SOCIAL ATTITUDES AND SECURITY DISCOURSES REGARDING ARTIFICIAL INTELLIGENCE, WITH A SPECIAL FOCUS ON THE EDUCATION SECTOR

Bálint Zoltán Biczó*

Dennis Gabor University, Budapest, Hungary

*Correspondence: biczo.zoltan.balint@gde.hu

DOI: 10.66538/DH.2025.1.2.25

Abstract

The exponential development of artificial intelligence (AI) and its widespread social embedding raise complex security and trust issues. This study examines the social attitudes, perceptions, and security discourses related to AI based on the international literature. Using the theoretical framework of the risk society, we analyze the mechanisms behind the formation of technological trust in various sectors (healthcare, finance, supply chains), with particular emphasis on the education sector, where large language models are creating new pedagogical paradigms. Our findings suggest that social acceptance is sector-dependent and is significantly influenced by the transparency, explainability, and the level of autonomy in decision-making of systems. The integration of large language models in education has a dual impact: it supports personalized learning but also creates challenges related to academic integrity and equality of opportunity. The concept of security in the AI context has expanded to include dimensions such as cybersecurity, data privacy, ethics, and social security. Media representations present a polarized image, influencing public opinion and trust. The study provides practical recommendations for the responsible development and application of AI systems, emphasizing the importance of multi-level governance mechanisms and continuous societal dialogue.

Keywords: artificial intelligence, social attitudes, security, risk society, educational integration, large language models

1. Introduction

The exponential development of artificial intelligence (AI) technologies and their increasingly widespread social embedding raise numerous questions regarding security, trust, and social acceptance. Understanding this complex interrelationship is crucial for the sustainable and responsible development and application of the technology. The aim of this study is to map the social attitudes and perceptions related to AI security, as well as the scientific discourses examining these, based on the international literature. I explore how social attitudes are shaped in different application areas of AI, which theoretical frameworks can help in understanding the risk perception and trust mechanisms related to AI, how the interpretation of security changes within the context of AI technologies, and what regulatory approaches are emerging for the development of secure AI systems. Special attention is given to the education sector, where the emergence of AI, particularly large language models, is creating new pedagogical paradigms and challenges.

2. The Perspective of the Risk Society

Ulrich Beck's theory of the risk society [1] provides a particularly relevant theoretical background for interpreting social attitudes and security discourses related to AI. According to Beck (1992), one of the defining features of modern societies is that technological development creates complex, human-induced risks that are global in scope and often unpredictable. This is especially true for AI technologies, as their potential impacts are system-level and difficult to predict in advance.

Beck's theory emphasizes that modern risks are not merely local or temporally limited, but rather transboundary and global in impact [1]. This concept can be directly applied to AI systems, whose effects are not confined by geographical or social borders. The phenomenon of reflexive modernization, in which society becomes increasingly aware of its own risk-creating activities [2], is also relevant for interpreting the social acceptance of AI.

3. Social Trust and Technological Acceptance

A fundamental prerequisite for the societal acceptance of technological innovation is the establishment of trust. Trust mechanisms related to AI systems operate on multiple levels: institutional (regulatory authorities, development companies), technological (operational reliability, transparency), and social (community norms, cultural attitudes) dimensions [3].

Transparency and explainability play a key role in the development of trust in AI systems. Explainable AI (XAI) approaches specifically aim to address this issue by developing technological solutions that allow for the interpretation and tracking of AI systems' decision-making processes [4]. Research shows that transparent and interpretable AI solutions significantly increase users' trust and the acceptance of technology across various sectors [5].

4. Social Attitudes towards AI in Different Application Areas

4.1. Healthcare and AI

Research indicates that society's opinions on AI are dual: there is both excitement and concern. According to a survey spanning ten countries across four continents, respondents are evenly split on whether AI will change the world as we know it or whether it should be heavily regulated [6]. This duality is particularly characteristic of healthcare applications, where people see the technology as "exciting," "useful," "worrying," and "futuristic" simultaneously [7].

Long-term trend studies have pointed out that while optimism generally dominates, growing concern is observed regarding the loss of control over AI, ethical issues, and the negative impact on labor. However, hopes related to the application of AI in healthcare and education have increased in recent years [8].

Willingness to accept AI varies significantly depending on the healthcare application. A representative study shows that the highest acceptance level was achieved by AI applications for monitoring home medication, while the lowest acceptance was found for AI systems that autonomously make decisions about therapy modes [9]. This suggests that people are more accepting of AI in a supportive role but are more reluctant when it comes to autonomous decision-making.

4.2. Finance and Economy

In the financial sector, social attitudes towards AI applications are heavily influenced by the security and reliability of the systems. AI systems developed for detecting credit card fraud show increasing acceptance as long as they can make interpretable and transparent decisions [10]. XAI approaches are particularly important in this area because the transparency of financial decisions not only builds user trust but also ensures regulatory compliance. Users are more willing to accept AI-based financial services if they understand the operational mechanisms and decision-making criteria of the system.

AI in financial systems can act as a double-edged sword: it may either increase systemic financial risk or stabilize the system, depending on endogenous responses and strategic complementarities. Experts argue that due to AI's quick responsiveness and complexity management, future financial crises may be more intense than previous ones [11].

In the financial services sector, the bias of algorithms and the range of ethical issues are of particular importance. AI systems often perpetuate the biases present in training data, leading to unfair credit evaluations or discriminatory financial services. These issues raise not only ethical but also regulatory concerns [12].

4.3. Supply Chains and Logistics

In the area of supply chains, the integration of AI and blockchain technologies opens new opportunities for increasing efficiency, transparency, and reliability (Salah et al., 2019b). Social attitudes in this field are closely linked to the security and ethical compliance of systems. An especially interesting area is the ethical dimension of halal supply chains, where the application of AI requires not only technological efficiency but also respect for religious and cultural values [13]. This illustrates that AI security issues cannot be separated from cultural and ethical dimensions.

5. Attitudes within Education: Challenges and Opportunities of Integrating Large Language Models

The education sector presents a particularly interesting field for examining the societal acceptance of AI technologies. The integration of large language models (LLMs) into higher education presents a complex challenge with significant impacts on both the attitudes and expectations of instructors and students.

5.1. Integration Strategy and Stakeholder Attitudes

A key element for the successful integration of LLMs is the development of a comprehensive strategy that takes into account the perspectives of all stakeholders. Licht identifies three critical phases in this process: the preparatory phase, the explanatory phase, and the implementation phase [14]. In the preparatory phase, it is essential to involve all stakeholders and apply transparent processes, with particular attention to avoiding polarization between "AI optimists" and "AI skeptics." In the explanatory phase, the reasoning behind decisions must be clearly communicated, while in the implementation phase, appropriate resources and continuous feedback mechanisms should be provided.

Creating legitimacy—namely, ensuring the transparency of the process, the involvement of stakeholders, and the understanding of arguments—is crucial for acceptance. This is particularly important in the higher education environment, where critical thinking and academic integrity are fundamental values [14].

5.2 Student Perspective

The perspective and experiences of students play a decisive role in the successful integration of LLMs. Zdravkova and colleagues [15] identified three main approaches among students: traditional methods (without LLMs), complete dependency on LLMs, and a hybrid approach that combines independent research with LLM-generated recommendations. Research shows that the hybrid approach proved to be the most popular, suggesting that students value the benefits of LLMs but do not wish to fully abandon traditional learning methods.

Grande and colleagues [16] conducted an educational experiment to examine how master's students in computer science used LLMs to discuss ethical dilemmas. The results showed that the use of LLMs helped students better understand ethical theories, offered new perspectives, and alleviated the burden of language difficulties. However, a challenge arose as some students placed too much trust in the responses from LLMs and did not critically engage with them. Occasionally, the LLMs provided impractical suggestions or refrained from making specific recommendations—which, interestingly, students positively evaluated, stating that ethical decisions should be made by people [16].

5.3. Challenges and Ethical Considerations

The integration of LLMs into education raises numerous challenges and ethical issues. Sasvári highlights the issue of maintaining academic integrity, especially the blurring of boundaries between original work and AI-generated content [17]. The author emphasizes copyright issues, the danger of deepening inequalities (access disparities), the potential for discrimination (reproduction of biases in AI training datasets), and the risk of education becoming impersonal.

Fowler and colleagues [18] identify additional general challenges, such as ensuring the accuracy and reliability of information, preserving academic honesty, and ensuring equal access for all students. These challenges are particularly important in the higher education environment, where academic standards and equality of opportunity are fundamental values.

5.4. Integration Strategies and Pedagogical Approaches

The successful integration of LLMs requires various pedagogical strategies. Karabacak and colleagues [19], in examining the potential applications of AI-based systems in medical education, highlight the benefits of realistic simulations, personalized feedback, and automated assessment. Claman and Sezgin [20], analyzing the role of LLMs and multimodal models in dental education, emphasize the importance of instructor supervision and the promotion of critical thinking .

Prompt-based learning, as demonstrated by Denny and colleagues [21], develops students' skills in interacting with LLMs and fosters critical thinking. Peláez-Sánchez [22] and colleagues interpret the relationship between LLMs and integrated educational systems through the Education 4.0 framework, emphasizing the importance of integrating cyberpedagogy, heutagogy (self-directed learning), and peeragogy (learning among peers).

Nagy [23] and Kasneci [24] emphasize the importance of continuous monitoring, feedback systems, and content validation. Since models' responses based on statistical patterns are not always up-to-date or accurate, multiple verification and validation are critical.

LLMs' Impact on Learning Outcomes

The impact of LLMs on learning outcomes presents a complex picture. While some studies have found a negative correlation between excessive use of LLMs and academic performance in tasks requiring critical thinking [22], other studies highlight the positive effects of LLMs, particularly in personalized learning experiences [25].

Studies show that the impact of LLMs largely depends on the mode of integration and the pedagogical goal. Successful implementation requires a clear institutional AI strategy, infrastructure development, staff training, ethical guidelines, and stakeholder involvement [18].

The Evolving Interpretation of Security

The concept of security is undergoing significant transformation with the rise of AI technologies. The traditional focus on physical security has expanded to include dimensions such as cybersecurity, data privacy, ethical security, and social security. The results of the Evolving Concept of Security (EvoCS) project highlight [26] that security challenges vary by region and time, requiring differentiated security policies and research and development strategies. This is especially important in the context of AI, where the potential risks and impacts of the technology can significantly differ across various social and economic environments.

The emergence of the multi-domain concept also signals the increasing complexity of interpreting security. New concepts, including artificial intelligence, integrate security considerations from different technological fields, requiring coordinated actions against potential adversaries [27].

5.5. Media Representations and Public Opinion

The media representation of AI systems significantly influences social perceptions and attitudes. Research shows that the media often presents a polarized image of AI, depicting it either as a utopian solution or as a dystopian threat [28]. The literature emphasizes that the representation of AI in the media directly influences young people's technological trust and acceptance, particularly based on images transmitted by the entertainment industry [29].

"Alarmist narratives" in media discourses often amplify societal fears using the tools of technological determinism [29]. This trend can be compared with other media discourses surrounding technological innovations, where narratives created by the media significantly shaped public opinion.

In shaping public opinion regarding AI, the communication of expert communities with the public plays a key role. Transparent communication and an honest presentation of risks can enhance societal trust, while the collaboration between technological development and risk communication experts is essential for effective societal dialogue.

6. AI Security and Governance

The transparency and interpretability of AI systems are crucial for building social trust. Explainable AI (XAI) approaches aim to address the "black box" problem, enabling users to understand the decision-making processes of AI-based systems [4]. In the field of cybersecurity, XAI allows security analysts to validate and interpret alerts generated by AI-based detection systems [5]. In the financial sector, the application of XAI not only strengthens user trust but also serves regulatory compliance.

The safe development and application of AI systems require both proactive and reactive governance mechanisms. Proactive approaches apply built-in defensive mechanisms, such as threat modeling and adversarial testing [3]. Reactive mechanisms, such as content detection and correction, address problems that arise in real-time.

Hybrid governance frameworks, which integrate both approaches, can strike a balance between scalability and operational efficiency while maintaining ethical integrity. Increasing emphasis is placed on fairness, accountability, and transparency in AI governance, as these are essential pillars for building societal trust.

Discussion

A comprehensive analysis of the literature reveals several key conclusions. Building social trust in AI systems is a complex process that requires technological transparency, appropriate regulation, and continuous societal dialogue. AI security issues cannot be separated from broader social, ethical, and cultural contexts, necessitating a holistic approach.

Ulrich Beck's theory of the risk society remains a relevant framework for interpreting the societal impacts of AI, especially within the context of global, transboundary risks and reflective modernization. The diversity of application areas requires differentiated security approaches, presenting a complex regulatory challenge.

In the education sector, the integration of LLMs holds significant potential but requires careful strategic planning, addressing ethical concerns, and actively involving instructors and students. Pedagogical methods must adapt to new technologies, while critical thinking and the development of core skills remain priorities.

Future research may include:

- Intercultural comparative studies on the cultural differences in AI attitudes
- Longitudinal impact studies on the societal effects of AI and the evolution of attitudes
- Development of interdisciplinary ethical frameworks
- In-depth analysis of media representations and narratives of AI
- Examination of participatory AI development methodologies
- Comparative analysis of different national and international regulatory approaches

In the context of education, further research is needed on the long-term effects of LLMs on learning outcomes and critical thinking, as well as best practices for ensuring equitable access and ethical use. The study of societal attitudes and security discourses regarding AI is critical for the responsible development and application of technology. As AI systems become increasingly embedded in social and economic processes, balancing technological development with societal considerations and continuously strengthening trust and security will become ever more important.

7. Acknowledgments

Project no. TKP2021-NVA-05 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the TKP 2021 funding scheme.

References

1. W. Leiss, U. Beck, M. Ritter, S. Lash, and B. Wynne, 'Risk Society, Towards a New Modernity', Canadian Journal of Sociology / Cahiers canadiens de sociologie, vol. 19, p. 544, Nov. 1995, doi: 10.2307/3341155.
2. U. Beck, W. Bonß, and C. Lau, 'The Theory of Reflexive Modernization', Theory, Culture & Society, vol. 20, Apr. 2003, doi: 10.1177/0263276403020002001.
3. A. Kaplan and M. Haenlein, 'Rulers of the world, unite! The challenges and opportunities of artificial intelligence', Business Horizons, vol. 63, no. 1, pp. 37–50, Jan. 2020, doi: 10.1016/j.bushor.2019.09.003.
4. A. Barredo Arrieta et al., 'Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI', Information Fusion, vol. 58, Dec. 2019, doi: 10.1016/j.inffus.2019.12.012.
5. D. Shin, 'The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI', International Journal of Human-Computer Studies, vol. 146, p. 102551, Feb. 2021, doi: 10.1016/j.ijhcs.2020.102551.
6. J. Seth, 'Public Perception of AI: Sentiment and Opportunity', Jul. 22, 2024, arXiv: arXiv:2407.15998. doi: 10.48550/arXiv.2407.15998.
7. 'Exciting, Useful, Worrying, Futuristic: Public Perception of Artificial Intelligence in 8 Countries | Request PDF', ResearchGate. Accessed: May 14, 2025. [Online]. Available: https://www.researchgate.net/publication/353602997_Exciting_Useful_Worrying_Futuristic_Public_Perception_of_Artificial_Intelligence_in_8_Countries
8. E. Fast and E. Horvitz, 'Long-Term Trends in the Public Perception of Artificial Intelligence', in Proceedings of the AAAI Conference on Artificial Intelligence, Feb. 2017. doi: 10.1609/aaai.v31i1.10635.
9. P. Smola, M. Wojcieszko, and M. Duplaga, 'The assessment of society's acceptance of artificial intelligence use in healthcare', Eur J Public Health, vol. 34, no. Suppl 3, p. ckae144.684, Oct. 2024, doi: 10.1093/eurpub/ckae144.684.
10. '(PDF) Blockchain for AI: Review and Open Research Challenges', ResearchGate. Accessed: May 14, 2025. [Online]. Available: https://www.researchgate.net/publication/330009592_Blockchain_for_AI_Review_and_Open_Research_Challenges
11. J. Danielsson and A. Uthemann, 'Artificial intelligence and financial crises', Jul. 24, 2024, arXiv: arXiv:2407.17048. doi: 10.48550/arXiv.2407.17048.
12. N. Mike, 'Előítélet, diszkrimináció és gazdasági egyenlőtlenségek : A mesterséges intelligencia etikai kihívásai', Máltai Tanulmányok, vol. 6, no. 4, pp. 72–85, 2024, doi: 10.56699/MT.2024.4.5.
13. 'Leveraging Blockchain Technology for Halal Supply Chains | Request PDF', ResearchGate. Accessed: May 14, 2025. [Online]. Available: https://www.researchgate.net/publication/325422967_Leveraging_Blockchain_Technology_for_Halal_Supply_Chains

14. K. de Fine Licht, Integrating Large Language Models into Higher Education: Guidelines for Effective Implementation. 2023. doi: 10.3390/cmsf2023008065.

15. K. Zdravkova, F. Dalipi, and F. Ahlgren, Integration of Large Language Models into Higher Education: A Perspective from Learners. 2024, p. 6. doi: 10.1109/SIIE59826.2023.10423681.

16. V. Grande, N. Kiesler, and M. A. Francisco Rodriguez, Student Perspectives on Using a Large Language Model (LLM) for an Assignment on Professional Ethics. 2024, p. 484. doi: 10.1145/3649217.3653624.

17. '(PDF) Chatbotok a tudományos írásban: Idézhető-e a Mesterséges Intelligencia?', ResearchGate. Accessed: May 14, 2025. [Online]. Available: https://www.researchgate.net/publication/385828039_Chatbotok_a_tudomanyos_irasban_Idezheto-e_a_Mesterseges_Intelligencia

18. S. Fowler, M. Korolkiewicz, and R. Marrone, 'First 100 days of ChatGPT at Australian Universities An analysis of policy landscape and media discussions about the role of AI in Higher Education', Journal of Learning Letters, pp. 1–8, Jan. 2023, doi: 10.59453/JMTN6001.

19. M. Karabacak, B. B. Ozkara, K. Margetis, M. Wintermark, and S. Bisdas, 'The Advent of Generative Language Models in Medical Education', JMIR Medical Education, vol. 9, p. e48163, Jun. 2023, doi: 10.2196/48163.

20. D. Claman and E. Sezgin, 'Artificial Intelligence in Dental Education: Opportunities and Challenges of Large Language Models and Multimodal Foundation Models', JMIR Medical Education, Sep. 2023, doi: 10.2196/52346.

21. P. Denny et al., Promptly: Using Prompt Problems to Teach Learners How to Effectively Utilize AI Code Generators. 2023. doi: 10.48550/arXiv.2307.16364.

22. '(PDF) The impact of large language models on higher education: exploring the connection between AI and Education 4.0', ResearchGate. Accessed: May 14, 2025. [Online]. Available: https://www.researchgate.net/publication/381460867_The_impact_of_large_language_models_on_higher_education_exploring_the_connection_between_AI_and_Education_40

23. '(PDF) ChatGPT és más nagy nyelvi modellek (LLM-ek) biztonsági kérdései, szervezeti és társadalmi hatásai', ResearchGate. Accessed: May 14, 2025. [Online]. Available: https://www.researchgate.net/publication/379782926_ChatGPT_es_mas_nagy_nyelvi_modellek_LLM-ek_biztonsagi_kerdesei_szervezeti_es_tarsadalmi_hatasai

24. E. Kasneci et al., 'ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education', Learning and Individual Differences, vol. 103, p. 102274, Jan. 2023, doi: 10.1016/j.lindif.2023.102274.

25. M. Thway, J. Recatala-Gomez, F. S. Lim, K. Hippalgaonkar, and L. W. T. Ng, 'Battling Botpoop using GenAI for Higher Education: A Study of a Retrieval Augmented Generation Chatbots Impact on Learning', Jun. 22, 2024, arXiv: arXiv:2406.07796. doi: 10.48550/arXiv.2406.07796.

26. M. Jovanović and J. Burbiel, THE EVOLVING CONCEPT OF SECURITY: A CRITICAL EVALUATION ACROSS FOUR DIMENSIONS. 2014.

27. C. Paparone, 'How we fight: A critical exploration of US military doctrine', Organization, vol. 24, pp. 516–533, Jul. 2017, doi: 10.1177/1350508417693853.

28. S. Cave et al., Portrayals and perceptions of AI and why they matter. 2018. doi: 10.17863/CAM.34502.

29. J. Brennen, P. N. Howard, and R. Nielsen, 'An industry-led debate: how UK media cover artificial intelligence', 2018. Accessed: May 14, 2025. [Online]. Available: https://www.semanticscholar.org/paper/An-industry-led-debate%3A-how-UK-media-cover-Brennen-Howard/61b590750625a2a83a22c3798c06e73d7ed42d33