

AI and Ethics Handbook

Larry Medsker
Editor

AI and Ethics Handbook

 Springer

Editor

Larry Medsker
Department of Physics
University of Vermont
Burlington, VT, USA

ISBN 978-3-032-12200-1 ISBN 978-3-032-12201-8 (eBook)

<https://doi.org/10.1007/978-3-032-12201-8>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2027

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Preface to AI and Ethics Handbook

AI and Ethics Handbook arrives at a pivotal moment, as artificial intelligence moves from specialized research into systems that shape daily life. No longer confined to laboratories or niche industries, artificial intelligence now influences education, healthcare, law enforcement, finance, governance, and even cultural life. With this reach comes profound questions: How should AI be governed? Who bears responsibility for its impacts? And how can societies ensure that they advance human flourishing rather than undermine it?

This volume offers a response. Written in clear, everyday language, it brings together global scholars and practitioners to illuminate the most pressing ethical challenges of our time. The Handbook is designed not only for academics and professionals but also for teachers seeking classroom tools, healthcare providers integrating AI into patient care, and community leaders confronting algorithmic bias in local contexts.

To orient readers, we highlight ten recurring themes that span the chapters of the volume. These themes serve as a road map to the debates, sectoral applications, and governance challenges explored throughout the Handbook. The Handbook focuses on how AI techniques, tools, and infrastructure are developing and where they may lead. It offers opportunities for academics, practitioners, policymakers, and the public to reflect on both the benefits and risks of AI, with the overarching goal of expanding its value while protecting individuals and communities from harm.

Organization of the Handbook

The chapters in this Handbook follow a consistent structure to guide readers. They begin with a background on the topic, an overview of the relevant AI technologies, and a discussion of key stakeholders and ethical issues. Authors then provide illustrative examples or case studies, concluding with references and an annotated bibliography for those wishing to explore further.

The overall design of the volume moves from foundations to applications and regional perspectives.

Part “AI and Ethics Foundations” addresses foundational concerns, including human-centered design, critical AI literacy, explainability, human-centered systems, and role of agents and AGI.

Part “Social Systems, Equity, and Long-Term Futures” explores broader social systems and equity issues, with chapters on workforce, systemic racism, digital rights, youth safety, and long-term futures.

Part “Applications Across Sectors” examines sector-specific challenges in health-care, finance, transportation, science, art, and other domains.

Part “Global Perspectives” presents global perspectives, emphasizing the ethical implications of AI across various regions, including Africa, India, Europe, and South America.

Together, these contributions provide a shared vocabulary and intellectual toolkit for understanding and guiding the future of AI. Ethics is not an afterthought; it is central to the design, deployment, and governance of technology.

At a Glance: Ten Themes of the Handbook

Ten points distill the central lessons from AI and Ethics Handbook. They trace the journey from ethical foundations and sector-specific challenges to regional perspectives, governance models, and future trajectories. Together, they provide readers with a road map for understanding not only today’s most pressing issues but also the emerging debates that will define the next decade of AI and ethics.

1. Accountability and Responsibility

AI systems often operate as “black boxes,” raising difficult questions about who is accountable when harm occurs. This Handbook emphasizes the need for transparent mechanisms and clear lines of responsibility across developers, institutions, and policymakers.

2. Bias and Fairness

Algorithmic systems frequently inherit and amplify social inequalities. Tackling bias is essential for equitable outcomes in critical domains such as hiring, credit scoring, and law enforcement and requires both technical remedies and structural reforms.

3. Misinformation and Epistemic Harm

Generative AI can fabricate persuasive but false content, undermining trust in knowledge systems and democratic discourse. Contributors highlight the urgency of addressing deepfakes, fabricated sources, and erosion of epistemic reliability.

4. Surveillance and Autonomy

Data-driven systems fuel pervasive surveillance, from targeted advertising to state monitoring. Ethical analysis must strike a balance between legitimate uses of data and protection of privacy, autonomy, and conditions for democratic self-determination.

5. AI in Warfare and Security

Autonomous weapons, cyberattacks, and automated disinformation present unprecedented ethical dilemmas. This Handbook probes the risks of militarized AI while exploring proposals for international governance and restraint.

6. Ethical Frameworks and Governance

From human rights approaches to value-sensitive design, diverse ethical frameworks offer tools for evaluating AI. Authors assess both philosophical foundations and practical governance models, including regional regulations and international initiatives.

7. Education and AI Literacy

Critical AI literacy is essential for citizens to recognize, question, and resist the harms caused by algorithms. The Handbook showcases innovative approaches to teaching ethics in universities, professional programs, and community settings worldwide.

8. Human Agency and Complementarity

AI will reshape human work and decision-making, but it should be designed to complement rather than replace human judgment. Case studies illustrate how human–AI collaboration can enhance medicine, science, and education when guided by ethical safeguards.

9. Hybrid and Future AI

The Handbook traces the long arc of AI, encompassing symbolic, statistical, and hybrid systems, and considers the futures that may emerge with general or agentic AI. These reflections highlight the limitations of current paradigms and underscore the importance of foresight.

10. Global and Plural Perspectives

AI is a global phenomenon with regionally specific impacts. Contributions from Africa, Asia, Latin America, Europe, and beyond emphasize that ethical analysis must be pluralistic, attentive to cultural diversity, historical inequities, and geopolitical realities.

Authors: Dedication and Statistics

AI and Ethics Handbook is a landmark publication by AI researchers and application experts who dedicated themselves to writing on the impacts of AI on society. The editor is grateful for the cooperation, dedication, and support of the authors in making this Handbook a success.

Dedication to Peter Smith

Our dear colleague and friend, Peter Smith, Professor of Computer Science at the University of Sunderland, passed away in 2025. Peter was a steadfast supporter of the Handbook project and a pioneering voice in the field of AI and disability studies. Peter was a mentor and advocate. He championed the inclusion of underrepresented voices in AI ethics, insisting that debates about fairness and accountability must always consider the lived realities of marginalized communities. His encouragement sustained many of the contributors to this Handbook, and his insights shaped its very structure. His absence is deeply felt, but his influence is everywhere in the Handbook and in the values we uphold in our writing.

Acknowledgements I am deeply grateful to Sumit Virmani, my former student and now colleague, whose tireless work made this project possible. He helped process 50 manuscripts, organized reviews, advised on the structure of the Handbook, and contributed as both a co-author and reviewer. His dedication and expertise were indispensable to the realization of this volume. I am very grateful for the support and patience of my family, Karen Medsker and Karla, Ruth, and Koi Jones.

I also thank my colleagues John MacIntyre and Rachel Moriarty, with whom I have shared a long journey in AI and ethics, culminating in the launch of the *AI and Ethics* journal at Springer Nature. Their insights and experiences are reflected throughout this Handbook. Colleagues Drs. J.P. Auffret, James Giordano, Anthony Rhem, Ryan Watkins, and Allan Weatherwax continued long-time collaborations through discussions and Handbook chapters. I greatly appreciate the input and discussion with Mia Carter.

I am especially pleased to acknowledge Omar Askar, Kevin Cruz, Chengmin Ding, Farhana Faruqe, Yuting Feng, and Zachary Wu, former data science and computer science students who are now valued colleagues, for their contributions to chapters and generous support during the project. Finally, I extend my appreciation to Jennifer Evans, who conceived the idea for this Handbook and encouraged us throughout the process. Her belief in the project was a source of constant motivation.

Springer Nature Policy on Authorship and Use of LLMs

The Handbook follows Springer Nature's policy: AI tools cannot be credited as authors, AI-generated images are not permitted, and any use of LLMs must be documented. Full policies are available online at <https://www.springernature.com/gp/policies/editorial-policies>. Human authorship and responsibility are central to Springer Nature and the Handbook's ethos.

Closing

The Handbook focuses on how AI techniques, tools, and infrastructure are developing and where they may lead. It offers opportunities for academics, practitioners, policymakers, and the public to reflect on both the benefits and risks of AI, with the overarching goal of expanding its value while protecting individuals and communities from harm.

This Handbook is the product of a remarkable collective effort. We are indebted to the more than 50 authors whose chapters span disciplines, sectors, and regions

and to the reviewers and editorial team whose guidance shaped the final volume. We especially acknowledge our colleague Peter Smith, whose scholarship and advocacy for inclusion and accessibility left a lasting mark on this project and whose memory we honor within these pages.

We hope that this Handbook serves as both a reference and an invitation: a reference for understanding the ethical foundations and dilemmas of AI and an invitation to join an ongoing, global conversation about shaping technologies in ways that advance justice, dignity, human flourishing, and a sustainable future.

Burlington, VT, USA

Larry Medsker

Contents

AI and Ethics Foundations

The Emergence of Critical AI Literacy 3
Tania Duarte, Kathryn Conrad, and Ismael Kherroubi Garcia

Responsible AI Principles 29
Ricardo Baeza-Yates

Responsible AI: From Theory to Practice 45
Richard Benjamins, Miguel Ángel Liébanas, Ricardo Palomo,
Idoia Salazar, and Laura González

Explainable AI 61
Farhana Faruqe and Muhammad Salar Khan

Preserving Privacy in the Age of Intelligent Machines 83
Brian Dean and Lorraine Kisselburgh

**Expanding the Ethics-By-Design Landscape: A Proposal for
Values-Based AI Innovation** 107
Olivia Gambelin

Human-Centered Design and Human-Centered AI 129
Farhana Faruqe

Ethics at the Interface: Ethical Design and User Experience for AI 145
Dorian Peters

Truth Distorted: AI’s Role in Misinformation and Disinformation 169
Arijit Goswami

AGI and Ethics 199
Kentaro Toyama

Agentic AI 221
Rebekah Tweed

**Practical Implementation of AI Governance and Guardrail Agents
for Enterprise Applications** 231
Duc Nghia Pham, Dickson Lukose, and Amy Shi-Nash

**Mitigating AI Bias: Escaping the Geometry of Bias via
Independence from Learned σ -Algebras** 243
Sergio Scandizzo and Chitro Majumdar

**Hybrid Intelligent Systems: Revisiting Ethical AI Architecture
in the Modern Era** 257
Larry Medsker and Sumit Virmani

Social Systems, Equity, and Long-Term Futures

AI and the Workforce of the Future 281
Jean-Pierre Auffret and Larry Medsker

**SYNTHComm: What AI Portends for Warfare...and Requires
of Coordinated Command and Control** 303
Elise Annett and James Giordano

**From the Neutrality Myth to Moral Design: AI’s Role in Systemic
Racism** 321
Larry Medsker

**Safety by Design: Prioritizing the Protection of Children
and Youth in the Development of Generative AI Systems** 337
Rebecca S. Portnoff, Tiffany Saade, Prasanna Vijayanathan,
Olga Scrivner, Candice Adler, Abel Peña Fernandez, Christian Pecora,
Olukayode Akanni, Iyabosola Busola Oronti, Amey Ambade,
Michael Simpson, Arijit Goswami, Beckett LeClair,
Mahendra Samarawickrama, and Ezekiel Aneke

**Growing Up Connected: Teaching Children and Youth
Healthy Tech Habits** 355
Dave Lauer, Rachel Lauer, Rhonda Lauer, and Dave Tomar

AI Data Centers’ Use of Natural Resources 383
Kevin LaGrandeur

**Safe Autonomy in the Age of Generative AI: Protocols for Risk
Mitigation and Ethical Oversight in Agentic Robotic Systems** 401
Olukayode Akanni

Artificial Intelligence and Ethics in Business Operations: Navigating Risk, Fairness, and Accountability in Finance, Human Resources, and Marketing	407
Sumit Virmani	
AI and Ethics Education	419
Bernhardt L. Trout	
Organizational AI Ethics Maturity: Concept and Application in Higher Education	439
J. Krijger and T. Thuis	
AI for People with Disabilities: Positive and Negative Ethical Impacts . .	453
Lluc Vila and Richard Benjamins	
The Diary of a Disabled PhD Student: Using AI for Inclusion, Creativity, and Practice-Based Publishing	475
Emma Astra	
The Ten Commandments of Computer Ethics in the Age of Artificial Intelligence	499
Ramon C. Barquin	
Applications Across Sectors	
AI Ethics in the Physical Sciences: Transparency, Bias, and the Transformation of Discovery	519
A. T. Weatherwax and Evan T. Weatherwax	
AI and the Belmont Report 50 Years on: Time for an Update	539
Colleen P. Lyons	
Bioethics and AI: How Bioethics Is Different	563
David N. Hoffman	
How AI Is Transforming Cybersecurity	583
Sumit Virmani	
The Power of Why: Leveraging Causal AI for Results, Trust, and Compliance	599
Cortnie Abercrombie and Frederik Hartmann	
LLMs in Research: Opportunities, Risks, and Ethics in the Age of AI-Augmented Science	625
Ryan Watkins and Heather Jorgenson	
AI and Ethics in Art	643
Grace Shoji	

AI and Movies 663
Joel Foreman

AI-Assisted Diplomatic Decision-Making During Crises: Challenges and Opportunities 685
Neeti Pokhriyal and Till Koebe

Governance Lens: An Integrated Solution for Data and AI Governance for Business 707
Chengmin Ding

AI as a Market Participant: The Rise of Autonomous Financial Agents 727
Sumit Virmani

Artificial Intelligence in Knowledge Management 741
Anthony J. Rhem

AI and Autonomous Technologies 767
Omar Askar

Ethical and Societal Implications of Automated Vehicles in the Age of AI 779
Yuting Feng and Zachary Wu

Ethical Considerations of Artificial Intelligence and Autonomous Vehicle Liability 805
Tom Karol, Lindsey Klarkowski, and Tony Cotto

Global Perspectives

Artificial Intelligence and Ethical Implications in Africa: Innovation, Inequality, and Sovereignty in the Digital Age 827
Olukayode Akanni and John Kamara

Artificial Intelligence and Ethical Implications in the Asia-Pacific Region 849
Jirat Boomuang

Artificial Intelligence and Ethics in Europe 869
Jarosław Kozak

India’s AI Trajectory: Innovation, Policy, and Ethics in the World’s Largest Democracy 891
Siddarth Razdan and Sumit Virmani

AI and Its Ethical Implications in South and Central America: Inequality, Innovation, and Ethical Governance 913
Edson Prestes and Kevin M. Cruz

**Rights-Respecting AI Governance: Big Tech Accountability
and Community-Led Solutions from the Global South. 927**
Olukayode Akanni

**Artificial Intelligence, Surveillance Neocolonial Capitalism,
and Indigenous Peoples: A Call for the Human Right to Culture 949**
Lutiana Valadares Fernandes Barbosa and Edson Prestes

Dataset Creation and Indigenous Communities 967
José Emilio García, Norma Elva Chávez, and Saiph Savage

Index. 983