Artificial Intelligence in Sociology: A Critical Review and Future Directions

CHUNFA ZHOU

College of Design and Innovation, Zhejiang Normal University, 688 Yingbin Avenue, Jinhua, Zhejiang, 321004 P. R. C.
Email: zcf@zjnu.cn

This study presents a critical review of the emerging field of Artificial Intelligence (AI) Sociology, examining the social implications and ethical considerations of AI technologies. Through a qualitative methodology incorporating a systematic literature review and thematic analysis, this research explores the intersection of AI and sociology, aiming to bridge the gap between technological advancement and societal impact. The study investigates key theoretical frameworks, including critical theory and actor-network theory, to analyse power relations, social stratification, and the dynamic interplay between AI and society. The findings reveal the multifaceted influence of AI on social structures, ethical challenges, and the need for interdisciplinary approaches to address the societal implications of AI.

Keywords: Artificial Intelligence(AI) Sociology, social implications of AI, ethical considerations in AI, technological impact on society, interdisciplinary AI research

INTRODUCTION

The rapid advancement of Artificial Intelligence (AI) technologies has brought about significant changes in various aspects of society (Brynjolfsson et al. 2014). As AI continues to permeate different domains, there is an increasing need to understand its social implications from a sociological perspective. While AI has often been explored from a technological or economic standpoint, there remains a limited understanding of its broader societal impact (Rezaev et al. 2023). However, many studies underscore the need for a sociological perspective in understanding the complex relationship between AI and society. It is crucial for sociologists to engage with these developments and contribute to shaping the future of AI in ways that promote social well-being and equity (Mühlhoff 2020). This literature review aims to bridge this research gap by providing a comprehensive examination of the emerging field of 'AI Sociology'. The primary purpose is to shed light on the social dimensions of AI and its implications for individuals, communities, and societal structures. It contributes to the ongoing dialogue on AI Sociology and informs future scholarship and policy-making in this rapidly evolving field.

RESEARCH METHODOLOGY AND METHODS

This study adopts a qualitative research methodology to critically review the intersection of Artificial Intelligence (AI) and sociology. The approach involved a comprehensive literature

review, thematic analysis, and attention to ethical considerations, providing insights into the social dimensions and ethical implications of AI. The primary method used was a literature review of scholarly articles, books, and conference proceedings. Databases such as Web of Science, Scopus, IEEE Xplore and Google Scholar were utilised to identify key publications. The literature was categorised via thematic analysis to uncover themes such as social stratification, ethical implications and the impact of AI on social structures. With respect to research ethics, this study advocates ethical frameworks ensuring an equitable AI access and nondiscrimination, drawing on the works of several prominent scholars (Bostrom et al. 2014; Jobin et al. 2019).

THEORETICAL FRAMEWORKS IN AI SOCIOLOGY

The study of AI Sociology draws upon a range of theoretical frameworks that provide analytical tools for understanding the complex relationship between AI and society. This section explores key theoretical frameworks used in AI Sociology and highlights their application in the field, supported by specific studies and recent empirical evidence.

Critical Theory

Critical theory serves as a fundamental perspective in AI Sociology, examining power relations and surveillance capitalism within the context of AI. It emphasises the social, economic and political dimensions of AI and investigates how power is distributed, exerted and reinforced through AI systems. Zuboff (2019) explores the concept of surveillance capitalism, highlighting how AI-driven data collection and analysis reshape power dynamics in society. Her work demonstrates how AI technologies enable unprecedented levels of surveillance and data extraction, leading to new forms of social control and manipulation. Zuboff argues that this new economic logic threatens individual autonomy and democratic values. Andrejevic (2020) investigated the implications of automated media systems for social interactions and cultural practices. His research reveals how AI-driven recommendation systems and content curation algorithms shape public discourse and individual preferences, potentially reinforcing existing power structures and limiting exposure to diverse perspectives. Critical theory also addresses issues of algorithmic bias, automated decision-making, and the concentration of power in the hands of technology corporations. Crawford et al. (2014) examine the legal and ethical implications of big data and AI-driven decision-making processes, highlighting the potential for discriminatory outcomes and violations of due processes. Their work underscores the need for robust accountability mechanisms in AI systems. Noble (2018) provides a critical analysis of search engine algorithms, demonstrating how AI systems can perpetuate and amplify racial and gender biases. Her research reveals the ways in which seemingly neutral technologies can reinforce societal inequalities and marginalise certain groups. Recent empirical studies have further substantiated these critical perspectives. For instance, Buolamwini et al. (2018) conducted a comprehensive study on facial recognition systems, revealing significant gender and racial biases in commercial AI products. Their work highlights the urgent need for diverse and representative datasets in AI development to mitigate discriminatory outcomes.

Actor-Network Theory

Actor-network theory (ANT) emphasises the agency of both human and non-human actors in shaping social phenomena, including the role of AI in society. It recognises the interconnectedness of human actors, technological artifacts and socio-technical systems, considering

them as networked entities that influence and shape social interactions and structures. Latour (2005) has contributed significantly to this field by exploring the agency of nonhuman actors, including AI systems, in shaping social networks, influencing decision-making processes and generating new social realities. His work provides a framework for understanding AI not as a standalone technology, but as part of a complex network of human and nonhuman actors. Building on ANT, recent studies have examined the intricate relationships between AI systems and human actors. For example, Vertesi (2019) conducted an ethnographic study of NASA's Mars Exploration Rover mission, demonstrating how human–robot interactions shape scientific practices and knowledge production. Her work illustrates the coconstitution of human and machine agency in complex sociotechnical systems. Mühlhoff (2020) applied ANT principles to analyse human–AI interactions in the context of machine learning systems. His research reveals how AI technologies capture and leverage human cognitive abilities, creating hybrid human–machine apparatuses that challenge traditional notions of agency and intelligence.

Social Construction of Technology

The social construction of technology (SCOT) perspective examines how social factors shape the development, adoption, and use of AI systems. It acknowledges that technologies are not solely determined by their technical features but are coconstructed through social processes, interests and values. Pinch et al. (1987) laid the groundwork for SCOT, demonstrating how social groups influence technological development through processes of negotiation and interpretation. Their framework has been applied to AI technologies, revealing the ways in which social factors influence the development and adoption of AI systems. Recent studies have extended SCOT to analyse AI development and implementation. For instance, Selbst et al. (2019) examine the social construction of 'fairness' in machine learning systems, highlighting how different stakeholders' interpretations and values shape the development of 'fair' AI algorithms. Their work demonstrated the importance of considering diverse perspectives in AI design and implementation. Rezaev et al. (2023) apply SCOT principles to analyse the emergence of human-centred AI, emphasising the role of societal values and cultural contexts in shaping AI technologies. Their research underscores the need for a nuanced understanding of how different social groups interpret and influence AI development.

Technological Determinism

Technological determinism posits that technology drives social change and shapes various aspects of society. In the context of AI Sociology, this framework explores how AI technologies impact social structures, relationships and cultural practices. Brynjolfsson et al. (2014) analyse the implications of AI-driven automation on employment, income distribution and social inequality. They argue that technological advancements, including AI, have profound consequences for labour markets, economic systems and social structures. Their work highlights both the potential benefits and challenges of AI-driven economic transformation. Ford (2015) extends this analysis, examining the potential for AI and robotics to displace human workers across various industries. His research raises important questions about the future of work, income distribution and social welfare in an AI-driven economy. Recent empirical studies have provided further insights into the deterministic effects of AI on society. Frey et al. (2017) conducted a comprehensive analysis of the susceptibility of various occupations to computerisation, estimating that nearly half of U. S. jobs are at risk of automation. Their

work has sparked debates about the need for policy interventions and educational reforms to address the potential societal impacts of AI-driven job displacement.

By applying these theoretical frameworks, researchers in the field of AI Sociology gain a deeper understanding of the multifaceted relationship between AI and society. Critical theory highlights power dynamics and surveillance capitalism, actor-network theory emphasises the agency of nonhuman actors, the social construction of technology perspective focuses on the social shaping of AI, and technological determinism examines the transformative potential of AI in driving social change. These frameworks, when combined with empirical studies, provide valuable insights into the sociological dimensions of AI and its implications for individuals, communities, and societal structures.

THEMATIC ANALYSIS OF THE LITERATURE

AI and Social Stratification

Artificial Intelligence (AI) technologies have a significant potential to impact social stratification, which refers to the hierarchical division of society based on factors such as socioeconomic status, race and gender. This theme explores the complex relationship between AI and social stratification, shedding light on how AI systems can perpetuate or challenge existing inequalities. To provide a comprehensive understanding, it is crucial to first define social stratification in the context of AI. Social stratification refers to the unequal distribution of resources, opportunities and privileges within a society, resulting in differentiated social positions and experiences (Grusky 2018). AI technologies have the potential to reinforce existing social hierarchies or disrupt them, depending on how they are designed, deployed and utilised. A critical review of relevant papers allows us to delve into the various dimensions of AI's influence on social stratification. Several key topics within this theme merit examination:

• **Digital divide:** The digital divide refers to the unequal access to and use of digital technologies, including AI. AI advancements can exacerbate existing disparities in access to technological resources, perpetuating inequalities in education, employment, and information access (DiMaggio et al. 2004). Recent studies have shown that the AI-driven digital divide is not only about access to technology but also about the skills and knowledge required to effectively utilise AI systems (van Dijk 2020).

• **Bias and algorithmic discrimination:** AI systems can perpetuate biases present in training data or encoding algorithms, resulting in discriminatory outcomes. Examples include racial and gender bias in facial recognition systems and biased decision-making processes in areas such as criminal justice and hiring (Noble 2018; O'Neil 2016). Buolamwini et al. (2018) conducted a groundbreaking study on gender and racial bias in commercial facial recognition systems, highlighting the urgent need for diverse and representative datasets in AI development.

• **Threats to democracy and fairness:** AI can influence democratic processes and fairness by shaping information flows, political discourse, and decision-making. The concentration of AI power in a few entities and the potential for manipulation raise concerns about democratic values and equal representation (Barocas et al. 2016). Recent research by Helbing et al. (2019) explored the potential of AI to manipulate public opinion and undermine democratic processes through targeted misinformation campaigns.

Prominent studies in this field include the works of Noble (2018), who critically examines the biases and inequalities perpetuated by AI and algorithms in her book 'Algorithms of Oppression'. O'Neil (2016) explored the impact of algorithms on social inequality in 'Weapons of Math Destruction', providing compelling examples of how AI-driven decision-making can reinforce existing social disparities. Barocas et al. (2016) discuss the implications of AI for fairness and accountability in their seminal paper 'Big Data's Disparate Impact', which has been influential in shaping discussions on algorithmic fairness. Recent empirical studies have further substantiated these concerns. For example, Obermeyer et al. (2019) revealed racial bias in a widely used algorithm for predicting health care needs, demonstrating how AI systems can perpetuate and amplify existing racial disparities in healthcare access. Similarly, Angwin et al. (2016) revealed racial bias in criminal risk assessment algorithms used in the U. S. justice system, highlighting the potential for AI to reinforce systemic racism. These studies highlight the need to address biases, ensure fairness and promote an equal access to AI systems to mitigate the potential exacerbation of social inequalities. Understanding these dynamics is crucial for fostering more inclusive and equitable AI technologies.

Role of AI in Social Change and Development

This study explores the dynamic relationship between AI and social change, focusing on how AI technologies contribute to the transformation of institutions, culture, and power structures. To provide a comprehensive understanding, it is important to first clarify the concepts of social change and development in the context of AI. Social change refers to the alteration in societal structures, values, norms and behaviours over time (Sztompka 1993). Development, on the other hand, encompasses broader aspects of social progress, including economic growth, human well-being, and sustainable practices (Sen 1999). AI, as a transformative technology, has the potential to significantly influence both social change and development processes. When examining the role of AI in social change and development, it is essential to analyse relevant studies that explore these dynamics. Several topics related to this topic warrant exploration:

• **AI's impact on institutions:** AI technologies are reshaping various institutions, such as education, healthcare and governance. For example, AI-powered education platforms are revolutionising learning methods, whereas AI-driven healthcare systems are enhancing diagnostic accuracy and treatment outcomes. Holmes et al. (2019) conducted a comprehensive review of AI applications in education, highlighting both the potential benefits and challenges of AI-driven personalised learning. In healthcare, Topol (2019) explored how AI is transforming medical practice from diagnosis to treatment planning and discussed the implications for patient care and healthcare systems.

• **The influence of AI on culture:** Culture encompasses shared beliefs, values, practices and traditions within a society. AI technologies are increasingly influencing cultural practices and interactions. For instance, AI algorithms shape online content recommendations, influencing individuals' access to information and exposure to diverse perspectives. Bozdag (2013) examines the impact of AI-driven personalisation algorithms on information diversity and the formation of filter bubbles. More recently, Manovich (2018) explored how AI is reshaping cultural production and consumption in the age of social media and big data.

• **AI's impact on power structures:** AI technologies have implications for power dynamics and social inequalities. The concentration of AI resources and decision-making algorithms

in the hands of a few entities can influence access to resources, the distribution of opportunities, and decision-making processes. Zuboff (2019) provides a critical analysis of the rise of 'surveillance capitalism' and its implications for power structures in her influential work 'The Age of Surveillance Capitalism'. Additionally, Crawford et al. (2018) offer a detailed examination of the power dynamics embedded in AI systems through their 'Anatomy of an AI System' project.

Recent empirical studies have provided further insights into the role of AI in social change and development. For instance, Acemoglu et al. (2020) analyse the impact of AI and automation on labour markets and wage inequality, offering a nuanced perspective on the economic implications of AI-driven social change. In the realm of governance, Vesnic-Alujevic et al. (2020) explore the use of AI in public services and its implications for citizen–state relationships and democratic processes.

Ethical Implications of AI in Society

Ethics plays a crucial role in the study of AI from a sociological perspective. Ethics in AI refers to moral considerations and principles guiding the design, development, and use of AI technologies (Floridi et al. 2021). Ethical issues arise because of the potential consequences of AI systems on individuals' well-being, social justice and human rights. Understanding and addressing these ethical implications are vital for the responsible development and deployment of AI. Several topics within this theme merit examination:

• **Fairness:** AI systems are expected to make fair decisions without exhibiting bias or discrimination based on factors such as race, gender or socioeconomic status. Ensuring fairness in AI requires a careful consideration of algorithmic design, data selection and model training (Barocas et al. 2016). Recent studies have shown that AI systems can perpetuate and even amplify existing societal biases. For example, Buolamwini et al. (2018) demonstrated a significant gender and racial bias in commercial facial recognition systems, highlighting the need for diverse and representative datasets in AI development.

• **Transparency:** Transparency in AI systems refers to the ability to understand the underlying decision-making processes and the factors influencing AI outcomes. Transparent AI systems promote accountability and trust by enabling individuals to comprehend and challenge algorithmic decisions (Doshi-Velez et al. 2017). The concept of 'explainable AI' has gained traction, with researchers developing methods to make complex AI models more interpretable (Adadi et al. 2018).

• **Accountability:** The accountability of AI systems pertains to the attribution of responsibility for their actions. As AI systems become more autonomous, issues of accountability arise, requiring mechanisms for addressing and redressing harm caused by AI technologies (Jobin et al. 2019). Dignum (2019) proposes a framework for responsible AI that emphasises the need for clear lines of accountability throughout the AI lifecycle, from development to deployment and use.

• **Privacy:** AI systems often rely on vast amounts of personal data, raising concerns about privacy and data protection. Safeguarding individuals' privacy rights is crucial in the design and deployment of AI technologies (Mittelstadt et al. 2016). The implementation of regulations such as the General Data Protection Regulation (GDPR) in the European Union has sparked discussions on the balance between data-driven innovation and privacy protection in AI development (Wachter et al. 2017).

Additionally, it is essential to consider the risks associated with advanced AI and superintelligence. Advanced AI systems with capabilities that surpass human intelligence pose potential risks and uncertainties, requiring careful ethical scrutiny (Bostrom et al. 2014). The field of AI safety research has emerged to address these long-term concerns, with a focus on ensuring that advanced AI systems remain aligned with human values and interests (Amodei et al. 2016). Moreover, the need for transparency and explainability in AI systems is increasingly recognised. Individuals affected by AI decisions should have access to information about how these decisions are made, fostering trust and enabling individuals to hold AI systems accountable (Whittlestone et al. 2019). This has led to the development of 'algorithmic auditing' techniques to assess the fairness and transparency of AI systems (Raji et al. 2020). The ethical implications of AI extend beyond individual technologies to broader societal impacts. For instance, the use of AI in surveillance and social control raises concerns about civil liberties and human rights (Zuboff 2019). The concept of 'algorithmic governance' has emerged to describe the increasing role of AI systems in shaping social and political processes, raising questions about democratic accountability and the distribution of power in society (Danaher et al. 2017). Furthermore, the global nature of AI development and deployment necessitates consideration of cross-cultural ethical perspectives. Different societies may have varying ethical priorities and values, which can influence the development and acceptance of AI technologies (Hagendorff 2020). Recent empirical studies have begun to address the gap in practical applications of AI ethics. For example, Morley et al. (2020) conducted a systematic review of AI ethics guidelines, identifying common principles and areas of divergence across different frameworks.

RESEARCH GAPS AND SOLVING STRATEGIES

Despite the growing body of literature on AI Sociology, there are notable research gaps that need to be addressed. This section aims to identify and discuss these gaps, highlighting areas that require further exploration. The corresponding solution strategies are then provided.

Need for More Empirical Studies on AI Social Impact

While theoretical discussions and conceptual frameworks have contributed significantly to the field of AI Sociology, there is a pressing need for more empirical studies that examine the actual social impact of AI technologies. Based on the previous content, the landscape of empirical research in this field is rapidly evolving, with new studies emerging frequently. However, a significant gap still exists in our understanding of how AI is experienced and shaped by individuals and communities in real-world contexts. To further bridge this gap, researchers should conduct rigorous empirical research using a variety of methods, including the following: (1) Survey Method: large-scale quantitative studies to assess public perceptions and experiences with AI technologies across different demographic groups; (2) Interviews and Focus Group Method: in-depth qualitative research to explore individual and community experiences with AI systems; (3) Ethnographic Study: long-term observational research to understand how AI technologies are integrated into and shape social practices over time; (4) Mixed-Methods Approach: combining quantitative and qualitative methods to provide a comprehensive understanding of the social impact of AI. By employing these empirical approaches, we can gain valuable insights into the nuanced dynamics between AI and society, moving beyond theoretical speculations to evidence-based understandings.

Limited Attention to Non-Western Contexts

Cultural, social and institutional factors may influence the interaction between AI technologies and society. Much of the existing literature on AI Sociology is concentrated within Western contexts, primarily exploring the impact of AI in Western societies. This leaves a significant gap in how AI is experienced and perceived in non-Western contexts. Recent studies have begun to address this gap, but more research is needed. For example, Arora (2019) examined the implications of AI for digital labour in India, highlighting the unique challenges and opportunities in a non-Western context. Similarly, Wang et al. (2023) explored the social implications of China's AI development strategies, offering insights into how cultural and political factors shape AI adoption and impact in a non-Western setting. To develop a more comprehensive understanding of the global implications of AI and avoid potential biases in our analysis, future research should (1) conduct comparative studies across different cultural contexts to identify similarities and differences in the societal impact of AI; (2) engage with local researchers and communities in non-Western countries to ensure culturally sensitive and relevant research approaches; (3) examine how different cultural values, social norms and governance structures influence the development, adoption and impact of AI technologies.

Need for More Robust Interdisciplinary Frameworks in AI Sociology

AI is a complex phenomenon that requires interdisciplinary perspectives to fully comprehend its societal implications. While some interdisciplinary work exists, such as the integration of critical theory and AI ethics (Birhane et al. 2019), there is a need for more comprehensive interdisciplinary frameworks that integrate insights from several relevant disciplines. To address this gap, researchers should (1) develop collaborative research projects that bring together experts from diverse fields, including sociology, computer science, ethics, anthropology and policy studies; (2) establish interdisciplinary research centres and programs focused on AI Sociology to foster long-term collaboration and knowledge exchange; (3) encourage cross-disciplinary training and education to equip researchers with the diverse skills needed to address complex AI-related societal challenges. By incorporating these interdisciplinary approaches, we can gain a more holistic understanding of AI's impact on society, addressing the multifaceted nature of AI and its interactions with different social systems. This interdisciplinary perspective can enhance the robustness of AI Sociology as a field of study.

Ethical Considerations in AI Sociology Research

It is crucial to address the ethical issues that arise in AI Sociology research. Future studies should (1) develop ethical guidelines specific to AI Sociology research, considering the unique challenges posed by studying the societal impact of AI; (2) address issues of data privacy and consent when collecting and analysing data related to AI use and impact; (3) consider the potential unintended consequences of AI Sociology research, such as reinforcing harmful stereotypes or exacerbating existing inequalities; (4) engage with affected communities in the research process, ensuring that their voices and perspectives are adequately represented. Addressing these research gaps is crucial for advancing the field of AI Sociology and ensuring a comprehensive understanding of the social implications of AI technologies.

CONCLUSIONS

Through an analysis of theoretical frameworks, thematic discussions, and identification of research gaps, we have gained valuable insights into the multifaceted relationship between AI and society. This review highlights several critical aspects of the social implications of AI, including ethical considerations, social stratification, and social change and development. The review has shown how AI technologies drive transformative processes, reshape institutions and impact power structures. This theme highlights the importance of considering both the potential benefits and challenges posed by AI-driven social change, calling for a nuanced understanding of the broader sociological implications of AI. By integrating theoretical frameworks, empirical studies and interdisciplinary perspectives, researchers have made substantial progress in understanding the complex interplay between technology and society.

Despite these advancements, the limitations of the research conducted thus far include the following: (1) Rapid technological advancements: The field of AI is evolving at a pace that often outstrips academic research. Future studies should aim to keep pace with these rapid changes and provide timely insights into emerging social challenges. (2) Need for diverse and inclusive research: The majority of existing literature focuses on Western contexts, leaving gaps in our understanding of AI's impact in non-Western societies and marginalised communities. Future research should strive for a more global perspective, considering diverse cultural, social and economic contexts. (3) Empirical studies: There is a growing need for empirical studies that examine the actual social impact of AI technologies in real-world contexts. Future research should prioritise rigorous empirical investigations to complement theoretical analyses. (4) Interdisciplinary approaches: The complex nature of the societal impact of AI necessitates interdisciplinary research approaches. Future studies should foster collaborations between sociologists, computer scientists, ethicists and policymakers to develop comprehensive frameworks for understanding and addressing AI challenges.

In summary, this literature review provides valuable insights into the social implications of AI technologies, highlighting ethical considerations, social stratification dynamics, and the transformative role of AI in society. By addressing the identified research gaps and embracing interdisciplinary approaches, future studies can contribute to responsible AI development, informed policymaking, and a more equitable and inclusive society.

References

1. Acemoglu, D.; Restrepo, P. 2020. 'The Wrong Kind of AI? Artificial Intelligence and the Future of Labour Demand', *Cambridge Journal of Regions, Economy and Society* 13(1): 25–35.
2. Adadi, A.; Berrada, M. 2018. 'Peeking Inside the Black-box: A Survey on Explainable Artificial Intelligence (XAI)', *IEEE Access* 6: 52138–52160.
3. Amodei, D.; Olah, C.; Steinhardt, J.; Christiano, P.; Schulman, J.; Mané, D. 2016. 'Concrete Problems in AI Safety', *arXiv preprint*. Available at: https://arxiv.org/abs/1606.06565
4. Andrejevic, M. 2020. *Automated Media*. New York: Routledge.
5. Angwin, J.; Larson, J.; Mattu, S.; Kirchner, L. 2016. 'Machine Bias', *ProPublica*, 23 May. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
6. Arora, P. 2019. *The Next Billion Users: Digital Life Beyond the West*. Harvard University Press.
7. Barocas, S.; Selbst, A. D. 2016. 'Big Data's Disparate Impact', *California Law Review* 104(3): 671–732.
8. Bostrom, N.; Yudkowsky, E. 2014. 'The Ethics of Artificial Intelligence', in *The Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press, 316–334.
9. Bozdag, E. 2013. 'Bias in Algorithmic Filtering and Personalization', *Ethics and Information Technology* 15(3): 209–227. Available at: https://doi.org/10.1007/s10676-013-9321-6

10. Brynjolfsson, E.; McAfee, A. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W. W. Norton & Company.

11. Buolamwini, J.; Gebru, T. 2018. 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', in *Conference on Fairness, Accountability, and Transparency*. PMLR, 77–91.

12. Crawford, K.; Joler, V. 2018. *Anatomy of an AI System: The Amazon Echo as an Anatomical Map of Human Labor, Data, and Planetary Resources*. AI Now Institute and Share Lab. Available at: https://anatomyof.ai/

13. Crawford, K.; Schultz, J. 2014. 'Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms', *Boston College Law Review* 55(1): 93–128.

14. Danaher, J.; Hogan, M. J.; Noone, C.; Kennedy, R.; Behan, A.; De Paor, A.; Felzmann, H.; Haklay, M.; Khoo, S. M.; Morison, J.; Murphy, M. H. 2017. 'Algorithmic Governance: Developing a Research Agenda Through the Power of Collective Intelligence', *Big Data & Society* 4(2). Available at: https://doi.org/10.1177/2053951717726554

15. Dignum, V. 2019. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer Nature.

16. DiMaggio, P.; Hargittai, E.; Celeste, C.; Shafer, S. 2004. 'From Unequal Access to Differentiated Use: A Literature Review and Agenda for Research on Digital Inequality', in *Social Inequality*, ed. K. Neckerman. New York: Russell Sage Foundation, 355–400.

17. Doshi-Velez, F.; Kim, B. 2017. 'Towards a Rigorous Science of Interpretable Machine Learning', *arXiv preprint* arXiv: 1702.08608.

18. Floridi, L.; Cowls, J. 2019. 'A Unified Framework of Five Principles for AI in Society', *Harvard Data Science Review* 1(1).

19. Floridi, L.; Cowls, J.; King, T. C.; Taddeo, M. 2021. 'How to Design AI for Social Good: Seven Essential Factors', in *Ethics, Governance, and Policies in Artificial Intelligence*, 125–151.

20. Ford, M. 2015. *Rise of the Robots*. New York: Basic Books.

21. Frey, C. B.; Osborne, M. A. 2017. 'The Future of Employment: How Susceptible are Jobs to Computerisation?', *Technological Forecasting and Social Change* 114: 254–280.

22. Grusky, D. B. 2018. *Social Stratification: Class, Race, and Gender in Sociological Perspective*. 4th edn. New York: Routledge.

23. Hagendorff, T. 2020. 'The Ethics of AI Ethics: An Evaluation of Guidelines', *Minds and Machines* 30(1): 99–120.

24. Helbing, D.; Frey, B. S.; Gigerenzer, G.; Hafen, E.; Hagner, M.; Hofstetter, Y.; van den Hoven, J.; Zicari, R. V.; Zwitter, A. 2019. 'Will Democracy Survive Big Data and Artificial Intelligence?', in *Towards Digital Enlightenment*, ed. D. Helbing. Cham: Springer, 73–98. Available at: https://doi.org/10.1007/978-3-319-90869-4_7

25. Holmes, W.; Bialik, M.; Fadel, C. 2019. *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Boston: Center for Curriculum Redesign.

26. Jobin, A.; Ienca, M.; Vayena, E. 2019. 'The Global Landscape of AI Ethics Guidelines', *Nature Machine Intelligence* 1(9): 389–399.

27. Latour, B. 2005. 'From Realpolitik to Dingpolitik or How to Make Things Public', in *Making Things Public: Atmospheres of Democracy*, eds. B. Latour and P. Weibel. Cambridge, MA: MIT Press, 14–41.

28. Manovich, L. 2018. 'Digital Traces in Context: 100 Billion Data Rows per Second: Media Analytics in the Early 21st Century', *International Journal of Communication* 12: 16.

29. Mittelstadt, B. D.; Allo, P.; Taddeo, M.; Wachter, S.; Floridi, L. 2016. 'The Ethics of Algorithms: Mapping the Debate', *Big Data & Society* 3(2).

30. Mühlhoff, R. 2020. 'Human-aided Artificial Intelligence: Or, How to Run Large Computations in Human Brains? Toward a Media Sociology of Machine Learning', *New Media & Society* 22(10): 1868–1884. Available at: https://doi.org/10.1177/1461444819885334

31. Noble, S. U. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

32. Obermeyer, Z.; Powers, B.; Vogeli, C.; Mullainathan, S. 2019. 'Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations', *Science* 366(6464): 447–453. Available at: https://doi.org/10.1126/science.aax2342

33. O'Neil, C. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

34. Pinch, T.; Bijker, W. E. 1987. 'The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other', in *The Social Construction of Technological Systems*, eds. W. E. Bijker, T. P. Hughes and T. Pinch. Cambridge, MA: MIT Press, 17–50.

35. Raji, I. D.; Smart, A.; White, R. N.; Mitchell, M.; Gebru, T.; Hutchinson, B.; Smith-Loud, J.; Theron, D.; Barnes, P. 2020. 'Closing the AI Accountability Gap: Defining an End-to-end Framework for Internal Algorithmic Auditing', in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 33–44.

36. Rezaev, A. V.; Tregubova, N. D. 2023. 'Looking at Human-centered Artificial Intelligence as a Problem and Prospect for Sociology: An Analytic Review', *Current Sociology*.

37. Selbst, A. D.; Boyd, D.; Friedler, S. A.; Venkatasubramanian, S.; Vertesi, J. 2019. 'Fairness and Abstraction in Sociotechnical Systems', in *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, 59–68.

38. Sen, A. 1999. *Development as Freedom*. Oxford: Oxford University Press.

39. Sztompka, P. 1993. *The Sociology of Social Change*. Oxford: Blackwell.

40. Van Dijk, J. 2020. *The Digital Divide*. John Wiley & Sons.

41. Vertesi, J. 2019. 'Shaping Science with the Past: How History Becomes Active in Working Relationships with Spacecraft', *Social Studies of Science* 49(1): 120–145.

42. Vesnic-Alujevic, L.; Stoermer, E.; Rudkin, J.; Scapolo, F.; Kimbell, L. 2020. *The Future of Government 2030+: A Citizen-centric Perspective on New Government Models*. Luxembourg: Publications Office of the European Union. Available at: https://doi.org/10.2760/145751

43. Wachter, S.; Mittelstadt, B.; Floridi, L. 2017. 'Why a Right to Explanation of Automated Decision-Making Does not Exist in the General Data Protection Regulation', *International Data Privacy Law* 7(2): 76–99. Available at: https://doi.org/10.1093/idpl/ipx005

44. Whittlestone, J.; Nyrup, R.; Alexandrova, A.; Dihal, K.; Cave, S. 2019. *Ethical and Societal Implications of Algorithms, Data, and Artificial Intelligence: A Roadmap for Research*. London: Nuffield Foundation.

45. Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

CHUNFA ZHOU

Dirbtinis intelektas sociologijoje: kritinė apžvalga ir ateities kryptys

Santrauka

Šiame tyrime pateikiama kritiška besiformuojančios dirbtinio intelekto (DI) sociologijos srities apžvalga, nagrinėjant DI technologijų socialines pasekmes ir etinius aspektus. Taikant kokybinę metodologiją, apimančią literatūros apžvalgą ir teminę analizę, šiame tyrime nagrinėjama DI ir sociologijos sankirta, siekiant sumažinti atotrūkį tarp technologinės pažangos ir poveikio visuomenei. Tyrime nagrinėjamos pagrindinės teorinės sistemos, įskaitant kritinę teoriją ir veikėjų bei tinklų teoriją, siekiant analizuoti galios santykius, socialinę stratifikaciją ir dinamišką DI ir visuomenės sąveiką. Išvados atskleidžia įvairiapusę DI įtaką socialinėms struktūroms, etinius iššūkius ir tarpdisciplininio požiūrio poreikį sprendžiant DI poveikį visuomenei.

Raktažodžiai: dirbtinio intelekto (DI) sociologija, socialiniai DI padariniai, etiniai DI aspektai, technologinis poveikis visuomenei, tarpdisciplininiai DI tyrimai