучат класификаторите да различават „нефтен разлив“ от „приличащи“, например район със слаб вятър, биогенен филм и т.н. Тези алгоритми за класификация използват методи като „случайна гора“ (random forest) и машина на опорни вектори. Първият е метод, който конструира множество дървета на решенията, много ефективен за тази задача поради своята устойчивост на шум и способността си да обработва многомерни пространства от характеристики. Вторият се стреми да намери оптималната хиперравнина, която максимално да разделя различните класове в многомерното пространство от характеристики. Този метод е особено ефективен при ясно разграничаване на характеристиките.

Описаният горе подход е силно зависим от качеството и релевантността на ръчно проектираните характеристики. Процесът често отнема повече време, специфичен е за дадена област и може да не се обобщава добре за нови сензори или сложни, нелинейни модели в данните. За SAR изображенията (предимно Sentinel-1) първата стъпка е радиометричното калибриране и корекцията на терена. Суровите числови данни се преобразуват в калибрирани стойности на коефициента на обратно разсейване на радара (σ°), като осигуряват физическа съгласуваност между изображенията. Корекцията на терена с помощта на цифров модел на релефа (DEM) премахва геометричните изкривявания. SAR изображенията също така са по своята същност засегнати от „петна“ – гранулиран шум, причинен от интерференция на кохерентни вълни. Адаптивни филтри (например Lee, Frost или Refined Lee филтри) се прилагат за потискане на петната, като същевременно се запазват ръбовете и характеристиките, които са от решаващо значение за откриване на хлъзгави граници. Последната стъпка е маскиране на сухите райони с помощта на глобални набори от данни за бреговата линия, за да се предотвратят фалшиви положителни резултати от спокойни вътрешни води или радарни сенки от топографията.

За оптични изображения (отново Sentinel-2, Landsat и др.) обработката започва с атмосферна корекция. Прилагат се алгоритми (например Sen2Cor, ACOLITE) за преобразуване на сиянието на горните слоеве на атмосферата в повърхностно отражение, като се премахват ефектите на разсейване и абсорбция от атмосферата. Следващата стъпка е маскиране на облаците и сенките от облаците. Пикселите на облаците трябва да бъдат надеждно идентифицирани и маскирани с помощта на специални алгоритми (например Fmask, s2cloudless) или прагове на лентите (например с използване на SWIR или облачни ленти). Като последна стъпка, слънчевият отблясък трябва да бъде филтриран. За сцени, засегнати от огледално отражение, може да се приложат емпирични (например с използване на NIR лентата) или физически модели за смекчаване на отблясъците, които могат да замъглят или имитират петролни сигнатури.

Изборът на данни за обучение играе решаваща роля. Моделите с контролиран изкуствен интелект изискват големи, висококачествени етикетирани набори от данни. Всяко изображение трябва да има съответстваща маска за „ground truth“, където пикселите са етикетирани като „нефт“, „вода“, „подобни“ или „земя“. Създаването на тези етикети е съществено затруднение. Те се създават чрез ръчна анотация, като за основа се използват исторически случили се нефтени разливи (например Deepwater Horizon, 2010 г., разливът на остров Маврикий, 2020 г.), или посредством полуавтоматично генериране чрез използване на традиционни алгоритми за откриване (например постоянна честота на фалшиви аларми – CFAR за SAR), с последващо експертно уточняване. Синтезът на данни при използване на безпилотни летателни апарати (UAV) осигурява проверка. Използването на прецизно геореферирани изображения и спектрални данни от безпилотни летателни апарати служи за валидиране и коригиране на етикети. Известни отворени набори от данни, като например тези от Службата за реагиране и възстановяване на Националната агенция за океаните и атмосферата на САЩ (NAOA) или анотираните сцени от Sentinel-1 в каталога за отворена наука на Европейската космическа агенция, са безценни споделени ресурси за обучение и сравнителен анализ.

Дълбокото обучение за откриване на нефтени разливи

Дълбокото обучение (DL), по-специално конволюционните невронни мрежи (CNN), ко-

ренно променя съществуващата парадигма, като изучава йерархични характеристики директно от суровите или минимално предварително обработени изображения, автоматизирайки процеса на проектиране на характеристики (Ronneberger et al. 2015; Chen et al. 2018).

Семантичните модели на сегментиране (пикселна класификация) са доминиращият DL подход за откриване на нефтени разливи, като на всеки пиксел в изображението се присвоява етикет „нефт“ или „вода“. U-Net архитектурата е де факто стандартът за сегментиране на биомедицински и екологични изображения. Нейната симетрична структура на енкодер-декодер, с прескачащи връзки, ѝ позволява да улавя както контекст на високо ниво („какво“ на петното), така и прецизна локализация („къде“ на неговите граници). Вариации като Attention U-Net или ResU-Net (с остатъчни блокове) са често срещани подобрения, които подобряват производителността при сложни петна. DeepLab (DeepLabv3+) използва разширени конволюционни методи за улавяне на многомащабна контекстуална информация без загуба на разделителна способност и структура енкодер-декодер за прецизиране на границите на обектите.

Трансформаторите за зрение (ViT) (Zheng et al. 2021; Dosovitskiy et al. 2020; Xie et al.) са нововъзникваща ИИ архитектура, която третира изображението като поредица от петна. Чрез прилагане на механизми за самовнимание ViT могат да моделират зависимости на дълги разстояния в цялата сцена, потенциално подобрявайки разбирането на морфологията на петната и взаимодействието им с фона. Те все по-често се прилагат в задачи за наблюдение на Земята, но изискват големи количества данни за обучение.

Моделите за откриване на обекти (Slick-Level Detection) включват рамки като Faster R-CNN и YOLO. Те са по-рядко срещани, но могат да се използват, когато задачата е формулирана като откриване на дискретни екземпляри на нефтен разлив, а не като извършване на класификация на плътни пиксели. Те извеждат ограничаващи рамки и оценки за достоверност за отделните разливи. Обучението на тези модели изисква значителна изчислителна мощност (обикновено графични процесори) и големи, разнообразни набори от данни с етикети, за да се предотврати преобучение. Основно предизвикателство остава дисбалансът в класовете – пикселите с нефт са значително по-малко от пикселите с вода. Техники като претеглени функции за загуба (например загуба на зарчета, фокална загуба) или увеличаване на данните (въртене, обръщане, добавяне на шум) са от съществено значение за справянето с това.

Постобработка и интеграция

Суровият краен резултат от DL модел рядко представлява напълно завършена и оперативно използваема карта на откритите обекти (нефтени разливи). Необходима е постобработка за усъвършенстване на резултата. Използваните методи включват морфологични операции: прилагане на „затваряне“ (разширение, последвано от ерозия) за запълване на малки дупки в открито петно или „отваряне“ (ерозия, последвана от разширение) за премахване на изолирани малки фалшиво положителни пиксели. Анализът на свързаните компоненти е процесът на идентифициране и етикетиране на отделни обекти на нефтеното петно, което позволява изчисляването на ключовите му показатели: обща площ, периметър, местоположение и ориентация на главната ос. За сателитни времеви серии от изображения проследяването на движението и еволюцията на петно през последователни наблюдения, с използване на алгоритми за съвпадение на формата или вероятностни алгоритми, позволява моделиране на дрейфа и идентифициране на източника.

Процесът на обработка на изображения с помощта на изкуствения интелект се е развил от многоетапен и зависим от човешко участие до по-самостоятелен и основан на данни подход за дълбоко обучение. Докато класическото машинно обучение (ML) остава обект на интерпретация, който несъмнено е полезен за специфични задачи, моделите за дълбоко обучение (DL), особено архитектурите, базирани на U-Net, демонстрират превъзходна точност и устойчивост при справянето със сложни, нелинейни модели на нефтени разливи в различни морски среди. Способността им да се възползват директно от нарастващия обем данни с отворен код, е ключов двигател за тяхното все по-често използване.

ПОДХОД ЗА ИЗГРАЖДАНЕ НА МНОГОПЛАТФОРМЕНА РАМКА ЗА НАБЛЮДЕНИЕ С ИЗКУСТВЕН ИНТЕЛЕКТ

По-горе са описани подробно компонентите на инструментариума за откриване на нефтени разливи. Интеграцията на тези елементи дава възможност за сравнение на възможностите, производителността и оптималните случаи на употреба на различните платформи и подходи с изкуствен интелект. Според съществуващата литература нито една платформа или алгоритъм не са универсално превъзходни; по-скоро бъдещето е в интелигентното обединяване на данни и модели от орбитални и въздушни платформи с отворен код в сплотена, управлявана от изкуствения интелект рамка за наблюдение.

Дефиниране на оперативна ниша чрез матрица на платформата и възможностите

Изборът на платформа за наблюдение диктува мащаба, разделителната способност и вида на получената информация, определяйки нейната ниша във веригата за мониторинг. Таблица 1 синтезира основните характеристики и последиците от изкуствения интелект на всяка основна платформа с отворен код.

Таблица 1. Сравнителен анализ на платформи с отворен код за откриване на нефтени разливи, задвижвани от изкуствения интелект

Платформа и основен сензор	Пространствена разделителна способност	Времева рамка	Ключови характеристики	Основни силни страни на ИИ метода	Основни ограничения и предизвикателства пред ИИ
Сателит: SAR (например Sentinel-1)	~10 – 40 м	12 дни (глобално)	Затихване на разсейването на Браг (грапавост на повърхността).	Работа при всякакви метеорологични условия, денем/нощем.	Разграничаване по сходство (слаб вятър, биогенни филми).
Сателит: Оптичен/Мултиспектрален (например Sentinel-2)	10 – 60 м	5 дни (глобално)	Спектрално поглъщане, отражение и термичен контраст.	Позволяват разграничаване на материалите и потенциално определяне на типа на нефта.	Изключителна зависимост от безоблачни, дневни условия.
БЛА: RGB/Мултиспектрален/ Термичен	1 см – 1 м	При поискване	Спектрални/термични сигнатури с висока резолюция и визуална морфология.	Несравнима пространствена детайлност за структура на гладки повърхности, оценка на дебелината (термична) и валидиране на наземни данни.	Изключително ограничено пространствено покритие. Кратка продължителност на полета. Регулаторни пречки.

Горната таблица разкрива естествена йерархия от мащаб и надеждност. Сателитните SAR системи действат като широкообхватна мрежа, сканирайки океаните непрекъснато, за да сигнализират за потенциални инциденти. Сателитните оптични данни осигуряват спектрално потвърждение, когато е възможно. Безпилотните летателни апарати (БЛА) след това действат като окончателен верификатор, разположен към идентифицирана от SAR цел, за да събират приложими, високоточни данни за реагиращите и да генерират безупречни анотации за подобряване на сателитните модели с изкуствен интелект.

Еволюция на алгоритмите от ръчно създадени функции до контекстно осъзнато обучение

Преходът от класическо машинно обучение към дълбоко обучение представлява повече от постепенно повишаване на точността; той означава преминаване от експлицитна, дефинирана от човека логика към имплицитно, управлявано от данни разпознаване на модели. Таблица 2 сравнява тези парадигми.

Таблица 2. Еволюция и сравнение на методите с изкуствен интелект за откриване на нефтени разливи. F1-оценката (0-1) е хармоничната средна стойност на прецизността и пълнотата, стандартен показател за небалансирани задачи за класификация, като например откриване на нефтени разливи.

Парадигма на алгоритъма	Примерни модели	Ключови характеристики	Типична точност (F1-оценка)	Идеален случай на употреба и източник на данни
Класическо машинно обучение (базирано на характеристики)	Случайна гора, машина с опорни вектори (SVM)	Разчита на ръчно извлечени характеристики (текстура, геометрия).	0,75 – 0,88	Добре дефинирани проблеми с ограничени, хомогенни данни. Ефективни за предварителна SAR класификация на тъмни петна.
Дълбоко обучение: Семантична сегментация	U-Net, DeepLabv3+	Изучава характеристики директно от пиксели. Отлично се справя с улавянето на сложни форми и контексти. Изисква големи набори от данни и GPU изчисления.	0,89 – 0,97	Най-съвременна технология за пикселна класификация както на SAR, така и на оптични данни.
Дълбоко обучение: Трансформатори на зрението	SegFormer, U-Net с трансформаторни блокове	Използва самовнимание, за да моделира глобален контекст в цялото изображение. По-малко зависи от локалните индуктивни отклонения на CNN.	Развиващи се (0,90+ в проучвания)	**Бъдеща посока.** За предсказващи задачи, като прогнозиране на дрейф на петрол или разграничаване на петрол от подобни, с използване на хидродинамичен контекст.
Хибридни/физически информирани модели	PINN-и, модели, интегриращи данни за вълни/вятър	Интегрира физически уравнения (например дрейф на нефт) или спомагателни данни във функцията за загуби на ИИ.	В процес на активно проучване	**Бъдеща посока.** За предсказващи задачи, като прогнозиране на дрейф на петрол или разграничаване от подобни явления.

Преходът към DL, особено към архитектури, подобни на U-Net, осигурява значителен скок в устойчивостта, особено при обработката на шумните и променливи сигнатури в SAR данните. Това обаче идва за сметка на проблема с „черната кутия“ – намаляване на интерпретацията, което е значителна пречка за регулаторното приемане и последващия анализ. Нарастващата тенденция не е да се изоставят класическите методи, а да се използват като интерпретируеми допълнения или да се интегрират физически принципи директно в DL архитектурите (например Physics-Informed Neural Networks), за да се създадат по-надеждни и физически последователни модели.

ЗАКЛЮЧЕНИЕ

Наблюдението на морските нефтени разливи се развива от традиционни ръчни методи към автоматизирани системи, базирани на изкуствения интелект и данни с отворен достъп. Сателитните сензорни платформи осигуряват широкомащабно и непрекъснато наблюдение, докато безпилотните летателни апарати предоставят проверка с висока разделителна способност на място. В основата на тази трансформация стоят моделите за дълбоко обучение, особено архитектури за семантична сегментация като U-Net, които позволяват бързо и точно откриване на нефтени разливи.

Най-големият потенциал се крие в интегрирането на различни източници на данни в единна интелигентна система. В такава среда сателитните наблюдения откриват потенциал-

ни разливи, дроновете извършват последваща проверка, а изкуственият интелект обединява информацията в обща оперативна картина в реално време. Получените полеви данни допълнително подобряват точността на моделите чрез непрекъснато обучение.

Въпреки постигнатия напредък остават предизвикателства, свързани с разграничаването на нефтени разливи от сходни явления, ограничените обучаващи данни и необходимостта от по-обясними и надеждни модели. Независимо от това комбинацията от отворени сателитни данни, дронови технологии и изкуствен интелект поставя основите на ново поколение системи за мониторинг, които позволяват по-бързо откриване, по-ефективна реакция и по-добра защита на морската среда.

ЛИТЕРАТУРА

BAI, Y.; LI, C., 2021. A review on the applications of unmanned aerial vehicles in oil spill detection and monitoring. *Journal of Marine Science and Engineering*, vol. 9, no. 11, p. 1225. Available from: https://doi.org/10.3390/jmse9111225

BREKKE, C.; SOLBERG, A. H., 2005. Oil spill detection by satellite remote sensing. *Remote Sensing of Environment*, vol. 95, no. 1, pp. 1–13. Available from: https://doi.org/10.1016/j.rse.2004.11.015

CHEN, L. C.; ZHU, Y.; PAPANDREOU, G.; SCHROFF, F.; ADAM, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 801–818. Available from: https://doi.org/10.1007/978-3-030-01234-2_49

CHUANMIN, H.; YINGCHENG, L.; SHAOJIE, S.; YONGXUE, L., 2021. Optical remote sensing of oil spills in the ocean: what is really possible? *Journal of Remote Sensing*, vol. 2021, Article ID 9141902. Available from: https://doi.org/10.34133/2021/9141902

DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBORN, D.; ZHAI, X.; UNTERTHINER, T.; HOULSBY, N. et al., 2020. *An image is worth 16×16 words: transformers for image recognition at scale*. arXiv preprint, arXiv:2010.11929. Available from: https://doi.org/10.48550/arXiv.2010.11929

DUGGAN, A.; ANDRADE, B.; AFLI, H., 2025. *Advancing Earth observation: a survey on AI-powered image processing in satellites*. arXiv preprint, arXiv:2501.12030. Available from: https://doi.org/10.48550/arXiv.2501.12030

ESPEDAL, H. A.; WAHL, T., 1999. Satellite SAR oil spill detection using wind history information. *International Journal of Remote Sensing*, vol. 20, no. 1, pp. 49–65. Available from: https://doi.org/10.1080/014311699213659

FINGAS, M.; BROWN, C. E., 2018. A review of oil spill remote sensing. *Sensors*, vol. 18, no. 1, p. 91. Available from: https://doi.org/10.3390/s18010091

JIAO, Z.; JIA, G.; CAI, Y., 2019. A new approach to oil spill detection that combines deep learning with unmanned aerial vehicles. *Computers & Industrial Engineering*, vol. 135, pp. 1300–1311. Available from: https://doi.org/10.1016/j.cie.2018.11.048

KRESTENITIS, M.; ORFANIDIS, G.; IOANNIDIS, K.; AVGERINAKIS, K.; VROCHIDIS, S.; KOMPATSIARIS, I., 2019. Oil spill identification from satellite images using deep neural networks. *Remote Sensing*, vol. 11, no. 15, p. 1762. Available from: https://doi.org/10.3390/rs11151762

LEIFER, I.; LEHR, W. J.; SIMECEK-BEATTY, D.; BRADLEY, E.; CLARK, R.; DENNISON, P.; JONES, B., 2012. State of the art satellite and airborne marine oil spill remote sensing: application to the BP Deepwater Horizon oil spill. *Remote Sensing of Environment*, vol. 124, pp. 185–209. Available from: https://doi.org/10.1016/j.rse.2012.03.024

PAHLEVAN, N.; SARKAR, S.; FRANZ, B. A.; BALASUBRAMANIAN, S. V.; HE, J., 2017. Sentinel-2 MultiSpectral Instrument (MSI) data processing for aquatic science applications: demonstrations and validations. *Remote Sensing of Environment*, vol. 201, pp. 47–56. Available from: https://doi.org/10.1016/j.rse.2017.08.033

RAISSI, M.; PERDIKARIS, P.; KARNIADAKIS, G. E., 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, vol. 378, pp. 686–707. Available from: https://doi.org/10.1016/j.jcp.2018.10.045

REICHSTEIN, M.; CAMPS-VALLS, G.; STEVENS, B.; JUNG, M.; DENZLER, J.; CARVALHAIS, N.; PRABHAT, 2019. Deep learning and process understanding for data-driven Earth system science. *Nature*, vol. 566, no. 7743, pp. 195–204. Available from: https://doi.org/10.1038/s41586-019-0912-1

RONNEBERGER, O.; FISCHER, P.; BROX, T., 2015. U-Net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer, pp. 234–241. Available from: https://doi.org/10.1007/978-3-319-24574-4_28

SURESH, L.; ABHILASH, S.; ADITHYA, M.; ANIRUDH, N.; CHANDAN, S., 2024. A survey on satellite image processing using machine learning. *International Journal of Innovative Science and Research Technology*, vol. 9, no. 11, pp. 2589–2594.

TORRES, R.; SNOEIJ, P.; GEUDTNER, D.; BIBBY, D.; DAVIDSON, M.; ATTEMA, E.; ROSTAN, F. et al., 2012. GMES Sentinel-1 mission. *Remote Sensing of Environment*, vol. 120, pp. 9–24. Available from: https://doi.org/10.1016/j.rse.2011.05.028

XIE, E.; WANG, W.; YU, Z.; ANANDKUMAR, A.; ALVAREZ, J. M.; LUO, P., 2021. SegFormer: simple and efficient

design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, vol. 34.
ZHENG, Z.; ZHONG, Y.; WANG, J.; MA, A., 2021. A deep learning based oil spill detection system using synthetic aperture radar imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 181, pp. 178–191. Available from: https://doi.org/10.1016/j.isprsjprs.2021.09.007
ZHENGSIN, Z.; ZHU, L., 2023. A review on unmanned aerial vehicle remote sensing: platforms, sensors, data processing methods, and applications. *Drones*, vol. 7, no. 6, p. 398. Available from: https://doi.org/10.3390/drones7060398

FROM ORBIT TO COASTLINE: AI-POWERED TOOLS FOR OIL SPILL DETECTION USING OPEN-SOURCE PLATFORMS

Abstract: *Rapid and accurate detection of marine oil spills is a critical global environmental challenge. Traditional monitoring methods, relying on ship patrols and aerial surveillance, are limited by cost, scope, and time. A paradigm shift is underway, fueled by the convergence of three transformative technologies: open-source satellite data, unmanned aerial vehicles (UAVs), and artificial intelligence (AI). This synthesis examines the current ecosystem of AI-powered tools that use freely available data from orbital platforms and drone-based imagery to automate and improve oil spill monitoring. The standard AI workflow is analyzed, from data preprocessing and classical machine learning to advanced deep learning segmentation models such as U-Net. Satellite systems provide all-weather surveillance, and unmanned aerial vehicles (UAVs) provide unparalleled high-resolution inspection and thickness estimation. Despite progress, key challenges remain, including distinguishing "lookalike" phenomena and the need for explainable AI in operational contexts. The future trajectory points to integrated, multi-platform monitoring systems where AI not only detects but also predicts spill dynamics, building a powerful open-access toolkit for marine ecosystem conservation.*
Keywords: *Oil Spill Detection; Artificial Intelligence; Unmanned Aerial Vehicles (UAV); Sentinel; Environmental Monitoring*

Vassil Mihalkov, PhD Student
ORCID: 0000-0003-3630-0040
Rakovski National Defense College
Sofia, Bulgaria
E-mail: vmihalkov@icloud.com

АКТУАЛНО
CURRENT TOPICS

COMMUNICATION AND INFORMATION PROCESSES AS A FACTOR FOR INCREASING JOB SATISFACTION IN THE EMERGENCY MEDICAL SERVICES

Dennis Jörn Richard Rahneberg
University of Library Studies and Information Technologies
https://doi.org/10.70300/BYID9777

Abstract: *Staff shortages, high turnover, and heavy workloads have a significant impact on the daily work of paramedics in Germany. At the same time, there is still uncertainty regarding the factors influencing the independent performance of medical procedures. Control by the medical director of the emergency medical services (ÄLRD) is particularly important for the structuring of powers of action. The extent to which this authority structure is important for job satisfaction in the emergency medical services and how it is communicated in practice was investigated using a mixed-methods research design based on guided expert interviews. The research project also provides insights into the job-specific influences on the job satisfaction of paramedics in order to contribute to professionalized human resource management and effective communication processes. The results of the research project showed that autonomy is very relevant to job satisfaction and that two different constructions of reality are hidden between the power of the labor market and the power of the ÄRLD. Suggestions for improving communication processes are made.*
Keywords: *job satisfaction; paramedics; skills shortage; human resource risk management; information management*

INTRODUCTION

The emergency medical services are also severely threatened by the shortage of skilled workers in the healthcare sector (PWC 2022). Staff turnover, combined with an already existing shortage of personnel, can develop into a very critical personnel risk for emergency medical services, which can then reach the limits of their capacity. The job satisfaction of emergency paramedics is particularly relevant in this context, as job satisfaction has been empirically linked to significant personnel risks. These include staff retention as the opposite of turnover (Chang et al. 2009; Fried et al. 2008; Jiang et al. 2012; Singh 2019), but also work performance (Jiang et al. 2012) and employee health (Faragher et al. 2005). In short, satisfied professionals are more likely to stay with their organisations, tend to perform better and have fewer health-related absences. Job satisfaction can therefore be regarded as a key control variable for human resource risks and an important contributor to occupational health management.

For successful human resource management in emergency medical services, the question therefore arises as to how the job satisfaction of emergency paramedics can be positively influenced. The meta-analytically based Job Characteristics Model emphasises the effect of autonomy (Kanning 2017), so that this variable and the way how autonomy is communicated and negotiated was placed at the centre of the dissertation research project.

From an information science perspective, it was of central interest to examine information in communication and negotiation as an expression of power structures (Hobohm 2022). The dualism of information science between realism and constructivism, moves within the field of tension that, according to the paradigm of realism, there is an objective reality, while constructivism envisages a social construction of reality and thus encompasses deviations from the respective reality (Hobohm 2022). Through the constructivist perspective of reality as a construction of reality and information as a manifestation of power structures, it could therefore be possible in the research project to trace the divergent construction of reality by the medical management (all decision-making power belongs to them) as a contrast to the construction of

reality by the personnel market (a shortage of skilled workers is causing a shift in power towards skilled workers). This could provide valuable insights into how communication processes could be improved.

To explore potential areas for improvement, it is helpful to first establish a basic understanding of communication as an exchange of information. Communication can be understood as an active process of information exchange (Zdravkovska and Haque 2023). The literature offers perspectives that view information and communication as paired categories or, more specifically, limit information processes to transmission, while communicative processes focus on meaning-making, social regulation, and goals (Pavlova and Paliy 2020; Zdravkovska and Haque 2023). Regardless of this distinction, the information process can be regarded as the foundation of communication due to its transmission function.

Of interest regarding the communicative possibilities in this context is how the flow of communication takes place. Flow models differ in terms of whether communication is linear – that is, one-way from sender to receiver – whether it is reciprocal (circular), or whether it takes place within a complex network (Chew and Ng 2021). Therefore, the following basic types of flow models can be distinguished (Beskaravainaya and Kharybina 2024; Chew and Ng 2021; Kumar and Sinha 2021):

– Linear models focus on information that flows one-way from the sender to the receiver, i.e., unidirectionally.

– Circular models represent interactivity, as they already incorporate feedback and operate in loops.

– Ecological models view all actors simultaneously as both senders and receivers operating within a dynamic system.

– Network and diffusion models enable an analysis of information dissemination between different nodes and are based on the perspective of many-to-many communication.

It becomes apparent that a linear process allows the sender to transmit information but does not provide for a response from the recipient. Only at the level of circular communication is an exchange possible that connects the constructed realities with one another.

Healthcare systems often focus on linear information flow (Johnson et al. 2024). The reasons for this can be seen, on the one hand, in the fact that healthcare organizations are often conceived as input-throughput-output machines; however, hierarchically structured perspectives also contribute to this (e.g., top-down communication from physicians to healthcare staff or a hierarchical, paternalistic model of the doctor-patient relationship) (Manojlovich et al. 2015; Shopnikolova 2025). It is also necessary to consider the numerous medical safety protocols for healthcare systems, which attribute errors primarily to missing information in documentation and aim to prevent them through a standardized, one-way transfer process (McCarthy et al. 2025).

In communication, this can become a problem, as efficiency is often already hindered by unanticipated response options, and no dialogical engagement can take place that would facilitate a shared understanding of meaning (Dartiguelongue and Cafiero 2021; Tu et al. 2024). A reduction to linear communication processes can thus be understood as an obstacle to learning health systems (Tu et al. 2024).

From a constructivist perspective, this can be conceived as fostering diverse realities. These realities will be examined through the lens of personnel decision-makers in emergency medical services in order to develop recommendations for improving communication based on these findings.

RESEARCH METHODOLOGY

The aim of the study is to examine for decision-makers in the emergency medical services in Germany whether and to what extent the autonomous freedom of action of employed paramedics within the meaning of Section 2a of the Paramedics Act (NotSanG) and the advance delegation of tasks contributes to employee satisfaction. This would attribute potential to freedom of action for coping with scarce human resources.

Using a mixed methods approach, eight decision-makers at management level (as district man-

agers, managing directors or management assistants) from rescue services in Baden-Württemberg were surveyed in Germany about the possible impact of authorisations on the job satisfaction of emergency paramedics. Based on their work and other characteristics common to the sample, such as involvement in personnel decisions, many years of personal experience in operational emergency services and their own professional qualifications as emergency paramedics, it was assumed that the sample had adequate expertise on the subject.

The qualitative data was evaluated using content analysis according to Mayring (Mayring and Fenzl 2019) and was supported by the QCAmap software to ensure compliance with the procedural rules. Due to the small sample size, the quantitative data could only be evaluated descriptively.

Limitations of the study must of course be considered in view of its design as pilot research (small, non-representative sample size and a qualitative research design focus).

RESULTS

The quantitative data initially showed that there are still emergency services (25%) that do not grant authorisation for more autnomomy despite the possibilities offered by Section 2a NotSanG. In 37.5% of cases, authorisation was granted in accordance with the standard instructions for Baden-Württemberg. 50% comminicated authorisation beyond this with the employees. However, the data revealed a significantly more restrictive situation with regard to the release of anaesthetics: 75% stated that the ÄLRD does not grant the possible autonomy of action for anaesthetics in their rescue area. This indicates clear restrictions that may limit pain management in emergency rescue services.

In order to investigate the extent to which approvals could actually be implemented in practice, questions were asked about equipment, medication and material enabling. The sample already showed extensive availability in many areas: 100% reported that a ventilator with CPAP function was available, as well as material for endotracheal intubation. However, only 25% reported having an ultrasound device in their ambulance. Seven of the eight interviewees confirmed that the materials and medications required for approved EVM were available.

In order to determine how the necessary medical expertise is transferred and monitored within the framework of quality assurance, organisational practices were examined: Annual reviews are intended to ensure knowledge of advanced care measures in the emergency department, which was the case in 87.5% of the cases examined. 75% of respondents stated that approvals are updated and adapted at least once a year. 12.5% also did this, but at longer intervals (2–5 years). In this highly patient-relevant field of activity, there has already been a fairly comprehensive development in terms of keeping the relevant skills of emergency paramedics up to date, which should, of course, always be accompanied by targeted practical monitoring.

The extent to which there are potential consequences under labour law if independent and medically acceptable emergency care is actually provided was also investigated. It was found that 87.5% did not expect any negative consequences and 0% expected to be dismissed as a result. However, despite providing medically acceptable emergency care in accordance with the law, 12.5% indicated that they were facing a warning. This indicates that, despite Section 2a of the NotSanG, there is still a risk of negative consequences under labour law in professional practice in Baden-Württemberg.

These working conditions are compounded by a drastic shortage of personnel in the sample: 100% of respondents rated staff turnover in their rescue stations as high, with it already impossible to fill all planned positions: 25% were only able to fill between 60–70% of their positions, while a further 37.5% of respondents were only able to fill 80–90% or <90%. All respondents agreed that there was high recruitment pressure.

As part of the qualitative content analysis, the development of the emergency paramedic profession over the last ten years was first examined: The dominant category identified was a conflict between training and reality:

“The new NFS are trained in three years for scenarios that only occur in a very small percentage of daily cases (= disappointment). Actions are sometimes carried out strictly according to

SAA and prescribed guidelines rather than based on perception, observation, consideration and the patient's needs." (Interview 6-M)

"Due to the difference between reality (often without NA in action and a high number of non-emergencies) and training (trained for the worst case), there is a need for reform in the current training." (Interview 8-S)

"At the same time, this level of training leads to an ever-widening gap between school and real life. This increases frustration among employees." (Interview 4-D)

The conflict with the Heilpraktikergesetz (German law regulating alternative medicine) and a possible extension and the importance of expanded care measures for a broader range of activities were also mentioned in connection with the development of the job profile:

"Unfortunately, the measures that were actually desired were not correctly incorporated into law, so emergency paramedics still come up against limitations under the Alternative Medical Practitioners Act and risk penalties." (Interview 2-B)

"This change was accompanied by the introduction of emergency paramedics, who, even after completing a full examination/state supplementary examination, were not allowed to perform any extended measures for many years. With the implementation of the extended care measures, including annual reviews and advance delegation in Baden-Württemberg, the range of activities has been significantly expanded." (Interview 3-B)

It was also pointed out that the job description of doctors and their associations is being curtailed.

When asked about their own position on medical measures, 7 out of 8 respondents rated them positively. Only one of the participants rejected them due to legal concerns and worries about compensation for damages.

The data revealed very different perspectives on how an expansion of freedom in the implementation of healing measures in one's own area of responsibility would affect the profession: Increased job satisfaction and staff retention were among the responses, as was a possible increase in employer attractiveness:

"Employee satisfaction would increase, as would the length of time they stay in their jobs. Only those who are allowed to do what they have learned and are good at will be satisfied with what they do." (Interview 1-B)

At the same time, however, there were also considerations that this was not necessary due to the insufficient implementation of existing options or that the extended care measures were already generous. It was also emphasised that legal adjustments are important and that consideration should be given to the extent to which all emergency paramedics really attach great importance to this, as the motivational effect could be limited and to what extent this is significant due to the parallel alerting of the emergency doctor.

In contrast to the extension of authorisations, respondents were also asked about their expectations if the freedom were to be abolished instead. It was precisely at this point that the survey data highlighted the link between authorisations, personnel risks and job satisfaction:

"This would certainly be frustrating for all former Ras and all NFS." (Interview 6-M)

"That would certainly send a fatal signal to many emergency paramedics and would undoubtedly lead to many of them seeking new careers outside the emergency services." (Interview 3-B)

"In the border region with Switzerland where we are located, this would result in a massive exodus of dedicated NFS staff. It would also lead to people moving to other professions." (Interview 8-S)

"Maintaining operating hours would no longer be possible in the NFR due to the threat of staff shortages." (Interview 8-S)

"Decline in quality" (Interview 6-M)

Only one of the interviewees assumed that there would be "no impact" (I2-B) if the freedom were to be removed. However, this was the interviewee who rejected the approvals due to the risks involved. This justifies considering the extent to which his area of responsibility would actually undergo any practical changes if it were to be abolished, if a distinction is to be made between opportunities not granted and opportunities abolished.

However, the research project provided numerous other insights into the positive and negative influences on the job satisfaction of emergency paramedics: In addition to job opportunities and expanded care measures, other autonomy characteristics such as freedom of decision-making and greater responsibility were also mentioned as positive factors. The need for autonomy among emergency paramedics was also significant in terms of shift planning and working hours, in order to better adapt these to their needs. In practice, this is difficult to achieve under the prevailing conditions of staff shortages, but if the available options are exploited to the full, it can make a very important contribution to job satisfaction, and thus to staff retention and ultimately to greater employer attractiveness. Considering the situation in the labor market, it is striking how little the communication potential of this professional group has been tapped to address such needs. Expressions of appreciation and incentives were also mentioned as positive factors, with an easy-to-implement fruit basket being cited as a frequent example. In terms of appreciation, support for staff and leadership were also identified as positive influences.

The data also indicated that areas of responsibility and equipment are relevant factors for satisfaction. In the data set, the categories of duty roster/working time management and equipment were the most frequently cited positive influences.

Remuneration as a positive influencing factor also pointed to a traditional starting point for satisfaction. However, the data also showed that improved control and interface management were structural factors contributing to positive job satisfaction.

For professional human resources management that strives to manage personnel risks, numerous starting points are thus identified, of which at least those that are feasible in practice can be pursued in the best possible way.

When it came to negative influences on job satisfaction, there was consensus that staff shortages and interface overload were significant factors. However, unnecessary call-outs were also mentioned frequently in the interviews. In addition to classic negative factors such as limited development opportunities, low social status, the working atmosphere and corporate culture, the importance of equipment was also highlighted here. The rescue station was also a factor: it should be noted that paramedics spend a lot of time there. Restrictive measures in the federal state and the medical on-call service were also cited as factors contributing to dissatisfaction, as were fears about expectations regarding medical measures.

CONCLUSIONS

Overall, the research project revealed serious personnel risks in the emergency medical services, which in many areas are also affected by health policy and legal factors and are therefore partially beyond the control of the emergency medical services themselves. When considering information as a manifestation of power (Hobohm 2022), the research results showed that medical management currently often still adheres to a restrictive demonstration of power as their own construction of reality, which contrasts with realityconstruction of the power relations among staff due to the shortage of skilled workers.

However, the research results also showed that emergency services have some potential for action to differentiate themselves positively as employers, become more successful in competing for personnel and reduce their personnel risks as far as possible within the scope of their capabilities. The design of the approvals can be seen as an important tool for emergency services in Germany to increase job satisfaction, retain staff and counteract the shortage of skilled workers in the German healthcare system.

The question of how to mediate between the reality constructs of medical management and the personnel market appears relevant. Improving communication could serve as an important starting point. When examining information flow models that focus on how information flows through systems or organizations, it becomes clear that no regular communication regarding staff concerns is planned between the medical management of the emergency medical services and the staff. Their communication processes occur primarily on an ad hoc basis. Occasion-based communication is largely initiated by the medical management or is structurally provided for in the form of com-

petency assessments. Staff needs and staff dissatisfaction have not yet been considered a formal occasion, nor is it typical for employees to initiate communication processes.

With regard to staff interests, what stands out in terms of the flow models is that these are not even intended as linear communication from staff to the Medical Management. The communication processes are thus still far from being circular. Therefore, when identifying opportunities for improvement, it is essential to establish an ongoing exchange regarding the staffing situation and staffing needs between human resources management, the staff, and medical management, in which all stakeholders can communicate on an equal footing. This can be seen as a significant step toward countering paternalistic mindsets in favor of effective communication.

Nevertheless, it should be noted that this pilot research project is only the beginning of the investigation of job satisfaction and autonomy in the German emergency medical services. In particular, quantitative research on larger nationwide samples can be considered desirable. Regression models should also be used to identify correlations between significant variables and their moderator variables. This would make a key contribution to the management of emergency services by enabling them to reflect on the impact of decisions in advance. In the same way, the establishment of new communication models is a pilot project that makes it possible to develop successful practical examples for emergency services.

REFERENCES

BESKARAVAINAYA, E. V.; KHARYBINA, T. N., 2024. Characteristics of information flow in scientific research. *Scientific and Technical Information Processing*, vol. 51, no. 3, pp. 206–214.

CHANG, C.-H.; ROSEN, C. C.; LEVY, P. E., 2009. The relationship between perceptions of organizational politics and employee attitudes, strain, and behavior: a meta-analytic examination. *Academy of Management Journal*, vol. 52, no. 4, pp. 779–801.

CHEW, S. Y.; NG, L. L., 2021. Models of communication process. In: CHEW, S. Y.; NG, L. L. (eds.). *Interpersonal interactions and language learning*. Cham: Springer International Publishing, pp. 17–26.

DARTIGUELONGUE, J. B.; CAFIERO, P. J., 2021. Communication in health care teams. *Archivos Argentinos de Pediatría*, vol. 119, no. 6, pp. 1–5.

FARAGHER, E. B.; CASS, M.; COOPER, C. L., 2005. The relationship between job satisfaction and health: a meta-analysis. *Occupational and Environmental Medicine*, vol. 62, no. 2, pp. 105–112.

FRIED, Y.; SHIROM, A.; GILBOA, S.; COOPER, C. L., 2008. The mediating effects of job satisfaction and propensity to leave on role stress-job performance relationships: combining meta-analysis and structural equation modeling. *International Journal of Stress Management*, vol. 15, no. 4, pp. 305–328.

HOBOHM, H.-C., 2022. Theorien der Informationswissenschaften. In: KUHLEN, R.; LEWANDOWSKI, D.; SEMAR, W.; WOMSER-HACKER, C. (eds.). *Grundlagen der Informationswissenschaft*. Berlin; Boston: De Gruyter, pp. 45–53. Available from: https://directory.doabooks.org/handle/20.500.12854/96472

JIANG, K.; LIU, D.; MCKAY, P. F.; LEE, T. W.; MITCHELL, T. R., 2012. When and how is job embeddedness predictive of turnover? A meta-analytic investigation. *The Journal of Applied Psychology*, vol. 97, no. 5, pp. 1077–1096.

JOHNSON, N. L.; VAN TIEM, J.; BALKENENDE, E.; JONES, D.; FRIBERG, J. E.; CHASCO, E. E.; MOECKLI, J.; STEFFENSMEIER, K. S.; STEFFEN, M. J. A.; ARORA, K.; RABIN, B. A.; REISINGER, H. S., 2024. Gaps in communication theory paradigms when conducting implementation science research: qualitative observations from interviews with administrators, implementors, and evaluators of rural health programs. *Implementation Science*, vol. 19, no. 1, p. 66.

KANNING, U. P., 2017. *Personalmarketing, employer branding und Mitarbeiterbindung*. Berlin; Heidelberg: Springer.

KUMAR, P.; SINHA, A., 2021. Information diffusion modeling and analysis for socially interacting networks. *Social Network Analysis and Mining*, vol. 11, no. 1, p. 11.

MANOJLOVICH, M.; SQUIRES, J. E.; DAVIES, B.; GRAHAM, I. D., 2015. Hiding in plain sight: communication theory in implementation science. *Implementation Science*, vol. 10, pp. 1–11.

MCCARTHY, S.; MOTALA, A.; LAWSON, E.; SHEKELLE, P. G., 2025. Use of structured handoff protocols for within-hospital unit transitions: a systematic review from Making Healthcare Safer IV. *BMJ Quality & Safety*, vol. 34, no. 10, pp. 680–690.

PAVLOVA, E.; PALIY, I., 2020. Global informational society: questions and perspectives (the problem of information systems as ideal models of interpersonal communication). *SHS Web of Conferences*, vol. 74, p. 02012.

PWC, 2022. *Fachkräftemangel im Gesundheitswesen: Auswege aus der drohenden Versorgungskrise*. [viewed 22 July 2025]. Available from: https://www.pwc.de/de/content/a893f304-8f55-402f-bd4e-2e080e4c45d0/pwc-fachkraeftemangel-im-gesundheitswesen-2022.pdf

SHOPNIKOLOVA, P. T., 2025. The need to change the doctor-patient communication model in contemporary medical practice. *International Journal of Innovative Research in Medical Science*, vol. 11, no. 9, pp. 333–337.
SINGH, D., 2019. A literature review on employee retention with focus on recent trends. *International Journal of Scientific Research in Science, Engineering and Technology*, pp. 425–431.
TU, S.-P.; GARCIA, B.; ZHU, X.; SEWELL, D.; MISHRA, V.; MATIN, K.; DOW, A., 2024. Patient care in complex sociotechnological ecosystems and learning health systems. *Learning Health Systems*, vol. 8, suppl. 1, pp. 1–9.
ZDRAVKOVSKA, S.; HAQUE, S., 2023. The theoretical basis of internal communication as an active process of information exchange. *Technium Social Sciences Journal*, vol. 43, pp. 252–262.

КОМУНИКАЦИОННИТЕ И ИНФОРМАЦИОННИТЕ ПРОЦЕСИ КАТО ФАКТОР ЗА ПОВИШАВАНЕ НА УДОВЛЕТВОРЕНОСТТА ОТ РАБОТАТА В СЛУЖБИТЕ ЗА СПЕШНА МЕДИЦИНСКА ПОМОЩ

Резюме: *Недостигът на персонал, високото текучество и голямата натовареност оказват значително влияние върху ежедневната работа на парамедиците в Германия. В същото време все още съществува несигурност относно факторите, влияещи върху самостоятелното изпълнение на медицинските процедури. Контролът от страна на медицинския директор на спешните медицински служби (ÄLRD) е особено важен за структурирането на правомощията за действие. Степента, в която тази структура на властта е важна за удовлетвореността от работата в спешните медицински служби, и това как тя се комуникира на практика, бяха проучени чрез изследователски дизайн със смесени методи, базиран на насочени експертни интервюта. Изследователският проект също така предоставя информация за специфичните за работата фактори, влияещи върху удовлетвореността от работата на парамедиците, с цел да допринесе за професионализирано управление на човешките ресурси и ефективни комуникационни процеси. Резултатите от изследователския проект показаха, че автономността е много важна за удовлетвореността от работата и че между властта на пазара на труда и властта на ÄRLD се крият две различни конструкции на реалността. Предложени са препоръки за подобряване на комуникационните процеси.*
Ключови думи: *удовлетвореност от работата; парамедици; недостиг на квалифицирани кадри; управление на риска в сферата на човешките ресурси; управление на информацията*

Денис Йорн Рихард Ранеберг, докторант
Университет по библиотекознание и информационни технологии
София, България
E-mail: rahneberg@aczepta.de

ИЗДАТЕЛ
Академично издателство
„За буквите – О писменехъ“
Университет по библиотекознание
и информационни технологии

ДИРЕКТОР
доц. д-р Диана Стоянова
бул. „Цариградско шосе“ № 119,
ет. 2, стая 213
София 1784, България
тел.: +359 879 14 83 85
е-поща: d.stoyanova@unibit.bg

Списание „Образование, научни изследвания и иновации“ излиза четири пъти годишно.

PUBLISHER
Academic Publisher
“Za Bukvite – O Pismeneh”
University of Library Studies and
Information Technologies

DIRECTOR
Assoc. Prof. Diana Stoyanova, PhD
119, Tsarigradsko Shosse Blvd.
fl. 2, room 213
Sofia 1784, Bulgaria
tel.: +359 879 14 83 85
E-mail: d.stoyanova@unibit.bg

Journal “Education, Scientific Research and Innovations” is published four times a year.