

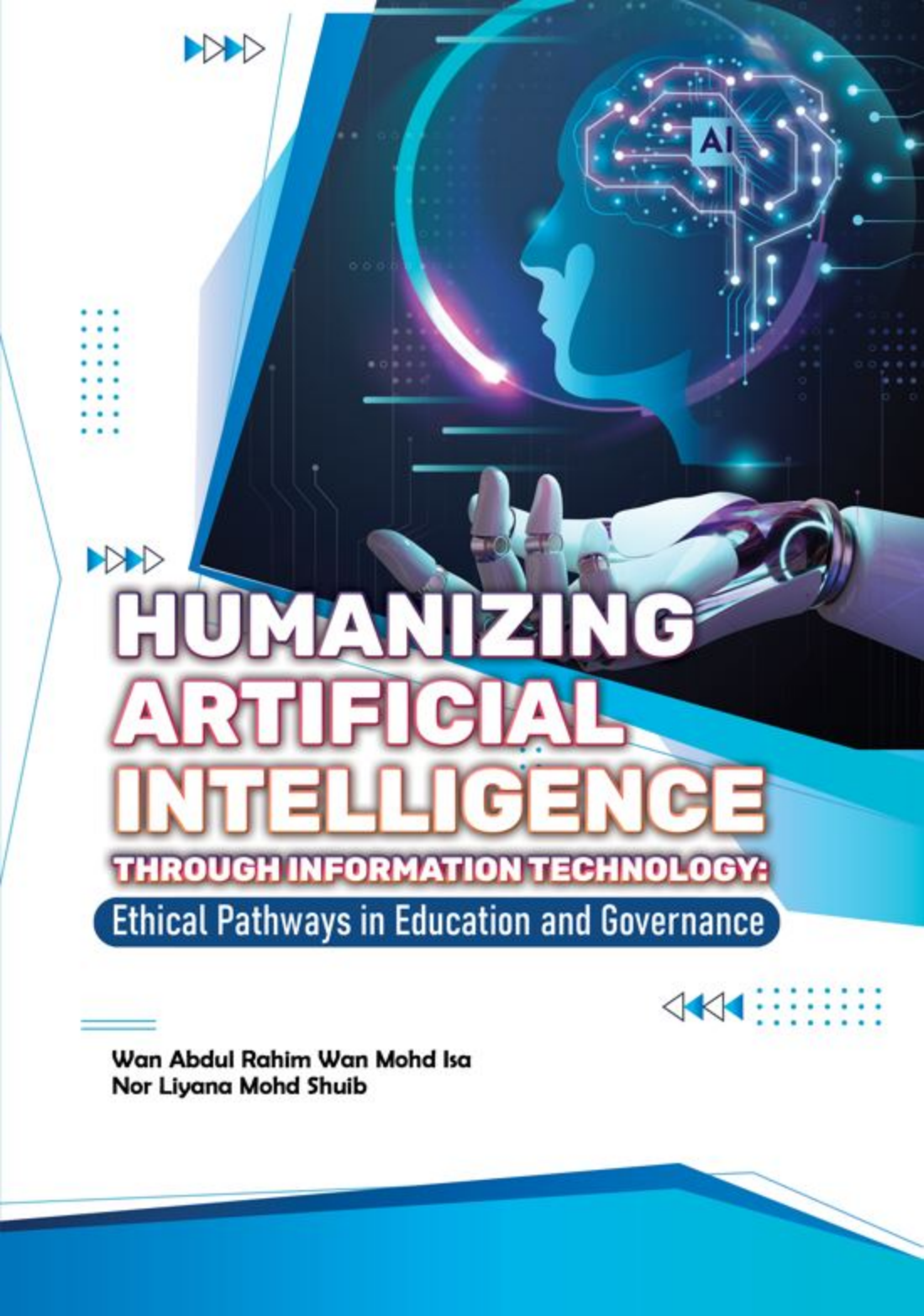
HUMANIZING ARTIFICIAL INTELLIGENCE THROUGH INFORMATION TECHNOLOGY: ETHICAL PATHWAYS IN EDUCATION AND GOVERNANCE

Authors:

Wan Abdul Rahim Wan Mohd Isa¹, Nor Liyana Mohd Shuib²

Abstract: Artificial Intelligence is fundamentally transforming society, yet its rapid ascent raises profound ethical questions regarding human agency. Humanizing Artificial Intelligence through Information Technology explores the critical intersection of AI, IT and human values, specifically within education and public governance. The authors argue that Information Technology serves as the essential infrastructure for embedding ethics into algorithmic systems. By contrasting machine-centered paradigms with a human-centered lens, this book emphasizes fairness, transparency, and accountability. It details how IT frameworks ranging from privacy-by-design to robust audit trails, effectively operationalize these vital principles. Readers will discover strategies for managing algorithmic bias, ensuring data sovereignty, and maintaining meaningful human control. Featuring comprehensive case studies and actionable best practices, this volume offers a strategic roadmap for educators, policymakers and technologist alike. This is an essential guide for building a sustainable, ethical future where AI technology successfully augments human potential and strengthens our diverse democratic institutions and societies globally.

Keywords: Algorithmic, modern society, paradigm, human agency



HUMANIZING ARTIFICIAL INTELLIGENCE

THROUGH INFORMATION TECHNOLOGY:

Ethical Pathways in Education and Governance



Wan Abdul Rahim Wan Mohd Isa
Nor Liyana Mohd Shuib

HUMANIZING ARTIFICIAL INTELLIGENCE

THROUGH INFORMATION TECHNOLOGY:

Ethical Pathways in Education and Governance

HUMANIZING ARTIFICIAL INTELLIGENCE

THROUGH INFORMATION TECHNOLOGY:
Ethical Pathways in Education and Governance

**Wan Abdul Rahim Wan Mohd Isa
Nor Liyana Mohd Shuib**

Penerbit UTHM
First Published 2026

Copyright reserved. Reproduction of any articles, illustrations and content of this book in any form be it electronic, mechanical photocopy, recording or any other form without any prior written permission from The Publisher's Office of Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, Johor is prohibited. Any negotiations are subjected to calculations of royalty and honorarium.

Author: Wan Abdul Rahim Wan Mohd Isa, Nor Liyana Mohd Shuib

Published by:

Penerbit UTHM
Universiti Tun Hussein Onn Malaysia
86400 Parit Raja,
Batu Pahat, Johor
Tel: 07-453 7051 / 7454
Fax: 07-453 6145
Website: <http://penerbit.uthm.edu.my>
E-mail: pemasaran.uthm@gmail.com
<http://e-bookstore.uthm.edu.my>

Penerbit UTHM is a member of
Majlis Penerbitan Ilmiah Malaysia (MAPIM)

Printed by:

ATTIN PRESS SDN. BHD.
No. 8, Jalan Perindustrian PP4,
Taman Perindustrian Putra Permai,
43300 Seri Kembangan, Selangor, Kuala Lumpur,
Phone number: 03- 8939 0660
Email: attinpress@gmail.com



Cataloguing-in-Publication Data

Perpustakaan Negara Malaysia

A catalogue record for this book is available
from the National Library of Malaysia

ISBN 978-629-490-306-7

TABLE OF CONTENTS

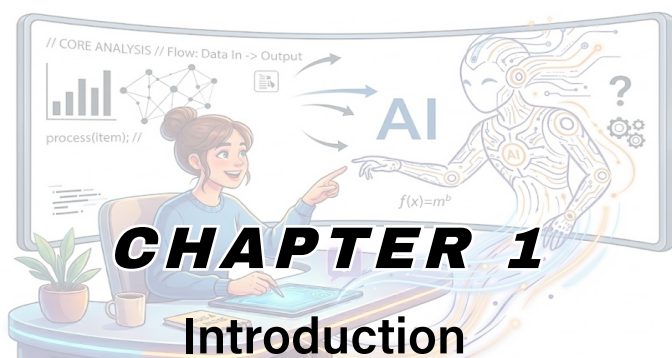
List of Tables	vii
List of Figures	ix
List of Abbreviations	xi
Preface	xv
Acknowledgements	xvii
 CHAPTER 1	 1
Introduction	
CHAPTER 2	11
Reframing AI through IT in a Human-Centered Lens	
CHAPTER 3	55
Ethical Foundations of AI in Education	
CHAPTER 4	89
IT as Enabler of Responsible AI	
CHAPTER 5	125
Governance Models for AI in the Public Sector	
CHAPTER 6	159
Building Trustworthy AI: Transparency and Explainability	
CHAPTER 7	197
Case Studies: Ethical Issues and Best Practices	

CHAPTER 8	261
Towards an Ethical AI Future in Education and Governance	
Glossary	289
Bibliography	293
Biography	331
Index	333

PREFACE

The rapid advancement of Artificial Intelligence (AI) technologies has fundamentally transformed the landscape of Information Technology (IT), reshaping how we interact with information systems, make decisions, and organize social institutions. From personalized learning platforms in our classrooms to algorithmic decision-making systems in public governance, AI has become deeply embedded in the fabric of modern society. Yet, as we stand at the precipice of this technological revolution, we face profound questions about the relationship between humans and machines, the preservation of human values, and the ethical responsibilities that must accompany such power. *Humanizing Artificial Intelligence through Information Technology: Ethical Pathways in Education and Governance* was born out of the recognition that technology is not value-neutral. Every algorithm and system design embodies choices that reflect specific priorities and power relationships. We argue that the process of "humanizing" AI requires a deliberate integration of ethical considerations, human-centered design principles, and robust governance frameworks throughout the AI lifecycle. The central premise of this research book is that AI systems must be designed to enhance rather than diminish human agency, dignity, and well-being. We posit that IT serves as the critical operational infrastructure that creates the capacity to either humanize or dehumanize AI, depending on the specific design choices made by IT professionals. Through the robust lens of IT, we explore how platforms, interfaces, and data architectures can be scientifically constructed to preserve human oversight and

support contextual decision-making. Our focus on the sectors of education and governance is intentional. These are domains where the humanization of AI is both urgent and challenging. In education, AI promises personalized learning but risks commodifying education and undermining teacher autonomy. In governance, AI offers administrative efficiency but raises critical concerns regarding algorithmic accountability and transparency. This research book employs a narrative review methodology to synthesize diverse literature, offering a comprehensive understanding of how AI can be governed to serve human values. The chapters are structured to guide the reader from conceptual foundations to practical application. We begin by reframing AI through a human-centered lens, contrasting this with machine-centered paradigms. We then delve into the ethical foundations specific to education, such as fairness and autonomy, and explore how IT infrastructure acts as a technical enabler for responsible AI through privacy-by-design and accountability mechanisms. Subsequent chapters analyze governance models in the public sector, the necessity of transparency and explainability, and real-world case studies that highlight both ethical pitfalls and best practices. Ultimately, humanizing AI is not merely a technical challenge but a social and ethical imperative. It requires the collective effort of technologists, educators, policymakers, and citizens to ensure that the choices made today regarding AI design and deployment shape a future that reflects our deepest values. We hope this research book serves as both a conceptual roadmap and a practical guide for those committed to building that future.

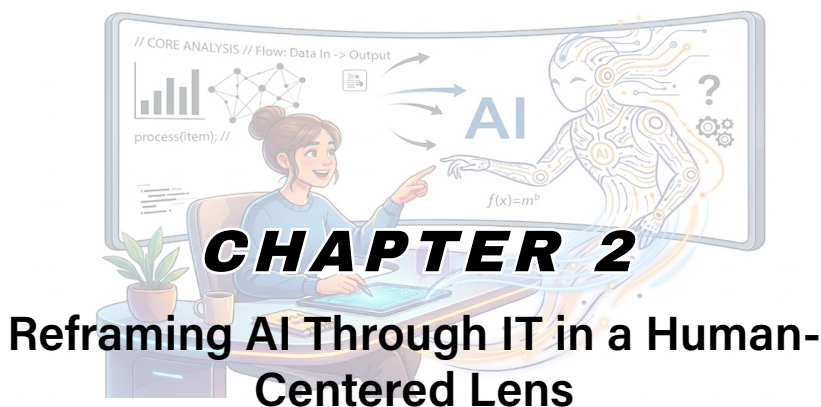


Introduction

The rapid advancement of Artificial Intelligence (AI) represents a paradigm shift within the broader field of Information Technology (IT), fundamentally transforming how we interact with digital ecosystems, make decisions, and organize social institutions. From personalized learning platforms in classrooms to algorithmic decision-making systems in public governance, AI has become deeply embedded in the fabric of modern society. However, this technological revolution raises profound questions about the relationship between humans and machines, the preservation of human values in automated systems, and the ethical responsibilities that accompany such powerful technologies.

This book explores the critical intersection of AI, IT, and human values, with particular emphasis on education and governance sectors. The central premise is that AI systems must be deliberately designed and governed to enhance rather than diminish human agency, dignity, and well-being. This process of “humanizing” artificial intelligence requires integrating ethical considerations, human-centered design principles, and robust governance frameworks throughout the AI lifecycle.

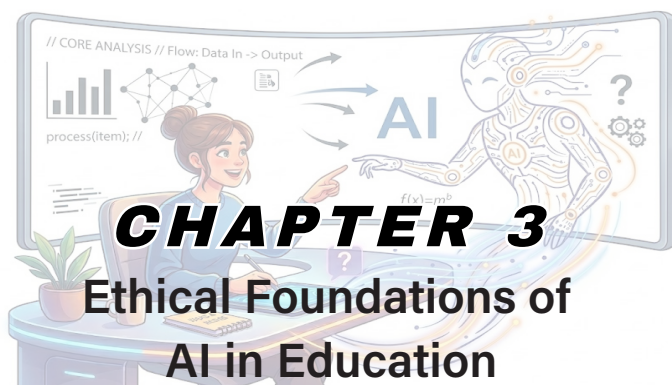
The concept of humanizing AI recognizes that technology is not value-neutral. Every algorithm, data structure, and system design embodies choices that reflect and reinforce values, priorities, and power relationships. When AI systems make decisions about student



Introduction

The rapid advancement of AI within the IT sector has fundamentally transformed the landscape of education and governance, presenting both unprecedented opportunities and complex ethical challenges. As AI systems become increasingly integrated into educational institutions and public sector operations, there is a growing recognition that purely technical approaches to AI development are insufficient to address the multifaceted needs of human societies (Dignum, 2018; Shneiderman, 2021). This realization has catalyzed a paradigm shift toward human-centered artificial intelligence, which prioritizes human well-being, agency, and values over mere technical optimization.

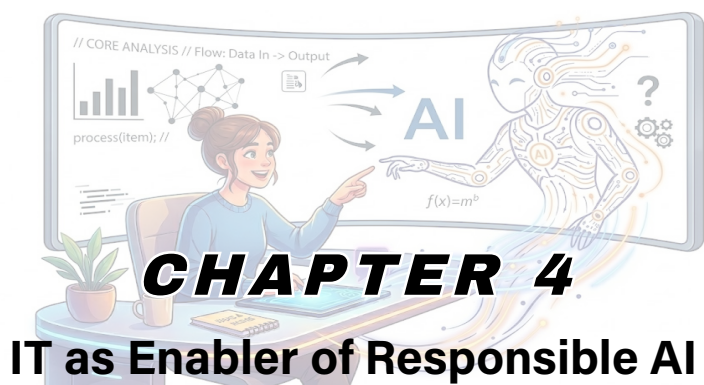
The concept of human-centered AI represents a fundamental reconceptualization of the relationship between humans and intelligent systems. Rather than viewing AI as a replacement for human intelligence or as autonomous decision-making entities, human-centered approaches position AI as an augmentative technology designed to enhance human capabilities while maintaining meaningful human control (Fischer, 2023; Garibay et al., 2023). This perspective acknowledges that AI systems operate within complex sociotechnical contexts where technical performance metrics alone cannot capture the full spectrum of human values, ethical considerations, and societal impacts that must inform responsible AI deployment.



Introduction

The integration of artificial intelligence into educational systems represents a transformative shift in IT, extending far beyond simple digitization to fundamentally reshape pedagogical practices, institutional governance, and the very nature of teaching and learning relationships (Holmes, Bialik, & Fadel, 2019). As educational institutions worldwide increasingly adopt AI-driven tools for personalized learning, automated assessment, intelligent tutoring systems, and administrative decision-making, critical questions emerge regarding the ethical implications of these technologies and their alignment with educational values and human rights (Zawacki-Richter, Marín, Bond, & Gouverneur, 2019). The ethical foundations of AI in education demand systematic examination not merely as an afterthought to IT implementation, but as a prerequisite for responsible innovation that serves the best interests of learners, educators, and society at large (Holmes & Tuomi, 2022).

The imperative for establishing robust ethical foundations in educational AI stems from multiple converging factors. First, the unique vulnerability and developmental nature of learners necessitate heightened ethical scrutiny compared to AI applications in other domains (Luckin, Holmes, Griffiths, & Forcier, 2016). Students, particularly younger learners, may lack the capacity to fully comprehend algorithmic decision-making processes or to advocate effectively for their rights when systems produce harmful

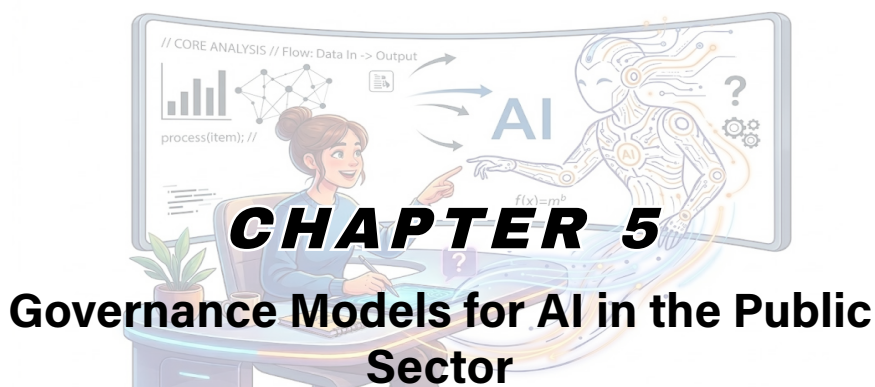


Introduction

The rapid advancement and widespread adoption of artificial intelligencesystemsacrosseducationalinstitutionsandgovernmental organizations have brought unprecedented opportunities alongside significant ethical challenges. As AI systems increasingly influence critical decisions affecting students, educators, citizens, and public services, ensuring these systems operate responsibly, transparently, and ethically has become paramount. IT infrastructure serves as the active architectural foundation that transforms abstract ethical principles into concrete, implementable mechanisms for responsible AI deployment.

This chapter examines how IT infrastructure, governance frameworks, and technical mechanisms enable responsible AI implementation in education and governance contexts. While previous chapters have established the theoretical foundations of human-centered AI and ethical frameworks, this chapter focuses specifically on the practical role of IT systems in operationalizing these principles. The relationship between IT infrastructure and responsible AI is not merely supportive but constitutive. Without robust IT foundations, ethical AI principles remain aspirational rather than operational.

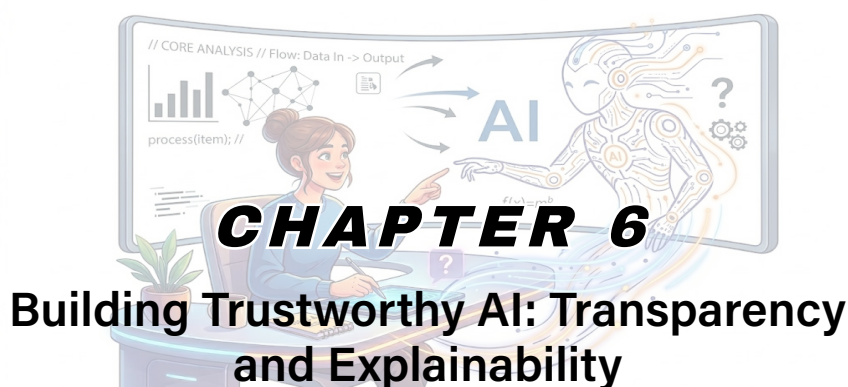
The concept of responsible AI encompasses multiple dimensions, including fairness, accountability, transparency, privacy protection, and human oversight. Each of these dimensions requires specific technical implementations supported by IT infrastructure.



Introduction

The rapid integration of AI into public sector operations has fundamentally transformed the landscape of IT governance, creating both unprecedented opportunities and complex regulatory challenges that demand systematic institutional responses (Novelli et al., 2024; Wang & Xie, 2025). As governments worldwide increasingly deploy AI systems in education, healthcare, social services, and administrative functions, the absence of coherent IT governance frameworks poses significant risks to public accountability, democratic values, and citizen rights (Busuioc, 2021; Robles & Mallinson, 2023). This chapter examines the emerging governance models for AI in the public sector, with particular emphasis on educational institutions and government administration, analyzing how IT governance standards can be integrated with ethical AI principles to create robust, accountable, and trustworthy systems.

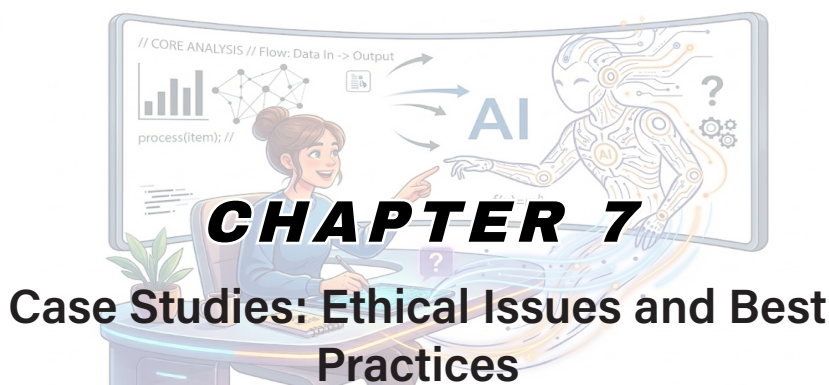
The governance of AI in public contexts differs fundamentally from private sector applications due to the unique characteristics of governmental operations, including mandatory service provision, democratic accountability requirements, protection of fundamental rights, and the imperative to serve all citizens equitably regardless of commercial considerations (Filgueiras, 2023; Kuziemski & Misuraca, 2020). Public sector IT governance must balance multiple objectives: promoting innovation and efficiency gains, protecting individual rights and public values, ensuring transparency and accountability,



Introduction

The rapid proliferation of artificial intelligence systems in educational institutions and governmental agencies has catalyzed fundamental questions about trustworthiness, accountability, and the ethical foundations of automated decision-making. AI systems increasingly influence critical decisions, from student assessment and course recommendations to resource allocation and policy implementation (Khosravi et al., 2022). Consequently, the imperative for transparency and explainability has emerged as a cornerstone of responsible IT architecture (Khosravi et al., 2022). This chapter examines the multifaceted dimensions of building trustworthy AI through enhanced transparency and explainability, with particular emphasis on educational and governance contexts where public accountability, ethical obligations, and stakeholder trust converge.

The concept of trustworthy AI extends beyond mere technical performance metrics to encompass broader considerations of fairness, accountability, interpretability, and alignment with human values. In educational settings, AI systems that recommend learning pathways, predict student success, or automate assessment must not only demonstrate accuracy but also provide comprehensible rationales that educators, students, and parents can understand, evaluate, and contest when necessary (Afroogh, Akbari, Malone, Kargar, & Alambeigi, 2024). Similarly, in governmental contexts, AI-driven decisions affecting resource allocation, service delivery,



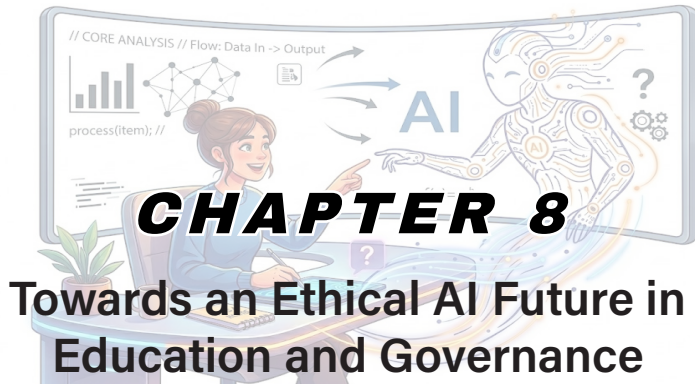
CHAPTER 7

Case Studies: Ethical Issues and Best Practices

Introduction

The integration of artificial intelligence into the IT ecosystems of educational institutions and government agencies has accelerated dramatically over the past decade, transforming how these organizations deliver services, make decisions, and interact with stakeholders. While AI technologies promise enhanced efficiency, personalization, and data-driven insights, their implementation has revealed significant ethical challenges that threaten fundamental values of fairness, transparency, accountability, and human dignity (Dignum, 2018; Floridi et al., 2018). The gap between aspirational ethical principles articulated in policy documents and the lived realities of AI deployment in practice has become increasingly apparent, necessitating a systematic examination of real-world cases to understand how IT infrastructure can either mitigate or exacerbate these issues (Mittelstadt, 2019; Jobin et al., 2019).

This chapter presents a critical analysis of case studies that illuminate the ethical landscape of AI deployment in education and governance contexts. By examining specific instances where AI systems have been implemented, evaluated, and sometimes reformed or discontinued, this chapter aims to extract actionable lessons for IT professionals. The case study approach is particularly valuable for understanding AI ethics because it captures the complex sociotechnical interactions, institutional dynamics, and contextual factors that shape outcomes in ways that abstract principles alone cannot address (Saxena et al., 2021; Kinder et al., 2023).



Introduction

The integration of AI into the IT infrastructure of education and governance represents one of the most transformative engineering shifts of the twenty-first century. As AI technologies become increasingly sophisticated and ubiquitous, their deployment in educational institutions and public sector organizations raises fundamental questions about human values, ethical principles, and the future of human-centered systems (Dignum, 2019; Holmes, Bialik, & Fadel, 2019). The preceding chapters of this book have examined the conceptual foundations of humanizing AI through information technology, explored ethical principles, analyzed governance frameworks, and investigated practical implementation challenges. This concluding chapter synthesizes these insights to chart pathways toward an ethical AI future in education and governance.

The urgency of establishing ethical frameworks for AI in education and governance stems from the profound impacts these systems have on human development, social equity, and democratic processes. Educational AI systems influence learning outcomes, shape cognitive development, and can either reinforce or mitigate existing inequalities (Holmes et al., 2022; Zawacki-Richter, Marín, Bond, & Gouverneur, 2019). Similarly, AI applications in public governance affect citizen rights, access to services, and the distribution of resources and opportunities (Wirtz, Weyerer, & Geyer,

BIBLIOGRAPHY

- Abudaqqa, F. (2025). Artificial intelligence for IT governance in Saudi Arabia: Opportunities, challenges, and future directions within COBIT 2019 and ISO/IEC 38500 frameworks. *European Scientific Journal, ESJ*, 21(25), 125. <https://doi.org/10.19044/esj.2025.v21n25p125>
- Acar, A., Aksu, H., Uluagac, A. S., & Conti, M. (2018). A survey on homomorphic encryption schemes: Theory and implementation. *ACM Computing Surveys*, 51(4), 1-35. <https://doi.org/10.1145/3214303>
- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, 6, 52138-52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- Adepoju, D. A. & Adepoju, A. G. (2025). Establishing ethical frameworks for scalable data engineering and governance in AI-driven healthcare systems. *International Journal of Research Publication and Reviews*, 6(4), 8710-8726. <https://doi.org/10.55248/gengpi.6.0425.1547>
- Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024). Trust in AI: Progress, challenges, and future directions. *Humanities & Social Sciences Communications*, 11(1568), 1-30. <https://doi.org/10.1057/s41599-024-04044-8>

- Aler Tubella, A., Mora-Cantalops, M., & Nieves, J. C. (2024). How to teach responsible AI in higher education: Challenges and opportunities. *Ethics and Information Technology*, 26(3). <https://doi.org/10.1007/s10676-023-09733-7>
- Almeida, P. G. R. d., Santos, C. D. d., & Farias, J. S. (2021). Artificial intelligence regulation: A framework for governance. *Ethics and Information Technology*, 23, 505-525. <https://doi.org/10.1007/s10676-021-09593-z>
- Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973-989. <https://doi.org/10.1177/1461444816676645>
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Asrifan, A., Sudirman, R., Rosmayanti, V., & Sulolipu, A. A. (2026). Navigating ethical AI: Ensuring Fairness in Educational Technologies. In V. Wang (Ed.), *Foundations and Frameworks for AI in Education* (pp. 27-52). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3373-2397-8.ch002>
- Atabey, A., Wilson, C., Urquhart, L. D., & Schafer, B. (2025). Fairness by Design: Cross-Cultural Perspectives from Children on AI and Fair Data Processing in their Education Futures. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1067, 1-20. <https://doi.org/10.1145/3706598.3714402>
- Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, 29(3), 3-30. <https://doi.org/10.1257/jep.29.3.3>

- Azmi, N. M. W., Zahibi, M. F. H. M., Tarudin, M. D. A., & Shahidan, N. A. (2025). AI and public governance in Malaysia: A literature review on ethical, legal, and operational challenges. *Journal of Mathematics and Computing Science*, 11 (1): 41-47. <https://doi.org/10.24191/jmcs.v11i1.8090>
- Babšek, M., Murko, E., & Aristovnik, A. (2025). Organisational AI readiness for public administration: A comprehensive review and framework for conceptual modelling. *International Journal of Economics and Business Administration*, XIII (3), 24-47. <https://doi.org/10.35808/ijeba/894>
- Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52-62. <https://doi.org/10.61969/jai.1337500>
- Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32, 1052-1092. <https://doi.org/10.1007/s40593-021-00285-9>
- Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 61-75). Springer. https://doi.org/10.1007/978-1-4614-3305-7_4
- Bang, M., & Medin, D. (2010). Cultural processes in science education: Supporting the navigation of multiple epistemologies. *Science Education*, 94(6), 1008-1026. <https://doi.org/10.1002/sce.20392>
- Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. MIT Press. <https://fairmlbook.org>
- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104, 671-732. <https://doi.org/10.15779/Z38BG31>

- Baumeister, R. F., & Leary, M. R. (1997). Writing narrative literature reviews. *Review of General Psychology*, 1(3), 311-320. <https://doi.org/10.1037/1089-2680.1.3.311>
- Bell, A., Nov, O., & Stoyanovich, J. (2023). Think about the stakeholders first! Toward an algorithmic transparency playbook for regulatory compliance. *Data & Policy*, 5, e12. <https://doi.org/10.1017/dap.2023.8>
- Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J. Mehta, S., Mojsilović, A., Nagar, S., Natesan Ramamurthy, K., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R. & Zhang, Y. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development*, 63(4/5), 4:1-4:15. <https://doi.org/10.1147/JRD.2019.2942287>
- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity Press.
- Bignami, F. (2022). Artificial intelligence accountability of public administration. *The American Journal of Comparative Law*, 70, i312-i346. <https://doi.org/10.1093/ajcl/avac012>
- Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability*, 21, 33-46. <https://doi.org/10.1007/s11092-008-9064-9>
- Biesta, G. J. J. (2010). Why 'what works' still won't work: From evidence-based education to value-based education. *Studies in Philosophy and Education*, 29, 491-503. <https://doi.org/10.1007/s11217-010-9191-x>
- Bin-Nashwan, S. A., Sadallah, M., & Bouteraa, M. (2023). Use of ChatGPT in academia: Academic integrity hangs in the balance. *Technology in Society*, 75, 102370. <https://doi.org/10.1016/j.techsoc.2023.102370>

- Binns, R. (2018). Algorithmic accountability and public reason. *Philosophy & Technology*, 31, 543-556. <https://doi.org/10.1007/s13347-017-0263-5>
- Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5454-5476). Online. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.485>
- Boninger, F., Molnar, A., & Murray, K. (2017). *Asleep at the switch: Schoolhouse commercialism, student privacy, and the failure of policymaking*. Boulder, CO: National Education Policy Center. <https://nepc.colorado.edu/publication/schoolhouse-commercialism-2017>
- Boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662-679. <https://doi.org/10.1080/1369118X.2012.678878>
- Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review*, 82(5), 977-1008. <https://doi.org/10.1177/0003122417725865>
- Breck, E., Cai, S., Nielsen, E., Salib, M., & Sculley, D. (2017). The ML test score: A rubric for ML production readiness and technical debt reduction. *2017 IEEE International Conference on Big Data (Big Data)*, Boston, MA, USA, 2017, pp. 1123-1132. <https://doi.org/10.1109/BigData.2017.8258038>
- Bukar, U. A., Sayeed, M. S., Razak, S. F. A., Yogarayan, S. & Sneesl, R. (2024). Decision-making framework for the utilization of generative artificial intelligence in education: A case study of ChatGPT. *IEEE Access*, 12, 95368-95389. <https://doi.org/10.1109/ACCESS.2024.3425172>
- Bulathwela, S., Pérez-Ortiz, M., Holloway, C., Cukurova, M., & Shawe-Taylor, J. (2024). Artificial intelligence alone will not

democratise education: On educational inequality, technosolutionism and inclusive tools. *Sustainability*, 16(2), 781. <https://doi.org/10.3390/su16020781>

Bulger, M., McCormick, P., & Pitcan, M. (2017). The legacy of inBloom. Data & Society Research Institute. <https://datasociety.net/library/the-legacy-of-inbloom/>

Bull, S., & Kay, J. (2016). SMILI☺: A framework for interfaces to learning data in open learner models, learning analytics and related fields. *International Journal of Artificial Intelligence in Education*, 26, 293-331. <https://doi.org/10.1007/s40593-015-0090-8>

Bullock, J. B. (2019). Artificial intelligence, discretion, and bureaucracy. *The American Review of Public Administration*, 49(7), 751-761. <https://doi.org/10.1177/0275074019856123>

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. In *Proceedings of Machine Learning Research*, 81, 77-91. <https://proceedings.mlr.press/v81/buolamwini18a.html>

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1-12. <https://doi.org/10.1177/2053951715622512>

Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, 81(5), 825-836. <https://doi.org/10.1111/puar.13293>

Calo, R. (2013). Against notice skepticism in privacy (and elsewhere). *Notre Dame Law Review*, 87(3), 1027-1072. <https://scholarship.law.nd.edu/ndlr/vol87/iss3/3/>

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parson, M., Raeroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J. &

- Hudson, M. (2020). The CARE principles for indigenous data governance. *Data Science Journal*, 19(43), 1-12. <https://doi.org/10.5334/dsj-2020-043>
- Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the 'good society': The US, EU, and UK approach. *Science and Engineering Ethics*, 24, 505-528. <https://doi.org/10.1007/s11948-017-9901-7>
- Cavoukian, A. (2009). Privacy by design: The 7 foundational principles. Information and Privacy Commissioner of Ontario, Canada. https://iapp.org/media/pdf/resource_center/pbd_implement_7found_principles.pdf
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20(38), 1-25. <https://doi.org/10.1186/s41239-023-00408-3>
- Christl, W., & Spiekermann, S. (2016). Networks of control: A report on corporate surveillance, digital tracking, big data & privacy. *Facultas*
- Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. *Washington Law Review*, 89(1). <https://digitalcommons.law.uw.edu/wlr/vol89/iss1/2>
- Coeckelbergh, M. (2020). Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Science and Engineering Ethics*, 26, 2051-2068. <https://doi.org/10.1007/s11948-019-00146-8>
- Conati, C., Barral, O., Putnam, V., & Rieger, L. (2021). Toward personalized XAI: A case study in intelligent tutoring systems. *Artificial Intelligence*, 298, 103503. <https://doi.org/10.1016/j.artint.2021.103503>
- Conijn, R., Kahr, P., & Snijders, C. (2023). The effects of explanations in automated essay scoring systems on student trust and

- motivation. *Journal of Learning Analytics*, 10(1), 37-53. <https://doi.org/10.18608/jla.2023.7801>
- Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. MIT Press. <https://doi.org/10.7551/mitpress/12255.001.0001>
- Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 1989(1), 139-167.
- Cukurova, M. (2024). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology*, 56(2), 469-488. <https://doi.org/10.1111/bjet.13514>
- Daggett, L. M. (2008). FERPA in the twenty-first century: Failure to effectively regulate privacy for all students. *Catholic University Law Review*, 58(1). <https://scholarship.law.edu/lawreview/vol58/iss1/4>
- Dai, R., Thomas, M. K. E., & Rawolle, S. (2025). The roles of AI and educational leaders in AI-assisted administrative decision-making: A proposed framework for symbiotic collaboration. *The Australian Educational Researcher*, 52, 1471-1487. <https://doi.org/10.1007/s13384-024-00771-8>
- Dankar, F. K., & Emam, K. E. (2013). Practicing differential privacy in health care: A review. *Transactions on Data Privacy*, 6(1), 35-67.
- Delacroix, S., & Lawrence, N. D. (2019). Bottom-up data trusts: Disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law*, 9(4), 236-252. <https://doi.org/10.1093/idpl/ipz014>
- Dencik, L., Hintz, A., Redden, J., & Treré, E. (2019). Exploring data justice: Conceptions, applications and directions. *Information*,

- Communication & Society, 22(7), 873-881. <https://doi.org/10.1080/1369118X.2019.1606268>
- Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56-62. <https://doi.org/10.1145/2844110>
- Dignum, V. (2018). Ethics in artificial intelligence: Introduction to the special issue. *Ethics and Information Technology*, 20, 1-3. <https://doi.org/10.1007/s10676-018-9450-z>
- Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way*. Springer. <https://doi.org/10.1007/978-3-030-30371-6>
- DiSalvo, B., Yip, J., Bonsignore, E., & DiSalvo, C. (Eds.). (2017). *Participatory design for learning: Perspectives from practice and research*. Routledge.
- Dragomir, A. (2025). The role of the European Union in shaping an ethical and legal framework for artificial intelligence (AI) in education. *SDGs Studies Review*, 6(studies), e035. <https://doi.org/10.37497/sdgs.v6istudies.35>
- Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2017). "Hey Google is it OK if I eat you?": Initial explorations in child-agent interaction. In *Proceedings of the 2017 Conference on Interaction Design and Children (IDC '17)*. Association for Computing Machinery, New York, NY, USA, 595-600. <https://doi.org/10.1145/3078072.3084330>
- Druin, A. (2002). The role of children in the design of new technology. *Behaviour & Information Technology*, 21(1), 1-25. <https://doi.org/10.1080/01449290110108659>
- Dwork, C. (2008). Differential privacy: A survey of results. In M. Agrawal, D. Du, Z. Duan, & A. Li (Eds.), *Theory and applications of models of computation*. TAMC 2008. Lecture Notes in Computer Science, vol 4978, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-79228-4_1

- Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. In Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS '12), Association for Computing Machinery, New York, NY, USA, 214–226. <https://doi.org/10.1145/2090236.2090255>
- Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3-4), 211-407. <https://doi.org/10.1561/04000000042>
- Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.
- European Parliament and Council. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, L119, 1-88.
- Federal Trade Commission. (2013). Complying with COPPA: Frequently asked questions. <https://www.ftc.gov/business-guidance/resources/complying-coppa-frequently-asked-questions>
- Feldman-Maggor, Y., Cukurova, M., Kent, C., & Alexandron, G. (2025). The impact of explainable AI on teachers' trust and acceptance of AI EdTech recommendations: The power of domain-specific explanations. *International Journal of Artificial Intelligence in Education*, 35, 2889–2922. <https://doi.org/10.1007/s40593-025-00486-6>
- Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, 6(1). <https://doi.org/10.1177/2053951719860542>

- Fiesler, C., Garrett, N., & Beard, N. (2020). What do we teach when we teach tech ethics? A syllabi analysis. In Proceedings of the 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20). Association for Computing Machinery, New York, NY, USA, 289–295. <https://doi.org/10.1145/3328778.3366825>
- Filgueiras, F. (2023). Artificial intelligence and education governance. *Education, Citizenship and Social Justice*, 19(3), 349–361. <https://doi.org/10.1177/17461979231160674>
- Firdaus, D. U. F. B., Roslan, N. A. B., Wan Mohamad Saferdin, W. F. H. B., Zulkarnain, I. F. B., Samasu, N. F. B. (2024). AI integration in Malaysian public administration for improved governance. *International Journal of Research and Innovation in Social Science*, VIII(IX), 3799–3812. <https://doi.org/10.47772/ijriss.2024.8090316>
- Fischer, G. (2023). A research framework focused on AI and humans instead of AI versus humans. *Interaction Design and Architecture(s) Journal - IxD&A*, 59 (2023-24), 17-36. <https://doi.org/10.55612/s-5002-059-001sp>
- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. C., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. The Berkman Klein Center for Internet & Society at Harvard University. <https://cyber.harvard.edu/publication/2020/principled-ai>
- Floridi, L. (2016). Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions. *Philosophical Transactions of the Royal Society A*, 374(2083), 20160112. <https://doi.org/10.1098/rsta.2016.0112>
- Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, 1, 261–262. <https://doi.org/10.1038/s42256-019-0055-y>

- Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.8cd550d1>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28, 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- Foucault, M. (1977). *Discipline and punish: The birth of the prison*. Vintage Books.
- Foulds, J. R., Islam, R., Keya, K. N., & Pan, S. (2020). An intersectional definition of fairness. *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, Dallas, TX, USA, 2020, pp. 1918-1921. <https://doi.org/10.1109/ICDE48307.2020.00203>
- Friedman, B., & Hendry, D. G. (2019). *Value sensitive design: Shaping technology with moral imagination*. MIT Press.
- Friedman, B., Kahn, P. H., Jr. & Borning, A. (2008). Value sensitive design and information systems. In K. E. Himma & H. T. Tavani (Eds.), *The handbook of information and computer ethics* (pp. 69-101). Wiley. <https://doi.org/10.1002/9780470281819.ch4>
- Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, 14(3), 330-347. <https://doi.org/10.1145/230538.230561>
- Gambo, I., Massenon, R., Lin, C.-C., Ogundokun, R. O., Agarwal, S., & Park, W. (2024). Enhancing user trust and interpretability in AI-driven feature request detection for mobile app reviews: An explainable approach. *IEEE Access*, 12, 114023-114045. <https://doi.org/10.1109/access.2024.3443527>
- Gangadharan, S. P., & Niklas, J. (2019). Decentering technology in discourse on discrimination. *Information, Communication*

- & Society, 22(7), 882-899. <https://doi.org/10.1080/1369118X.2019.1593484>
- Gardner, J., Brooks, C., & Baker, R. (2019). Evaluating the fairness of predictive student models through slicing analysis. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK19)*. Association for Computing Machinery, New York, NY, USA, 225-234. <https://doi.org/10.1145/3303772.3303791>
- Garibay, O. O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., Falco, G., Fiore, S. M., Garibay, I., Grieman, K., Havens, J., Jirotko, M., Kacorri, H., Karwowski, W., Kider, J., Konstan, J. A., Koon, S., López-González, M., Maifeld-Carucci, I., McGregor, S., Salvendy, G., Shneiderman, B., Stephanidis, C., Strobel, C., Holter, C. T. & Xu, W. (2023). Six human-centered artificial intelligence grand challenges. *International Journal of Human-Computer Interaction*, 39(3), 391-437. <https://doi.org/10.1080/10447318.2022.2153320>
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64-71. <https://doi.org/10.1007/s11528-014-0822-x>
- Gasser, U., & Almeida, V. A. F. (2017). A layered model for AI governance. *IEEE Internet Computing*, 21(6), 58-62. <https://doi.org/10.1109/MIC.2017.4180835>
- Giannakos, M. N., Horn, M., & Cukurova, M. (2025). Learning, design and technology in the age of AI. *Behaviour & Information Technology*, 44(5), 883-887. <https://doi.org/10.1080/0144929x.2025.2469394>
- Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19(1), 121-127. <https://doi.org/10.1136/amiajnl-2011-000089>

- Green, B. (2021). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, 45, 105681. <https://doi.org/10.1016/j.clsr.2022.105681>.
- Green, B., & Chen, Y. (2019). Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 90–99. <https://doi.org/10.1145/3287560.3287563>
- Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: Secrets of the trade. *Journal of Chiropractic Medicine*, 5(3), 101-117. [https://doi.org/10.1016/S0899-3467\(07\)60142-6](https://doi.org/10.1016/S0899-3467(07)60142-6)
- Grimmelikhuijsen, S (2023). Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making. *Public Administration Review* 83(2), 241–262. <https://doi.org/10.1111/puar.13483>
- Grosz, B. J., Grant, D. G., Vredenburgh, K., Behrends, J., Hu, L., Simmons, A., & Waldo, J. (2019). Embedded ethics: Integrating ethics across CS education. *Communications of the ACM*, 62(8), 54-61. <https://doi.org/10.1145/3330794>
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5). <https://doi.org/10.1145/3236009>
- Gursoy, D., Chi, O. H., Lu, L., & Nunkoo, R. (2019). Consumers acceptance of artificially intelligent (AI) device use in service delivery. *International Journal of Information Management*, 49, 157-169. <https://doi.org/10.1016/j.ijinfomgt.2019.03.008>
- Gürses, S., Troncoso, C., & Diaz, C. (2011). Engineering privacy by design. *Computers, Privacy & Data Protection*, 14(3).

- Ha, T., & Kim, S. (2023). Improving trust in AI with mitigating confirmation bias: A two-stage approach with debiasing information and explanation. *International Journal of Human-Computer Interaction*, 40(24), 8562–8573. <https://doi.org/10.1080/10447318.2023.2285640>
- Hafen, E., Kossmann, D., & Brand, A. (2014). Health data cooperatives—citizen empowerment. *Methods of Information in Medicine*, 53(2), 82–86. <https://doi.org/10.3414/ME13-02-0051>
- Hanna, A., Denton, R., Smart, A., & Smith-Loud, J. (2020). Towards a critical race methodology in algorithmic fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 501–512. <https://doi.org/10.1145/3351095.3372826>
- Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. In *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS'16)*. Curran Associates Inc., Red Hook, NY, USA, 3323–3331.
- Harrison, T. M., & Luna-Reyes, L. F. (2020). Cultivating trustworthy artificial intelligence in digital government. *Social Science Computer Review*, 40(2), 494–511. <https://doi.org/10.1177/0894439320980122>
- Heyder, T., Passlack, N., & Posegga, O. (2023). Ethical management of human-AI interaction: Theory development review. *The Journal of Strategic Information Systems*, 32(3), 101772. <https://doi.org/10.1016/j.jsis.2023.101772>
- Hoel, T., & Chen, W. (2016). Privacy-driven design of learning analytics applications - Exploring the design space of solutions for data sharing and interoperability. *Journal of Learning Analytics*, 3(1), 139–158. <http://dx.doi.org/10.18608/jla.2016.31.9>

Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.

Holmes, W., Persson, J., Chounta, I.-A., Wasson, B., & Dimitrova, V. (2022). Artificial intelligence and education: A critical view through the lens of human rights, democracy and the rule of law. Council of Europe.

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. G., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 32, 504-526. <https://doi.org/10.1007/s40593-021-00239-1>

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542-570. <https://doi.org/10.1111/ejed.12533>

Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Paper 600, 1-16. <https://doi.org/10.1145/3290605.3300830>

Hummel, P., Braun, M., Tretter, M., & Dabrock, P. (2021). Data sovereignty: A review. *Big Data & Society*, 8(1). <https://doi.org/10.1177/2053951720982012>

Hutchinson, B., & Mitchell, M. (2019). 50 years of test (un)fairness: Lessons for machine learning. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 49-58. <https://doi.org/10.1145/3287560.3287600>

IEEE. (2019). *Ethically aligned design - A vision for prioritizing human well-being with autonomous and intelligent systems (1st ed.)*.

- IEEE. https://standards.ieee.org/ieee/White_Paper/10594/
- Ifenthaler, D., Majumdar, R., Gorissen, P., Judge, M., Mishra, S., Raffaghelli, J. & Shimada, A. (2024). Artificial intelligence in education: Implications for policymakers, researchers, and practitioners. *Technology, Knowledge and Learning*, 29, 1693-1710. <https://doi.org/10.1007/s10758-024-09747-0>
- Jenkins, R., Hammond, K., Spurlock, S., & Gilpin, L. (2023). Separating facts and evaluation: Motivation, account, and learnings from a novel approach to evaluating the human impacts of machine learning. *AI & Society*, 38, 1415-1428. <https://doi.org/10.1007/s00146-022-01417-y>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389-399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kamiran, F., & Calders, T. (2012). Data preprocessing techniques for classification without discrimination. *Knowledge and Information Systems*, 33, 1-33. <https://doi.org/10.1007/s10115-011-0463-8>
- Karran, A. J., Charland, P., Trempe-Martineau, J., Lopez de Arana, A. O. D., Lesage, A.-M., Sénécal, S., & Léger, P.-M. (2025). Multi-stakeholder perspective on responsible artificial intelligence and acceptability in education: A vignette-based investigation. *npj Science of Learning*, 10 (44), 1-12. <https://doi.org/10.1038/s41539-025-00333-2>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>

- Katell, M., Young, M., Dailey, D., Herman, B., Guetler, V., Tam, A., Bintz, C., Raz, D., & Krafft, P. M. (2020). Toward situated interventions for algorithmic equity: Lessons from the field. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 45–55. <https://doi.org/10.1145/3351095.3372874>
- Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, 22(14), 2081-2096. <https://doi.org/10.1080/1369118X.2018.1477967>
- Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers & Education: Artificial Intelligence*, 3, 100074. <https://doi.org/10.1016/j.caeai.2022.100074>
- Kinder, T., Stenvall, J., Koskimies, E., Webb, H., & Janenova, S. (2023). Local public services and the ethical deployment of artificial intelligence. *Government Information Quarterly*, 40(4), 101865. <https://doi.org/10.1016/j.giq.2023.101865>
- Kizilcec, R. F., & Lee, H. (2022). Algorithmic fairness in education. In W. Holmes & K. Porayska-Pomsta (Eds.), *The Ethics of Artificial Intelligence in Education*. Routledge. <https://doi.org/10.4324/9780429329067>
- Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. In C. H. Papadimitrou (Ed.) *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, Leibniz International Proceedings in Informatics, Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany, 43, 43:1–43:23. <https://doi.org/10.4230/LIPIcs.ITCS.2017.43>
- Kluttz, D. N., Kohli N., Mulligan D. K. (2020) *Shaping Our Tools: Contestability as a Means to Promote Responsible*

- Algorithmic Decision Making in the Professions. In: Werbach K. (Ed.) *After the Digital Tornado: Networks, Algorithms, Humanity*. Cambridge University Press, 137-152. <https://doi.org/10.1017/9781108610018>
- Kodakandla, P. (2024). Unified data governance: Embedding privacy by design into AI model pipelines. *International Journal of Novel Research and Development*, 9(10), d247-d255.
- Krishnamoorthy, S. (2024). Interpretable classifier models for decision support using high utility gain patterns. *IEEE Access*, 12, 121456-121472. <https://doi.org/10.1109/access.2024.3455563>
- Kroll, J. A. (2018). The fallacy of inscrutability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180084. <https://doi.org/10.1098/rsta.2018.0084>
- Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, 165(3), 633-705. <https://scholarship.law.upenn.edu/penn-law-review/vol165/iss3/3>
- Kronblad, C., Essén, A., & Mähring, M. (2024). When justice is blind to algorithms: Multilayered blackboxing of algorithmic decision-making in the public sector. *MIS Quarterly*, 48(4), 1637-1662. <https://doi.org/10.25300/MISQ/2024/18251>
- Kukutai, T., & Taylor, J. (Eds.). (2016). *Indigenous data sovereignty: Toward an agenda*. ANU Press. <https://doi.org/10.22459/CAEPR38.11.2016>
- Kumar, R., Sporn, K., Waisberg, E., Ong, J., & Paladugu, P., Vadhera, A. S., Amiri, D., Ngo, A., Jagadeesan, R., Tavakkoli, A., Loftus, T., & Lee, A. G. (2025). Navigating healthcare AI governance: The comprehensive algorithmic oversight and stewardship framework for risk and equity. *Health Care Analysis*. <https://doi.org/10.1007/s10728-025-00537-y>

- Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy*, 44(6), 101976. <https://doi.org/10.1016/j.telpol.2020.101976>
- Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management Science*, 65(7), 2966-2981. <https://doi.org/10.1287/mnsc.2018.3093>
- Lee, A. H. S., Shankararaman, V., & Ouh, E. L. (2023). Vision paper: Advancing of AI explainability for the use of ChatGPT in government agencies—Proposal of a 4-step framework. In 2023 IEEE International Conference on Big Data (BigData), Sorrento, Italy, 2023, pp. 5852-5856. IEEE. <https://doi.org/10.1109/bigdata59044.2023.10386797>
- Lestari, M., Wijaya, A. F., Puspita, M. A., & Vicky. (2025). Designing ambidextrous AI governance for digital transformation in fintech sector. *Jurnal Teknologi dan Manajemen Industri Terapan*, 4(2), 522-529. <https://doi.org/10.55826/jtmit.v4i3.1042>
- Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, 61(10), 36-43. <https://doi.org/10.1145/3233231>
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1-16. <https://doi.org/10.1145/3313831.3376727>
- Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., & Jacquet, A. (2024). Responsible AI pattern catalogue: A collection of best practices for AI governance and engineering. *ACM Computing Surveys*, 56(7), 173:1-173:35. <https://doi.org/10.1145/3626234>
- Lu, Y., Wang, D., Chen, P., & Zhang, Z. (2024). Design and evaluation of trustworthy knowledge tracing model for intelligent tutoring

- system. *IEEE Transactions on Learning Technologies*, 17, 1661-1676. <https://doi.org/10.1109/tlt.2024.3403135>
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. London: Pearson.
- Mahusin, N., Sallehudin, H., & Satar, N. S. M. (2024). Malaysia public sector challenges of implementation of artificial intelligence (AI). *IEEE Access*, 12, 121035-121051. <https://doi.org/10.1109/ACCESS.2024.3448311>
- Mason, J., Peoples, B. E., & Lee, J. (2020). Questioning the scope of AI standardization in learning, education, and training. *Journal of ICT Standardization*, 8(2), 107-122. <https://doi.org/10.13052/jicts2245-800x.822>
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
- Mayson, S. G. (2019). Bias in, bias out. *The Yale Law Journal*, 128(8), 2218-2300.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6). <https://doi.org/10.1145/3457607>
- Metcalf, J., Moss, E., & Boyd, D. (2019). *Owning ethics: Corporate logics, silicon valley, and the institutionalization of ethics*. Social Research: An International Quarterly, Johns Hopkins University Press, 86(2), 449-476. <https://doi.org/10.1353/sor.2019.0022>
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 220-229. <https://doi.org/10.1145/3287560.3287596>

- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1, 501-507. <https://doi.org/10.1038/s42256-019-0114-4>
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716679679>
- Moore, P. V., & Robinson, A. (2016). The quantified self: What counts in the neoliberal workplace. *New Media & Society*, 18(11), 2774-2792. <https://doi.org/10.1177/1461444815604328>
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, 26, 2141-2168. <https://doi.org/10.1007/s11948-019-00165-5>
- Moss, E., Watkins, E. A., Singh, R., Elish, M. C., & Metcalf, J. (2021). Assembling accountability: Algorithmic impact assessment for the public interest. Data & Society Research Institute. <https://datasociety.net/library/assembling-accountability-algorithmic-impact-assessment-for-the-public-interest/>
- Mouta, A., Pinto-Llorente, A. M., & Torrecilla-Sánchez, E. M. (2024). Uncovering blind spots in education ethics: Insights from a systematic literature review on artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 34, 1166-1205 <https://doi.org/10.1007/s40593-023-00384-9>
- Muller, M., Wolf, C. T., Andres, J., Desmond, M., Joshi, N. N., Ashktorab, Z., Sharma, A., Brimijoin, K., Pan, Q., Duesterwald, E., & Dugan, C. (2021). Designing Ground Truth and the Social Life of Labels. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 94, 1-16. <https://doi.org/10.1145/3411764.3445402>

- Naja, I., Markovic, M., Edwards, P., & Cottrill, C. (2021). A semantic framework to support AI system accountability and audit. In: Verborgh, R., et al. *The Semantic Web. ESWC 2021. Lecture Notes in Computer Science*, vol 12731. Springer, Cham. https://doi.org/10.1007/978-3-030-77385-4_10
- Nassar, M., Salah, K., Ur Rehman, M. H., & Svetinovic, D. (2020). Blockchain for explainable and trustworthy artificial intelligence. *WIREs Data Mining and Knowledge Discovery*, 10(1), e1340. <https://doi.org/10.1002/WIDM.1340>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, 28, 4221–4241. <https://doi.org/10.1007/s10639-022-11316-w>
- Ninaus, M., & Sailer, M. (2022). Closing the loop - The human role in artificial intelligence for education. *Frontiers in Psychology*, 13, 956798. <https://doi.org/10.3389/fpsyg.2022.956798>
- Nissenbaum, H. (1996). Accountability in a computerized society. *Science and Engineering Ethics*, 2, 25-42. <https://doi.org/10.1007/BF02639315>
- NIST. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
- Noddings, N. (2013). *Caring: A relational approach to ethics and moral education* (2nd ed.). University of California Press.

- Novelli, C., Hacker, P., Morley, J., Trondal, J., & Floridi, L. (2024). A robust governance for the AI Act: AI Office, AI Board, Scientific Panel, and National Authorities. *European Journal of Risk Regulation*, 16(2), 566-590. <https://doi.org/10.1017/err.2024.57>
- Ntoutsis, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdli, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., Broelemann, K., Kasneci, G., Tiropanis, T., & Staab, S. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1356. <https://doi.org/10.1002/widm.1356>
- OECD. (2019). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments. <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group
- Onoja, J. P., Ajala, O. A., & Ige, A. B. (2022). Harnessing artificial intelligence for transformative community development: A comprehensive framework for enhancing engagement and impact. *GSC Advanced Research and Reviews*, 11(3), 158–166. <https://doi.org/10.30574/gscarr.2022.11.3.0154>
- Ooge, J., Kato, S., & Verbert, K. (2022). Explaining recommendations in e-learning: Effects on adolescents' trust. In *Proceedings of the 27th International Conference on Intelligent User Interfaces (IUI '22)*. Association for Computing Machinery, New York, NY, USA, 93–105. <https://doi.org/10.1145/3490099.3511140>
- Pangrazio, L., & Sefton-Green, J. (2020). The social utility of 'data literacy'. *Learning, Media and Technology*, 45(2), 208-220. <https://doi.org/10.1080/17439884.2020.1707223>

- Papadakis, T., Christou, I. T., Ipektsidis, C., Soldatos, J., & Amicone, A. (2024). Explainable and transparent artificial intelligence for public policymaking. *Data & Policy*, 6, e10. <https://doi.org/10.1017/dap.2024.3>
- Pareja, H. P. (2025). Artificial intelligence as a driver of digital government transformation: Applications and trends in government operations in the Philippines. *Policy and Governance Review*, 9(3), 302-317. <https://doi.org/10.30589/pgr.v9i3.1257>
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Pentland, A. (2014). *Social physics: How good ideas spread—The lessons from a new science*. Scribe Publications.
- Petriashvili, I., & Piotrowski, S. J. (2025). Role of AI in combating corruption in the U.S. federal government: Case study 2019-2024. *Public Integrity*, 1-8. <https://doi.org/10.1080/10999922.2025.2554424>
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Publishing. <https://doi.org/10.1002/9780470754887>
- Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J., & Weinberger, K. Q. (2017). On fairness and calibration.
- In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 5684–5693.
- Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(22), 1-13. <https://doi.org/10.1186/s41039-017-0062-8>
- Prinsloo, P., & Slade, S. (2015). Student privacy self-management: Implications for learning analytics. In *Proceedings of*

- the Fifth International Conference on Learning Analytics And Knowledge (LAK '15). Association for Computing Machinery, New York, NY, USA, 83–92. <https://doi.org/10.1145/2723576.2723585>
- Prinsloo, P., & Slade, S. (2017). Chapter 4: Ethics and learning analytics: Charting the (un)charted. In C. Lang, G. Siemens, A. Wise, & D. Gasevic (Eds.), *Handbook of learning analytics* (pp. 49-57). SOLAR. <https://doi.org/10.18608/hla17>
- Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology*, 20, 5-14. <https://doi.org/10.1007/s10676-017-9430-8>
- Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19)*. Association for Computing Machinery, New York, NY, USA, 429–435. <https://doi.org/10.1145/3306618.3314244>
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 33–44. <https://doi.org/10.1145/3351095.3372873>
- Rakova, B., Yang, J., Cramer, H., & Chowdhury, R. (2021). Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1-23. <https://doi.org/10.1145/3449081>
- Rastrollo Suárez, J. J., & Ferraro, A. E. (2024). How to employ artificial intelligence in public administration? Analysis and discussion of the Ibero American charter on artificial intelligence in civil

- service. *Eastern Journal of European Studies*, 15(2), 127-146.
<https://doi.org/10.47743/ejes-2024-0207>
- Rawls, J. (1971). *A theory of justice*. Harvard University Press.
- Rebholz, S., Libbrecht, P., & Müller, W. (2025). Designing trustworthy AI in higher education. *IntechOpen*. <https://doi.org/10.5772/intechopen.1009134>
- Reich, J., & Ito, M. (2017). *From Good Intentions to Real Outcomes: Equity by Design in Learning Technologies*. Irvine, CA: Digital Media and Learning Research Hub.
- Reidenberg, J., Russell, N. C., Kovnot, J., Norton, T. B., Cloutier, R., & Alvarado, D. (2013). Privacy and cloud computing in public schools. Center on Law and Information Policy. 2. <https://ir.lawnet.fordham.edu/clip/2>
- Reidenberg, J. R., & Schaub, F. (2018). Achieving big data privacy in education. *Theory and Research in Education*, 16(3), 263-279. <https://doi.org/10.1177/1477878518805308>
- Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). Algorithmic impact assessments: A practical framework for public agency accountability. AI Now Institute.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. Association for Computing Machinery, New York, NY, USA, 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- Ridley, M. (2024). Human-centered explainable artificial intelligence: An Annual Review of Information Science and Technology (ARIST) paper. *Journal of the Association for Information Science and Technology*, 76(1), 98-120. <https://doi.org/10.1002/asi.24889>

- Robles, P., & Mallinson, D. J. (2023). Artificial intelligence technology, public trust, and effective governance. *Review of Policy Research*, 42(1), 11-28. <https://doi.org/10.1111/ropr.12555>
- Robles, P., & Mallinson, D. J. (2025). Advancing AI governance with a unified theoretical framework: A systematic review. *Perspectives on Public Management and Governance*, 8(4), 213-227. <https://doi.org/10.1093/ppmgov/gvaf013>
- Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 26, 582-599. <https://doi.org/10.1007/s40593-016-0110-3>
- Rubel, A., & Jones, K. M. L. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society*, 32(2), 143-159. <https://doi.org/10.1080/01972243.2016.1130502>
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1, 206-215. <https://doi.org/10.1038/s42256-019-0048-x>
- Rudolf, G., & Kovač, P. (2024). The role of automated decision-making in modern administrative law: Challenges and data protection implications. *Central European Public Administration Review*, 22(2), 83-108. <https://doi.org/10.17573/cepar.2024.2.04>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68-78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Saenz, A. D., Mass General Brigham AI Governance Committee, Centi, A., Ting, D., You, J. G., Landman, A., & Mishuris, R. G. (2024). Establishing responsible use of AI guidelines: A comprehensive case study for healthcare institutions. *npj Digital Medicine*, 7(348). <https://doi.org/10.1038/s41746-024-01300-8>

- Sadiq, S., Aryani, A., Demartini, G., Hua, W., Indulska, M., Burton-Jones, A., Khosravi, H., Benavides-Prado, D., Sellis, T., Someh, I., Vaithianathan, R., Wang, S., & Zhou, X. (2022). Information resilience: The nexus of responsible and agile approaches to information use. *The VLDB Journal*, 31, 1059-1084. <https://doi.org/10.1007/s00778-021-00720-2>
- Salako, A. O., Fabuyi, J. A., Aideyan, N. T., Selesi-Aina, O., Dapo-Oyewole, D. L., & Olaniyi, O. O. (2024). Advancing information governance in AI-driven cloud ecosystem: Strategies for enhancing data security and meeting regulatory compliance. *Asian Journal of Research in Computer Science*, 17(12), 66-88. <https://doi.org/10.9734/ajrcos/2024/v17i12530>
- Sandhu, R. S., Coyne, E. J., Feinstein, H. L., & Youman, C. E. (1996). Role-based access control models. *Computer*, 29(2), 38-47. <https://doi.org/10.1109/2.485845>
- Santoni de Sio, F., & Van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI*, 5. <https://doi.org/10.3389/frobt.2018.00015>
- Sato, E., Shyyan, V., Chauhan, S., Christensen, L. (2024). Putting AI in Fair: A Framework for Equity in AI-driven Learner Models and Inclusive Assessments. *Journal of Measurement and Evaluation in Education and Psychology*, 15(Special Issue), 263-281. <https://doi.org/10.21031/epod.1526527>
- Savandha, S. D., Amelia, & Safii, N. M. (2025). Balancing student success and algorithmic fairness: A human-centered approach to educational technology. *Cristal International Journal of Interdisciplinary Research*, 1(2), 65-76. <https://doi.org/10.64042/k50x6q17>
- Saxena, D., Badillo-Urquiola, K., Wisniewski, P. J., & Guha, S. (2021). A framework of high-stakes algorithmic decision-making for the public sector developed through a case study of child-welfare. *Proceedings of the ACM on Human-*

- Computer Interaction, 5(CSCW2), 348:1-348:41. <https://doi.org/10.1145/3476089>
- Schuler, D., & Namioka, A. (Eds.). (1993). Participatory design: Principles and practices. CRC Press.
- Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54-63. <https://doi.org/10.1145/3381831>
- Sclater, N. (2016). Developing a Code of Practice for Learning Analytics. *Journal of Learning Analytics*, 3(1), 16-42. <https://doi.org/10.18608/jla.2016.31.3>
- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 59-68. <https://doi.org/10.1145/3287560.3287598>
- Selvaratnam, R., & Venaruzzo, L. (2024). Human-centered Approach to the Governance of AI in Higher Education: Principles of International Practice. *Journal of Ethics in Higher Education*, (5), 79-102. <https://doi.org/10.26034/fr.jehe.2024.6864>
- Selwyn, N. (2019). Should robots replace teachers? AI and the future of education. Polity Press.
- Shang, B. (2025). Artificial intelligence drives the governance reform of graduate education: Theoretical logic, practical dilemma and innovation breakthrough. *International Journal of Education and Social Development*, 3(2), 20-24. <https://doi.org/10.54097/xvqxmy43>
- Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, 36(6), 495-504. <https://doi.org/10.1080/10447318.2020.1741118>

- Shneiderman, B. (2021). Tutorial: Human-centered AI: Reliable, safe and trustworthy. In Companion Proceedings of the 26th International Conference on Intelligent User Interfaces (IUI '21 Companion). Association for Computing Machinery, New York, NY, USA, 7–8. <https://doi.org/10.1145/3397482.3453994>
- Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 31-40.
- Sigfrids, A., Leikas, J., Salo-Pöntinen, H., & Koskimies, E. (2023). Human-centricity in AI governance: A systemic approach. *Frontiers in Artificial Intelligence*, 6 (976887). <https://doi.org/10.3389/frai.2023.976887>
- Simjanoska Misheva M., Shahpaski D., Dobрева J., Bukovec D., Gjorgjioski B., Nikolov M., Frtunikj D., Lameski P., Aliu A., Mishev K. & Gams M. (2025). AI Act Compliance Within the MyHealth@EU Framework: Tutorial. *Journal of Medical Internet Research*, 27 (e81184). <https://doi.org/10.2196/81184>
- Snipp, C. M. (2016). What does data sovereignty imply: What does it look like? In T. Kukutai & J. Taylor (Eds.), *Indigenous data sovereignty: Toward an agenda* (pp. 39-56). ANU Press. <https://doi.org/10.22459/CAEPR38.11.2016.03>
- Solove, D. J. (2013). Introduction: Privacy self-management and the consent dilemma. *Harvard Law Review*, 126(7), 1880-1903.
- Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568-1580. <https://doi.org/10.1016/j.respol.2013.05.008>
- Swamy, V., Du, S., Marras, M., & Kaser, T. (2023). Trusting the explainers: Teacher validation of explainable artificial intelligence for course design. In LAK23: 13th International Learning Analytics and Knowledge Conference (LAK2023). Association for Computing Machinery, New York, NY, USA, 345–356. ACM. <https://doi.org/10.1145/3576050.3576147>

- Swauger, S. (2020). Software that monitors students during tests perpetuates inequality and violates their privacy. MIT Technology Review.
- Swist, T., Buckingham Shum, S., & Gulson, K. N. (2024). Co-producing AIED ethics under lockdown: An empirical study of deliberative democracy in action. *International Journal of Artificial Intelligence in Education*, 34, 670–705. <https://doi.org/10.1007/s40593-023-00380-z>
- Tan, M. X. Y., Qu, Y., & Wang, J. (2025). Student perceptions of generative artificial intelligence regulations: A mixed-methods study of higher education in Singapore. *Higher Education Quarterly*, 79(3), e70038. <https://doi.org/10.1111/hequ.70038>
- Tan, S. Y., & Taeihagh, A. (2021). Governing the adoption of robotics and autonomous systems in long-term care in Singapore. *Policy and Society*, 40(2), 211–231. <https://doi.org/10.1080/14494035.2020.1782627>
- Te'eni, D., Yahav, I., & Schwartz, D. (2025). What it takes to control AI by design: Human learning. *AI & Society*, 41:237–250. <https://doi.org/10.1007/s00146-025-02401-y>
- Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What should every child know about AI? *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01), 9795–9799. <https://doi.org/10.1609/aaai.v33i01.33019795>
- Tronto, J. (1993). *Moral boundaries: A political argument for an ethic of care*. Routledge.
- Tsai, Y. S., Perrotta, C., & Gašević, D. (2019). Empowering learners with personalised learning approaches? Agency, equity and transparency in the context of learning analytics. *Assessment & Evaluation in Higher Education*, 45(4), 554–567. <https://doi.org/10.1080/02602938.2019.1676396>

- Tsai, Y.-S., Whitelock-Wainwright, A., & Gašević, D. (2020). The privacy paradox and its implications for learning analytics. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 230–239. <https://doi.org/10.1145/3375462.3375536>
- Tsarouhas, P., & Grigoriadis, K. (2025). Building trust in AI for public administration: A strategic framework for transparency, XAI, participation, and digital literacy. In *2025 7th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (ICHORA)*, Ankara, Türkiye, 2025, pp. 1-9. <https://doi.org/10.1109/ichora65333.2025.11017116>
- Tsoi, A. H., Gartner, G., Cotten, S. W., Kim, J., Nazarian, J., Thomas, J., McSwain, S. D., Ahmadi-Moosavi, R., Rimal, R. (2025). Establishing and implementing a responsible artificial intelligence framework: A 1-year review. *Journal of the American Medical Informatics Association*, 32(11), 1778–1784. <https://doi.org/10.1093/jamia/ocaf147>
- Tun, H. M., Naing, L., Malik, O. A., & Rahman, H. A. (2025). Navigating ASEAN region artificial intelligence (AI) governance readiness in healthcare. *Health Policy and Technology*, 14(2), 100981. <https://doi.org/10.1016/j.hlpt.2025.100981>
- UNESCO. (2022). Recommendation on the ethics of artificial intelligence. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.
- Van Berkel, N., Skov, M. B., & Kjeldskov, J. (2021). Human-AI interaction: Intermittent, continuous, and proactive. *Interactions*, 28(6), 67-71. <https://doi.org/10.1145/3486941>
- Van Berkel, N., Tag, B., Goncalves, J., & Hosio, S. (2020). Human-centred artificial intelligence: A contextual morality

- perspective. *Behaviour & Information Technology*, 41(3), 502-518. <https://doi.org/10.1080/0144929X.2020.1818828>
- Van Nuenen, T., Ferrer, X., Such, J. M., & Cote, M. (2020). Transparency for Whom? Assessing Discriminatory Artificial Intelligence. *IEEE Computer*, 53(11), 36-44. <https://doi.org/10.1109/MC.2020.3002181>
- Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4), 97-112. <https://doi.org/10.9785/crl-2021-220402>
- Veale, M., & Brass, I. (2019). Administration by algorithm? Public management meets public sector machine learning. In K. Yeung & M. Lodge (Eds.), *Algorithmic Regulation*. Oxford University Press. <https://doi.org/10.1093/oso/9780198838494.003.0006>
- Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, Paper 440, 1-14. <https://doi.org/10.1145/3173574.3174014>
- Verma, S., & Rubin, J. (2018). Fairness definitions explained. In *Proceedings of the International Workshop on Software Fairness (FairWare '18)*. Association for Computing Machinery, New York, NY, USA, 1-7. <https://doi.org/10.1145/3194770.3194776>
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11 (233). <https://doi.org/10.1038/s41467-019-14108-y>

- Vogl, T. M., Seidelin, C., Ganesh, B., & Bright, J. (2020). Smart technology and the emergence of algorithmic bureaucracy: Artificial intelligence in UK local authorities. *Public Administration Review*, 80(6), 946-961. <https://doi.org/10.1111/puar.13286>
- Voigt, P., & Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR): A practical guide*. Springer. <https://doi.org/10.1007/978-3-319-57959-7>
- Vossoughi, S., & Vakil, S. (2018). Toward what ends? A critical analysis of militarism, equity and STEM education. In A. Ali & T. L. Buenavista (Eds.), *Education at war: The fight for students of color in America's public schools* (1st ed., pp. 117-140). Fordham University Press.
- Wachter, S., & Mittelstadt, B. (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review*, 2019(2), 494-620. <https://doi.org/10.7916/cblr.v2019i2.3424>
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, 7(2), 76-99. <https://doi.org/10.1093/idpl/ixp005>
- Wang, D., Bian, C., & Chen, G. (2024). Using explainable AI to unravel classroom dialogue analysis: Effects of explanations on teachers' trust, technology acceptance and cognitive load. *British Journal of Educational Technology*, 55(6), 2530-2556. <https://doi.org/10.1111/bjet.13466>
- Wang, X., & Xie, F. (2025). Global artificial intelligence governance research in the digital and intelligent era: Advances, trends and countermeasures. *Journal of Knowledge Management*, 30(1), 30-68. <https://doi.org/10.1108/jkm-01-2025-0006>
- Warschauer, M., & Matuchniak, T. (2010). New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. *Review of Research in Education*, 34(1), 179-225. <https://doi.org/10.3102/0091732X09349791>

- Watters, A. (2021). *Teaching machines: The history of personalized learning*. The MIT Press. <https://doi.org/10.7551/mitpress/12262.001.0001>
- Weller, A. (2019). Transparency: Motivations and challenges. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, & K.-R. Müller (Eds.), *Explainable AI: Interpreting, explaining and visualizing deep learning*. Lecture Notes in Computer Science, 11700. Springer, Cham. https://doi.org/10.1007/978-3-030-28954-6_2
- Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S. M., Richardson, R., Schultz, J., & Schwartz, O. (2018). *AI Now Report 2018*. AI Now Institute.
- Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). Ethical and societal implications of algorithms, data, and artificial intelligence: A roadmap for research. London: Nuffield Foundation.
- Williams, R., Park, H. W., & Breazeal, C. (2019). A is for artificial intelligence: The impact of artificial intelligence activities on young children's perceptions of robots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Paper 447, 1–11. <https://doi.org/10.1145/3290605.3300677>
- Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice*. SAGE Publications Ltd. <https://doi.org/10.4135/9781529714920>
- Williamson, B. (2021). Making markets through digital platforms: Pearson, edu-business, and the (e)valuation of higher education. *Critical Studies in Education*, 62(1), 50–66. <https://doi.org/10.1080/17508487.2020.1737556>
- Williamson, B., Bayne, S., & Shay, S. (2020). The datafication of teaching in higher education: Critical issues and perspectives.

- Teaching in Higher Education, 25(4), 351-365. <https://doi.org/10.1080/13562517.2020.1748811>
- Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223-235. <https://doi.org/10.1080/17439884.2020.1798995>
- Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—Applications and challenges. *International Journal of Public Administration*, 42(7), 596-615. <https://doi.org/10.1080/01900692.2018.1498103>
- Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The dark sides of artificial intelligence: An integrated AI governance framework for public administration. *International Journal of Public Administration*, 43(9), 818-829. <https://doi.org/10.1080/01900692.2020.1749851>
- Xu, R., Baracaldo, N., Zhou, Y., Anwar, A., & Ludwig, H. (2019). HybridAlpha: An efficient approach for privacy-preserving federated learning. In *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security (AISeC'19)*. Association for Computing Machinery, New York, NY, USA, 13-23. <https://doi.org/10.1145/3338501.3357371>
- Wadipalapa, R. P., Katharina, R., Nainggolan, P. P., Aminah, S., Apriani, T., Ma'rifah, D., & Anisah, A. L. (2024). An ambitious artificial intelligence policy in a decentralised governance system: Evidence from Indonesia. *Journal of Current Southeast Asian Affairs*, 43(1), 65-93. <https://doi.org/10.1177/18681034231226393>
- Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2). <https://doi.org/10.1145/3298981>

- Yeung, K. (2017). 'Hypernudge': Big data as a mode of regulation by design. *Information, Communication & Society*, 20(1), 118-136. <https://doi.org/10.1080/1369118X.2016.1186713>
- Zafar, M. B., Valera, I., Rodriguez, M. G., & Gummadi, K. P. (2017). Fairness constraints: Mechanisms for fair classification. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) 2017*, 962–970.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators?. *International Journal of Educational Technology in Higher Education*, 16, 39. <https://doi.org/10.1186/s41239-019-0171-0>
- Zech, H. (2021). Liability for AI: Public policy considerations. *ERA Forum* 22, 147–158. <https://doi.org/10.1007/s12027-020-00648-0>
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: PublicAffairs.

HUMANIZING ARTIFICIAL INTELLIGENCE

THROUGH INFORMATION TECHNOLOGY:

Ethical Pathways in Education and Governance

Artificial Intelligence is fundamentally transforming society, yet its rapid ascent raises profound ethical questions regarding human agency. *Humanizing Artificial Intelligence through Information Technology* explores the critical intersection of AI, IT and human values, specifically within education and public governance. The authors argue that Information Technology serves as the essential infrastructure for embedding ethics into algorithmic systems. By contrasting machine-centered paradigms with a human-centered lens, this book emphasizes fairness, transparency, and accountability. It details how IT frameworks ranging from privacy-by-design to robust audit trails, effectively operationalize these vital principles. Readers will discover strategies for managing algorithmic bias, ensuring data sovereignty, and maintaining meaningful human control. Featuring comprehensive case studies and actionable best practices, this volume offers a strategic roadmap for educators, policymakers and technologists alike. This is an essential guide for building a sustainable, ethical future where AI technology successfully augments human potential and strengthens our diverse democratic institutions and societies globally.



UTHM
Universiti Tun Hussein Onn Malaysia

Penerbit

ISBN 978-629-490-306-7



9 786294 903067

**PENERBIT
UTHM**

