
*Rutaba Saleem**

BOOK REVIEW: AI AND THE HUMAN RIGHTS OF MINORITIES

Quintavalla, Alberto – Tempermán, Jeroen: *Artificial Intelligence and Human Rights*. Oxford: Oxford University Press. 2023.
ISBN 978-0-19-288248-6. 641 pages. Price: 200 USD.

Introduction

The edited volume *AI and Human Rights*, brought together by Alberto Quintavalla and Jeroen Tempermán, offers a timely examination of how artificial intelligence (AI) intersects with fundamental human rights across different areas of life. Its principal objective is to provide a systematic exploration of both the risks and opportunities generated by AI technologies, using international human rights law as the central evaluative framework. The book, therefore, strives to map the legal, ethical, and societal dimensions of AI's impacts on rights ranging from liberty and security to racial equality.

The book provides a comprehensive analytical view of the interaction between AI technologies and first generation (civil and political rights), second generation (economic, social, and cultural rights) and third generation (economic, social, and cultural rights) of human rights. Additionally, the editors predict that, unlike earlier generations—focused on liberty, equality, and solidarity – the coming fourth wave is about empowerment in digital ecosystems. It is not just about avoiding harm; it is about ensuring dignity, control, and fairness in how AI interacts with humans, including persons belonging to national minorities.

While the volume covers a broad array of human rights, this book review focuses on nine chapters that most directly engage with issues

* International Relations BA student at Corvinus University of Budapest.

of national, linguistic, ethnic, and religious minorities. These include contributions on liberty and security, freedom of expression and assembly, racial discrimination, facial recognition, digital authoritarianism, privacy, and the governance of AI risks.

The following review discusses each of the nine chapters in turn, then reflects on the book's methodological soundness, contextual positioning, intended audience, and overall strengths and weaknesses.

Chapter 2: AI Life Cycle and Human Rights – Risks and Remedies (Martina Šmuclerová, Luboš Král, and Jan Drchal)

Šmuclerová, Král, and Drchal open with a systematic analysis of the AI life cycle, from business conception and data preparation through modelling, evaluation, and deployment. Their central claim is that human rights risks are embedded throughout these stages and must be anticipated rather than remedied only after violations occur. The chapter adopts an interdisciplinary approach, drawing from legal analysis and technical expertise, and builds on research funded by the Czech Republic's Technological Agency.

For minority protection, the discussion of biased or incomplete data is particularly significant. The authors show that if training data are skewed towards majority populations, AI systems will replicate and amplify these biases, leading to discriminatory outcomes. Examples include predictive policing systems disproportionately targeting minority neighbourhoods or hiring algorithms privileging dominant-group profiles. The chapter also emphasises transparency and explainability as prerequisites for fair trial rights, a crucial safeguard for minorities who may otherwise lack the power to contest algorithmic decisions.

The remedies suggested, including human rights risk assessments embedded into every phase of the AI life cycle and the classification of certain systems as "human rights-critical", are both pragmatic and normatively grounded. For minorities, these mechanisms are essential to prevent systemic exclusion and to guarantee equal treatment in contexts such as employment, healthcare, or criminal justice.

Chapter 3: AI and the Right to Liberty and Security (Valentina Golunova)

Golunova examines how AI applications in law enforcement and criminal justice affect the right to liberty and security, a cornerstone of international human rights law. She identifies four main technological domains: automated facial recognition (AFR), social media monitoring (SMM), predictive policing, and recidivism risk assessment tools. Each is presented as offering efficiency gains in crime detection and prevention, but also poses severe risks of arbitrariness and discrimination.

From a minority rights perspective, this chapter is particularly salient. The examples of AFR misidentifying individuals of certain ethnic backgrounds at disproportionately high rates demonstrate the dangers minorities face in everyday encounters with law enforcement. Social media monitoring tools, while sometimes defended as less intrusive, have already been used to surveil minority activists, as illustrated by the monitoring of Black Lives Matter groups in the United States. Predictive policing systems, such as those trialled in Los Angeles and the Netherlands, are shown to reinforce historical patterns of over-policing in minority neighbourhoods. Finally, recidivism assessment tools like COMPAS often embed racially biased data, leading to harsher sentencing outcomes for minority defendants.

Golunova's chapter makes clear that AI's entanglement with liberty and security is not a neutral development but one that risks undermining the procedural and substantive guarantees meant to shield minorities from arbitrary detention and unequal treatment. The call for stronger international cooperation and regulatory oversight is thus highly pertinent.

Chapter 5: AI and Freedom of Expression (Giovanni De Gregorio and Pietro Dunn)

De Gregorio and Dunn shift focus to the domain of online speech. They analyse how AI-driven content moderation systems, widely used

by social media platforms, affect the enjoyment of freedom of expression. Their analysis identifies both the promises (scalability, quick removal of harmful content) and the dangers (opacity, over-removal, and chilling effects) of automated moderation.

For minority communities, the stakes are particularly high. On the one hand, minorities often rely on digital platforms to sustain cultural expression, transmit linguistic heritage, and mobilise politically. On the other hand, automated filters frequently misclassify minority languages, slang, or political dissent as “hate speech” or “terrorism-related content.” De Gregorio and Dunn provide examples of indigenous and minority voices being disproportionately silenced due to algorithmic errors or over-cautious content removal.

The chapter’s relevance lies in its insistence that freedom of expression must not be assessed only in terms of state censorship but also in terms of private power exercised by tech companies. For minorities, who already face structural disadvantages, algorithmic governance of speech risks further narrowing the space for cultural and political participation.

Chapter 6: AI and Freedom of Assembly (Margaret Warthon)

Warthon explores how AI technologies affect the right to freedom of peaceful assembly, guaranteed in numerous human rights instruments. She highlights two main developments: the use of AI surveillance tools to monitor protests and the deployment of predictive analytics to pre-empt or control demonstrations.

The chapter documents cases where AI-driven surveillance has been used to deter participation in protests, including by identifying attendees through facial recognition or social media scraping. Minority communities, often at the forefront of political mobilisation for rights, are especially vulnerable to such practices. For example, ethnic or linguistic minority groups demanding autonomy or recognition may be branded as “security threats,” and their protests disproportionately monitored or suppressed.

Warthon warns that algorithmic pre-emption of assemblies undermines the essence of democratic participation. By predicting the likelihood of protests and deploying police accordingly, authorities risk treating minority gatherings as inherently suspicious. The chapter thus links AI technologies to the possible erosion of civic space, with direct implications for the capacity of minorities to advocate collectively for their rights.

Chapter 9: Facial Recognition in the Global North (Natalia Menéndez González)

Menéndez González provides a detailed study of facial recognition technologies (FRTs) in the Global North, tracing their adoption in public and private sectors. She analyses the legal frameworks, especially data protection and anti-discrimination laws, that seek to govern their use.

Her chapter is particularly relevant for minority protection because it underscores the structural biases embedded in FRTs. Empirical studies cited in the chapter demonstrate that error rates are significantly higher, among others, for darker-skinned individuals. In practical terms, according to the author, this means people belonging to certain social groups are disproportionately misidentified, subjected to wrongful stops, or denied access to services.

Menéndez González also situates the debate within broader discourses of trust in democratic governance. While democratic states often claim to deploy FRTs responsibly, the author shows that even in these settings, safeguards are insufficient. For minorities in the Global North, the lack of robust regulation risks perpetuating exclusion and surveillance. The chapter concludes with a call for stronger legislative interventions, echoing proposals for moratoria or bans on certain uses of facial recognition.

Chapter 10: Digital Authoritarianism in the Global South (Malcolm Katrak and Ishita Chakrabarty)

Katrak and Chakrabarty turn the lens to the Global South, where they explore the phenomenon of “digital authoritarianism.” They argue that AI is increasingly used by states to entrench authoritarian practices, from mass surveillance to censorship. Unlike in the Global North, where regulatory debates dominate, in many Global South contexts, weak institutions and limited oversight exacerbate the risks.

The chapter’s relevance to minorities lies in the fact that in authoritarian settings, ethnic and religious minorities are often among the primary targets of digital repression. The authors cite examples of AI-powered surveillance in India, where Muslim communities have been disproportionately monitored, and in China, where Uyghur Muslims are subjected to extensive biometric tracking. In such contexts, AI is not merely a technical tool, but a political weapon used to enforce majoritarian control and suppress dissent.

By drawing attention to the global asymmetries in how AI interacts with rights, the chapter highlights the urgent need for international solidarity and norm-setting. For minority protection scholars, it offers a sobering reminder that technological governance cannot be detached from geopolitical and socio-political realities.

Chapter 11: Data Control and Privacy (Bart van der Sloot)

Van der Sloot’s contribution examines the right to privacy and the issue of data control in the age of AI. He analyses how personal data is collected, processed, and repurposed, and asks whether existing data protection regimes, such as the EU’s General Data Protection Regulation (GDPR), are sufficient.

For minorities, privacy is more than an individual right: it is also a collective safeguard. Historical experience shows that states often use data collection to monitor, marginalise, or even persecute minority groups. Van der Sloot warns that AI-driven data processing,

particularly when opaque or centralised, magnifies these dangers. The risks are compounded when sensitive attributes such as ethnicity, religion, or language are inferred from seemingly neutral datasets.

The chapter suggests that stronger frameworks for collective data rights and community-based consent mechanisms may be necessary. From a minority rights perspective, this aligns with calls for “data sovereignty” that would give communities greater control over how information about them is used.

Chapter 13: Racial Discrimination (Louis Koen and Kgomotso Mufamadi)

Koen and Mufamadi provide a focused discussion on racial discrimination in AI systems. They analyse how algorithmic bias reproduces historical inequalities and perpetuates exclusion in areas such as employment, housing, and law enforcement.

The chapter’s empirical grounding is especially valuable. The authors detail how training datasets often reflect majority-dominant norms, producing outputs that disadvantage racial minorities. For example, credit-scoring algorithms may systematically underrate applicants from minority neighbourhoods, while hiring tools reproduce gendered and racialised hierarchies.

From a minority protection standpoint, this chapter is central. It not only diagnoses the problem but also highlights the cumulative effects of discrimination across multiple rights: the right to work, the right to housing, the right to non-discrimination, and even the right to political participation. Koen and Mufamadi argue that combating algorithmic racism requires not just technical fixes but also structural changes in data governance and accountability.

Chapter 29: Governing Risks and Benefits of AI (Kostina Prifti, Alberto Quintavalla, and Jeroen Temperman)

The closing chapter by Prifti, Quintavalla, and Temperman provides a broad reflection on how to govern the risks and benefits of AI. It synthesises insights from the preceding chapters and situates them within global governance debates.

For minority protection, the authors' insistence on participatory governance is particularly salient. They argue that AI regulation must include diverse voices, especially those most affected by technological harms. The chapter also stresses the importance of international cooperation, noting that unilateral approaches are insufficient in the face of globalised data flows and cross-border AI applications.

By framing AI governance as a balancing act between risks and benefits, the chapter underpins that minority protection cannot be an afterthought. It must be integrated into the design of governance frameworks from the outset.

Methodological Soundness

The volume demonstrates methodological pluralism, combining doctrinal legal analysis with empirical examples and interdisciplinary collaboration between legal scholars and technical experts. For example, Šmuclerová, Král, and Drchal's integration of computer science into legal analysis strengthens their arguments, while Koen and Mufamadi's use of empirical case studies grounds their critique of racial bias.

That said, not all chapters engage equally with empirical data, and some rely heavily on normative assertions. Still, the diversity of sources, ranging from international treaties to case law and technical reports, ensures a robust evidentiary foundation.

Context and Motivation

The book emerges against the backdrop of intense global debates over the societal impacts of AI. Politically, it engages with concerns about surveillance, discrimination, and democratic backsliding. Scholarly debates about the regulation of “high-risk AI” systems, such as those in the European Union’s AI Act, form another part of its context. The editors clearly aim to influence both academic and policy conversations, positioning the volume as a resource for understanding how AI reshapes core human rights, including the human rights of persons belonging to minorities.

This context is particularly critical for the field of minority protection. As states increasingly deploy AI in areas such as policing, welfare provision, and border control, people belonging to minorities face a disproportionate exposure to risks. The book situates these challenges within broader struggles over equality, justice, and accountability. Among the key sources it engages with is Frank Pasquale’s influential work on the “black box society.” Pasquale argues that algorithms often operate as “black boxes”, i.e. their internal logic, data inputs, and decision-making criteria remain opaque both to those directly affected and to society at large. This opacity produces an imbalance of power, with significant consequences for the protection and exercise of human rights.

Intended Audience and Usefulness

The volume is well-suited for a diverse readership. Academics in law, political science, and technology studies will find it rich in conceptual analysis. Policymakers and regulators can draw on its practical recommendations for risk assessment and governance. Civil society organisations working on minority rights will benefit from the empirical cases and normative arguments highlighting how AI technologies affect vulnerable groups.

For scholars focused on minority protection, the book is particularly valuable as it links abstract human rights debates with concrete technologies that shape the lived experiences of minorities.

Strengths and Weaknesses

Strengths:

- Comprehensive coverage of human rights dimensions of AI.
- Interdisciplinary collaboration that bridges law and technology.
- Empirical grounding in real-world cases, from predictive policing in the US to surveillance of minorities in the Global South.
- Strong normative orientation towards human rights protection, with explicit recognition of risks for minorities.

Weaknesses:

- Uneven empirical engagement across chapters; some contributions remain more descriptive than analytical.
- Limited attention to positive case studies where AI has been used to empower minorities (e.g., language revitalisation).
- While the concluding governance chapter stresses participation, the volume as a whole could have more consistently incorporated minority perspectives as co-authors or case-study framers.

Conclusion

AI and Human Rights is a timely contribution to the growing literature on technology and rights. By mapping the risks and remedies of AI across multiple domains, the volume demonstrates both the breadth of challenges and the need for principled governance. For the field of minority protection, the chapters reviewed here highlight how AI can

exacerbate long-standing inequalities, from discriminatory policing to suppression of cultural expression.

The book's main achievement lies in demonstrating that AI is not a neutral tool, but a socio-technical system deeply entangled with structures of power. For minorities, this means that without deliberate safeguards, AI risks becoming another instrument of exclusion. At the same time, the book offers hope that through interdisciplinary cooperation, participatory governance, and rigorous rights-based frameworks, AI can be steered towards inclusion and justice. For scholars, practitioners, and policymakers concerned with the human rights of minorities, this volume is a valuable read.