

Legal Study On Ethical Issues In The Use Of Artificial Intelligence For Legal Decisions: Critical Literature Review

Imam Hanafi^{1*}, Kaharuddin Syah², Loso Judijanto³, Irma Rachmawati Maruf⁴, Ihat Subihat⁵
Imamhanafi@ub.ac.id^{1*}, movingkahar69@gmail.com², losojudijantobumn@gmail.com³,
Irma.rachmawati@unpas.ac.id⁴, ihat.bali@gmail.com⁵
Universitas Brawijaya^{1*}, Universitas Muhammadiyah Palu², IPOSS Jakarta³, Universita Pasundan⁴,
Universitas Wiralodra Indramayu⁵

Article Info

Received: 2023-02-20

Revised: 2024-02-20

Accepted: 2024-06-20

Keywords:

Algorithmic Bias, Data Privacy, Human Agency, Transparency Imperatives, Socio-Economic Impacts, Global Perspectives .

Abstract

This comprehensive research embarks on meticulously exploring the intricate ethical nuances at the convergence of artificial intelligence (AI) and legal decision-making. Through an exhaustive literature review, the study meticulously navigates the complexities woven into algorithmic bias, the multifaceted dimensions of data privacy concerns, the profound implications on human agency, imperatives surrounding transparency, the socio-economic impacts stemming from the integration of AI, and the global perspectives that cast a profound influence on this intricate landscape. The synthesis of these insights reveals a dynamic interplay between the rapid evolution of technological capabilities and the intricate ethical considerations that underpin responsible AI integration into legal frameworks. The study underscores the need for ongoing interdisciplinary discourse, urging scholars, practitioners, and policymakers to engage in a continuous dialogue to ensure that ethical frameworks evolve in tandem with the relentless progression of AI technology. The conclusion advocates for a flexible and adaptive ethical framework poised to navigate the evolving ethical horizon, thereby ensuring AI's judicious and equitable integration into legal decision-making.

I. Introduction

The ethical considerations surrounding AI in legal decision-making extend to issues of accountability and transparency, as the inherent complexity of AI algorithms, often treated as "black boxes," poses challenges in understanding decision-making processes (Floridi et al., 2021; Mittelstadt et al., 2016). Legal practitioners, judges, and the public may need a clear understanding of the decision-making process to hold AI systems accountable. This makes transparency essential for maintaining trust in the legal system (Diakopoulos, 2016; Selbst & Barocas, 2018). Balancing the protection of proprietary algorithms with transparency becomes a delicate ethical challenge that necessitates careful attention (Wachter et al., 2017).

Furthermore, integrating AI into legal processes raises ethical questions about its impact on employment within the legal profession. AI systems streamlining tasks traditionally performed by legal professionals prompt concerns about job displacement and potential power concentration in those controlling AI technologies (Barocas & Selbst, 2016). Ethical considerations extend to the socio-economic implications of AI implementation, leading to discussions on the need for retraining programs, ethical guidelines for AI-assisted legal work, and policies to mitigate adverse effects on employment (Crawford & Calo, 2016).

The global nature of AI technology adds an international dimension to the ethical discourse, with legal systems varying across jurisdictions (Jobin et al., 2019). The deployment of AI in legal decision-making may have disparate impacts on different societies, raising questions about cultural sensitivity, the universality of ethical principles, and the potential for AI to exacerbate existing disparities in legal systems globally (Veale et al., 2018). The development of ethical frameworks for AI in the legal domain must consider these cross-cultural nuances to ensure a harmonized and equitable approach.

Ethical considerations also extend to the accessibility of AI-powered legal tools, where a risk exists that only well-funded entities or wealthy individuals may have access to advanced AI legal services (O'Neil, 2016). This raises questions about equal access to justice and the potential for AI to exacerbate existing inequalities within the legal system (Citron & Pasquale, 2014). Legal scholars and policymakers must address these accessibility concerns to prevent the entrenchment of a two-tiered legal system where only specific individuals or organizations benefit from AI advancements.

The ongoing evolution of AI technology adds a temporal dimension to the ethical discourse, where rapid advancements may outpace the ability of legal and ethical frameworks to adapt (Angwin et al., 2016). Ethical considerations must extend to continuous monitoring, assessment, and adaptation of regulations to keep pace with technological developments (Holzinger et al., 2017). The dynamic nature of AI requires a forward-looking approach to anticipate and address emerging ethical challenges proactively.

In conclusion, the multifaceted ethical considerations surrounding AI in legal decision-making, encompassing issues of bias, data privacy, human agency, accountability, transparency, employment, international implications, accessibility, and the temporal evolution of technology, highlight the need for a holistic understanding (Mittelstadt et al., 2019). This understanding is crucial for legal scholars, practitioners, and policymakers navigating the complex terrain of integrating AI into the legal domain. The ongoing interdisciplinary collaboration and dialogue on these issues should ensure that ethical guidelines and regulatory frameworks evolve with the rapid advancements in AI technology.

This literature review critically engages with the ethical dimensions of AI in legal decision-making, examining the current discourse on algorithmic bias, data privacy, the erosion of human agency, and regulatory considerations (Zarsky, 2016). The ethical implications discussed underscore the need for ongoing scholarly inquiry, collaborative efforts among stakeholders, and the development of ethical guidelines to ensure AI's responsible and equitable integration in legal decision-making processes (Jobin et al., 2019; Mittelstadt et al., 2019).

2. Research Method

The amalgamation of artificial intelligence (AI) and legal decision-making has become a focal point in academic discourse, reflecting the escalating integration of technological advancements into the legal landscape. This literature review systematically examines the ethical considerations of using AI in legal decisions, shedding light on pivotal themes and insights from various scholarly works.

One primary focus in the literature revolves around the pervasive issue of algorithmic bias in legal decision-making. Substantial research probes into the nuanced ways AI systems, often trained on historical datasets, may inadvertently perpetuate existing biases (Floridi, 2019). This raises concerns about replicating and exacerbating systemic inequities, challenging principles of equal protection under the law. The review extensively explores strategies for bias detection and mitigation, emphasizing the need for a comprehensive approach to ensure fairness and equity (Mittelstadt et al., 2019).

Another critical dimension in the literature concerns data privacy and confidentiality in AI-supported legal decision-making. Scholars scrutinize the substantial volumes of sensitive information processed by AI systems, raising concerns about threats to individual privacy rights (Diakopoulos, 2016). The review delves into the ethical considerations faced by legal professionals and policymakers, highlighting the delicate balance needed to harness AI's analytical power while safeguarding personal information. Discussions encompass evolving perspectives on informed consent, data anonymization, and the role of encryption in fortifying privacy protections (Selbst & Barocas, 2018).

The literature also delves into ethical quandaries surrounding human agency and the

potential dehumanization of legal processes due to AI intervention. Questions arise regarding accountability, empathy, and the broader societal implications of relying on algorithmic judgments (Angwin et al., 2016). The review critically engages in ongoing debates regarding the delicate equilibrium between AI efficiency gains and preserving human-centric values in the legal system.

Transparency and accountability challenges form another layer of ethical considerations explored in the literature. The intricate nature of AI algorithms introduces complexities in understanding decision-making processes (Barocas & Selbst, 2016). Scholars emphasize the imperative of transparency to uphold justice principles, examining discussions on explainability and interpretability. The review underscores the ethical necessity of ensuring that AI decisions are comprehensible to legal professionals, judges, and the public, fostering trust in the legal system.

Furthermore, the literature considers socio-economic implications and employment concerns from AI integration into legal processes. Scholars explore the potential implications, including job displacement within the legal profession, prompting ethical discussions about the concentration of power and the responsibility for job retraining programs (O’Neil, 2016). From an international and cultural perspective, the literature review acknowledges the diverse nature of global legal systems (Jobin et al., 2019). Scholars scrutinize the potential disparate impacts of AI on different societies, considering questions of cultural sensitivity and the universality of ethical principles. The review emphasizes the need for globally harmonized approaches to ensure equity and fairness in AI integration into legal decision-making processes.

In conclusion, this extensive literature review explores the critical ethical dimensions surrounding using AI for legal decisions. The multifaceted themes discussed in the literature, supported by citations such as Floridi (2019), Mittelstadt et al. (2016), Diakopoulos (2016), Selbst and Barocas (2018), Angwin et al. (2016), and Barocas and Selbst (2016), serve as a robust foundation for ongoing scholarly inquiry, offering valuable insights into the evolving relationship between AI and the legal domain.

Table 1. Key Ethical Considerations in AI-Driven Legal Decision-Making

Ethical Consideration	Focus in Literature	Notable Authors
Algorithmic Bias	Impact on fairness and equity	Floridi (2019), Mittelstadt et al. (2016)
Data Privacy	Balancing utility and protection	Diakopoulos (2016), Selbst & Barocas (2018)
Human Agency and Dehumanization	Accountability, empathy, and societal implications	Angwin et al. (2016)
Transparency and	Challenges in understanding	Barocas & Selbst (2016)

Ethical Consideration	Focus in Literature	Notable Authors
Accountability	decision-making processes	
Socio-economic Impacts	Job displacement and ethical responsibilities	O'Neil (2016)
International and Cultural Perspectives	Diverse impacts and cultural considerations	Jobin et al. (2019)

Created, 2023

3. Results and Discussion

Algorithmic Bias as a Persistent Ethical Challenge:

The literature review underscores algorithmic bias's persistent and pervasive challenge within AI-driven legal decision-making (Florida, 2019; Mittelstadt et al., 2016). The intricate dynamics of this issue become increasingly apparent as AI systems, in their quest for predictive accuracy, inadvertently reflect and perpetuate historical biases present in training datasets (Florida, 2019). As these systems analyze vast and diverse datasets, the risk of incorporating pre-existing biases becomes more pronounced, potentially leading to unintended disparate impacts on various demographic groups (Mittelstadt et al., 2016). This necessitates a thorough examination of how AI algorithms are trained and a heightened emphasis on ongoing monitoring and robust mitigation strategies to ensure the technical accuracy and the ethical dimensions of fairness and equity in legal decision-making (Florida, 2019).

Data Privacy: A Balancing Act Between Utility and Protection

The ethical discourse surrounding AI applications in legal decision-making extends into the intricate realm of data privacy, presenting a delicate and multifaceted challenge for legal professionals and policymakers alike (Diakopoulos, 2016; Selbst & Barocas, 2018). The literature review unveils the inherent tension between harnessing the analytical power of AI for legal advancements and the imperative to safeguard individuals' privacy rights (Diakopoulos, 2016). Striking a delicate balance becomes even more intricate as perspectives on informed consent evolve and discussions around data anonymization practices and the role of encryption as robust privacy protection mechanisms gain prominence (Selbst & Barocas, 2018). These nuanced considerations underscore the evolving nature of ethical challenges, demanding careful calibration of privacy safeguards to align with the advancing landscape of AI applications in the legal domain (Diakopoulos, 2016).

Human Agency and the Peril of Dehumanization:

One of the striking findings from the literature review is the profound ethical concern related to the potential erosion of human agency in legal processes, as AI systems assume more significant decision-making responsibilities (Angwin et al., 2016; Barocas & Selbst, 2016).

This revelation prompts a deeper exploration into the accountability, empathy, and broader societal implications of a legal system increasingly reliant on algorithmic judgments (Angwin et al., 2016). The delicate equilibrium that must be preserved between the efficiency gains offered by AI and the preservation of human-centric values becomes a central ethical dilemma (Barocas & Selbst, 2016). As the literature sheds light on the impact on legal professionals, litigants, and the overarching public perception of justice, it unravels a complex web of ethical considerations that require nuanced and adaptive approaches to maintain the integrity of the legal system (Barocas & Selbst, 2016).

Transparency and Accountability Challenges

The complexity inherent in AI algorithms introduces formidable challenges related to transparency and accountability, emerging as a pivotal finding with profound implications (Barocas & Selbst, 2016; Wachter et al., 2017). The opacity of specific AI systems, often called "black boxes," amplifies the difficulty for legal professionals, judges, and the public in understanding the decision-making processes (Barocas & Selbst, 2016). The literature highlights the ethical imperative of transparency as a cornerstone in upholding principles of justice (Wachter et al., 2017). The need for explainability and interpretability in AI systems becomes even more pronounced, emphasizing the importance of ensuring that AI decisions are not only accurate but comprehensible, fostering a sense of trust and accountability in the legal system (Barocas & Selbst, 2016; Wachter et al., 2017).

Socio-Economic Impacts and Employment Concerns

Delving into the socio-economic implications of integrating AI into legal processes, the literature reveals a multifaceted landscape that demands an ethical lens (O'Neil, 2016; Barocas & Hardt, 2019). Beyond the efficiency gains and streamlined legal tasks, discussions gravitate toward the potential displacement of jobs within the legal profession (O'Neil, 2016). This ethical consideration extends beyond the immediate technological implications, prompting a nuanced exploration of the concentration of power, the ethical responsibility for implementing job retraining programs, and strategies to mitigate adverse effects on employment (Barocas & Hardt, 2019). As legal scholars and policymakers grapple with these socioeconomic concerns, the ethical imperative to ensure responsible and equitable integration of AI into the legal domain becomes increasingly evident (O'Neil, 2016).

International and Cultural Perspectives:

The literature review emphasizes the significance of international and cultural perspectives in unraveling the ethical complexities surrounding AI in legal decision-making (Jobin et al., 2019; Calders et al., 2013). Recognizing the diversity of legal systems globally, the findings underscore the potential disparate impacts of AI deployment on different societies

(Jobin et al., 2019). This prompts in-depth explorations into cultural sensitivity, discussions on the universality of ethical principles, and the compelling need for globally harmonized approaches to ensure equity and fairness (Calders et al., 2013). The ethical considerations extend beyond technical implementations, requiring a broadened perspective incorporating cross-cultural nuances, underscoring the importance of interdisciplinary collaboration in developing global ethical guidelines (Jobin et al., 2019).

Temporal Evolution and Adaptive Regulatory Frameworks

An overarching finding of paramount importance is the temporal dimension inherently embedded in the ethical discourse surrounding AI and legal decision-making (Wachter et al., 2017; Chouldechova & Roth, 2018). As the literature reveals, the rapid evolution of AI technology outpaces existing legal and ethical frameworks, necessitating a dynamic and adaptive approach to regulation (Wachter et al., 2017). Policymakers face the formidable challenge of fostering innovation while safeguarding against potential harms (Chouldechova & Roth, 2018). This finding elevates the temporal evolution of technology to the forefront of ethical considerations, stressing the imperative for continuous monitoring, assessment, and the proactive adaptation of regulatory frameworks to keep pace with technological advancements (Wachter et al., 2017; Chouldechova & Roth, 2018). The forward-looking nature of regulatory frameworks becomes an ethical imperative in anticipating and addressing emerging challenges in a rapidly evolving landscape.

Table 2: Key Ethical Findings in AI-Driven Legal Decision-Making

Findings	Description	Implication	Evidence
Algorithmic Bias	A persistent challenge in AI-driven legal decisions is risking perpetuating historical biases (Floridi, 2019; Mittelstadt et al., 2016).	Necessitates ongoing monitoring and mitigation for fairness and equity (Floridi, 2019).	(Floridi, 2019; Mittelstadt et al., 2016)
Data Privacy	The ethical balance between AI utility and privacy protection demands careful safeguards calibration (Diakopoulos, 2016; Selbst & Barocas, 2018).	Requires nuanced policy sculpting considering evolving perspectives (Diakopoulos, 2016; Selbst & Barocas, 2018).	(Diakopoulos, 2016; Selbst & Barocas, 2018)
Human Agency	Ethical concern about the potential erosion of human agency in legal processes with increasing AI decision-making (Angwin et al., 2016; Barocas & Selbst, 2016).	Prompts nuanced and adaptive approaches to balance efficiency gains and human-centric values (Barocas & Selbst, 2016).	(Angwin et al., 2016; Barocas & Selbst, 2016)

Findings	Description	Implication	Evidence
	Barocas & Selbst, 2016).		
Transparency Challenges	Formidable challenges in AI transparency are highlighted by opacity and difficulties in understanding (Barocas & Selbst, 2016; Wachter et al., 2017).	Emphasizes the ethical need for explainability and interpretability (Barocas & Selbst, 2016; Wachter et al., 2017).	(Barocas & Selbst, 2016; Wachter et al., 2017)
Socio-Economic Impacts	Multifaceted landscape with ethical considerations on job displacement and concentration of power (O'Neil, 2016; Barocas & Hardt, 2019).	Urges responsible and equitable AI integration and ethical responsibilities for job retraining (O'Neil, 2016).	(O'Neil, 2016; Barocas & Hardt, 2019)
International Perspectives	Significance of global and cultural perspectives in addressing ethical complexities (Jobin et al., 2019; Custers et al., 2013).	Demands globally harmonized approaches and interdisciplinary collaboration (Jobin et al., 2019; Custers et al., 2013).	(Jobin et al., 2019; Custers et al., 2013)
Temporal Evolution	Temporal dimension in AI evolution outpacing regulations, emphasizing the need for adaptive frameworks (Wachter et al., 2017; Chouldechova & Roth, 2018).	Stresses the ethical imperative of continuous monitoring, assessment, and proactive adaptation (Wachter et al., 2017).	(Wachter et al., 2017; Chouldechova & Roth, 2018)

Created, 2023

Discussion

Delving into the labyrinth of ethical considerations surrounding integrating artificial intelligence (AI) into legal decision-making reveals complexities that necessitate nuanced exploration and profound contemplation. The amalgamation of technological prowess and legal processes presents many challenges, each demanding a deep understanding and thoughtful discourse (Florida, 2019; Mittelstadt et al., 2016).

The persistent concern of algorithmic bias is at the forefront of this ethical discourse. The literature extensively illuminates the intricate dance between AI algorithms and historical biases embedded in training datasets (Florida, 2019). The profound implications of inadvertently perpetuating systemic inequities require scrutiny that transcends algorithm design's technical intricacies. The very essence of fairness and equity in legal decisions hinges on unraveling and mitigating these biases. This discussion emphasizes the depth of engagement required to navigate the delicate balance between predictive accuracy and the

preservation of justice principles (Mittelstadt et al., 2016).

Moving further into the ethical terrain, the intricacies of data privacy unfurl with both subtlety and urgency. The balance sought between the utility derived from AI applications and the imperative to protect individual privacy rights represents a philosophical tightrope walk (Diakopoulos, 2016; Selbst & Barocas, 2018). Evolving perspectives on informed consent, nuanced data anonymization practices, and the evolving role of encryption all contribute to the ethical dialogue. This discussion underscores the depth of reflection required in sculpting policies that traverse the delicate intersections of technological advancement and individual rights (Diakopoulos, 2016; Selbst & Barocas, 2018).

Human agency, a fundamental tenet of legal systems, emerges as a poignant theme in the ethical narrative. The nuanced exploration into the potential erosion of human agency in legal processes reveals a profound ethical dilemma (Angwin et al., 2016; Barocas & Selbst, 2016). As AI systems assume decision-making responsibilities, accountability, empathy, and societal implications are scrutinized. This discussion invites a profound introspection into the soul of legal proceedings, posing fundamental questions about the coexistence of technological efficiency and human-centric values. It beckons legal scholars and practitioners to engage in a discourse that transcends the pragmatic to delve into the philosophical underpinnings of justice (Angwin et al., 2016; Barocas & Selbst, 2016).

Transparency and accountability, pillars of a just legal system, become amplified in their significance within AI. The opacity of specific AI systems, colloquially known as "black boxes," challenges the fabric of understanding in legal decision-making (Barocas & Selbst, 2016; Wachter et al., 2017). This discussion invites a deep dive into the complexities of transparency, exploring the ethical imperative of explainability and interpretability. It highlights the depth of understanding required to ensure that AI decisions are accurate and comprehensible, fostering trust and accountability in the legal system (Barocas & Selbst, 2016).

Socio-economic impacts, often relegated to a secondary consideration, demand a profound examination in the ethical discourse (O'Neil, 2016). Beyond the allure of efficiency gains, the potential displacement of jobs within the legal profession unfolds as a complex ethical consideration. This discussion transcends the technical aspects, reflecting on the concentration of power, ethical responsibilities for job retraining programs, and strategies to mitigate adverse socio-economic effects. It draws attention to the depth of ethical engagement required to ensure a responsible and equitable integration of AI into the legal domain (O'Neil, 2016; Selbst et al., 2019).

The international and cultural dimensions complicate the ethical dialogue (Jobin et al., 2019). Acknowledging the diversity of legal systems globally, this discussion underscores the need for cultural sensitivity and a nuanced exploration of ethical principles that transcend borders. The call for globally harmonized approaches demands a deep understanding beyond technical implementations. It requires an interdisciplinary collaboration that appreciates cross-cultural nuances and emphasizes the depth of ethical considerations that extend to a global scale (Jobin et al., 2019).

The discussion delves into the essence of technological governance in contemplating the temporal evolution of AI and the adaptive nature of regulatory frameworks (Chouldechova & Roth, 2018). The rapid evolution of AI technology outpacing existing regulations demands a

depth of foresight from policymakers. This discussion stresses the profound ethical imperative of continuous monitoring, assessment, and proactive adaptation of regulatory frameworks. It requires a depth of understanding that recognizes the forward-looking nature of ethical considerations in the face of emerging technological challenges (Wachter et al., 2017; Chouldechova & Roth, 2018).

In sum, the depth of the ethical discussion surrounding AI in legal decision-making transcends the technical intricacies. It invites scholars, practitioners, and policymakers to engage in a profound exploration that navigates the intersection of technology and justice. The depth of understanding required spans philosophical reflections, cultural nuances, and anticipatory governance. This discussion is an ongoing dialogue that demands a continuous deepening of our comprehension of the ethical complexities embedded in the evolving landscape of AI in the legal domain (Thurman et al., 2019).

4. Conclusion

In conclusion, the synthesis of artificial intelligence (AI) and legal decision-making heralds a transformative era marked by unprecedented challenges and ethical intricacies. The extensive literature review has provided a profound exploration of these complexities, inviting a deep understanding of the multifaceted dimensions that define the ethical landscape at this critical intersection.

The relentless pursuit of algorithmic perfection, while aiming for predictive accuracy, demands vigilant introspection into the persistent issue of algorithmic bias. The ethical imperative lies in unraveling the intricacies of biased datasets and crafting mitigation strategies that ensure fairness and equity in legal decisions. This acknowledgment underscores the conclusion that the path forward requires not just technical expertise but a profound commitment to the principles of justice.

The delicate dance between data privacy and the utility derived from AI applications casts a spotlight on the need for a nuanced ethical framework. The evolving perspectives on informed consent, data anonymization, and encryption necessitate an ongoing dialogue that transcends the binary realm of protection versus innovation. The conclusion is that ethical considerations must evolve with technological advancements, striking a harmonious balance between progress and safeguarding individual rights.

Human agency, a cornerstone of legal systems, emerges as a focal point demanding ethical introspection. The potential erosion of human-centric values after AI decision-making prompts a conclusion beyond pragmatic efficiency concerns. It calls for a profound reevaluation of the philosophical underpinnings of justice, emphasizing the symbiotic relationship between technological advancement and the preservation of accountability, empathy, and societal values.

Transparency and accountability, crucial for the legitimacy of legal systems, demand a conclusion that recognizes the ethical imperative of explainability in AI systems. The opacity inherent in specific algorithms challenges the fabric of understanding in legal decision-making. The call is for a commitment to transparency that transcends technical accuracy, fostering a comprehensible and just legal system in the age of AI.

Socio-economic impacts, often relegated to secondary considerations, draw a conclusion that compels a deeper ethical engagement. Beyond the allure of efficiency gains, ethical responsibilities for job displacement and strategies to mitigate adverse effects become paramount. The conclusion underscores the imperative for a responsible and equitable integration of AI, recognizing the profound societal implications.

The international and cultural dimensions add a conclusion that extends the ethical discourse beyond national boundaries. Acknowledging the diversity of legal systems globally, the call is for a global perspective that appreciates cultural nuances and seeks universally harmonized approaches. It emphasizes the depth of ethical considerations that demand interdisciplinary collaboration on a global scale.

In contemplating the temporal evolution of AI and adaptive regulatory frameworks, the conclusion is drawn that technology governance requires a continuous and forward-looking approach. Policymakers must navigate the evolving landscape with agility, anticipating and proactively addressing emerging challenges. The conclusion is a recognition that ethical considerations must be as dynamic as the technology they seek to govern.

The conclusion drawn from this comprehensive exploration of the ethical complexities surrounding AI in legal decision-making calls for a profound understanding that goes beyond the technical. It invites ongoing discourse, interdisciplinary collaboration, and a commitment to ethical frameworks that evolve with the relentless march of technological progress. As we stand at the precipice of a new era in the intersection of AI and the law, the conclusion is a recognition that the ethical horizon is vast, demanding continuous exploration, contemplation, and ethical stewardship.

References

- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. *ProPublica*.
- Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review*, 104(3), 671–732.
- Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big data & society*, 3(1), 2053951715622512.
- Calders, T., Schermer, B., Zarsky, T., & Custers, B. (Eds.). (2013). *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases*. Springer.
- Chouldechova, A., & Roth, A. (2018). The frontiers of fairness in machine learning. *arXiv preprint arXiv:1810.08810*.
- Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. *Wash. L. Rev.*, 89, 1.
- Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature*, 538(7625), 311–313.
- Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56–62.
- Florida, L. (2019). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines*, 29(4), 689–707.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2021). An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Ethics, governance, and policies in artificial intelligence*, 19–39.

- Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What Do We Need to Build Explainable AI Systems for the Medical Domain? arXiv preprint arXiv:1712.09923.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 2053951716679679.
- Mittelstadt, B. D., Russell, C., & Wachter, S. (2019). I am explaining Explanations in AI. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 279-288).
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books.
- O'neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.
- Selbst, A. D., & Barocas, S. (2018). The Intuitive Appeal of Explainable Machines. *Fordham Law Review*, 87(3), 1085–1134.
- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, January). Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 59-68).
- Thurman, N., Lewis, S. C., & Kunert, J. (2019). Algorithms, automation, and news. *Digital journalism*, 7(8), 980-992.
- Veale, M., Binns, R., & Edwards, L. (2018). Algorithms that remember: model inversion attacks and data protection law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180083.
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why does a right to explain automated decision-making does not exist in the general data protection regulation? *International Data Privacy Law*, 7(2), 76-99.
- Zarsky, T. (2016). The Trouble with Algorithmic Decisions: An Analytic Roadmap to Examine Efficiency and Fairness in Automated and Opaque Decision Making. *Science, Technology, & Human Values*, 41(1), 118–132.