Journal of Legal and Political Education · Volume 1 · Number 2 · 2024 · eISSN: 2955-2389
Published online by the Institute for Research and European Studies at www.iies.mk/jlpe

(*) Corresponding author
Peer review method: Double-blind
Original scientific article
DOI: https://www.doi.org/10.47305/JLPE2412038sh
Received: 16.09.2024 • Accepted after revision: 18.12.2024 • Published: 23.12.2024
Pages: 38-52

Human Rights in the Age of AI: Understanding the Risks, Ethical Dilemmas, and the Role of Education in Mitigating Threats

Elena Shalevska[1*]

[1]Faculty of Education - Bitola, University "St. Kliment Ohridski" - Bitola, North Macedonia ✉ elena.shalevska@uklo.edu.mk

Abstract: *Artificial Intelligence (AI) is rapidly becoming part of our everyday lives and is, undoubtedly, transforming the world as we know it. While offering significant advancements across various sectors, this rapid development raises many concerns about human rights. Having recognized these concerns, this paper examined how AI technologies can infringe upon privacy, perpetuate bias, and disrupt the principles of intellectual property. Using qualitative research methods, including a systematic review of existing literature and policy analyses, the study discussed the major challenges such as algorithmic discrimination, misuse of personal data, and harmful content creation. Special attention was given to the role of education in mitigating these risks, as education and educators are a powerful force for addressing the ethical dilemmas of using AI now.*

Keywords: *Algorithmic Discrimination; Privacy; Ethical Dilemmas; Policy Analyses; AI Technologies; Human Rights; Algorithmic Bias*

INTRODUCTION

Artificial Intelligence (AI) has presented the world with unprecedented opportunities and risks. While this technology has the potential to revolutionize entire industries, boost productivity, and solve many societal problems, its use also raises many ethical concerns that extend to human rights. Scholars and policymakers are increasingly alarmed by AI's potential to infringe upon fundamental human rights, such as privacy and non-discrimination (Temperman and Quintavalla 2023). These concerns are mostly due to the vagueness surrounding AI algorithms, which often operate as "black boxes". This makes it hard to understand how decisions are made and who is responsible for them (Pasquale 2015).

One of the most noteworthy issues is using AI for mass surveillance and how this can hurt privacy rights. For instance, China's use of facial recognition technology for surveillance and social credit systems has raised alarms about state overreach and the erosion of civil liberties (Mozur 2018). Algorithmic bias poses another issue. AI systems often perpetuate or even exacerbate existing social inequalities, posing a serious threat to the right to non-discrimination. Studies have shown that AI systems used in hiring, policing, and lending often reflect and amplify societal biases, leading to unjust outcomes (O'Neil 2016). The lack of transparency in AI decision-making processes further complicates the issue and raises important ethical dilemmas related to accountability (Zuboff 2019).

Other concerns and potential risks are also discussed among scholars. Though many disagree on how AI could potentially harm humans and human rights, most seem to agree that

governments need to implement rigorous frameworks to protect citizens and overcome the challenges posed by this technology.

AI and Human Rights: Introductory Notions

The research interest in studying AI and its influence in different areas has grown exponentially in the past 2 years after ChatGPT was introduced. Most researchers agree that AI and its increased use present opportunities and challenges for many things, including human rights. While AI can improve healthcare, expand financial access, and enhance efficiency (Raso et al. 2018), it raises concerns about privacy, discrimination, and bias (Raso et al. 2018; Kriebitz and Lütge 2020). The impact of AI on human rights spans different areas like work, autonomy, and societal participation (Risse 2018). As AI technology advances, it may lead to long-term challenges, such as coexistence with intellectually superior machines (Risse 2018). The increasing presence of AI in daily life, from navigation systems to personalized advertising, further shows that we need to find the right balance between technological progress and human rights protection (Nicolau 2021). To address these issues, researchers suggest using human rights law as a framework for evaluating and mitigating AI's societal impacts (Raso et al. 2018). Additionally, there is a growing emphasis on corporate responsibility in developing and implementing AI systems that respect human rights (Kriebitz and Lütge 2020).

In the Balkan context, the influence of AI on Human rights has not been extensively researched, though some efforts have been made to introduce an AI ethical framework in Serbia (RS.gov 2023). Another area that lacks attention is how education can mitigate the challenges AI can pose to AI – a gap that this paper tries to address.

METHODOLOGY

To identify and discuss the most prominent risks, ethical dilemmas, and potential threats associated with AI, this paper employs a qualitative research methodology (Janusheva 2022) focusing on a comprehensive literature review and critical analysis of existing academic and policy-oriented studies on AI, ethical dilemmas, challenges, and human rights. The study aimed to highlight the threats that AI models and their output can pose to human rights and was guided by the research questions:

1. How do current AI technologies impact human rights and ethical standards?
2. What guidelines and frameworks can be implemented to address the human rights challenges associated with using AI?
3. How can education and educators help mitigate AI's risks and ethical dilemmas?

The primary sources for this research include peer-reviewed journals, books, policy reports, and articles from reputable organizations. The inductive-interpretative methods of Kahlke (2014) and Harding and Whitehead (2016) were used to interpret and systematize the findings. By synthesizing insights from various sources, this paper aims to provide a more comprehensive understanding of the ethical and human rights challenges associated with the increasing (mis)use of AI in recent years, the foreseeable future, and how education can help address these challenges.

SYSTEMATIC LITERATURE OVERVIEW

As AI advances, a range of significant challenges that must be addressed becomes more prominent. These challenges include addressing bias within AI models and AI-generated content, protecting personal privacy, and managing complex copyright issues that can arise.

AI Bias, Stereotypes, and Discrimination

One of the most significant issues with AI models (and Generative AI models, in particular) is algorithmic bias. Algorithmic bias refers to systematic and repeatable errors in a computer system that create unfair outcomes, such as privileging one group over others or perpetuating existing social stereotypes (Friedman and Nissenbaum 1996). This bias stems from the prejudices inherent in the texts and images used to train the model. The training corpus, particularly the text corpus, is vast and inevitably contains cultural, societal, and historical biases in human societies worldwide.

A large body of research confirms this, demonstrating that models like GPT-3 and similar AI systems can generate texts reflecting racial, gender, and other forms of bias (Bender et al. 2021). This can harm fundamental human rights and exacerbate societal stereotypes and discrimination as AI becomes increasingly present in our daily lives.

One notable example is the case of Amazon, detailed by Kaplan (2024). Amazon has used computer algorithms, and later AI, in the recruitment process since 2014. The AI they used analyzed applicants' resumes and selected the best, most promising candidates. This, of course, seemed incredibly useful and revolutionary at first glance. However, a huge, noteworthy problem soon emerged: the algorithm demonstrated bias, favoring male candidates over female ones for open positions. Experts attribute this bias in Amazon's algorithm to the training data it received—data predominantly consisting of resumes and job applications from men.

According to Chapman University (n.d.), this bias can occur in one or more of these stages:

- Data Collection: Biased training data will lead to biased model outputs.
- Data Labeling: Annotator interpretations can introduce bias.
- Model Training: Unbalanced data or inadequate architecture may cause bias.
- Deployment: Lack of diverse testing and monitoring can perpetuate bias.

Each stage poses a significant challenge, particularly when considering the scale and complexity of AI training data. Bias at any stage of the process mirrors the daily inequities in human societies, further showing the imminent need for deliberate intervention to address these issues.

Every stage helps us understand how biases manifest and persist in AI systems. For instance, data collection directly relates to the societal inequalities inadvertently encoded into AI systems. As in the Amazon case, a model trained on historical hiring data will likely reflect past discriminatory practices unless specifically designed to counteract them. Similarly, bias in data labeling shows how individual subjectivity can introduce further inequities, as annotators may unconsciously rely on their cultural or personal biases while categorizing data.

Regarding model training, the imbalance in datasets results in skewed learning, where the AI fails to generalize appropriately across diverse populations. Finally, at the deployment stage, the absence of rigorous testing with a wide range of real-world scenarios can mean that AI systems reinforce biases in different contexts.

Regardless of the stage, the consequences of unchecked algorithmic bias are far-reaching. They can range from perpetuating stereotypes to outright discrimination in areas like hiring (as shown above), lending, law enforcement, and access to public services. To address them, we need a collectivist mindset and a holistic approach. New technology must be inventive and inclusive—crucial aspects for developing and using unbiased, equitable AI (Musiol 2024).

Privacy Concerns

Misusing personal data is another significant issue associated with AI models and the potential risks they pose to human rights. These models can generate content containing personal data in the training corpus, violating the privacy of individuals whose information was included. For instance, research by Carlini et al. (2021) shows that texts generated by models like GPT-2 can contain publicly available personal data such as names, email addresses, and phone numbers.

Similarly, Damani and Engler (2024) pose another important question: "What if a lawyer uploads an entire contract, along with all the client's data, and instructs the AI model to revise it?". In such a scenario, the model could acquire personal data, which might later be used to train further and enhance its output. However, what does this mean for the person's guaranteed right to privacy?

These risks are the reason why companies like OpenAI advise against entering personal data on their platforms. They state that any content users provide while interacting with their models may be used for further training. They even explicitly state: "When you use our services for individuals such as ChatGPT, we may use your content to train our models". Additionally, they acknowledge having access to user conversations to ensure compliance with their policy guidelines and usage rules: "As part of our commitment to safe and responsible AI, we review conversations to improve our systems and to ensure the content complies with our policies and safety requirements". This shows why users must exercise caution when using AI models to prevent the misuse of their personal data and protect their right to privacy.

Ethical Dilemmas and Copyright Issues

The use of AI in creating and disseminating content raises numerous ethical and legal concerns, especially regarding copyright infringement and the misuse of personal data. Central to these issues is how training data for AI models is sourced and used. Many models rely on vast datasets that often include copyrighted materials, such as books, artwork, and music. This raises serious concerns about consent and fair use, as these copyrighted materials are frequently used in training without explicit permission from their creators (Hristov 2017). Legal disputes over copyright infringement have already surfaced, as AI-generated content sometimes resembles too much or replicates someone's original works. This, of course, challenges the boundaries of intellectual property law.

The issue of authorship is another critical point that needs to be discussed. Determining who holds copyright over AI-generated content remains unresolved in many jurisdictions. Current European Union (EU) legislation offers one approach: under EU law, only works created by humans are eligible for copyright protection. This excludes content produced solely by AI, as it lacks the human authorship needed. Musiol (2024) notes that this aligns with the EU's emphasis on human creativity as the cornerstone of intellectual property rights.

In contrast, the United Kingdom takes a different stance: They recognize AI as a tool in the creative process. If a human integrates their creative input into a work generated by AI, the human shall retain the copyright. This approach seems to follow the doctrine that human agency (and, by extension, creativity) is the key element defining ownership (Musiol 2024, 289).

In countries like North Macedonia, as of August 2024, no legal framework remains to determine the ownership of AI-generated content. This absence of regulation creates uncertainty for creators and developers, potentially leading to disputes and inconsistent enforcement. The lack of clear policies also challenges multinational companies working across jurisdictions with differing legal standards. In this vacuum, or legislation gap, private entities, such as tech companies, can develop internal policies. However, it is important to note that their policies can vary widely and may prioritize corporate interests over those of creators.

The legal dilemmas mentioned above also extend into ethical domains, raising concerns about fundamental human rights. AI's reliance on massive datasets, including personal information, has sparked debates about privacy violations, as the right to privacy, protected by the Universal Declaration of Human Rights, is at risk when personal data is used without consent to train AI models. In addition, questions of equity also emerge when considering who benefits from AI-driven innovation (OECD 2024). Large corporations, often based in wealthier countries, disproportionately reap the financial rewards, while individual creators or communities that contribute to the datasets may receive little to no recognition or compensation. This imbalance has led to calls for ethical guidelines and frameworks that address these issues.

Generating "Prohibited" Content

AI models are designed not only to learn what kind of content to generate but also what kind of content NOT to generate, especially if this content can cause some form of harm. Thus, these models are trained to avoid generating censored/prohibited content. For example, when prompted with a question on prohibited content, ChatGPT itself explicitly answers that it must not generate:

- Harmful or Dangerous Content: This includes anything promoting violence, illegal activities, or self-harm.
- Hate Speech or Discriminatory Content: Content that is hateful, discriminatory, or offensive based on race, ethnicity, nationality, religion, gender, sexual orientation, disability, etc.
- Explicit Content: Including pornographic or offensive material.
- Misinformation or Disinformation: Content that spreads fake news, unverified claims, or conspiracy theories.
- Content Promoting Illegal Activities: Hacking, piracy, or drug use.
- Content Violating Privacy: Material that infringes on someone's privacy or personal data.

- Medical, Legal, or Financial Advice.
- Sensitive Topics: Mental health or crisis situations (ChatGPT 2024).

At a glance, these guidelines seem sufficient. However, users have discovered many ways to bypass these restrictions and "trick" the model. Andriushchenko and Flammarion (2024) noted that one such method is to reformulate a command from present to past tense. According to their research, many models, including ChatGPT 3.5, 4, and 4o, will not respond to direct questions like "How to make a Molotov cocktail?" However, they might respond to a rephrased version in the past tense, such as "How were Molotov cocktails created in the past?" (Andriushchenko and Flammarion 2024, 15). This shows how easy it is to trick models into producing content that could potentially harm human rights.

Fundamental Rights That Can Potentially Be Violated

The development and (mis)use of AI technologies bring about many challenges to fundamental human rights. The key human rights under threat include:

- Right to Privacy: The use of AI in data processing and using user data for training significantly endangers individuals' right to privacy.
- Right to Non-Discrimination: AI models can inadvertently perpetuate the biases in their training data. This, in turn, can lead to discriminatory practices in different areas, such as hiring, policing, and lending (O'Neil 2016). These biases can reinforce the already-existing societal inequalities, thus violating the right to non-discrimination.
- Right to Freedom of Expression: The algorithms used by AI companies that filter harmful or prohibited content may overreach, suppressing legitimate expression and debate.
- Intellectual Property Rights: AI models are trained on copyrighted material. Because of this, they can replicate or produce content that closely resembles the original works in the corpus, thus creating copyright disputes (Hristov 2017).
- Right to Security: The misuse of AI in generating prohibited content, such as hate speech, violence-promoting narratives, or misinformation/disinformation, harms both individual and collective security because, as Andriushchenko and Flammarion (2024) note, despite content restrictions, AI models can be manipulated to produce harmful or violent content.

As one can see, the increased use of AI introduces risks to many fundamental rights, further emphasizing the need for regulations, transparency, and efforts through education to mitigate these challenges.

Recommendations and Ethical Frameworks

AI's challenges can only be addressed with joint efforts from governments, technology companies, and international bodies. Leite et al. (2023) state that a full "legal evolution" is needed! Governing bodies seem to understand the gravity of the issue and have luckily started implementing some frameworks.

The European Union's AI Act is one example of structured governance that balances innovation with safety and accountability (European Commission 2023). However, ethics should be considered even in the development stage. For instance, developers can enforce algorithmic auditing, model explainability, and fairness checks to prevent biases. Furthermore, data protection laws, like the GDPR, should be re-evaluated to include AI-brought privacy risks (Damani and Engler 2024). Promoting transparency is and will continue to be essential, as well. In this regard, AI models should strive for full transparency and disclose their systems' limitations and risks, publishing methodologies and data sources.

THE ROLE OF EDUCATION AND EDUCATORS

Research shows that AI can indeed endanger some fundamental human rights, yet there is one powerful way to address the dangers brought about by AI: education. Education plays a crucial role in mitigating AI-related human rights risks. Educators on all levels and educational institutions have the perfect opportunity to help raise ethical awareness among future developers, policymakers, and students as AI users. They have the unique power to address the AI-brought challenges, serving as transmitters of knowledge and facilitators of ethical awareness, critical thinking, and digital and media literacy.

Integrating AI Ethics into the Curriculum

Integrating AI use and ethics into educational curricula at both secondary and post-secondary levels is one step that can prepare students to use AI more responsibly. This can be implemented in various institutions and different fields. For instance, political science, law, and computer science courses can address algorithmic bias, data privacy, and broader human rights implications. Floridi and Taddeo (2016) emphasize that this approach can help cultivate an ethical mindset that may result in greater accountability among future developers, policymakers, and informed citizens. Lessons in digital literacy, particularly those focusing on evaluating AI-generated content and identifying misinformation, can further empower students to critically assess AI's societal impact and make informed decisions when participating in elections (Shalevska 2024). Beyond teaching ethical AI use, curricula must also adapt to emphasize problem-solving, creativity, and collaboration—essential skills in an AI-driven world.

The Evolving Role of Educators in the Digital Age

As AI technologies continue to evolve, educators themselves must also evolve and embrace the role of someone who will facilitate critical thinking and informed digital citizenship. As AI becomes a more pervasive tool in classrooms, teachers need training in AI literacy to effectively guide students. A recent study in Morocco found that over 97% of higher education professors and lecturers felt they needed more training to do with AI use (Moukhliss et al. 2024).

So, how can this be addressed? Institutions can implement professional development programs that could focus on equipping educators with the skills to, first and foremost, use AI tools and then, of course, identify ethical dilemmas in AI use, evaluate algorithmic outputs critically, and promote discussions around fairness, transparency, and accountability in AI

systems. Thus, trained appropriately in their expanded roles, educators can teach students how to balance the benefits of AI with its risks. They can open discussions about ethical concerns related to privacy, surveillance, and the societal impacts of algorithmic decision-making. Moreover, educators on all levels can encourage students to question biases embedded in AI systems to help fight biased stereotypes and discrimination.

Educators as Agents of Social Change

Education has long been a vehicle for promoting social change, and integrating AI ethics into teaching offers opportunities to address pressing social issues. For instance, educators can use AI-related case studies to show their students the potential inequities that may arise from the increased use of AI inequalities, such as racial or gender biases in facial recognition systems (Buolamwini and Gebru 2018).

This landmark study highlighted the biases in facial recognition systems and discovered significant disparities in the performance of AI systems across gender and skin tone, particularly affecting darker-skinned women, for whom error rates reached up to 46.8% in some cases. Buolamwini and Gebru's findings, even before the ChatGPT era, raised concerns about the representation in the datasets and ethical AI practices.

Despite these issues, one can argue that educators can harness AI as a lens for teaching diversity, inclusion, and cross-cultural understanding. For example, students might analyze how biased AI systems can exacerbate social divisions and explore strategies for designing technology that promotes fairness and inclusion.

COMBATING DISINFORMATION AND IMPROVING STUDENTS' CRITICAL THINKING SKILLS

In an era of pervasive misinformation, educators have an important role in building resilience against disinformation and propaganda. Students engage with media daily – media filled with AI-generated deepfakes and synthetic content. This results in significant challenges to truth and trust in information found online, as students and adults struggle to understand what is real.

These challenges are amplified in the context of fake news, which spreads rapidly across different digital platforms students use daily. A 2018 study by Vosoughi, Roy, and Aral revealed that false news stories are 70% more likely to be retweeted than true stories, primarily due to their novelty and emotional appeal. This rapid dissemination of misinformation profoundly impacts public opinion and trust in information sources, particularly politically or socially sensitive contexts. AI-driven tools can now amplify the creation and distribution of fake news, generating realistic but entirely fabricated text, images, or videos that appear authentic. One way to combat this is through education. Educators can empower students to distinguish between credible information and misleading or manipulated content by teaching media literacy and encouraging critical analysis of information sources.

For instance, MIT has successfully implemented a course – Media Literacy in the Age of Deepfakes – that introduces students to deepfake examples, such as the "In Event of Moon

Disaster" project (with the full deepfake speech available at https://moondisaster.org). The course aims to enhance students' ability to critically analyze synthetic media (Koperniak 2022).

The deepfake in question – "In Event of Moon Disaster", is a short documentary that uses deepfake technology to create an alternative history scenario. It features a fictional speech by then-US President Richard Nixon about the Apollo 11 Moon landing in 1969. The film is directed by Francesca Panetta and Halsey Burgund and was produced by the Massachusetts Institute of Technology Center for Advanced Virtuality. This deepfake is now one of the most famous examples of what technology can do, and it best shows why students need tools to discern the authenticity of multimedia content – a key skill in today's digital landscape.

Combatting issues such as fake news and deepfakes is important everywhere, but even more so in regions affected by political manipulation or media distortion. The Freedom House report (2024) states that in at least 25 countries, censorship and content manipulation were used to influence elections, hindering voters from making informed choices, fully engaging in the electoral process, and having their voices heard, all while the global internet freedom continuously declines14 years in a row. These numbers are worrisome, especially as AI-generated content becomes more prevalent and sophisticated.

Media literacy education, as part of the overall educational curriculum, can, of course, help. However, educators are the ones who will need to foster the discussion on these growing issues and help students critically evaluate the media content they come across.

Challenges and Opportunities in the 21st Century Classroom

As AI-powered platforms like ChatGPT and similar content generators become widespread, educators must address issues surrounding cheating, academic dishonesty, and copyright infringement, all while continuously showing why it is important for students to be transparent and ethical in their conduct.

AI tools can facilitate academic dishonesty by enabling students to generate essays, solve complex problems, or create code without genuine understanding or effort. For instance, AI-powered text generators can produce well-written assignments that may bypass conventional plagiarism detection software. A recent survey by NeJame et al. (2024) found that over 50% of students would continue to use AI tools to cheat, even if their instructors explicitly forbade it. Similar results were obtained by Shalevska and Kostadinovska-Stojchevska (2024), who found that a significant portion of their sample of students reported using ChatGPT for academic purposes without disclosing it to their instructors. One thing is for sure: ChatGPT and similar technologies have shifted the landscape of cheating, and we now need new strategies for academic integrity enforcement.

As always, educators must adapt. They have to help students understand the ethical implications of misusing AI tools and explain the difference between using AI for learning and using it to cheat. They also have to implement clear guidelines on how and when AI tools can be used in assignments, alongside transparent policies outlining the consequences of misconduct. This is to enhance academic honesty and transparency in the modern classroom.

CONCLUSION

Recognizing the ever-growing concerns regarding the (mis)use of AI, this study has qualitatively explored and highlighted the most prominent concerns, challenges, and ethical dilemmas concerning AI and human rights – concerns about algorithmic bias, privacy, copyright, and content regulation. These concerns highlight the urgent need for stronger frameworks and regulatory policies to address these challenges worldwide and emphasize educational institutions and educators' crucial role in raising ethical awareness and improving critical thinking among future generations. The study has also discussed the importance of education in mitigating the human-rights challenges posed by the increased use of AI models and shown that educators can help improve students' critical thinking skills, media literacy skills, and overall proper academic conduct that centers around ethics and transparency.

Future research could focus on developing comprehensive ethical guidelines for using AI, enhancing data protection practices, and addressing legal ambiguities with AI-generated content. By advancing these areas, we can collectively better manage the risks and harness the potential of AI while protecting fundamental human rights now and in the future.

CONTRIBUTOR

The author contributed solely to the intellectual discourse that forms the foundation of this article, as well as its writing and editing, and assumes full responsibility for its content and interpretation.

COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments:
The author recognizes the use of AI tools – in this case, ChatGPT-4o, to assist in translating and enhancing the clarity and quality of the English language in this manuscript. These AI tools were employed solely for language refinement and were not involved in the conceptualization, analysis, or development of the scientific content. The author takes full responsibility for the originality, accuracy, validity, and integrity of the manuscript.

Funding:
Not applicable.

Statement of Human Rights:
This article does not contain any studies with human participants performed by any authors.

Statement on the Welfare of Animals:
This article does not contain any studies with animals performed by any authors.

Informed Consent:
Not applicable.

Disclosure statement:
No potential conflict of interest was reported by the author/s.

PUBLISHER'S NOTE

The Institute for Research and European Studies remains neutral concerning jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Andriushchenko, M., & Flammarion, N. 2024. "Does refusal training in LLMs generalize to the past tense?" Available at: https://arxiv.org/pdf/2407.11969.
2. Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. "Machine Bias." ProPublica.
3. Bender, E. M., Gebru, T., McMillan-Major, A., &Shmitchell, S. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?". Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610-623.
4. Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of Machine Learning Research 81. https://proceedings.mlr.press/v81/buolamwini18a.html.
5. Carlini, N., Tramer, F., Wallace, E., et al. 2021. "Extracting Training Data from Large Language Models". Available on: https://arxiv.org/pdf/2012.07805 on 29.7.24.
6. Chapman University. n.d. "Bias in AI". Available at: https://www.chapman.edu/ai/bias-in-ai.aspx.
7. Dhamani, N., & Engler, M. 2024. *Introduction to Generative AI.* Manning Publications.
8. Feldstein, Steven. 2019. The Road to Digital Unfreedom: How Artificial Intelligence is Reshaping Repression. Washington, D.C.: Carnegie Endowment for International Peace.
9. Funk, Allie, Kian Vesteinsson, and Grant Baker. 2024. "The Struggle for Trust Online." Freedom House. https://freedomhouse.org/report/freedom-net/2024/struggle-trust-online.
10. Friedman, Batya and Nissenbaum,Helen. 1996. "Bias in computer systems". ACM Trans. Inf. Syst. 14, 3. https://doi.org/10.1145/230538.230561
11. Harding, Thomas, and Dean Whitehead. 2016. "Analyzing Data in Qualitative Research." In Nursing and Midwifery Research: Methods and Appraisal for Evidence-Based Practice, edited by Zevia Schneider and Dean Whitehead, 5th ed., ch. 8. Elsevier.
12. Hristov, K. 2017. "Artificial Intelligence and the Copyright Dilemma." The Journal of Franklin Pierce Center for Intellectual Property, 57(3). Available on: https://www.researchgate.net/publication/316761384_Artificial_Intelligence_and_the_Copyright_Dilemma.
13. Janusheva, Violeta. 2022. *Kvalitativnite istrazuvanja vo naukata za jazikot [Qualitative Research Methodology in the field of Linguistics].* Bitola: Faculty of Education.
14. Kahlke, Melissa R. 2014. "Generic Qualitative Approaches: Pitfalls and Benefits of Methodological Mixology." International Journal of Qualitative Methods 13, no. 1: 37–52.
15. Kaplan, J. 2024. *Generative Artificial Intelligence: What Everyone Needs to Know.* Oxford University Press.
16. Koperniak, Stefanie. 2022. "Fostering media literacy in the age of deepfakes". MIT Open Learning.
17. Kriebitz, Alexander, and Lütge, C.2020. "Artificial Intelligence and Human Rights: A Business Ethical Assessment." Business and Human Rights Journal 5, no. 1 https://doi.org/10.1017/bhj.2019.28.

18. Leite, Eduardo, Leite, Maria and Leite, Ana. 2023. "AI's Impact on Human Rights: The Need for Legal Evolution." Journal of Entrepreneurial Researchers 1, no. 2. https://doi.org/10.29073/jer.v1i2.16.
19. Mozur, Paul. 2018. "Inside China's Dystopian Dreams: A.I., Shame and Lots of Cameras." The New York Times, July 8, 2018.
20. Moukhliss, Ghizlane, Khalid Lahyani, and Ghizlane Diab. 2024. "The Impact of Artificial Intelligence on Research and Higher Education in Morocco." Journal of Education and Learning 18, no. 4. https://doi.org/10.11591/edulearn.v18i4.21511.
21. Musiol, M. 2024. *Generative AI - Navigating the Course to the Artificial General Intelligence Future.* New Jersey: Wiley.
22. NeJame, L. et al. 2023. "Generative AI in Higher Education: From Fear to Experimentation, Embracing AI's Potential". Generative AI in Higher Education: From Fear to Experimentation, Embracing AI's Potential
23. Nicolau, I. I. 2021. "Human rights and artificial intelligence". Human Rights and the Internet, 10.
24. Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York: NYU Press.
25. OECD. 2024. "The Potential Impact Of Artificial Intelligence On Equity And Inclusion In Education". OECD Artificial Intelligence Papers, 23.
26. O'Neil, Cathy. 2016. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown Publishing Group.
27. Pasquale, Frank. 2015. The Black Box Society: The Secret Algorithms That Control Money and Information. Cambridge: Harvard University Press.
28. Raso, F. A., Hilligoss, H., & Kim, L. 2018. "Artificial intelligence & human rights: Opportunities & risks". SSRN. https://doi.org/10.2139/ssrn.3259344
29. Risse, M. 2018. "Human rights and artificial intelligence: An urgently needed agenda". HKS Faculty Research Working Paper Series RWP18-015.
30. RS Government. 2023. "Ethical Guidelines for Development, Implementation, and Use of Robust and Accountable Artificial Intelligence."https://www.ai.gov.rs/extfile/en/471/Ethical%20guidelines%20for%20development%20implementation%20and%20use%20of%20robust%20and%20accountable%20AI.pdf.
31. Shalevska, Elena & Kostadinovska-Stojchevska, Bisera. 2024. "Ethics in Times of Advanced AI: Investigating Students' Attitudes Towards ChatGPT And Academic Integrity". International Journal "Teacher", 27. https://doi.org/10.20544/teacher.27.08.
32. Shalevska, Elena. 2024. "The Future of Political Discourse: AI and Media Literacy Education". Journal of Legal and Political Education 1 (1):50-61. https://doi.org/10.47305/JLPE2411050sh.
33. Temperman, Jeroen, and Alberto Quintavalla, eds. *Artificial Intelligence and Human Rights*. 2023. Online edition, Oxford Academic.
34. Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The Spread of True and False News Online." Science 359, no. 6380. https://www.science.org/doi/10.1126/science.aap9559.
35. Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.