Horizon Scanning Series

The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing

Regulation

This input paper was prepared by Olivia Erdélyi

Suggested Citation

Erdélyi, O (2018). Regulation. Input paper for the Horizon Scanning Project "The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing" on behalf of the Australian Council of Learned, www.acola.org.

Considerations on Regulating Artificial Intelligence in the Domestic and Global Context

Olivia J Erde´lyi
School of Law
University of Canterbury

August 5, 2018

1 Background

Artificial intelligence (AI) is becoming increasingly pervasive in society. On one side, various instantiations of AI bring breakthroughs in the solution of previously insurmountable problems. On the other side, however, it is important to be aware of the highly disruptive nature of these technologies and their potential to impact economic welfare and inequality, threaten our most basic freedoms, upset our fundamental, established social values, and overall destabilize the international community through what is commonly referred to as AI-race — a dangerous competition for technological superiority between different academic and industry stakeholders developing AI, as well as nations and regional groupings. Such a race poses a number of significant risks from encouraging corner-cutting on safety and governance considerations to creating and exacerbating conflict situations [2]. In order to optimally harness AI's benefits and address its potential risks — preferably proactively rather than retroactively and in a manner beneficial to all humanity — it is indispensable to develop adequate policies in relation to AI technologies at the earliest possible stage.

That said, regulating AI is a formidable task for at least three reasons: First, AI is surrounded by daunting uncertainties given that our capabilities to even understand the internal workings of existing AI technologies, let alone anticipate the ways in which they will evolve and impact humanity in the future are severely limited.

Second, the manifold challenges posed by the use of AI are individually complex and inextricably intertwined. The corollary of this observation is that high-level policy approaches taking due account of multidisciplinary imperatives and spanning distinct policy domains are needed to avoid unintended side-effects. As straightforward as it seems, this crucial consideration still tends to remain clouded from the siloed vision of various disciplines.

Third, and relatedly, the effective solution of AI-related challenges necessitates large-scale, previously unexperienced levels of collaboration both among different constituencies within nation states and internationally. On the domestic level, the need for collaboration is mostly grounded in the fact that for designing truly effective AI policies,

governments depend on the substantial and mutually complementing expertise of various industry and academic stakeholders. Moreover, in light of the far-reaching impact of such policies on all aspects of individual human lives, their political sustainability requires early involvement of and continuing support from the public.

International policy coordination, on the other hand, is imperative both from a legal and eco- nomic perspective. Legally speaking, internationally coordinated policy action is a quintessential prerequisite for ensuring the authority and legitimacy of the emerging body of law governing AI. The reason for this is that AI is a phenomenon with global impact, i.e., its regulation involves externalities transcending national boundaries. In such cases, domestic policymakers are hardly able and often also unwilling to control the effects of their actions in foreign jurisdictions. Equally importantly, isolated domestic policy initiatives tend to conflict with each other, significantly complicating cross-border interactions. Such problems — ultimately attributable to the discrepancy between the transnational nature of a problem and the national character of the law governing it — eventually create pressures for transnational regulation, which is subject to a set of complex, recursive, multi- directional processes that crucially affect norms authority [4].

From an economic perspective, aligning the scope of regulatory coordination with the reach of externalities is advisable to ensure that the proliferation of AI is welfare enhancing, rather than further aggravating already pronounced, worldwide problems of economic inequality, and hence enjoys as broad as possible social and political support [5].

These already powerful arguments are further compounded by AI's much more rapid pace of in- novation and considerably greater disruptive potential compared to previous waves of technological innovation.

Against this background, AI policy initiatives must be coordinated in consistent domestic and international regulatory frameworks to avoid conflicts through fragmentation and maximize efficiency. To date, unsurprisingly, such a regulatory framework is missing whether at national, regional, or international level. AI policies are developed by largely uncoordinated efforts of various academic and industry groupings as well as first government initiatives, the regulatory purviews of agencies involved in the process are not clearly delineated, issues of regulatory architecture design have not yet been addressed, and AI applications are at best sporadically regulated. The current AI governance landscape exhibits many academic and/or industry based non-governmental entities mostly working on a stand-alone basis, although sometimes in a more or less coordinated fashion. As AI continues to be deployed in safety-critical settings, policymakers around the globe will need to get involved and design suitable AI policies, preferably supported by the expertise of these organizations. The next two sections will outline proposals for AI policy coordination in both a global and domestic context.

2 A Global Approach to Regulating AI

The following recommendation on a possible global AI regulatory and governance framework is based on the paper of Erdélyi and Goldsmith [3] and the references cited

therein. Marchant and Wal- lach [6] cultivate a similar idea stressing the importance of coordinated ethical and legal oversight of emerging technologies through what they termed governance coordination committees (GCCs) established either on the national or international level as appropriate depending on the issue area addressed.

At the heart of Erde´lyi and Goldsmith's proposal is to establish a new intergovernmental organization (IGO) — possibly named International Artificial Intelligence Organization (IAIO) — which could serve as a forum for intergovernmental coordination and support national policymakers in the development of AI policies. The goal is to ensure internationally consistent AI policy approaches by directly engaging governments in policy debates before they lock in on particular and with all likelihood differing positions, which may lead to path dependencies, spark conflicts, and are difficult to renege without political damage. Ideally, the IAIO should complement and collaborate with the diverse array of non-governmental entities involved in AI research and development, so that common approaches are informed by their valuable expertise. Such collaboration could take a number of forms, including consultation processes, commissioning of expert opinions, and such stakeholders' formal involvement in the organization.

With a view to facilitate international consensus and best accommodate the uncertainties sur- rounding AI as well as its fast-paced development, the IAIO should start out as an informal agency relying on soft-law instruments. In this capacity, it could play an instrumental role in developing widely accepted, unified standards and best practices. Over time and given sufficient consensus in the international community, a move towards more formalized cooperation could be envisaged, potentially turning the IAIO into a formal organization with regulatory and perhaps also conflict resolution powers. Even though this latter prospect is probably rather remote, historically diverse institutional choices in other areas of international cooperation can serve as useful guides, suggesting that many different settings can be successful.

A further consideration — one potentially interesting from a New Zealand and Australian perspective — is that, similarly to many other key IGOs, the IAIO should be hosted by a country widely considered as neutral in order to provide for a safe environment, limiting avenues for political conflict and encouraging a climate of mutual tolerance and appreciation. Apart from Switzerland, which is already home to many important organizations, New Zealand and Australia are also very good candidates for such a role given their relatively small size and their amicable relationship to other countries. Hence, the New Zealand and Australian Governments — either jointly or separately — should look into whether they are interested in steering the global AI debate by serving as host. If yes, one way to move forward could be to convene an initial, informal government-level conference to discuss the idea.

3 Domestic AI-Regulatory Frameworks

With the increasing ubiquity of AI-enabled systems in a growing number of critical domains — such as military, health care, financial services, criminal justice system, and

transportation, to mention just a few examples — adequate domestic AI policies to facilitate the transition of Australia and New Zealand in an AI-driven economy are urgently needed. Given the vast amount of work and specialized expertise needed to formulate sustainable AI policies across diverse policy domains, however, it is unrealistic to expect the Australian and New Zealand governments (or that of any other country for that matter) to rise up to this challenge on their own. At the same time, numerous academic and industry stakeholders invested with AI research and development in dispersed areas possess exactly the kind of invaluable expertise governments need to inform policy initiatives.

What is more, recognizing society's shared responsibility to shape AI mindful of the critical impact of our decisions on the life of future generations, many of these parties are also very keen on getting involved in policy debates and supporting government work as best they can. Therefore, to over- come the practical impediments stemming from governments' limited regulatory capacities and to capitalize on this climate of enthusiasm before it lapses, it is suggested that governments approach AI regulation adopting strategies that exhibit some degree of self-regulation. Doing so would have the auxiliary benefit, that governments would, at least to some extent, de facto collaborate with leading tech companies (Amazon, Apple, Google, IBM, Microsoft, and their piers) in the design and implementation of AI policies. This, in turn would send the general public an important signal, counterbalancing the increasingly present and latently hostile rhetoric warning against a dangerous concentration of power in the hands of a handful of tech giants, which admittedly dominate AI research and development given that their capacities by far outweigh that of any government or academic institution. While such fears have to be addressed with due caution, it would be a mistake to view these firms solely as a threat to society that has to be controlled at any cost. Apart from the fact that complete control over these powerful parties is unrealistic, this would lead to a mutually detrimental, adversarial relationship between tech firms on one side, and governments and society on the other side, ignoring that, with appropriate safeguards in place, society depends on these firms to drive technological innovation.

Such an approach would also be in line with the prevailing decentred understanding of regulation, which envisages a radically changed government involvement in regulatory and governance processes. The essence of that understanding is the recognition that due to the complexity of mod- ern policy and governance matters, the state no longer has the ability to exercise power and control over all segments of society and, consequently, should not have the monopoly to do so. Rather regulation is a process of convoluted multi-stakeholder interactions, in which both the construction of knowledge and the exercise of power is shared between different autonomous social actors and government stakeholders, who are mutually interdependent co-producers of regulation. In such an environment, where no single party has the knowledge and power to effectively solve and regulate complex and continuously changing problems, the state is but one of many actors responsible for regulation and governance. Accordingly, regulatory strategies should be hybrid, i.e., unite govern- mental and non-governmental parties, multi-faceted, i.e., consist of a dynamically changing mix of strategies, and indirect, i.e., a flexible and sensitive process of steering, coordinating, balancing, and influencing with a view to

create patterns of interaction that best serve the public interest. This dynamic responsiveness and deliberate diversity is also key to the success of self-regulatory arrangements and to preventing them from becoming nothing more than disappointing mirrors of traditional state-driven command and control regulation, i.e., rules backed by sanctions, displaced to another location. For a good overview of the relevant regulatory literature, see [1].

Regarding the possible institutional configuration of an AI self-regulatory framework: In New Zealand, the Artificial Intelligence Forum of New Zealand (AIFNZ) is an organization uniting industry, academia, and government with the overarching objective to secure New Zealand's future prosperity in the impending AI age. To this end, the AIFNZ works toward raising the level of awareness and capabilities of AI in New Zealand and contributing to the social and political debate on AI's broader implications for society. It is member of and closely works together with the Partnership on AI (PAI) - an international industry consortium established to study and formulate best practices on AI technologies, to advance the publics understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences on people and society. Being a conduit between all these national and international stakeholders, the AIFNZ is uniquely positioned to take on the role of a self-regulatory association (SRA) and contribute to shaping both domestic and international AI policies using industry insights and best practices. I am not currently aware of an equivalent organization in Australia, but see no major obstacles to establishing one, given consensus to do so.

Mindful of the superiority of an incremental approach, the SRA could start out with a limited mandate, merely as a forum for collaboration and information exchange between government, businesses, academia, and the general public. Industry could furnish valuable information on what firms would like to get out of AI, what their needs are (in terms of human capital, organizational optimization, etc.), what AI policies they see necessary, and back up policy considerations with firm-level data. Academia could tailor their research to service these needs — a somewhat less obvious yet vital way of boosting domestic economies, as in reality firms often grapple with problems for which researchers have readily available solutions. Businesses and society will also need to be continuously informed and educated about the current state of AI research and development and their impact on their organizations and lives, in order to help them form responsible and sustainable AI strategies.

On the medium term, it could then be turned into a full-blown SRA involving all or any given combination of rule-making, monitoring, enforcement, and dispute resolution functions — again, a lower degree of institutional formality and the use of softer legal instruments would be better able to accommodate the uncertainties inherent in AI policymaking and hence probably more effectively further the SRA's acceptance and overall efficiency. While concrete organizational and power con- figurations, collaboration arrangements, and the degree of necessary state oversight (if any) could be worked out at a later time, including an element of self-regulation in domestic AI regulatory frame- works would yield valuable synergies, allowing governments to rely on industry's and

academia's expertise and capabilities in the design of AI policies.

Presuming a positive experience with the proposed self-regulatory model in the domestic arena, Australia and New Zealand could initiate a similar conversation at the international level to endow the PAI with self-regulatory capacities. Such a step could bridge the time until meaningful international collaboration on an intergovernmental level is established.

References

1. J. Black. Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a Post- Regulatory World. Current Legal Problems, 54(1):103–146, 2001.
2. S. Cave and S. S. O´ hE´ igeartaigh. An AI Race for Strategic Advantage:Rhetoric and Risks. In Proceedings of the 2018 AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society. AAAI/ACM, 2018. Forthcoming. Available at: http://www.aies-conference.com/wp-content/papers/main/AIES 2018 paper 163.pdf.
3. O. J. Erdelyi and J. Goldsmith. Regulating AI: A Global Solution. In Proceedings of the 2018 AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society. AAAI/ACM, 2018. Forthcoming. Available at: http://www.aies-conference.com/wp-content/papers/main/AIES 2018 paper 13.pdf.
4. T. Halliday and G. C. Shaffer. Transnational Legal Orders. In T. Halliday and G. C. Shaffer, editors,Transnational Legal Orders, pages 3–72. Cambridge University Press, 2015.
5. A. Korinek and J. E. Stiglitz. Artificial Intelligence and Its Implications for Income Distribu- tion and Unemployment. In A. K. Agrawal, J. Gans, and A. Goldfarb, editors, The Economics of Artificial Intelligence: An Agenda. University of Chicago Press. Forthcoming. Available at: http://papers.nber.org/books/agra-1.
6. W. Wallach and G. Marchant. An Agile Ethical/Legal Model for the International and National Gover- nance of AI and Robotics. In Proceedings of the 2018 AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society. AAAI/ACM, 2018. Forthcoming. Available at: http://www.aies-conference.com/wp-content/papers/main/AIES 2018 paper 77.pdf.