International Articles Journal
A Multidisciplinary Journal
Issue 1, March 2025

مجلة المقالات الدولية
مجلة متعددة التخصصات
العدد 1، مارس 2025

Machine Learning and Human Rights:
Balancing Benefits and Risks

Abdelhaq BELFFQIH
Professor of Public Law and Political Science
Sidi Mohamed Ben Abdellah University, Fes

Abdellah Alami
PhD Researcher
Sidi Mohamed Ben Abdellah University, Fes

عبد الحق بلفقيه
أستاذ القانون العام والعلوم السياسية
جامعة سيدي محمد بن عبد الله ، فاس

عبد الله علمي
باحث بسلك الدكتوراة
جامعة سيدي محمد بن عبد الله ، فاس

Abstract :

This article explores the nuanced interplay between Machine Learning (ML) and human rights, emphasizing the need for a balanced approach amid evolving technology. ML's potential to empower global human rights initiatives is examined through applications like sentiment analysis and predictive policing, while acknowledging associated risks such as privacy concerns and algorithmic biases. The narrative underscores the importance of ethical guidelines, transparency, and comprehensibility in ML algorithms, aligning with human rights principles. It advocates for a human-centric development approach, focusing on augmenting human capabilities rather than replacement, and calls for inclusive development to ensure universal ML benefits. The article concludes by highlighting the imperative for collaborative efforts among technologists, policymakers, and human rights advocates to navigate this intricate intersection effectively.

المستخلص:

يكشف هذا المقال التفاعل الدقيق بين التعلم الآلي وحقوق الإنسان، مؤكدًا على ضرورة اتباع نهج متوازن في ظل التكنولوجيا المتطورة. ويتناول إمكانية تمكين التعلم الآلي لمبادرات حقوق الإنسان العالمية من خلال تطبيقات مثل تحليل المشاعر والرقابة التنبؤية، مع مراعاة المخاطر المرتبطة بها، مثل مخاوف الخصوصية والتحيزات الخوارزمية. ويؤكد المقال على أهمية المبادئ الأخلاقية والشفافية وسهولة الفهم في خوارزميات التعلم الآلي، بما يتماشى مع مبادئ حقوق الإنسان. ويدعو إلى نهج تنموي يركز على الإنسان، ويركز على تعزيز القدرات البشرية بدلًا من استبدالها، ويدعو إلى تنمية شاملة لضمان فوائد التعلم الآلي الشاملة. ويختتم المقال بتسليط الضوء على ضرورة تضافر الجهود بين خبراء التكنولوجيا وصانعي السياسات والمدافعين عن حقوق الإنسان لتجاوز هذا التقاطع المعقد بفعالية.

Keywords:
Machine Learning; Human Rights; Ethical Guidelines; Inclusive Development.

الكلمات المفتاحية:
التعلم الآلي؛ حقوق الإنسان؛ المبادئ الأخلاقية، التنمية الشاملة.

Introduction:

The dynamic evolution of technology is swiftly reshaping the global landscape, presenting novel avenues for human rights defenders to leverage innovative tools to their benefit, with machine learning emerging as a prominent example[1].

Machine Learning (ML) stands as a subfield within the realm of Artificial Intelligence (AI), wherein machines are endowed with the capability to autonomously acquire knowledge from data and previous encounters. This process involves the discernment of patterns, facilitating the formulation of predictions with minimal human intervention[2]. In other words, machine learning has emerged as a powerful force shaping various aspects of our lives, from personalized recommendations to autonomous vehicles. However, as this technology evolves, questions surrounding its impact on human rights have become increasingly complex. Striking a balance between the potential for harmony and the risk of dissonance is crucial in ensuring that the benefits of machine learning are harnessed responsibly.

Amidst the promise of innovation, the integration of machine learning into societal frameworks also raises ethical and human rights concerns. Instances of algorithmic bias, privacy violations, and unequal access to technology highlight the potential for harm when machine learning systems are improperly designed or implemented. These challenges necessitate a deeper examination of accountability, fairness, and transparency in the development of ML technologies. As stakeholders strive to harness the benefits of machine learning while minimizing its adverse effects, the dialogue surrounding ethical guidelines and regulatory frameworks becomes increasingly vital.

Defining Machine Learning

Machine learning methodologies empower computers to function autonomously without requiring explicit programming instructions. Instead of relying on a predefined set of rules, these systems utilize algorithms that process and analyse vast amounts of data, enabling them to recognize intricate patterns and correlations. When supplied with novel data, machine learning applications can adapt and evolve, continuously refining their decision-making processes[3]. This ability to learn from experience rather than rigid programming makes machine learning systems highly versatile and capable of handling complex, dynamic tasks. Over time, they mature through iterative learning, improving their accuracy and efficiency as they encounter more data[4]. By dynamically adjusting to new information, machine learning technologies can provide more precise predictions and insights, demonstrating a capacity for self-improvement. This continuous adaptation is what distinguishes machine

[1] Ferré, F. (1998). *Philosophy of technology*. Prentice Hall.

[2] Berk, R. (2017). An impact assessment of machine learning risk forecasts on parole board decisions and recidivism. Journal of Experimental Criminology, 13(2), 193. https://doi.org/10.1007/s11292-017-9292-4

[3] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.

[4] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press.

learning from traditional computational approaches, expanding its applications across diverse fields like healthcare, finance, and cybersecurity.[5]

Machine learning extracts meaningful insights from extensive datasets by employing algorithms to discern patterns, engaging in an iterative learning process. These algorithms employ computational techniques to directly learn from data, eschewing reliance on predetermined equations as modeling frameworks.[6]

The adaptive enhancement of machine learning algorithm

The performance of machine learning algorithms improves adaptively as the quantity of available data increases during the learning process. This enhancement occurs because more data allows the algorithms to recognize patterns more accurately, refining their predictions and decision-making. Deep learning, a specialized subfield of machine learning, takes this concept further by enabling computers to mimic human-like attributes, such as learning from examples. This approach is designed to replicate how humans recognize patterns and make decisions based on prior experiences. As a result, deep learning models often outperform traditional machine learning algorithms in tasks that involve large datasets, demonstrating superior accuracy and efficiency. The ability of deep learning to process vast amounts of data and learn complex representations allows it to excel in applications such as image recognition, natural language processing, and autonomous systems.[7]

The conceptual roots of machine learning can be traced back to World War II, when early computational models such as the Enigma Machine were used to perform complex calculations and decode encrypted messages. These early machines laid the groundwork for the later development of machine learning by demonstrating the power of automated computation. However, the automated application of intricate mathematical algorithms to increasingly vast and diverse datasets is a relatively recent development, driven by advances in computing power and data availability.[8]

In recent years, the emergence of big data, the Internet of Things (IoT), and pervasive computing technologies has propelled machine learning into a central role in addressing complex problems across various sectors. The explosive growth of data generated by digital devices, sensors, and online activities has created a demand for sophisticated algorithms capable of analyzing and deriving insights from this massive volume of information. Today, machine learning has evolved into a vital tool for tackling challenges in numerous fields, driving innovations that were once thought impossible.[9]

For instance, in computational finance, machine learning models are increasingly used for tasks such as credit scoring and algorithmic trading, helping financial institutions make more accurate decisions and predict market trends with remarkable precision. In computer vision, machine learning enables machines to process and

[5] Flach, P. (2012). *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*. Cambridge: Cambridge University Press.

[6] - Eileen, D. and Megan M.M. Artificial Intelligence and Human Rights. Journal of Democracy, Volume 30, Number 2, 2019, pp. 115-126.

[7] Alpaydin, E. (2020). *Introduction to Machine Learning* (4th ed.). Cambridge, MA: MIT Press.

[8] Mitchell, T. M. (1997). *Machine Learning*. New York, NY: McGraw-Hill.

[9] Jordan, M. I., & Mitchell, T. M. (2015). "Machine Learning: Trends, Perspectives, and Prospects." *Science*, 349(6245), 255-260.

understand visual data, leading to advancements in facial recognition, motion tracking, and object detection. These applications are transforming industries from security and surveillance to autonomous vehicles. In computational biology, machine learning techniques have made significant contributions to fields like DNA sequencing, brain tumor detection, and drug discovery, providing new insights into medical research and enabling more personalized treatments. The automotive, aerospace, and manufacturing industries are also benefiting from machine learning through applications such as predictive maintenance, where algorithms predict equipment failures before they occur, minimizing downtime and enhancing efficiency. Furthermore, in natural language processing, machine learning has revolutionized voice recognition systems, enabling technologies like virtual assistants and automated translation services to understand and process human language more effectively. As machine learning continues to evolve, its impact on these and other domains will only grow, unlocking new possibilities for innovation and problem-solving.[10]

Harmony: Empowering Human Rights Efforts

Machine learning emerges as a distinctive catalyst, presenting an unparalleled opportunity to bolster human rights initiatives on a global scale. Its potential spans a spectrum of applications, ranging from the proactive prediction and prevention of human rights abuses to the nuanced analysis of extensive datasets, thereby informing judicious policy decisions. This technological paradigm holds the promise of profoundly advancing the safeguarding of fundamental rights through its capacity to discern patterns and generate insights[11].

Innovations within the realm of machine learning, such as sentiment analysis and predictive policing, are directed towards the creation of safer environments. These applications aim not only to identify potential threats but also to preclude their occurrence, thereby contributing to the overarching goal of human rights protection. Simultaneously, automated translation services facilitated by machine learning foster communication and understanding among diverse communities, transcending language barriers and promoting inclusivity in the discourse surrounding human rights.

Industry leaders like Google have been instrumental in spreading knowledge and enhancing understanding of machine learning concepts and methodologies. Through a wide range of interactive learning tools, courses, and educational games, Google strives to make machine learning more accessible and less intimidating to a broad audience. These resources cater to individuals from various backgrounds, enabling them to explore the fundamentals of this transformative technology at their own pace. By offering a user-friendly approach to machine learning, Google aims to demystify the complexities of algorithms and data analysis, empowering users to engage with the technology confidently.

In addition to Google's efforts, online platforms such as R2D3 contribute significantly to making machine learning more inclusive. R2D3 offers introductory resources in over 10 languages, ensuring that non-English-

[10] Russell, S., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach* (3rd ed.). Pearson.

[11] Goldin, C., & Katz, L. F. (2008). *The race between education and technology*. Belknap Press of Harvard University Press.

speaking practitioners have access to essential learning materials. This dedication to inclusivity reflects a growing recognition of the global need for machine learning education, enabling individuals from diverse linguistic and cultural backgrounds to participate in the rapidly evolving field of data science. These initiatives collectively promote broader participation in machine learning and contribute to the ongoing democratization of technology.[12]

As the landscape of machine learning continues to evolve, practitioners are poised to navigate its domain with heightened efficacy, buoyed by a burgeoning array of cost-effective resources. These encompass a spectrum of offerings, including services, products, libraries, and hardware. Notable among these resources are Tensorflow, DeepLens, AWS Machine Learning, and Google's Cloud AI, collectively contributing to the democratization of machine learning accessibility[13]. This democratization, in turn, holds the potential to empower a broader cadre of practitioners, irrespective of geographical or financial constraints, in their pursuit of leveraging machine learning for the advancement of human rights initiatives.

Dissonance: Risks and Challenges

However, the deployment of ML in various contexts also introduces risks to human rights. Concerns arise in areas such as privacy infringement, algorithmic bias, and the potential misuse of AI for surveillance. Biased algorithms may perpetuate discrimination, leading to unintended consequences and exacerbating existing social inequalities. Striking the right balance is a delicate task, requiring careful consideration of ethical principles and robust regulatory frameworks[14].

To put it differently, machine learning is still in its developmental phases, yet it has already found applications in certain realms of human rights work. Machine learning programs demonstrate the potential to enhance the identification of human rights abuses, optimize existing systems, and proactively prevent perilous situations. Within the context of human rights endeavors, practitioners frequently grapple with the need to categorize reports, evidence, and other data. Machine learning tools play a pivotal role in reducing the time required for such tasks, exemplified by the implementation of tools capable of sentence classification, adaptable to the specific research questions pertinent to human rights defenders[15].

Advancements in machine learning have also given rise to tools utilizing video analysis, facilitating the detection of objects, sound, speech, text, and event types. This capability enables users to execute semantic queries within video collections, discerning the unfolding events[16]. In addition, these tools have the capability to document

[12] Schmid Huber, J. (2015). *Deep Learning in Neural Networks: An Overview*. Neural Networks, 61, 85-117.

[13] - Shurkin N. Joel, "Expert Systems: The Practical Face of Artificial Intelligence," in science, vol. 227, no. 4687, The MIT Press, Cambridge, MA, p. 78.

[14]Richardson, R., Schultz, J. M., & Crawford, K. (2018). Dirty data, bad predictions: How civil rights violations impact police data, predictive police systems, and justice. New York University Law Review, 94(1), 192–233. https://doi.org/10.2139/ssrn.3333423

[15] Mayer-Schönberger, V., & Cukier, K. (2014). *Big data: A revolution that will transform how we live, work, and think* (Reprint ed.). Eamon Dolan/Mariner Books.

[16] Michaels, A. C. 2019. Artificial Intelligence, Legal Change, and Separation of Powers. University of Cincinnati Law Review, pp. 70.

human rights violations, predict judicial proceedings, and function as open-source computer vision systems for large video datasets.

At present, video analysis is increasingly used in conflict zones to verify the authenticity of footage, providing reliable evidence for cases of war crimes. By analyzing video data, these tools help identify key events and individuals involved, contributing to accountability and justice. As conflicts continue around the world, these technologies play a critical role in ensuring that visual evidence can be trusted and used in legal contexts. The ability to process and authenticate videos in real-time is transforming how evidence is gathered and utilized in human rights investigations.[17]

In addition to ML's growth, concerns arise about its potential utilization in manners conflicting with human rights standards. A salient apprehension pertains to the potential misuse of facial recognition, which could pose risks to human rights practitioners and vulnerable populations. In response to this risk, initiatives such as Harvard Law School's EqualAIs have been developed. This project subtly alters images in ways imperceptible to the human eye, thereby preventing identification through other machine learning technologies and mitigating the potential threats associated with facial recognition misuse[18].

Navigating the Path Forward Ethical Guidelines and Transparency:

Implementing clear ethical guidelines and ensuring transparency in ML algorithms are essential steps. This involves making the decision-making processes of these algorithms understandable and accountable to avoid unintended biases. In December 2018, a group of experts drew up the "Draft Ethics Guidelines for Trustworthy A.I.[19] With this document, the European Commission warned of the risks associated with A.I., despite its considerable advantages, and recognizes the need for an anthropocentric approach to A.I.[20] This is the only approach capable of guaranteeing the dignity and autonomy of people, who must always be given the power to supervise machines.21 It has become obvious that even the Council of Europe purviews and warns against the risk of social discrimination provoked by algorithms.

A significant ethical issue emerges when considering the role of machine learning in assisting, or potentially replacing, human involvement in judicial decision-making. Critics argue that legal systems, which are constructed by human beings, play an essential role in maintaining social order and delivering justice. These critics contend that while machine learning can improve the efficiency of procedural tasks, it should not be entrusted with decisions that influence crucial aspects of individuals' lives and broader societal dynamics. The idea of relinquishing such important decisions to automated systems raises concerns about the loss of human judgment, empathy, and understanding.[22] Consequently, there is an ongoing debate about how to best incorporate machine

[17] Dean, J., & Ghemawat, S. (2008). "MapReduce: Simplified Data Processing on Large Clusters." *Communications of the ACM*, 51(1), 107-113.

[18] Jasanoff, S. (2016). *The ethics of invention: Technology and the human future*. W. W. Norton & Company.

[19] European Commission's High-Level Expert Group on Artificial Intelligence. (2021). Draft ethics guidelines for trustworthy AI. Retrieved January 11, 2025, from http://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai

[20] - Ibid, p. 13.

[21] - Ibid, p. 14

[22] Binns, R. (2018). *Fairness in Machine Learning: Lessons from Political Philosophy*. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1-14.

learning into the judicial process without undermining the human element that is vital to achieving fair and just outcomes.

To address these concerns, efforts are being made to ensure that machine learning systems are designed with ethical considerations in mind. A key aspect of these efforts is ensuring that these technologies incorporate values such as fairness, transparency, and diversity. One of the main objectives is to align the application of machine learning with established human rights frameworks, ensuring that these tools do not inadvertently perpetuate biases or injustices. For instance, algorithms must be carefully scrutinized to ensure that they do not reinforce discriminatory patterns in decision-making, particularly in areas such as sentencing, parole, or bail decisions. Ensuring that these systems operate in a manner consistent with human rights is a critical step in maintaining public trust and accountability in the judicial process.[23]

Furthermore, many advocates suggest that existing regulations, such as the European Union's General Data Protection Regulation (GDPR), could serve as a useful guideline for the ethical deployment of machine learning in the legal domain. The GDPR provides a comprehensive framework for data protection and privacy, which can be adapted to address the unique ethical challenges posed by machine learning technologies. By adopting such regulations, proponents argue that the integration of machine learning into the judicial system could be done in a way that respects individual rights while leveraging the technology's potential to enhance efficiency and accuracy.[24] These combined efforts seek to strike a delicate balance, ensuring that while machine learning can augment legal processes, the essential role of human judgment in matters of personal and societal importance remains intact.

Charting the Course ahead Inclusive Development

Prioritizing inclusive development ensures that the benefits of ML are accessible to all. Addressing biases in training data and involving diverse voices in the development process can help mitigate the risk of unintentional discrimination[25]. In this sense, the prioritization of inclusive development in the context of machine learning is paramount to ensuring the universal accessibility of its benefits. The recognition and rectification of biases within training datasets represent crucial measures in this pursuit. Incorporating diverse voices in the developmental processes not only fosters a more comprehensive understanding of varied perspectives but also serves as a proactive strategy to mitigate the risk of unintentional discrimination. The concerted commitment to inclusivity in the developmental trajectory of machine learning systems contributes not only to ethical imperatives but also enhances the robustness and fairness of the technology, aligning it more closely with principles of equity and justice.

Human-Centric Approach

[23] Wachter, S., Mittelstadt, B., & Floridi, L. (2017). "Transparent, Explainable, and Accountable AI for Justice Systems." *Nature Machine Intelligence*, 2(11), 739-742.

[24] Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin's Press.

[25]Tegmark, M. (2017). Life 3.0: Being human in the age of artificial intelligence. Knopf.

Placing humans at the center of machine learning (ML) development underscores the critical importance of enhancing human capabilities rather than attempting to replace them altogether. This approach serves as a safeguard against the potential risks of dehumanization that may arise when technology becomes overly dominant in decision-making processes. By focusing on human-centered development, machine learning technologies are designed to support and augment the skills, knowledge, and intuition that individuals already possess, rather than supplanting them.[26] This philosophy aims to ensure that technology is used to improve the well-being of people, fostering a positive relationship between humans and the tools they create.

Moreover, this human-centered approach deliberately emphasizes the cultivation of a symbiotic relationship between humans and technology, rather than one where technology is seen as a substitute for human judgment. In this framework, machine learning systems are designed to complement human decision-making, elevating the capabilities of individuals while adhering to ethical guidelines and societal values. By prioritizing the enhancement of human abilities, this strategic direction seeks to align the development of machine learning with the greater good, ensuring that technological progress does not outpace our ability to manage it responsibly and ethically. This focus on human augmentation reflects a commitment to ensuring that technological advancements serve humanity, rather than overshadowing it.[27]

In adopting this paradigm, a foundational rationale is established to mitigate the potential pitfalls associated with the dehumanization of technological landscapes[28]. As machine learning systems evolve, the intrinsic role of human agency remains pivotal, serving as a crucial bulwark against the erosion of fundamental human qualities and the potential alienation of individuals from the technological advancements permeating various facets of contemporary life. This approach thus envisions a harmonious coexistence where machine learning acts as a tool to enhance and amplify human abilities rather than supplanting them.

Moreover, this human-centric orientation in machine learning development is entrenched in a broader commitment to ensure that technological progress unequivocally contributes to the well-being of individuals.[29] The symbiosis between humans and machine learning technologies is envisioned not merely as a matter of utility but as an ethical imperative, as the integration of technological advancements is strategically orchestrated to align with societal values, equity, and justice. By placing humans at the nexus of machine learning development, the overarching objective is to harness the potential of technology to uplift, empower, and cater to the diverse needs and aspirations of individuals within the broader social milieu.

In essence, the deliberate choice to prioritize inclusive human involvement in the trajectory of machine learning development transcends a mere technological strategy; it constitutes a comprehensive ethos. This ethos envisions a future where technological advancements, particularly in machine learning, seamlessly integrate with

[26] Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York, NY: W.W. Norton & Company.

[27] Sullivan, M., & Zeng, J. (2019). "Human-Centered Artificial Intelligence: A Critical Review and Agenda." *ACM Computing Surveys*, 52(6), 1-27.

[28] Osoba, O. A., & Welser, W. (2017). An intelligence in our image: The risks of bias and errors in artificial intelligence. RAND Corporation. https://www.rand.org/pubs/research_reports/RR1744.html

[29] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). "Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions." *Future Generation Computer Systems*, 29(7), 1645-1660.

human endeavors, enriching the human experience while safeguarding against dehumanization, and unequivocally serving the collective well-being of individuals[30].

Conclusion

The intricate nexus between machine learning and human rights delineates a complex interplay, necessitating a judicious equilibrium between the symphony of innovation and the potential dissonance of inherent risks. This delicate balance, critical for the conscientious evolution of both realms, can be achieved through the adoption of robust ethical guidelines, a steadfast commitment to inclusivity, and an unwavering adherence to a human-centric approach. These pillars form the bedrock upon which society can not only harness the transformative power of machine learning but also fortify the safeguarding of fundamental human rights.

Ethical considerations serve as the linchpin in this intricate dance between machine learning and human rights. By establishing and adhering to comprehensive ethical guidelines, the trajectory of machine learning development becomes inherently tethered to principles that prioritize fairness, transparency, and accountability. This ethical scaffold not only mitigates the potential risks associated with the unchecked proliferation of machine learning but also serves as a moral compass that guides the integration of innovative technologies within the broader societal fabric.

In parallel, the emphasis on inclusivity emerges as a pivotal element in this intricate tapestry. Prioritizing inclusivity entails a deliberate effort to involve diverse voices in the development and deployment of machine learning systems. By doing so, a multifaceted perspective is injected into the discourse, minimizing the risk of bias and ensuring that the benefits of machine learning are accessible to a broad spectrum of society. This commitment to inclusivity not only bolsters the ethical foundations of machine learning but also aligns with the overarching principles of human rights by promoting equality and preventing discriminatory practices.

Crucially, a human-centric approach stands as the keystone in this symbiotic relationship. The integration of machine learning into societal frameworks necessitates a conscientious acknowledgment of the indispensable role of human agency. By ensuring that machine learning technologies augment human capabilities rather than replace them, society can guard against dehumanization and cultivate an environment where technological advancements genuinely serve the well-being of individuals.

Amidst this dynamic interplay, fostering a collaborative dialogue assumes paramount significance. The convergence of technologists, policymakers, and human rights advocates in a shared discourse becomes imperative to navigate the complexities of this evolving landscape. Such collaborative efforts facilitate the formulation of policies that not only regulate the ethical use of machine learning but also address emerging challenges and potential risks in a proactive and adaptive manner. The coalescence of expertise from these

[30] Kaplan, J. (2016). *Artificial intelligence: What everyone needs to know* (1st ed.). Oxford University Press.

disparate spheres ensures a holistic approach to shaping a future where machine learning and human rights coexist harmoniously.

To bring it all together, the intricate relationship between machine learning and human rights necessitates a nuanced and multifaceted approach. Through the judicious adoption of ethical guidelines, prioritizing inclusivity, maintaining a human-centric ethos, and fostering collaborative dialogues, society can navigate this intricate interplay to harness the transformative potential of machine learning while safeguarding the foundational principles of human rights.

References

- Berk, R. (2017). An impact assessment of machine learning risk forecasts on parole board decisions and recidivism. *Journal of Experimental Criminology, 13*(2), 193. https://doi.org/10.1007/s11292-017-9292-4
- Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). https://doi.org/10.1145/3173574.3173884
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.
- Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM, 51*(1), 107–113. https://doi.org/10.1145/1327452.1327492
- Eileen, D., & Megan, M. M. (2019). Artificial intelligence and human rights. *Journal of Democracy, 30*(2), 115–126.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Ferré, F. (1998). *Philosophy of technology*. Prentice Hall.
- Flach, P. (2012). *Machine learning: The art and science of algorithms that make sense of data*. Cambridge University Press.
- Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems, 29*(7), 1645–1660. https://doi.org/10.1016/j.future.2013.01.010
- Goldin, C., & Katz, L. F. (2008). *The race between education and technology*. Belknap Press.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Jasanoff, S. (2016). *The ethics of invention: Technology and the human future*. W. W. Norton & Company.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415
- Kaplan, J. (2016). *Artificial intelligence: What everyone needs to know*. Oxford University Press.
- Mayer-Schönberger, V., & Cukier, K. (2014). *Big data: A revolution that will transform how we live, work, and think*. Eamon Dolan/Mariner Books.
- Michaels, A. C. (2019). Artificial intelligence, legal change, and separation of powers. *University of Cincinnati Law Review, 70*.
- Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.
- Osoba, O. A., & Welser, W. (2017). *An intelligence in our image: The risks of bias and errors in artificial intelligence*. RAND Corporation.

- Richardson, R., Schultz, J. M., & Crawford, K. (2018). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review, 63*.
- Russell, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (3rd ed.). Pearson.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003
- Shurkin, N. J. (1985). Expert systems: The practical face of artificial intelligence. *Science, 227*(4687), 78.
- Sullivan, M., & Zeng, J. (2019). Human-centered artificial intelligence: A critical review and agenda. *ACM Computing Surveys, 52*(6), 1–27. https://doi.org/10.1145/3359613
- Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Knopf.
- The European Commission's High-Level Expert Group on Artificial Intelligence. (2021). *Draft ethics guidelines for trustworthy AI*. http://ec.europa.eu/digital-single-market/en/news/draftethics-guidelines-trustworthy-ai
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Transparent, explainable, and accountable AI for justice systems. *Nature Machine Intelligence, 2*(11), 739–742. https://doi.org/10.1038/s42256-020-00246-2
- Alpaydin, E. (2020). *Introduction to machine learning* (4th ed.). MIT Press.