

## Initiatives in AI Governance

Summary by Jovana Jankovic based on a paper by Lawrence Zhang

### *Introduction*

We've already observed extensive real-world examples of the kinds of harms and problems artificial intelligence tools and systems can cause. These include, but are not limited to: bias replication, algorithmically-driven digital addiction, labour displacement, and harms arising from mass surveillance or autonomous vehicles and weapons systems.

Yet, despite high-profile failures and the inherent uncertainties of incorporating new technologies, more and more organizations across both the private and public sectors are turning towards AI to solve complex challenges that this technology can uniquely tackle in unprecedented ways.

While voluntary AI ethics principles have been extensively discussed and drafted across sectors and industries worldwide—such as by the OECD, UNESCO, the EU, Google, Microsoft, to name a few, as well as many academic researchers and technologists—very few concrete governance initiatives have actually been implemented. And it's doubtful that ethics protocols alone are sufficient in ensuring the development of ethical AI. With little consensus on whose AI ethics should guide development and deployment, and how, policy challenges will continue to go unmet.

There's, of course, a concern within industry and public discourse that governance will stifle innovation and prevent AI from attaining success, particularly as the technology is still in its infancy. However, some argue that if AI is already in wide use across the globe, and it has indisputable consequential impacts on the way almost every one of us conducts our lives, then it's certainly mature enough to be governed.

Still, in cases in which AI is already governed, it's being done so through policies and laws designed for an earlier era of technology, creating a mismatch that results in both inadequate governance and stifled development and deployment of AI—particularly as it might be used in socially beneficial domains. We simply don't have adequate conceptions of notions like data collection, legal liability, trade secret disclosure, and risk that pertain to the specific characteristics of AI tools and technologies.

To wit: economists Anton Korinek & Joseph E. Stiglitz have shown that AI's benefits are actually determined by the quality of their market-structuring regulatory environment. This means that the lack of a specific regulatory environment may in fact be hindering both financial and time investment into AI. For example: data collection rules may hinder the aggregation and analysis of data that is necessary for tackling public health problems with

AI, and unclear regulation may prevent technologists from being confident about the ways in which they could be permitted to evolve and improve AI tools and methods.

This executive summary of a paper by Schwartz Reisman researcher Lawrence Zhang examines AI governance initiatives that have been tangibly implemented by policymakers around the world, with the aim of informing further work in this area. By clearly identifying the specific challenges that need to be overcome and matching a proportionate solution, governance can ameliorate harmful effects without creating undue hardship on those responsible for researching, developing, and deploying AI.


## ***Policy Instruments***

Successful policy instrument must incorporate the work and voices of a wide variety of actors, sectors, and interventions. Though many conceptions of policy instruments exist, this paper will use the relatively simple “coercive, remunerative, normative” powers classification. Introduced by Amitai Etzioni in *Comparative Analysis of Complex Organizations*, policy instruments are divided into:

- **Coercive powers:** These are specific rules, regulations, or laws, often set by governments, with various types of clear sanctions for failures to comply. Due to the increasing pace of social and economic change, traditional coercive regulation (ie. “do this specific action”) has in recent decades been replaced with outcomes-, performance- and risk-based approaches (ie. “achieve these standards of safety through negotiable means and methods”) with involvement from third party regulators, regulatory markets, and industry self-regulation.
- **Remunerative powers:** These comprise economic incentives/disincentives or markets which act to channel good behaviour. Examples include tradeable permits, economic property rights, government procurement, subsidies, tax credits, levies, or user fees. While these give firms greater leeway and choice as to how much or whether to comply, they also come with increased uncertainty and risk of failure than direct regulation.
- **Normative powers:** These involve exhortation and voluntary measures aimed at establishing social norms that promote social welfare. Thus type of regulation relies on communicating data, facts, knowledge, arguments, and moral appeals to incite voluntary compliance. In collaborative environments, normative powers can work well by setting a higher bar for common industry practices—and firms can then advertise their corporate responsibility and good reputation. This is, however, the most lenient form of governance.

Many of the following types of commonly used policy instruments have not been considered for use in the field of AI governance:

Figure 1

Policy Instruments		
<b>Control</b> 	<b>Coercive</b>	Public ownership
		Legislation
		Command and control regulation
		Risk-based regulation
		Outcome-based regulation
		Inspection and testing
		Mandatory impact assessments
		Regulatory markets
		International agreements
		Mandatory reporting
<b>Remunerative</b>	Taxation	
	Licensing (corporate or professional)	
	User fees	
	Tradable permits	
	Government procurement	
	Subsidies and grants	
	Tax incentives	
<b>Normative</b>	Standards and third-party certification	
	Non-binding guidance	
	Voluntary program	
	Advisory bodies	
	Self-regulation	
	Principles	
	Labelling	
	Public awareness campaigns	
<b>Choice</b>		

Of course, all three approaches have to work cohesively to function properly. For example, remunerative powers rarely work without coercive powers at their foundations to give them credence and authority, and norm-setting can't always be successful without corollary market incentives.

The Brookings Institution’s Darrell M. West notes that AI governance must include both “horizontal and vertical rules,” with horizontal rules referring to AI challenges such as privacy that apply across all sectors, while vertical rules referring to AI risks that might differ between areas like retail and national defense. Other scholars have also called for “layered” models for AI governance, in which the best tool for each individual component of any problem is used.

## AI Governance Initiatives Around the World

Across AI ethical principles from all sectors, there is global convergence on five general principles: *transparency, fairness, non-maleficence, responsibility, and privacy*. However, abstract principles are not necessarily actionable, nor do they provide concrete guidance for developing specific governance mechanisms.

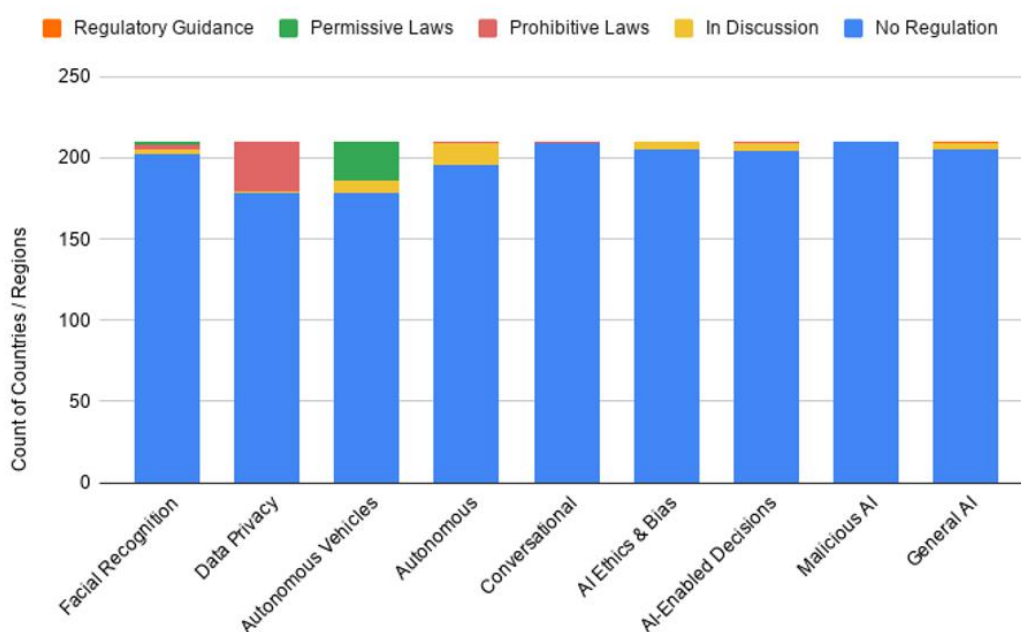
There's no shortage of brilliant ideas for how governance might be implemented, from *regulatory sandboxes* (a sort of 'test environment' that runs trials of new methods to compare with existing regulatory frameworks) to so-called "*society-in-the-loop approaches*" (a kind of algorithmic social contract between human stakeholders, mediated by machines). But the practical implementation of ideas like these is far sparser.

Almost all governments around the world have adopted "wait and see" approaches to governance of AI, with a few policies in place overseeing specific use cases.

Figure 2

### Country and Regional AI Laws and Regulation

Copyright © Cognilytica 2020



Tangible tools explored by governments and civil society in the past few years towards ensuring responsible AI include:

- a. **Standards:** Established by NGOs who provide implementable guidelines and offer certification, standards are enforced by management systems, training, documentation, delegation of responsibilities, and internal performance audits. See, for example: [AI standards compiled by non-profit AI Global](#)—including 46 International Organization for Standardization (ISO) standards. Other standards

bodies include Canada's CIO Strategy Council and the European Telecommunications Standards Institute.

- b. Government Strategies, Directives, and Action Plans:** While these do not explicitly outline accountability or actionable items, they're a good indicator of things like: how governments might situate AI development within their political agendas, what their willingness to invest in STEM is, how they envision application of AI in national industries, and what their plans for building laws and ethical norms for AI are. See: [list of government strategies, directives, and action plans worldwide](#).
- c. Regulatory Guidance:** While they don't outline binding requirements, guidance documents convey a regulator's thinking and promote compliance through exhortation. To date, only two governments have published AI regulatory guidance: Canada's algorithmic impact assessment tool (2018) and the UK's guidance on AI auditing (2020).
- d. Legislation and Regulation:** AI-related topics legislated to date include autonomous vehicles, data privacy, facial recognition, and lethal autonomous weapons systems.
- 24 countries have permissive legislation allowing the testing and conditional use of autonomous vehicles, while eight other countries are currently in the midst of legislative proceedings to enable autonomous vehicle testing and usage.
  - Machine learning techniques create a massive demand for data, putting pressure on existing approaches to data governance. 30 countries have data protection laws that restrict sharing or exchange of data without consent, 27 of which are European.
  - Perhaps the most controversial application of AI, facial recognition and related computer-vision technologies have been either restricted or outright banned in some United States jurisdictions. Both the US and the EU are considering limitations on facial recognition technology. Conversely, China and Zimbabwe's legislation now allows facial recognition in ways not otherwise previously permitted by law.
  - After an April 2018 UN meeting on Lethal Autonomous Weapons Systems (LAWS), 13 countries, primarily in Europe, Africa, and Latin America, have discussed a ban on LAWS in their respective legislatures. Belgium has gone the furthest, passing a non-binding resolution to prohibit the use of LAWS by their armed forces.
  - Other noteworthy pieces of AI governance legislation include Idaho's laws on pre-trial data being open to public inspection, auditing, and testing, as well as California's prohibition on deepfakes and misleading chatbots. The GDPR also grants individuals the right for autonomously-made decisions to be reviewed by a natural person. And existing laws (e.g. tort liability, protections against fraud and defamation) can be used to address some misuse of AI. But to date, no jurisdiction has passed legislation governing malicious AI, AI bias, predictive policing tools, or general AI use.

Figures 3.1, 3.2, and 3.3 below show AI and ML as hot topics in legislatures in recent years.

Figure 3.1

AI and ML mentions in U.S. Congress (1995-2019)

Source: U.S. Congressional Record website, the McKinsey Global Institute.

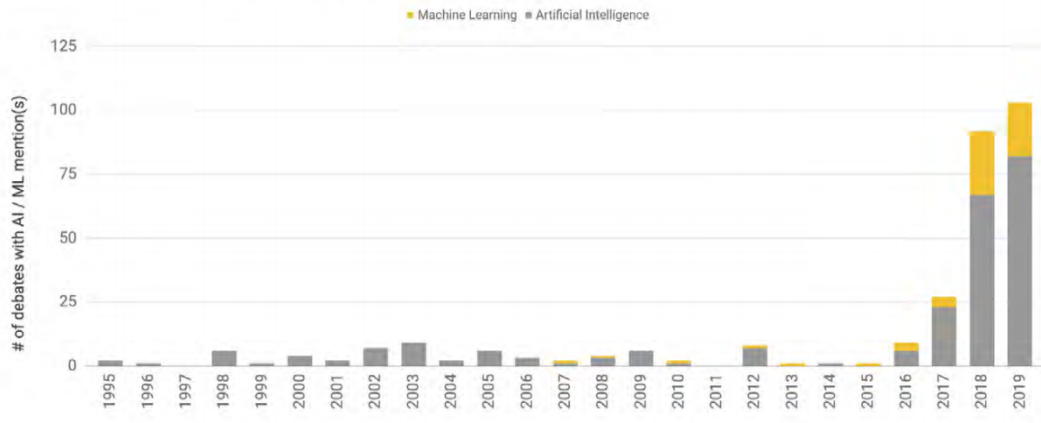


Figure 3.2

AI and ML mentions in U.K. Parliament (1980-2019)

Source: Parliament of UK website, the McKinsey Global Institute.

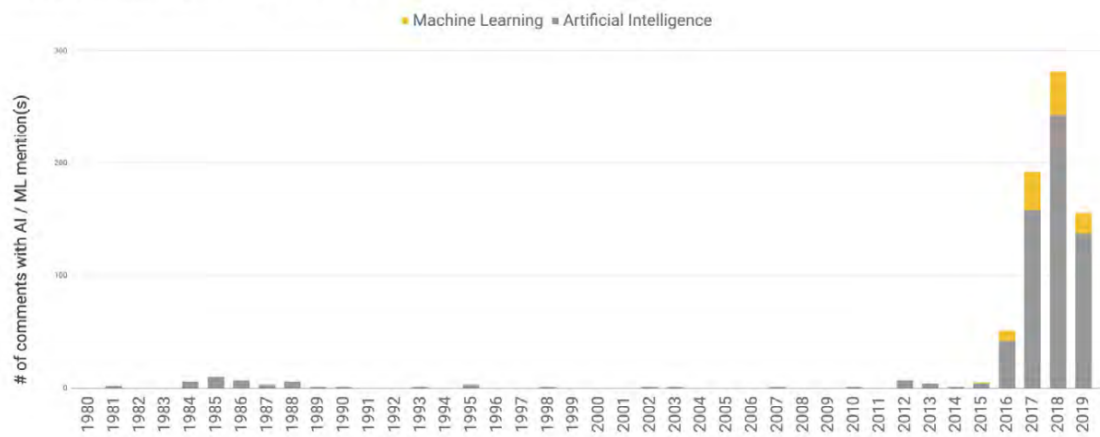
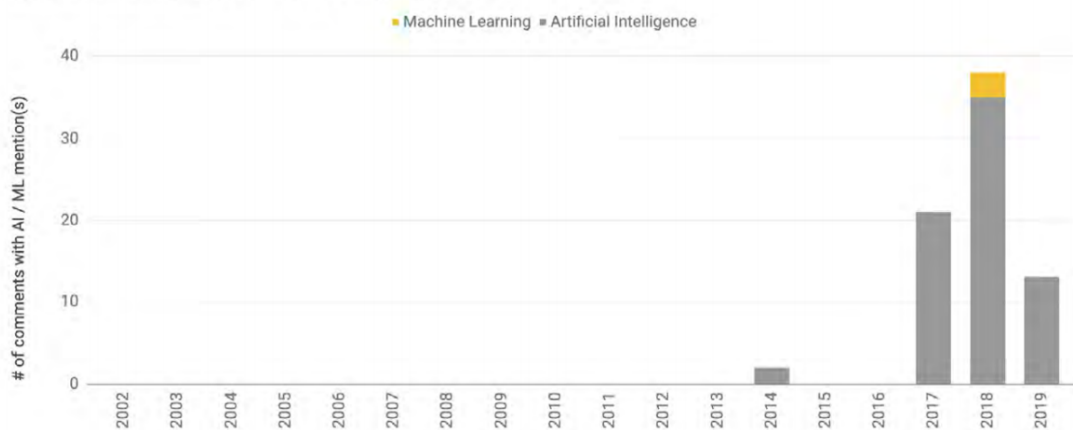


Figure 3.3

AI and ML mentions in Canadian Parliament (2002-2019)

Source: Parliament of Canada website, the McKinsey Global Institute.



- e. **Pilot Projects:** A number of pilot projects have been launched and prove useful as a testing ground for how best to approach AI governance.
- The UK's Financial Conduct Authority launched a regulatory sandbox pilot in 2015 which allowed firms to develop and test new financial and fintech products without undergoing rigorous authorization.
  - The US's Food & Drug Administration (FDA) is currently piloting a Digital Health Software Program aimed at overseeing and regulating software-based medical devices. The FDA provides reputable firms with certification prior to the release of new products by focusing on the software or technology developer instead of the product—similar to how traditional medical devices are inspected—hoping to ensure rapid approvals of new AI-powered medical technologies.

## Conclusion

The case for instituting tangible AI governance initiatives grows stronger every day—but there are far more suggestions and discussions on AI governance than actual laws and regulations.

New legislation has largely focused on particular use cases, such as permitting the testing of autonomous vehicles or banning facial recognition in law enforcement. As the wide-ranging applications of AI become increasingly ubiquitous, governments will soon be compelled to implement governance mechanisms. Given the current pace of both AI use and governance, the landscape will look significantly different in a few years' time.

In a realm where there are both high benefits and high risks, it is imperative that governments consider which policy instruments are appropriate for particular situations, to ensure that both innovation and the public interest are maximized.

[Click here to read the full paper on Initiatives in AI Governance.](#)



The [Schwartz Reisman Institute](#) aims to deepen our knowledge of technologies, societies, and what it means to be human by integrating research across traditional boundaries and building human-centred solutions that really make a difference. We want to make sure powerful technologies truly make the world a better place—for everyone. Comprising diverse areas of inquiry, from machine learning, computer engineering, epistemology, systems theory, and ethics to legal design, systems of governance, and human rights, our research agenda and solutions stream cross traditional boundaries and are fundamentally inspired by a commitment to reinventing from the ground up.