DESIGNING FOR DIGNITY: HUMAN-CENTRED AI GOVERNANCE AND THE FUTURE OF DIGITAL RIGHTS

O. Jayarevathi, M.A., M.L., Assistant Professor, Government Law College, Vellore

ABSTRACT

As artificial intelligence (AI) systems increasingly mediate access to information, employment, credit, healthcare, and public services, the question of how digital rights should be conceived and protected has moved from the margins of technology policy to its centre. This paper examines the concept of a human-centred AI society -- one in which the design, deployment, and governance of AI systems are oriented around human dignity, autonomy, and equitable participation rather than narrowly around technical performance or commercial efficiency. Drawing on human rights theory, comparative AI governance frameworks, and the sociotechnical critique of algorithmic systems, the paper argues that existing digital rights -- privacy, non-discrimination, due process, and freedom of expression -- are being reshaped and, in some cases, eroded by the scale, opacity, and automation characteristic of contemporary AI. The paper surveys the emerging global governance landscape, including the EU Artificial Intelligence Act, the Council of Europe's Framework Convention on Artificial Intelligence, the UNESCO Recommendation on the Ethics of Artificial Intelligence, the OECD AI Principles, and national approaches in India and the United States, to assess how far these instruments operationalize human-centred commitments in practice. Building on this analysis, the paper proposes institutional and design-level building blocks for a human-centred AI society, including participatory design, independent algorithmic auditing, public digital literacy, and dedicated oversight institutions such as AI ombudspersons and regulatory sandboxes. The paper concludes that a human-centred AI future depends not on any single regulatory instrument but on the sustained alignment of law, technical design, and public accountability around the protection of human dignity in an increasingly automated society.

Keywords: human-centred AI, digital rights, AI governance, human dignity, algorithmic accountability, AI ethics, technology policy.

1. Introduction

The twenty-first century's digital transformation has been accompanied by a parallel transformation in the conceptual vocabulary of rights. Where earlier generations of human rights discourse centered on civil, political, economic, and social entitlements, the pervasive integration of artificial intelligence (AI) into everyday governance, commerce, and communication has given rise to a distinct, though closely related, body of concerns often described collectively as "digital rights": the right to privacy in data-driven environments, the right to be free from algorithmic discrimination, the right to meaningful explanation of automated decisions, and the right to participate in the institutions that govern increasingly automated public life.[1]

This paper takes as its point of departure the concept of a "human-centred AI society," understood not as a slogan but as a substantive design and governance commitment: that AI systems should be built, deployed, and regulated in ways that place human dignity, autonomy, and welfare above narrow technical or commercial optimization.[2] The gap between this normative aspiration and the observed practice of AI deployment -- documented extensively in the sociotechnical literature on algorithmic bias, opaque automated decision-making, and the disproportionate burden of AI-driven harms on already marginalized populations -- is the central problem this paper addresses.[3]

This paper proceeds as follows. Section 2 reviews the theoretical and normative literature on human rights, human-centred design, and their application to AI. Section 3 develops the normative foundations of a human-centred AI society around dignity, autonomy, and equity. Section 4 examines specific digital rights placed at risk by AI systems. Section 5 surveys the global AI governance landscape. Section 6 proposes institutional and design-level building blocks for realizing a human-centred AI society. Section 7 offers illustrative case studies, Section 8 discusses tensions between innovation, sovereignty, and rights protection, Section 9 addresses limitations and future research, and Section 10 concludes.

[1]Latonero, M. (2018). Governing artificial intelligence: Upholding human rights & dignity. Data & Society Research Institute.
[2]Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1).
[3]Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

2. Review of Literature:

2.1 Human Rights Foundations and Emerging Digital Rights

The Universal Declaration of Human Rights, adopted in 1948, articulates foundational commitments to dignity, equality, privacy, and due process that predate digital technology but have proven remarkably adaptable as normative anchors for contemporary digital rights discourse.[4] Civil society organizations and scholars have increasingly argued that existing human rights instruments apply directly to AI-mediated contexts, rather than requiring an entirely new rights framework, though they acknowledge that the scale, automation, and opacity of AI systems create novel enforcement and accountability challenges not contemplated by mid-twentieth-century human rights instruments.[5]

2.2 Human-Centred AI: Origins and Core Principles

The concept of human-centred AI emerged from a convergence of human-computer interaction research, AI ethics scholarship, and policy advocacy, converging on the claim that AI systems should augment rather than displace human agency and should be designed with explicit attention to their social and psychological effects on users and affected communities.[6] A systematic review of AI ethics guidelines by Jobin, Ienca, and Vayena found substantial global convergence around five principles -- transparency, justice and fairness, non-maleficence, responsibility, and privacy -- while also identifying significant divergence in how these principles are interpreted and operationalized across jurisdictions and industry sectors.[7]

2.3 Critiques of Techno-Solutionism and the Limits of Ethics Guidelines

A substantial body of critical scholarship cautions against treating voluntary ethics principles as a substitute for enforceable rights protections. Crawford's political-economic analysis of AI systems argues that AI's material infrastructure -- from mineral extraction to data labor to energy consumption -- is bound up with concentrations of corporate and state power that

[4]Universal Declaration of Human Rights, G.A. Res. 217 (III) A, U.N. Doc. A/RES/217(III) (Dec. 10, 1948).
[5]Access Now. (2018). Human Rights in the Age of Artificial Intelligence. Access Now Publications.
[6]Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication No. 2020-1.
[7]Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399.

voluntary ethical commitments do little to constrain.[8] Selbst and colleagues similarly argue that many AI fairness interventions fail because they abstract algorithmic systems away from the social and institutional context in which they operate, a critique directly relevant to any effort to build genuinely human-centred AI governance rather than narrowly technical fixes.[9]

This critique of "ethics washing" -- the deployment of high-minded ethical language to forestall binding regulation -- has been echoed by scholars who observe that the proliferation of corporate AI ethics boards and voluntary principle statements has, in many documented instances, coincided with continued deployment of contested AI systems with little material change in practice.[10] These critiques do not counsel abandoning human-centred design principles altogether, but they do suggest that principles must be paired with enforceable legal obligations and independent verification mechanisms of the kind examined in Sections 5 and 6, rather than left to voluntary industry self-governance.

2.4 Measuring Convergence: Meta-Analyses of AI Ethics Principles

Beyond individual critiques, several meta-analytic studies have systematically compared the substantive content of AI ethics guidelines issued by governments, corporations, and civil society organizations worldwide. Hagendorff's review of more than twenty major guideline documents found broad rhetorical convergence around principles such as accountability and fairness, but substantially less agreement on how those principles should be weighed against one another when they conflict, and comparatively little attention across the surveyed documents to structural questions of power, labor, and environmental cost.[11] This finding is significant for the present paper's argument because it suggests that formal convergence on human-centred language, of the kind documented by Jobin, Ienca, and Vayena, should not be mistaken for convergence on substantive governance outcomes, reinforcing the case for

[8]Crawford, K. (2021). Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press.

[9]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 (pp. 59–68). ACM.

[10]Wagner, B. (2018). Ethics as an escape from regulation: From ethics-washing to ethics-shopping? In E. Bayamlioglu, I. Baraliuc, L. A. W. Janssens, & M. Hildebrandt (Eds.), Being Profiled: Cogitas Ergo Sum (pp. 84–89). Amsterdam University Press.

[11]Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines, 30(1), 99–120.

examining binding legal instruments, rather than principle statements alone, in Section 5.

3. Foundations of a Human-Centred AI Society

3.1 Human Dignity as a Normative Anchor

Human dignity, though notoriously difficult to define with precision, functions across constitutional and human rights traditions as a foundational value from which more specific rights are derived, and offers a normative anchor for AI governance precisely because it resists reduction to quantifiable metrics of efficiency or accuracy that dominate technical AI development.[12] A dignity-centered approach to AI governance asks not only whether a system is accurate or efficient, but whether its deployment respects the personhood, autonomy, and social standing of those subject to it.

3.2 Autonomy, Agency, and Meaningful Human Oversight

Human-centred AI requires preserving meaningful human agency in decisions that significantly affect individuals' lives, a principle operationalized in several regulatory frameworks as a requirement of meaningful human oversight over automated decision-making systems, rather than oversight that is merely nominal or performed after the fact.[13]

3.3 Equity, Inclusion, and Access to AI's Benefits

A human-centred conception of AI society must also address distributive questions: who has access to the benefits of AI-driven productivity gains, and who disproportionately bears its costs. Noble's analysis of search-engine algorithms demonstrates how ostensibly neutral AI systems can reproduce and amplify existing social hierarchies, particularly along lines of race and gender, underscoring that equity cannot be treated as a secondary consideration in AI system design.[14] The equity dimension of human-centred AI also has a pronounced global dimension: the concentration of frontier AI research and infrastructure in a small number of wealthy jurisdictions raises the risk that the Global South will experience AI predominantly as

[12]Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. Philosophical Transactions of the Royal Society A, 376(2133), 20180080.
[13]Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Art. 14.
[14]Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press.

a site of data extraction and low-wage annotation labor rather than as a genuine participant in shaping AI's benefits, a concern that distributive theories of digital justice increasingly foreground.[15]

3.4 Solidarity and Collective Digital Rights

While much digital rights discourse is framed in individualist terms -- an individual's privacy, an individual's right to explanation -- a human-centred AI society also requires attention to collective and solidarity-based rights, since many AI-driven harms, such as the disparate-impact discrimination discussed in Section 4.2 or the environmental costs of large-scale model training, are experienced by groups and communities rather than by isolated individuals.[16] Recognizing collective digital rights -- for instance, a community's right to contest the deployment of a surveillance system in a public space, independent of any single resident's individual consent -- offers a conceptual bridge between the dignity-centered framework developed here and the participatory governance mechanisms proposed in Section 6.

4. Digital Rights at Risk in the AI Era

4.1 The Right to Privacy and Data Protection

AI systems' dependence on large-scale data collection places sustained pressure on the right to privacy, particularly where inference and profiling capabilities allow sensitive attributes to be derived from ostensibly non-sensitive data, a concern examined extensively in the data protection literature and addressed, with varying degrees of success, by instruments such as the GDPR's purpose-limitation and data-minimization principles.

4.2 The Right to Non-Discrimination and Algorithmic Fairness

Machine learning systems trained on historical data risk encoding and perpetuating existing patterns of discrimination, a problem well documented in domains ranging from credit scoring to predictive policing to hiring algorithms.[17] The technical fairness literature has produced

[15]Birhane, A. (2020). Algorithmic colonization of Africa. Scripted, 17(2), 389–409.
[16]Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. Big Data & Society, 4(2), 1–14.
[17]Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Machine Learning: Limitations and Opportunities. fairmlbook.org.

multiple, sometimes mutually incompatible, formal definitions of algorithmic fairness, complicating the translation of the legal right to non-discrimination into enforceable technical requirements.

4.3 The Right to Explanation and Due Process

Where AI systems inform consequential decisions in domains such as social welfare eligibility, employment, or criminal justice, due process concerns require that affected individuals have meaningful access to the basis of those decisions and a genuine opportunity to contest them, a requirement that opaque, high-dimensional machine learning models can make difficult to satisfy in practice.

4.4 Digital Labor Rights and Automation

AI-driven automation raises distinct digital rights concerns for workers, both through direct displacement of routine tasks and through the algorithmic management of labor -- automated scheduling, performance monitoring, and task allocation -- that can undermine worker autonomy and bargaining power even where formal employment is preserved.[18] The International Labour Organization has highlighted generative AI's particular potential to reshape task composition across a wide range of occupations, calling for governance frameworks that address both job displacement and the quality and dignity of AI-mediated work that remains.[19]

4.5 Freedom of Expression, Content Moderation, and Information Rights

AI-driven content moderation and recommendation systems shape the practical scope of freedom of expression online, both by removing or down-ranking content flagged as violating platform policy and by determining, through opaque ranking algorithms, which lawful content individual users are most likely to encounter, a dynamic that has drawn sustained scrutiny from freedom-of-expression scholars concerned about the delegation of speech-governance functions to largely unaccountable private algorithmic systems.[20] The right to receive and

[18]Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change, 114, 254–280.
[19]International Labour Organization. (2023). World Employment and Social Outlook: The Value of Essential Work -- With a Special Focus on Generative AI. ILO Publications.
[20]Llansó, E. J. (2020). No amount of "AI" in content moderation will solve filtering's prior-restraint problem. Big Data & Society, 7(1), 1–6.

impart information, protected under most human rights instruments, is therefore not only a matter of formal censorship but also of the design choices embedded in AI-driven ranking and moderation systems, which the human-centred framework developed in this paper treats as squarely within the scope of digital rights protection rather than as a purely private commercial matter.

5. The Global AI Governance Landscape

5.1 The European Union's Artificial Intelligence Act

The EU AI Act adopts a risk-based regulatory architecture, prohibiting certain AI applications deemed to pose unacceptable risk to fundamental rights, imposing extensive obligations on providers of high-risk AI systems, and establishing lighter transparency obligations for lower-risk applications, representing the most comprehensive binding AI-specific regulatory framework adopted to date.[21]

5.2 The Council of Europe Framework Convention on AI

The Council of Europe's Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, opened for signature in 2024, represents the first international treaty explicitly binding signatory states to ensure that AI systems are designed, developed, and applied in a manner consistent with human rights, democracy, and the rule of law throughout their lifecycle.[22]

5.3 UNESCO, OECD, and the G7 Hiroshima Process

The UNESCO Recommendation on the Ethics of Artificial Intelligence, adopted by 193 member states in 2021, articulates a comprehensive set of values and principles -- including human dignity, environmental sustainability, and diversity -- intended to guide national AI policy even in the absence of binding legal force.[23] The OECD AI Principles, adopted in 2019 and subsequently endorsed by the G20, similarly commit signatory governments to AI that is innovative, trustworthy, and respectful of human rights and democratic values, while the G7

[21]Regulation (EU) 2024/1689, Arts. 5–6.
[22]Council of Europe. (2024). Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, CETS No. 225.
[23]UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. UNESCO Publishing.

Hiroshima process has produced voluntary international guiding principles specifically targeting developers of advanced, frontier AI systems.[24]

5.4 National Approaches: Comparative Snapshot

National AI governance strategies remain highly heterogeneous. India's NITI Aayog has articulated a "Responsible AI for All" approach emphasizing broad-based access, inclusion, and equitable distribution of AI's benefits across a large and diverse population, reflecting distinct developmental priorities from those emphasized in wealthier jurisdictions.[25] The United States, in the absence of comprehensive federal AI legislation, has relied on a combination of sectoral regulation and non-binding guidance, including the White House's Blueprint for an AI Bill of Rights, which articulates five aspirational protections -- safe and effective systems, protection from algorithmic discrimination, data privacy, notice and explanation, and human alternatives -- without establishing enforceable legal obligations comparable to the EU AI Act.[26]

5.5 Regional and Multilateral Initiatives Beyond the Transatlantic Axis

Much comparative AI governance scholarship has focused disproportionately on the United States and the European Union, but a fuller account of the global landscape must also consider regional and multilateral initiatives emerging elsewhere. The African Union's Continental Artificial Intelligence Strategy, adopted in 2024, explicitly frames AI governance around inclusive development, cultural and linguistic diversity, and safeguards against the entrenchment of existing global digital inequities, reflecting priorities distinct from the risk-based, compliance-oriented framing of the EU AI Act.[27] Within Southeast Asia, ASEAN's Guide on AI Governance and Ethics emphasizes voluntary, business-friendly principles intended to support regional economic integration while avoiding the compliance burdens associated with binding regulation, illustrating a third distinct governance posture situated between the EU's binding rights-based model and the United States' predominantly voluntary

[24]Organisation for Economic Co-operation and Development. (2019). Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449.
[25]NITI Aayog. (2021). Responsible AI for All: Approach Document for India. Government of India.
[26]The White House Office of Science and Technology Policy. (2022). Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People.
[27]African Union. (2024). Continental Artificial Intelligence Strategy. African Union Development Agency.

approach.[28]

These divergent regional postures matter for the human-centred framework developed in this paper because they suggest that any workable account of global AI governance convergence must accommodate substantial variation in regulatory instruments and enforcement mechanisms, even where there is meaningful convergence, as documented in Sections 2.2 and 2.4, around underlying normative commitments to human dignity and equitable access.

6. Building Blocks for a Human-Centred AI Society

6.1 Participatory and Inclusive AI Design

Human-centred AI governance requires moving affected communities from the position of passive data subjects to active participants in the design and evaluation of AI systems that govern their lives, a shift that participatory design methodologies from human-computer interaction research offer concrete tools to operationalize, including community advisory boards and co-design workshops conducted prior to deployment.

6.2 Algorithmic Accountability and Independent Auditing

Independent algorithmic auditing -- conducted by regulators, accredited third parties, or affected-community representatives rather than solely by the developing organization -- provides a mechanism for verifying compliance with fairness, safety, and transparency obligations that voluntary self-assessment cannot reliably deliver, and is increasingly incorporated as a formal requirement for high-risk systems under frameworks such as the EU AI Act.

6.3 Digital Literacy and Public Empowerment

Meaningful digital rights protection depends not only on institutional safeguards but on the public's capacity to understand and exercise those rights, making sustained investment in AI and digital literacy programs -- particularly for populations at greatest risk of AI-driven harm -- a necessary complement to formal legal protections rather than an optional add-on.

[28]Association of Southeast Asian Nations. (2024). ASEAN Guide on AI Governance and Ethics. ASEAN Secretariat.

6.4 Institutional Mechanisms: AI Ombudspersons and Regulatory Sandboxes

Dedicated institutional mechanisms, including AI ombudsperson offices empowered to receive and investigate individual complaints, and regulatory sandboxes that allow novel AI systems to be tested under close regulatory supervision prior to full-scale deployment, offer complementary pathways for embedding human-centred oversight directly into the AI governance ecosystem rather than relying exclusively on after-the-fact litigation.[29]

6.5 Multistakeholder Governance and Civil Society Participation

Beyond formal state institutions, human-centred AI governance benefits from sustained multistakeholder participation -- including civil society organizations, academic researchers, affected-community representatives, and labor unions -- in standard-setting processes that have historically been dominated by government and industry actors alone.[30] The multistakeholder Internet governance model, developed over several decades in institutions such as the Internet Governance Forum, offers an imperfect but instructive precedent for how AI governance bodies might formally incorporate non-state and non-corporate voices, particularly in the design of algorithmic auditing standards and in ongoing review of high-risk AI system classifications under frameworks such as the EU AI Act.

7. Illustrative Applications

7.1 Algorithmic Decision-Making in Public Welfare Systems

Automated eligibility-determination systems deployed in public welfare administration illustrate the due-process and dignity stakes of AI governance particularly starkly, since erroneous or opaque automated denials of essential benefits can produce severe material harm to already vulnerable populations, reinforcing the case for mandatory human review and accessible explanation in high-stakes public-sector AI applications.

7.2 Predictive Policing and Criminal Justice

AI-driven risk-assessment tools used in policing and sentencing contexts have drawn sustained

[29]Regulation (EU) 2024/1689, Art. 57.

[30]Cihon, P., Maas, M. M., & Kemp, L. (2020). Fragmentation and the future: Investigating architectures for international AI governance. Global Policy, 11(5), 545–556.

criticism for reproducing racially disparate patterns present in historical crime data, illustrating the equity concerns discussed in Section 3.3 and underscoring the necessity of independent auditing mechanisms discussed in Section 6.2.

7.3 AI in Educational Assessment

The use of automated systems for student assessment and university admissions decisions raises human-centred design concerns analogous to those in welfare administration, particularly where such systems are deployed with limited transparency to the students whose educational trajectories are directly affected by algorithmic outputs.

7.4 Biometric Surveillance and Facial Recognition in Public Spaces

The deployment of AI-driven facial recognition and biometric surveillance systems in public spaces illustrates several of the digital rights concerns examined throughout this paper simultaneously: privacy, since biometric identifiers are collected and processed at population scale without individualized consent; non-discrimination, since facial recognition systems have been repeatedly documented to exhibit substantially higher error rates for women and people with darker skin tones; and collective and solidarity-based rights, since the harms of ambient biometric surveillance accrue to entire communities present in a monitored space rather than to isolated individuals.[31] The EU AI Act's near-categorical prohibition on real-time remote biometric identification in publicly accessible spaces for law enforcement purposes, subject to narrow exceptions, reflects one regulatory attempt to resolve this tension in favor of dignity and privacy protections, while several other jurisdictions continue to permit substantially broader biometric surveillance deployment, illustrating comparative governance divergence discussed in Section 5.

8. Discussion: Tensions Between Innovation, Sovereignty, and Rights

Governments pursuing rapid AI-driven economic growth sometimes frame stringent rights-based regulation as a competitive disadvantage relative to jurisdictions with lighter-touch governance regimes, a tension visible in ongoing debates over the compliance costs associated

[31]Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of the Conference on Fairness, Accountability, and Transparency, 81, 77–91.

with the EU AI Act's risk-based obligations.[32]

At the same time, national AI sovereignty concerns -- the desire of individual states to develop independent AI capacity rather than depend entirely on foreign technology providers -- can pull in a different direction from harmonized international human rights standards, complicating the prospects for the kind of coordinated global governance that a genuinely human-centred AI society may ultimately require. This paper does not resolve these tensions but argues that the building blocks proposed in Section 6 -- participatory design, independent auditing, digital literacy, and dedicated oversight institutions -- are largely compatible with a wide range of national economic strategies, since they function primarily as accountability infrastructure rather than as blanket restrictions on AI development.

A further tension concerns the pace of technological change relative to the pace of legal and institutional reform. Regulatory sandboxes and ombudsperson offices require time to establish and staff, while AI applications are deployed and iterated on timescales of months rather than years, risking institutional oversight that perpetually lags behind the systems it is meant to govern.[33]

9. Limitations and Directions for Future Research

This paper's analysis is comparative and doctrinal rather than empirical, and does not measure the real-world effectiveness of the institutional mechanisms proposed in Section 6 as they are only beginning to be implemented across various jurisdictions. Future research should evaluate the operation of early AI ombudsperson offices and regulatory sandboxes as empirical evidence accumulates, and should examine how participatory design methodologies can be scaled beyond small pilot projects to large, high-stakes public-sector AI deployments.

Further work is also needed on the political economy of AI governance harmonization, particularly on whether the divergent priorities reflected in EU, U.S., Indian, African Union, and ASEAN approaches surveyed in Sections 5.4 and 5.5 can be reconciled within a coherent international framework, or whether a human-centred AI society will instead be realized, if at all, through a patchwork of regionally distinct governance regimes with varying levels of

[32]Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a 'right to explanation'. AI Magazine, 38(3), 50–57.
[33]Cihon, P., Maas, M. M., & Kemp, L. (2020). Fragmentation and the future: Investigating architectures for international AI governance. Global Policy, 11(5), 545–556.

enforceability.

10. Conclusion

Building a human-centred AI society is not a matter of adopting a single law, standard, or technical fix, but of sustaining alignment across legal frameworks, institutional design, and everyday technical practice around the protection of human dignity, autonomy, and equity. This paper has argued that existing digital rights -- privacy, non-discrimination, due process, and labor protections -- are being actively reshaped by the scale and opacity of AI systems, and that the emerging global governance landscape, from the EU AI Act to the Council of Europe's Framework Convention to the UNESCO and OECD principles, reflects a genuine, if uneven, convergence toward rights-respecting AI governance.

Realizing that convergence in practice will depend on the institutional and design-level building blocks examined in Section 6: participatory design processes that give affected communities genuine voice, independent algorithmic auditing that verifies compliance beyond self-assessment, sustained investment in public digital literacy, and dedicated oversight institutions capable of translating abstract rights commitments into accessible, enforceable protection. A human-centred AI society, in this sense, is less a fixed destination than an ongoing institutional commitment to keep human dignity at the center of an increasingly automated world.

References

Access Now. (2018). Human Rights in the Age of Artificial Intelligence. Access Now Publications.

African Union. (2024). Continental Artificial Intelligence Strategy. African Union Development Agency.

Association of Southeast Asian Nations. (2024). ASEAN Guide on AI Governance and Ethics. ASEAN Secretariat.

Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Machine Learning: Limitations and Opportunities. fairmlbook.org.

Birhane, A. (2020). Algorithmic colonization of Africa. Scripted, 17(2), 389–409.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of the Conference on Fairness, Accountability, and Transparency, 81, 77–91.

Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. Philosophical Transactions of the Royal Society A, 376(2133), 20180080.

Cihon, P., Maas, M. M., & Kemp, L. (2020). Fragmentation and the future: Investigating architectures for international AI governance. Global Policy, 11(5), 545–556.

Council of Europe. (2024). Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, CETS No. 225.

Crawford, K. (2021). Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press.

Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI.

Berkman Klein Center Research Publication No. 2020-1.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1).

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change, 114, 254–280.

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a 'right to explanation'. AI Magazine, 38(3), 50–57.

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines, 30(1), 99–120.

International Labour Organization. (2023). World Employment and Social Outlook: The Value of Essential Work -- With a Special Focus on Generative AI. ILO Publications.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399.

Latonero, M. (2018). Governing artificial intelligence: Upholding human rights & dignity. Data & Society Research Institute.

Llansó, E. J. (2020). No amount of "AI" in content moderation will solve filtering's prior-restraint problem. Big Data & Society, 7(1), 1–6.

NITI Aayog. (2021). Responsible AI for All: Approach Document for India. Government of India.

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press.

Organisation for Economic Co-operation and Development. (2019). Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Art. 14.

Regulation (EU) 2024/1689, Art. 57.

Regulation (EU) 2024/1689, Arts. 5–6.

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 (pp. 59–68). ACM.

Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. Big Data & Society, 4(2), 1–14.

The White House Office of Science and Technology Policy. (2022). Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People.

UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. UNESCO Publishing.

Universal Declaration of Human Rights, G.A. Res. 217 (III) A, U.N. Doc. A/RES/217(III) (Dec. 10, 1948).

Wagner, B. (2018). Ethics as an escape from regulation: From ethics-washing to ethics-shopping? In E. Bayamlioglu, I. Baraliuc, L. A. W. Janssens, & M. Hildebrandt (Eds.), Being Profiled: Cogitas Ergo Sum (pp. 84–89). Amsterdam University Press.