ELEMENT ^AI

Submission to the Australian Human Rights Commission's "Human Rights and Technology Discussion Paper

Background

Element AI is a global artificial intelligence (AI) solutions provider headquartered in Montreal, Canada, with offices in Toronto, London, Singapore and Seoul. From its inception, Element AI's founders and executive leadership have been committed to assisting international and national regulatory efforts undertaken to ensure the design, development and deployment of AI is safe, lawful, ethical, and, importantly, grounded in human rights.

To this end, representatives from Element AI have participated in several important international policy-making efforts related to AI, including the OECD's *AI Principles* and the European Commission's *Ethics Guidelines for Trustworthy AI*. More recently, Element AI convened a multi-stakeholder workshop attended by a group of international experts to develop practical recommendations on using the international human rights law framework to govern AI, leading to the publication of a report, entitled "[Closing the Human Rights Gap in AI Governance](#)". Many of the report's conclusions and recommendations are supportive of the principles and proposals outlined in the Australian Human Rights Commission's discussion paper, and have been extensively relied on in this submission.

Element AI would like to express its broad support for the work of the Australian Human Rights Commission's Human Rights and Technology Project as well as the discussion paper with respect to it has invited comments. The international human rights law framework provides the normative foundation for the governance of emerging technologies, such as AI, and should be used to guide the evolution of any national legislative, regulatory or policy reform aimed at ensuring its responsible use. We strongly encourage government institutions in Australia and in other countries to take up the findings and proposals contained in this discussion paper, and to apply them to their own national contexts.

In particular, we wish to express supposed for Proposals 1, 5, 7, 9, 12, 14 and 19 of the discussion paper, with respect to which we make the following additional comments and observations.

Comments on Proposals

Proposal 1: The Australian Government should develop a National Strategy on New and Emerging Technologies. This National Strategy should: (a) set the national aim of promoting responsible innovation and protecting human rights (b) prioritise and resource national leadership on AI (c) promote effective regulation— this includes law, coregulation and self-regulation (d) resource education and training for government, industry and civil society.

In addition to the above, we recommend that national AI strategies embed a human rights-based approach into industrial policy, for instance through the creation of tailored direct spending programs to help ensure that the design and technological foundations of rights-respecting AI, such as transparency, explainability and accountability, are firmly established in key sectors.

To these ends, national AI strategies should adopt policies and programs designed to:

- incentivize research and development for work on responsible AI (e.g., explainability) in the private sector through tailored direct spending programs, or other financial incentives;
- ensure that procurement of AI systems by public institutions place a strong emphasis on technological and design solutions, in particular, to promote transparency, explainability and accountability of AI systems;
- provide for new discretionary spending programs to help expand existing research initiatives rooted in privacy by design, to human rights by design; and
- invest in action-oriented research and tools to assist human rights by design policies and practices.

Proposal 5: The Australian Government should introduce legislation to require that an individual is informed where AI is materially used in a decision that has a legal, or similarly significant, effect on the individual's rights.

and

Proposal 7: The Australian Government should introduce legislation regarding the explainability of AI-informed decision making. This legislation should make clear that, if an individual would have been entitled to an explanation of the decision were it not made using AI, the individual should be able to demand: (a) a non-technical explanation of the AI-informed decision, which would be comprehensible by a lay person, and (b) a technical explanation of the AI-informed decision that can be assessed and validated by a person with relevant technical expertise.

The human rights framework highlights that the transparency, explainability and accountability of AI systems is not only a desirable ethical achievement, but a

prerequisite to upholding existing legal obligations. For instance, the lack of transparency of the use and operation of such systems can make it difficult to determine whether a breach of human rights has occurred, pre-empting the ability to seek redress. Similarly, the reliance on a recommendation, decision or insight provided by an AI system that is not explainable is at odds with human rights and administrative law, which in many cases require that an individual be provided with reasons for a decision made against them; the transparency and explainability of AI systems is therefore a legal prerequisite to their accountability.

Governments and businesses have a responsibility to ensure that the circumstances around the use of an AI system to inform such decisions are clearly communicated to affected individuals. Governments should clarify that a general right to an explanation already exists when an AI system informs a decision that has a significant effect on a person's rights, financial interests, personal health or well-being. This can be accomplished most effectively by legislation as proposed.

Proposal 12: Any standards applicable in Australia relating to AI-informed decision making should incorporate guidance on human rights compliance.

and

Proposal 14: The Australian Government should develop a human rights impact assessment tool for AI-informed decision making, and associated guidance for its use, in consultation with regulatory, industry and civil society bodies. Any 'toolkit for ethical AI' endorsed by the Australian Government, and any legislative framework or guidance, should expressly include a human rights impact assessment.

In light of the human rights risks posed by AI systems, governments should begin a phased-approach to making human rights due diligence (HRDD) and human rights impact assessments (HRIA) a regulatory requirement. First, governments should support the development of model frameworks for HRDD and HRIA that identify what factors must be included in these types of assessments (e.g. adequate consequences if identified risks of adverse human rights impacts are not duly mitigated and addressed). The creation of model frameworks for HRDD and HRIA could help mitigate the risk of them becoming rubber-stamping exercises, and would be necessary to facilitate uptake and implementation. Second, because of the context specific nature of these types of assessments, model HRDD and HRIA would also need to be adapted to the particular sectors in which AI is being deployed.

Third, while the clearest case for requiring HRDD and HRIA is in the context of the procurement of AI systems by public institutions, including all levels of government and the courts, a risk-based approach should guide their use in the private sector. Such an approach should be scalable and proportionate to the risk profile of the particular industry and application of the technology. In particular, implementation in the private

sector should focus on high-risk applications of AI or instances where human rights violations, such as discrimination, could have a significant effect on a person's financial interests, personal health or well-being, or where minors are affected. As such, the use of AI systems in financial services, the education system, human resources and healthcare should be considered as priority areas. While risk assessments can be used to help identify, control or minimize risks, they may also help identify instances in which the risks to human rights and people's well-being are simply too high, and lead to a decision to place a moratorium or prohibition on deployment.

The imposition of due diligence requirements or impact assessments is particularly appropriate to emerging technologies like AI, which can be a moving target for regulators. Rather than making the technology itself the object of regulation, this approach helps ensure companies develop a systematic and auditable way of controlling and minimizing risks, while allowing space for innovation. Any requirement to perform HRDD and HRIA grounded in international and domestic human rights law should be supplemented by reporting requirements and audits performed by an appropriate government entity. The European Union Directive 2014/95/EU on non-financial reporting, which requires the reporting of human rights impact assessments by large undertakings, could serve as a helpful example in this regard. Reporting requirements and audits of HRDD and HRIA could introduce greater transparency into how such assessments are conducted, incentivizing more meaningful accountability for technology companies.

As more technology companies perform HRIA and HRDD, standards should be developed by industry and professional bodies to further entrench respect for human rights in the design, development and deployment of AI. The IEEE included "human rights" as its first "General Principle" in guidance it produced Ethically Aligned Design, making explicit reference to the international human rights framework and the relevance of the UN Guiding Principles to the design, development and deployment of AI. Standard development organizations such as ISO, and national chapters, could adopt a human rights-based approach to the work currently being conducted in ISO/IEC JTC 1/SC 42 - Artificial intelligence, in particular, with respect to the development of an explainability framework for AI. Standards could eventually be incorporated by reference into legislation, enhancing accountability through statutory enforcement schemes.

Beyond helping to promote legal compliance, the business case for embedding HRDD grounded in the Universal Declaration of Human Rights through the UN Guiding Principles is also attractive. For instance, the UN Guiding Principles can help provide:

- A proactive risk management framework that merges legal, human rights and ethical considerations that captures a comprehensive portrait of potential harm (including risk of liability, as many universal human rights are incorporated into domestic legislation and regulation, e.g. privacy and discrimination).

- A realistic forecast of the true costs of the design, development and deployment of AI, by highlighting the technological imperatives of rights-respecting AI (e.g. human-centered design, transparency, explainability and accountability).

- Access to a growing pool of resources and guidance on due diligence best practices (including international jurisprudence, national case law, best practices in HRDD and HRIAs).

- An international framework that can be scaled across global operations; and

- In some cases, companies that have satisfied due diligence requirements may be able to raise these as a defense in the event of liability claims.

Proposal 19*: The Australian Government should establish an AI Safety Commissioner as an independent statutory office to take a national leadership role in the development and use of AI in Australia. The proposed AI Safety Commissioner should focus on preventing individual and community harm, and protecting and promoting human rights. The proposed AI Safety Commissioner should: (a) build the capacity of existing regulators and others regarding the development and use of AI (b) monitor the use of AI, and be a source of policy expertise in this area (c) be independent in its structure, operations and legislative mandate (d) be adequately resourced, wholly or primarily by the Australian Government (e) draw on diverse expertise and perspectives (f) determine issues of immediate concern that should form priorities and shape its own work.*

National AI strategies have committed substantial funds towards AI research and talent development to grow the AI sector; a capacity building effort is now needed to empower governments with the personnel and institutional arrangements needed to ensure that AI is designed, developed and deployed in a manner that protects human rights.

Government policymakers and regulators must be equipped with personnel capable of monitoring and assessing AI's domain-specific risks, especially regarding the rights, health and financial interests of those affected. As such, a dedicated capacity building effort within government is necessary to accelerate understanding of how the existing legislative and regulatory framework can be applied to ensure respect for human rights, and identify potential gaps where adjustments may be necessary, such as regarding the transparency, explainability and accountability of AI systems.

For instance, departments of justice must understand the risk of algorithmic bias in predictive policing and judicial-decision making; officials administering the public benefits system require expertise on algorithmic bias and discrimination; immigration, tax and revenue departments must ensure that the integration of AI respects human rights and principles by administrative law (which depend on the transparency, explainability and accountability of AI systems). The protection of internet users from the risk of online hate speech or the spread of disinformation requires expertise on internet governance, content moderation techniques and the freedom of opinion and expression.

Policy makers contemplating the reform of privacy legislation should understand how AI-driven methods of analysis on social media can be used to infer and generate sensitive information about people that they have neither provided nor confirmed, turning the consent-based approach to privacy protection on its head.

An independent AI Safety Commissioner could research, monitor, assess, report on, and provide advice to government and industry about risk management in AI. The commissioner should be an agile source of policy expertise, education and capacity building and could undertake the following tasks:

- support the development of Model Frameworks for human rights due diligence and human rights impact assessments, including sector-specific codes of conduct in priority sectors (as outlined above).

- undertake research and publish reports on the social impact and human rights risks posed by AI systems. Scholars, civil society and international organizations have published research outlining AI's human rights risks. The Commissioner's office should build on this work to provide society with a comprehensive overview of the human rights risks triggered by AI systems, and empower individuals with knowledge of the recourses available to them (e.g., under anti-discrimination law or human rights codes);

- ensure that departments and ministries understand how AI may affect their respective regulatory roles, notably through the human rights lens;

- issue notices or advisories on AI applications that pose high risks to human rights, in coordination with regulators. Similar to the function performed by aviation, tax or securities authorities, this type of action could help raise social awareness with respect to high risk uses of AI systems, expectations for compliance, while conditioning the market to the risk of liability;