

AI & EQUALITY INPUT COMMENT

To OHCHR Call: The Use of Artificial Intelligence and the UN Guiding Principles on Business and Human Rights

We present our answers for three of the posed questions (blue). Details of the authors and relevant references can be found at the end of the document (not in word count).

What do you consider are the main human rights risks linked to the procurement and deployment of AI systems by States and in which area?

We will detail six key risks.

1. Risk of Procuring the Wrong Types of Systems

Whilst AI systems are increasingly deployed in the delivery of government services, their objectives often have a primary focus on detecting fraud (AlgorithmWatch, 2020; Booth, 2024) or assigning risk scores to individuals or households (Kawakami et al., 2022; Cheng et al., 2022). If governments maintain their legitimacy in attempts to provide the best possible service for their citizens, this focus seems out of place. Instead, we might now be able to focus on providing more effective and easier to access welfare services. This requires a paradigm shift to a needs-based functionality for citizens and frontline workers, for example, moving away from an initial conceptual focus on technical capabilities and fraud detection. Which issues do these workers encounter? Can AI systems contribute to an effective solution to address their needs? At which level and area of AI deployment are workers and citizens supported, without risk of labour displacement (i.e. level of human-in-the-loop, and human agency thereby)?

Only after these needs have been identified, should governments then seek out potential suppliers of the identified technologies. We hope that governments focus on deploying AI technologies for the enhancement of their citizens' human rights - instead of mere focus on not infringing rights. We see a real opportunity to create more inclusive, citizen-led, and participatory development practices and thus technologies that elevate and support the human dignity of the citizens they serve.

2. Risk of Losing Digital Sovereignty

We are concerned about the further monopolization of AI infrastructure, i.e., that a small number of suppliers create and distribute both hardware and software required for AI deployment. Paired with a lack of participation of diverse stakeholders in digital

infrastructure development, this further concentrates power - ultimately risking the digital sovereignty of governments (OECD, 2024; van der Vlist et al., 2024).

We additionally want to stress this risk for systems connected to foreign aid in particular, including UN programmes and humanitarian aid: it is likely that systems with which the donor is more familiar with or that are produced by a manufacturer from the donor country are preferred in the procurement process. This might lead to systems that are not well suited for the local context (see chatbot Karim example below), compete with or disincentivise local startups / projects, and create further dependency on foreign technology.

There are different priorities around the world underpinning digital public infrastructure and digital sovereignty strategies (see e.g. Eaves et al., 2024). International and regional cooperation aimed at establishing public digital infrastructure and technology stacks with shared goals and values could help to pool resources and offer public interest alternatives to private vertically integrated offerings.

3. Risk of Perpetuating Bias

Certainly, a well-established risk is the potential of AI systems to perpetuate historical biases and marginalizations of specific communities. Instead of repeating these insights, we want to add further nuance to this discourse. First, the current focus of AI systems in welfare that focus on fraud detection pose an increased risk of perpetuating biases for historically marginalised communities, e.g. due to biased data collection (e.g., increased policing of some postcodes increase the rate with which criminal acts from specific populations are discovered, see Zilka et al., 2023).

Additionally, procurement processes themselves have the risk to perpetuate who supplies technologies: if procurement processes require substantial financial resources or infrastructure, smaller local suppliers of technology might be unable to complete the required steps. By no means do we want to argue for more loose procurement processes, however, we urge for a focus on early community engagement instead of computing- and resource-intensive technical evaluations.

4. Risk of Reduced Transparency or Training to Conduct Sufficiently Rigorous Procurements

We are concerned that the employees procuring or deploying AI systems have insufficient expertise of the capabilities and/or risks of AI systems. The lack of an in-depth understanding of the general risks of AI systems, the interactions with their deployment contexts more generally, as well as the importance of involving affected communities, potentially false claims of suppliers and the risks an AI system might pose for direct system users and relevant citizens cannot be uncovered (Casilli et al., 2024). Thus, unsafe AI systems might unintentionally be procured or deployed (related to our concerns in question 2, i.e., the deployment of 3rd party systems by companies, e.g., AI-based recruitment systems).

5. Lack of Accountability and Required Legal Frameworks.

Further - spanning all aspects of Human Rights - we are concerned about the lack of effective redressal mechanisms. This includes both the detection of (e.g., discrimination), as well as the taking of action to compensate for the resulting harm after its detection. Currently, it is often impossible for citizens to know whether and how AI systems are used (Hickok & Hu, 2024). As a result, they cannot challenge, contest, or reverse AI-based decisions that impact their lives. We advocate for more clear legal frameworks that enable accountabilities and the citizens' right to know!

6. Data Privacy & Security

In the context of government procurement of AI systems, we are especially concerned about AI use for surveillance, monitoring, and unsolicited data collection. With potential deployments in contexts such as welfare, policing, healthcare, or transport, such systems pose significant risks to data privacy and security. We are specifically sceptical towards the linkage of unique identification IDs (based on biometric data and facial recognition techniques) to welfare services, the distribution of free goods, or in other ways to the access of citizen services as currently happening in India (see following references below: Online Indians, 2022; Business Standard, 2022; Bhatia & Bhabha, 2017).

The potential misuse or misrepresentation of this data could have severe consequences for individuals, including wrongful denial of services, discrimination, or even surveillance-based harassment. These risks are exacerbated when citizens are forced to engage with systems that collect and store personal data, without the ability to opt-out or safeguard their privacy. These systems could cause significant harm and violate fundamental rights such as privacy, equality, and freedom from coercion (Prasanna et al., 2023). We strongly question whether the benefits of implementing such systems can outweigh potential disadvantages.

JOINT ANSWER:

Are there any policies, regulations or frameworks taken at the national, regional and international levels to address the human rights risks linked to the procurement and/or deployment of AI by States?

AND Are there any emerging positive business practices that include human rights requirements when procuring and deploying AI? Please provide examples.

We perceive the following to be examples of beneficial practices to reduce risks to human rights or even help to promote their fulfilment. These are useful across Governments and businesses, why we pulled both questions together.

Across all examples, we want to stress the ecosystem of law already in place that might be applied to procuring and deploying AI. EU Laws such as the Digital Services Act, General Data Protection Regulation, Product Liability Directive, and even the Corporate Sustainability Reporting Directive as examples already create an ecosystem that can be harnessed. Instead of reinventing the wheel, we must close the patches between existing regulations or enable their policing in the AI sphere.

1. Requiring the consideration of Human Rights from the outset of system development.

We believe that Human Rights-respecting AI cannot start at the procurement stage. Instead, it must start at the very outset of development, i.e. when a system's objective is selected. States and businesses could require proof from suppliers that they created AI systems following a Human Rights-based approach. An example of such an approach is our AI & Equality Framework. For each stage of the development AI lifecycle, we list essential reflection points and actions (and often entire mindsets) that are required to create AI technology to enhance Human Dignity. [Here is a link to our community](#) platform of 450+ researchers from 57 countries.

2. Banning Specific Technologies

Facial Recognition Bans in Certain Cities: Some local governments (e.g., San Francisco, Boston) have banned the use of facial recognition technology by government agencies to prevent surveillance and protect citizens' rights to privacy.

3. Gender-responsive procurement (GRP)

There is a growing and settled body of research in the analogue environment that defines benefits and details specific means to achieve gender-responsive procurement (GRP) defined as “sustainable selection of services, goods, and works that considers impact on gender equality going beyond cost management, leveraging purchasing and sourcing opportunities to support social and economic progress at all stages and tiers of procurement and the supply chain” (UN Women). This research and guidance for organizations can and should be extrapolated into procurement frameworks that address the digital realm with hard targets outlining roles and responsibilities required to apply these principles.

The [World Bank Group's](#) gender lens on public-private partnerships is specific to physical infrastructure, yet is immensely transferable to digital public infrastructure (DPI) and other digital initiatives. It is imperative to not reinvent the wheel and lose hard-learned lessons from the analogue. Additional strong GRP frameworks come from the [International Labour Organization](#); [International Trade Centre](#); [Chatham House](#); [UN OPS](#); [UN Women WEPS Women's Empowerment Principles](#) Assessment Tool. According to the 2023 Women's Empowerment Principles Gender Gap Analysis Tool (focussed on the analogue), only 4% of companies set procurement targets for female-owned businesses and less than 5% provide training to women to compete for formal bid

processes. New industries and innovation could be jump-started by catalysing GRP in the digital realm.

4. Involvement of affected communities and domain experts, see next question below.

At a government level, this should include public consultation for AI policy and regulations such as this one.

5. Thorough Human Rights Impact Assessments.

Such assessments are soon required by the EU AI Act or the US's Algorithmic Accountability Act that seeks to require companies to assess the impact of their algorithms on privacy, security, and fairness, and to mitigate discriminatory effects, including biases in AI systems. Such assessments are further incentivised by the European Convention on Human Rights: it clarifies that AI systems used by state actors are subject to the human rights protections guaranteed under this convention, such as the right to privacy (Article 8), non-discrimination (Article 14), and the prohibition of inhuman or degrading treatment (Article 3). We consider the *Human Rights, Democracy, and the Rule of Law Impact Assessment for AI Systems* ([HUDERIA](#)) currently under development by the Alan Turing Institute (2022) to be one of the most in-depth examples. The assessment helps to anticipate risks and harms to Human Rights in collaboration with affected communities and other stakeholders. In our own AI & Equality framework, we integrate this approach throughout the AI development cycle to enable technology creation that has Human Rights at its core from the conception of the technology. More in the next point.

How can businesses and States meaningfully engage with relevant stakeholders, including potentially affected right holders and workers, to identify and address adverse human rights impacts related to the procurement and deployment of AI? Please provide examples.

Stakeholder Involvement has the ability to localise abstract responsible AI principles and help to make them concrete and actionable (Sadek & Kallina, 2024). Additionally, they are a tool to re-balance the decision power of which AI systems are developed and how - a power that is currently skewed towards large tech companies and away from the communities that will be affected by resulting systems (Sadek & Kallina, 2024; Kallina, 2024). The following presents actions that can help to harness these benefits for the development of AI systems.

1. Involving affected communities in the development of new technologies.

Again, here we want to point towards the HUDEIRA of the Alan Turing Institute and its integration into our AI & Equality framework (see 4.) in question above). However, we believe that stakeholder engagement has to go beyond consultations during the

assessment of harms prior to deployment only. Instead, we believe that the creation of AI should start with the communities that it is meant to serve. Instead of the small group of people that creates or invests in technology projects, we should open the discourse about what type of AI systems should be built - and which not - to the affected publics. Concrete guidance on how to do this can be found in our AI & Equality online course and framework as well as in Kallina & Singh (2024, referenced below). Examples of actions that governments can adopt more broadly are inclusive outreach actions (see e.g. Wang & Liang, 2024) or Human Rights Impact Assessments as discussed above (see also Kolfshoeten & Shachar, 2023).

2. Enabling Stakeholder involvement after deployment.

Companies and especially governments should enable affected communities to contest AI-based decisions that impact their lives or are performed on their data. This includes the *right to know*: only if one is aware that AI is used, one is able to criticise it - and thus contest, circumvent, or reverse AI-based decisions, or to search for redress and compensation. In addition to the right to know, clear accountability structures require clear avenues for feedback and complaints, e.g. through a button, online form, or formalised process. These avenues must be accessible, easy to find, and highly usable to truly allow for stakeholder interventions.

3. Open dialogues through international forums

Such forums would create increased awareness on possible impacts of AI systems and the risks to human rights. Further, these endeavours would be supported by opportunities for upskilling and reskilling more broadly as they pertain to lifelong learning to make such technology more accessible to the global community and workforce with regard for human rights, dignity, and the ethical impact of such tools. This OHCHR consultation could become a great example of such. This can include and lead to **public consultation** for AI policy and regulations, so that citizens can transform their knowledge and awareness into actions.

4. Grants enabling independent research on the impact of technologies, as well as how to regulate and police them.

Examples are capacity building grants indirectly filling the skill gap for enhanced readiness assessments such as [this example](#).

5. Enabling and facilitating local technology creation, serving local needs.

We recommend a deliberate focus on international collaborations through technology transfers, open-source technologies etc. so that all regions are enabled to create technologies and tailor them for their local needs. Although localization became one of the most critical agendas after the 2016 World Humanitarian Summit (WHS), localization practices in the field and engaging local actors are still limited. For example, the AI chatbot Karim was launched to provide psychological support to Syrian refugees in Lebanon, a collaboration of the WHO and a company from San Francisco. Unfortunately,

the responses were not truly adapted to the context (Solon, 2016; Madianou, 2021). Testing and trying AI applications without meaningful, people-centered research with affected communities, can create harmful outcomes even if well-intended.

6. Multi-disciplinary teams procuring and deploying AI

Due to the above reasons, we require multi-disciplinary teams with diverse backgrounds and skill sets in the procurement and deployment of AI, and especially in such complex and/or vulnerable social situations. Examples are:

- **Humanitarian and Human Rights Experts:** Human Rights Lawyers, Humanitarian Organizations, Ethicists and Philosophers
 - **Affected and Targeted Groups (Lived Experience /communities):** Communities Affected by AI Systems (e.g., low-income, marginalized, or ethnic minority groups), Disability Advocacy Groups, Minority and Marginalized Communities, Workers and Labor Unions, Consumers and Citizens
 - **Governments and Policymakers:** Regulators, Policy Experts, International Organizations (e.g., UN, OECD, World Economic Forum)
 - **Academics and Researchers:** Sociologists and Social Scientists, Medical and Public Health Experts, Environmental Scientists, interdisciplinary experts
 - **Media and Communication Specialists:** Journalists and Media Organizations, Public Relations Experts
 - **Civil Society and Advocacy Groups:** Privacy Advocacy Groups (e.g., EFF, ACLU), Consumer Rights Organizations, Gender Equality and Social Justice Groups, Faith-Based Organizations
 - **Independent Oversight Bodies:** Independent Monitoring Bodies (e.g., AI ethics committees, watchdog organizations), Ombudsman Offices
 - **Psychologists and Behavioural Scientists:** Psychologists (Specialists in human behaviour, cognition, and emotions, working to understand how AI systems impact mental health, decision-making, and user interaction.)
 - **Legal Experts,** including intellectual property lawyers
-

AI & EQUALITY Authors

All authors are part of the AI & Equality Community (link to [platform](#) and [website](#))

Emma Kallina

4th year PhD student at the University of Cambridge, UK, focusing on stakeholder involvement for more responsible AI development in practice.

Caitlin Kraft-Buchman

CEO/Founder of Women At The Table, co-founder <A+> Alliance for Inclusive Algorithms, and leader of the <AI & Equality> Human Rights Toolbox initiative.

Alexis Speck Glennon

DSW, LCSW-R is a clinical social worker and social work educator who focuses on tech justice and impact of technology on society.

Geetika Bhatnagar

Marketing and Communications Professional and an Advocate for Goodtech, DEI, Sustainability, and Climate Action.

Özge Caglar

Psychologist and Humanitarian Aid Worker. Ongoing research on AI-driven technologies to support GBV survivors in a post-earthquake context in Türkiye.

Fawole John Oyeyemi

DBA Candidate at Liberty University, focusing on International Business Management and Contextual Adaptation of Global AI Ethics frameworks in Multinational Corporations.

Cinthya Leonor Vergara Silva

Management Control Engineer with a Master's in Business Engineering, currently pursuing a PhD in Data Science. Specializing in AI/ML project management, causal inference, pricing, explainable AI (UXAI) and fairness metrics, with a focus on developing transparent and fair AI systems. Member of a research Group for Trustworthy Systems at Universidad Adolfo Ibáñez.

Arthit Suriyawongkul

PhD candidate at ADAPT Centre, Trinity College Dublin, focusing on semantic technology to support AI accountability in the software supply chain.

Ann Borda

Ethics Fellow at The Alan Turing Institute (UK) involved in the development of public sector guidance on AI ethics, safety and governance, including a Human Rights Impact Assessment for public sector agencies.

Stephanie Lunn

Assistant Professor of Computing and Engineering Education at Florida International University, United States of America, who focuses on the ethics of AI systems and their potential impact on society.

IMPORTANT REFERENCES

AlgorithmWatch. 2020. How Dutch activists got an invasive fraud detection algorithm banned.

AlgorithmWatch on April 2020. Accessed on 04.12.2024 at

<https://algorithmwatch.org/en/syri-netherlands-algorithm/>

Anna Kawakami, Venkatesh Sivaraman, Hao-Fei Cheng, Logan Stapleton, Yanghui Cheng, Diana Qing, Adam Perer, Zhiwei Steven Wu, Haiyi Zhu, and Kenneth Holstein. 2022. Improving Human-AI Partnerships in Child Welfare: Understanding Worker Practices, Challenges, and Desires for Algorithmic Decision Support. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 52, 1-18. <https://doi.org/10.1145/3491102.3517439>

Bhatia, A., & Bhabha, J. (2017). India's Aadhaar scheme and the promise of inclusive social protection. *Oxford Development Studies*, 45(1), 64-79. <https://www.tandfonline.com/doi/abs/10.1080/13600818.2016.1263726>

Booth, Robert for The Guardian. 2024. DWP algorithm wrongly flags 200,000 people for possible fraud and error. *The Guardian* June 2024, accessed on 04.12.2024 at

<https://www.theguardian.com/society/article/2024/jun/23/dwp-algorithm-wrongly-flags-200000-people-possible-fraud-error>

Business Standard Web Team (2022), Aadhaar number mandatory for availing of govt subsidies and benefits, accessed on 06.01.25 at

https://www.business-standard.com/article/current-affairs/aadhaar-number-mandatory-for-availing-government-subsidies-benefits-122081601102_1.html

Casilli, A. A., Tubaro, P., Cornet, M., Ludec, C. L., Torres-Cierpe, J., & Braz, M. V. (2024). Global Inequalities in the Production of Artificial Intelligence: A Four-Country Study on Data Work.

Cui, R., Li, J., & Zhang, D. J. (2020). Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on Airbnb. *Management Science*, 66(3), 1071-1094.

Eaves, D., Mazzucato, M. and Vasconcellos, B. (2024). Digital public infrastructure and public value: What is 'public' about DPI? UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2024-05). Available at: <https://www.ucl.ac.uk/bartlett/public-purpose/>

Hao-Fei Cheng, Logan Stapleton, Anna Kawakami, Venkatesh Sivaraman, Yanghui Cheng, Diana Qing, Adam Perer, Kenneth Holstein, Zhiwei Steven Wu, and Haiyi Zhu. 2022. How Child Welfare Workers Reduce Racial Disparities in Algorithmic Decisions. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 162, 1-22. <https://doi.org/10.1145/3491102.3501831>

Hickok, Merve, and Evanna Hu. "Don't Let Governments Buy AI Systems That Ignore Human Rights." *Issues in Science and Technology* 40, no. 3 (Spring 2024): 37-41 <https://doi.org/10.58875/ACOG7116>

Kallina, E., & Singh, J. (2024, October). Stakeholder Involvement for Responsible AI Development: A Process Framework. In Proceedings of the 4th ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (pp. 1-14). <https://dl.acm.org/doi/abs/10.1145/3689904.3694698>

Leslie, D., Burr, C., Aitken, M., Katell, M., Briggs, M., & Rincon, C. (2022). Human rights, democracy, and the rule of law assurance framework for AI systems: A proposal. <https://arxiv.org/abs/2202.02776> - see also <https://rm.coe.int/cai-2024-16rev2-methodology-for-the-risk-and-impact-assessment-of-arti/1680b2a09f>

Madianou, M. (2021). Nonhuman humanitarianism: when 'AI for good' can be harmful. *Information, Communication & Society*, 24(6), 850-868.

Miri Zilka, Riccardo Fogliato, Jiri Hron, Bradley Butcher, Carolyn Ashurst, and Adrian Weller. 2023. The Progression of Disparities within the Criminal Justice System: Differential Enforcement and Risk Assessment Instruments. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23). Association for Computing Machinery, New York, NY, USA, 1553–1569.

<https://doi.org/10.1145/3593013.3594099>

OECD (2024), “AI, data governance and privacy: Synergies and areas of international co-operation”, OECD Artificial Intelligence Papers, No. 22, OECD Publishing, Paris, <https://doi.org/10.1787/2476b1a4-en>.

Online Indians (2022), Aadhaar Made Mandatory for 15 Government Schemes, accessed on 06.01.25 at <https://www.onlineindians.in/aadhaar-mandatory/>

Pozzi, G., & De Proost, M. (2024). Keeping an AI on the mental health of vulnerable populations: reflections on the potential for participatory injustice. *AI and Ethics*, 1-11.

Prasanna, S., Ameen, A., Saurav, K., Singh, D. B., Tripathy, A., Sharma, P., ... & Sanjaya, K. (2023). Empowerment And Equality Navigating Human Rights Law In A Complex World. Institute of Legal Education. <https://www.researchgate.net/publication/377466246>

Sadek, M., Kallina, E., Bohné, T., Mougnot, C., Calvo, R. A., & Cave, S. (2024). Challenges of responsible AI in practice: scoping review and recommended actions. *AI & SOCIETY*, 1-17. <https://link.springer.com/article/10.1007/s00146-024-01880-9>

Solon, O. (2016). Karim the AI delivers psychological support to Syrian refugees. *The Guardian*, accessed on 06.01.25 at <https://www.theguardian.com/technology/2016/mar/22/karim-the-ai-delivers-psychological-support-to-syrian-refugees>

UN Women. Gender-responsive procurement, accessed on 06.01.2025 at <https://www.unwomen.org/en/about-us/procurement/gender-responsive-procurement>

van der Vlist, F., Helmond, A., & Ferrari, F. (2024). Big AI: Cloud infrastructure dependence and the industrialisation of artificial intelligence. *Big Data & Society*, 11(1). <https://doi.org/10.1177/20539517241232630>

van Kolschooten, H., & Shachar, C. (2023). The Council of Europe's AI Convention (2023–2024): Promises and pitfalls for health protection. *Health Policy*, 138, <https://doi.org/10.1016/j.healthpol.2023.104935>

Wang, S., & Liang, Z. (2024). What does the public think about artificial intelligence? An investigation of technological frames in different technological contexts. *Government Information Quarterly*, 41(2). <https://doi.org/10.1016/j.giq.2024.101939>

Zhang, S., Mehta, N., Singh, P. V., & Srinivasan, K. (2021). Frontiers: Can an artificial intelligence algorithm mitigate racial economic inequality? An analysis in the context of Airbnb. *Marketing Science*, 40(5), 813-820.