The Future of AI Regulation

Originally published at https://www.iflexion.com/blog/ai-regulation

Published: January 27, 2021 | By Martin Anderson

Disruptive technologies tend to arrive in a blizzard of related developments and innovations, far in advance of any regulations that may eventually govern them and initially striking fear and foreboding into governments and peoples alike.

It was so for the printing press[1], the industrialization of drug production in the 19th century[2], and, more recently, the emergence of GPU-powered cryptocurrency, which has stormed ahead of regulatory oversight as a threat to governmental rule and traditional economic models[3].

And now, after more than fifty years of false starts as described in our machine learning overview, the current boom in artificial intelligence has gained enough credibility and market traction to similarly challenge lawmakers, threaten historic systems of production and consumption, and embed itself into a society that's struggling to understand its workings — and which lacks laws modern enough to address the possible significance and reach of an emerging 'algorithmic age'.

In this article we'll examine some of the approaches and solutions that various governments are taking to develop meaningful legislation, along with the central issues that are driving public and industry pressure for increased regulation and oversight of AI software development.

The Nascent State of AI Legislation

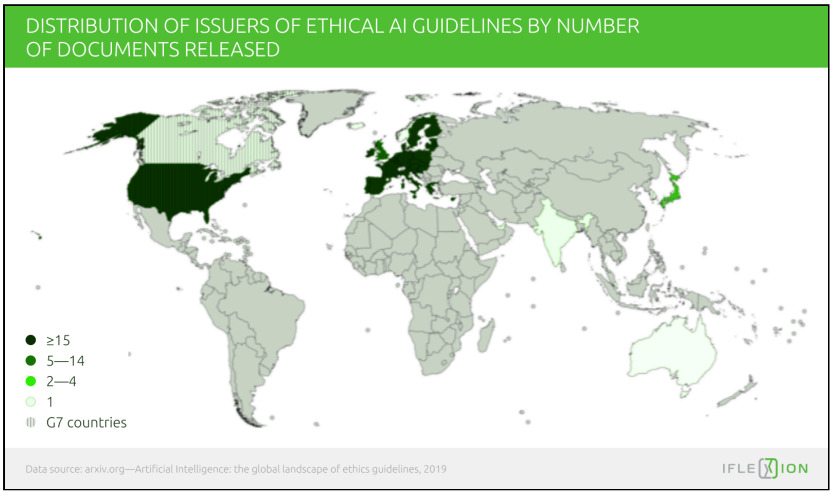
Nearly all democratic governments are currently on the back foot regarding the regulation of AI, since the technologies under discussion are proliferating either from the private sector, where regulatory interference has long fallen out of favor[4], or from China, which has not relinquished state control of its national technological base in the same way that the west has[5] (see below).

A resolution of this tension is necessary, partly because post-facto regulation tends to be driven by public demand for state action after damagingly controversial incidents, but also because the lack of legal clarity can be an inhibiting factor for long-term investment[6,7].

A Global Wave of Ethical Treatises On AI

Notwithstanding that commercial interests may prevail over ethical consensus, we can perhaps discern the trends of future machine learning regulation from the 100+ ethical guideline documents that have emerged from academia, governments, and government-affiliated think-tanks over the last five years[8].

Most of these guidelines are from the west, with a quarter proposed by the USA, nearly 17% from the UK, and at least 19 recommendations from the European Union.



In descending order of prevalence among the ethical reports and roadmaps studied, the core topics common to all are:

- Transparency

- Justice & Fairness
- Non-Maleficence
- Responsibility
- Privacy
- Beneficence
- Freedom and Autonomy
- Trust
- Sustainability
- Dignity
- (Social) Solidarity

Fears That AI Regulation Will Impede Innovation

In general, the governmental guidelines and working papers that have emerged so far express high levels of concern that premature regulation may stifle innovation in the machine learning and automation sector.

In 2020, a White House draft memorandum on guidance for AI regulation concluded that *'Federal agencies must avoid regulatory or non-regulatory actions that needlessly hamper AI innovation and growth'*[9].

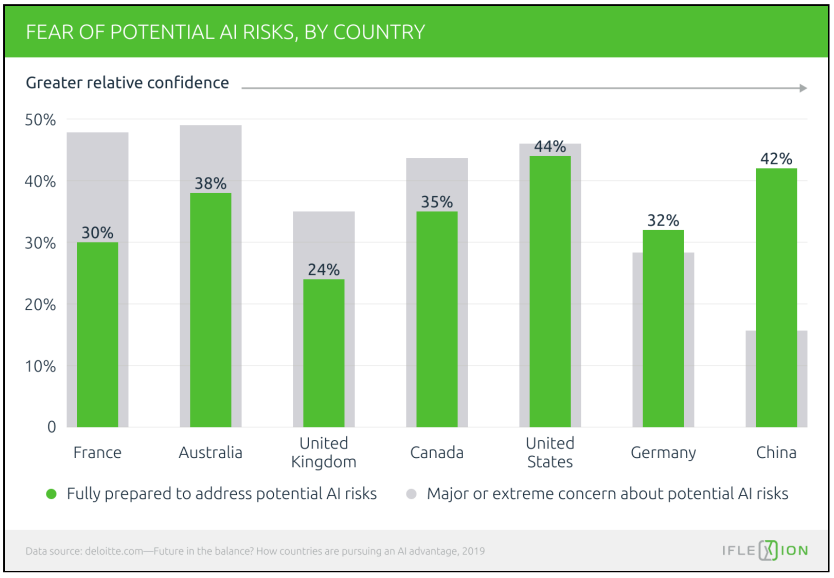
Likewise, the UK AI Council's roadmap report[10] typifies the UK's longstanding enthusiasm to gain ground in the AI sector, while maintaining a very circumspect approach to the development of new legislation around AI.

In the press, the UK's departure from the European Union has been seized as an opportunity to abandon the EU's greater commitment to AI regulation in favor of the more *laissez-faire* policies of the US[11] , an approach that has been criticized as irresponsible[12].

Pressure from China

This fractured landscape of ethical and legislative timidity might seem surprising, since, except for China, the major AI powers all signed an OECD accord for a global governance framework in 2019[13].

In reality, the long-term stability of China's political administration, together with its leadership in AI venture capital investment[14], a deeply state-driven economy, and an avowed determination to lead the world in AI development by 2030[15], is bringing competing democratic nations to a 'guilty envy', which increasingly sees oversight and regulation as a significant competitive disadvantage[16].



China's vanguard position in AI rests not so much in its academic acumen (though this is formidable) as in its ability to generate and exploit massive levels of data from the most-surveilled population in the world[17].

In contrast to China, democratic economies must negotiate successive waves of public and popular skepticism and terror[18] around AI, as controversies evolve from alarming headlines into a growing wave of petitions and media pressure around the consequences of AI-driven automation.

Possible Mechanisms for AI Regulation

Most of the aforementioned ethical guideline papers propose potential regulatory methods and mechanisms that are likely to hinder machine learning processes and widespread adoption in some way.

These include:

- In-built data quality evaluation
- Explainable AI (not easy to achieve)
- More granular and public data provenance
- Audits from the government and private sector
- Attribution of liability
- Increased levels of transparency that might be unpalatable to a commercial project[19,20,21]

While some maintain that AI legislation based on current ethical guideline recommendations would have a chilling effect on investment[22,23], or even lead to a new AI winter, others contend that, as with previous incursions by new technologies, regulation will simply focus and vindicate the best implementations[24].

Will AI legislation have a chilling effect on investment, or will regulation simply focus and vindicate the best implementations?

Web-scraping and AI

A great deal of the current impetus in machine learning has been fueled by the ability of automated systems to ingest data from the internet into machine learning systems for segmentation (in the case of images and videos), classification, and eventual processing inside neural networks.

Where a website uses some form of obfuscation to impede data-gathering systems, robotic process automation (RPA) can utilize AI-driven image-to-text transliteration, among various other techniques, to access and re-digitize the data for use in training machine learning models.

Theft, Plagiarism, or 'Education'?

OpenAI's headline-grabbing Generative Pre-Trained Transformer 3 (GPT-3) autoregressive language model was constructed and distilled chiefly from the CommonCrawl public data repository, a petabyte-level database derived from publicly available documents on the internet, in much the same way as the private databases that power Google and other search engines.

Though the CommonCrawl agent, and other mining agents, can be specifically blocked by configuring a robots.txt file on your server, there is nothing to stop another organization being less respectful of your data: in 2019 the US Court of Appeals denied LinkedIn's petition to prevent an external company from using web-scraping techniques to distil and incorporate publicly available data from the LinkedIn network[25]. This ruling (amongst similar cases listed below) effectively legitimized the public domain as a free training ground for data-hungry AI systems[26].

Whether or not courts rule that screen-scraping and AI-driven transliteration of screenshots are legal (or whether scrapers can exploit data on a logged-in account[27]) is relatively moot: when high volumes of data are run through a machine learning system, the neural network will effectively 'absorb' the individual data points into a non-attributable 'general wisdom', obscuring all the original sources.

For instance, in the case of GPT-3, a series of scientific articles by one particular author might have helped the AI to develop lucid viewpoints on certain scientific topics. But since GPT-3 has completely integrated that information into a wider knowledge base (and is unlikely to simply 'paste' the source text back into an unattributed response), there is no tangible evidence of the original author's contribution to the GPT-3 output.

Journalism has internal codes of conduct regarding this kind of uncredited and unpaid appropriation, but as of yet there is no reliable method for re-identifying data that has contributed to a machine learning algorithm, notwithstanding any existing legislation that could penalize such usage as a 'derivative' work. Unless the contributing datasets are constrained to be publicly available, there is no effective proof of appropriation, or legal recourse.

Laws Supporting Data Miners Over Copyright Holders

Even where the output of an algorithm *does* clearly indicate which source material has been copied into a contributing dataset, legislation is tending to favor machine learning rather than aggrieved content creators or copyright holders.

In 2016, the Supreme Court of the USA finally ratified the District Court's decision that Google may legally scan copyrighted books in order to provide search results via the algorithms of its Book Search service, ending 11 years of litigation by the Author's Guild[28].

In 2019, the European Union officially legalized[29] the induction of copyrighted material into machine learning datasets for non-profit organizations. As with the EU's general reliance on GDPR to police downstream use by AI systems, there are no obvious enforcement mechanisms that could prevent commercial entities from doing likewise (beyond the use of prohibitory metadata tags, which crawler bots may or may not respect[30]).

In certain countries, machine learning datasets may have more effective legal protection than the many sources that populated them, depending on the level of investment and effort that went into creating the dataset[31]. The EU[32], the UK, and Russia[33] all offer 'database rights' to compilers.

Deepfakes

The issue of dataset material 'evaporating' into the latent space of a neural network also affects one of the most public drivers for AI legislation — the phenomenon of deepfake video and audio, discussed at length in our recent post on the future of emotion recognition software.

Deepfake output is created from high volumes of sample image or sound data, generally from openly available (though not necessarily non-copyrighted) material on the internet, such as social networks, tube sites, or podcast output. By the time the model is trained on the data, it is impossible to identify any specific contributing image or sound from the faked output.

Due to negative public response to deepfake videos[34], instances of fraud[35], and fear of electoral manipulation[36] (ultimately unfounded[37]), deepfake content has spawned some of the earliest concrete legislation and/or legal initiative around AI, amongst which:

- China banned video deepfakes outright in January of 2020[38].
- Texas passed a 2019 law to ban political deepfake video content, but excluded other types of deepfake content due to issues around 'free speech'[39].
- In the wake of the failure[40] of the H.R. 3230 deepfake bill in Congress in 2019, two major new bills were passed into law[41] at the start of 2021, intended in part to inform and generate new legislation governing deepfakes.

Varying international legislation around image rights has the potential to make deepfake image content illegal by default[42], in the spirit of a GDPR-style 'transitive property' of prohibition. However, this would challenge a number of cherished American statutes around fair use of public figures in satire and the media, and has ramifications that extend beyond the misuse of AI.

Legally Defining 'AI'

Semantics are a major obstacle to new legislation, and the recent flurry of ethical papers do not provide a consensus on what 'AI' actually is.

Some evade the issue by referring to an 'algorithm', a term which covers everything from an *if else* loop in JavaScript to output from the 175 billion parameters of GPT-3[43].

Others overleap present regulatory needs to address the more distant and philosophical possibility of an AI singularity making undesirable decisions in a military or government context.

If it proves impossible to define the scope for the term 'AI', subsequent legislation may either devolve into a patchwork of laws for individual use cases (as is happening with deepfakes, self-driving vehicles, and facial recognition databases), or else AI will be left in the care of older laws until use cases emerge where those laws are proven not to apply. At that point, the procrastination and ambivalence of lawmakers might become obvious or even indictable.

If it proves impossible to define 'AI', subsequent legislation may either devolve into a patchwork of laws for individual use cases or else AI will be left in the care of older laws.

The Emergence of Scofflaw AI

Perhaps the greatest emerging controversy of the current machine learning boom is the extent to which critical information about our lives and habits can be obtained by stealth, when tracking technologies allow our data to cross from one domain to another — and the extent to which a reliance on existing laws is powerless to prevent this from happening.

Obtaining Privileged Health Data by Stealth

Besides the global outrage at the misuse of data mining and machine learning in the Cambridge Analytica scandal, the highest level of concern currently centers around healthcare and general insurance providers, who, while constrained by intractable industry regulations, are often able to access 'forbidden' data via the analytics systems of third-party providers.

In the US, for example, the Health Insurance Portability and Accountability Act (HIPAA) requires patient consent for the disclosure of medical information, but AI developers are not covered under these regulations, and are not prohibited from contributing to collaborative systems that provide insight-by-stealth into an individual's risk factors[44], enable re-identification[45], or reveal race, medical history, genetic predisposition to disease and other sensitive data points that may affect their treatment.

The Enforceability of GDPR for AI Systems

The EU's dependence[46] on GDPR as a statutory mechanism to prevent data misuse in AI systems has been criticized as unrealistic, partly because the provenance of an algorithm's contributing data is obscured in the trained model's output, but mostly because the existing oversight and enforcement mechanisms do not currently live up to the ideals of the regulation[47].

Additionally, the EU's mandated 'Right to Explanation' about algorithmic decisions requires AI technology that does not yet exist[48] , and the development of it is not the highest priority in a global scientific community currently embarking on an 'AI cold war'[49].

Conclusion

AI legislation outside of China is currently driven by optics and disaster management, as a ruminative west gauges the balance between the perceived value and viability of new AI systems and the volume of popular resistance that they might provoke.

In the meantime, there is a general tendency to let existing laws govern the regulation of machine learning systems and their derived algorithms, even where such laws may prove to have limited scope for the task.

However, one 2020 study from RAND[50], which examined 5,240 articles dealing with regulatory gaps in AI governance, takes the more optimistic view that minor amendments to existing legislation will be adequate to cover the legal ramifications of machine learning systems as use cases emerge.

Nonetheless, the report concludes that *'AI will continue to push the boundaries of public policy for the foreseeable future'*.

There is a general tendency to let existing laws govern the regulation of machine learning systems and their derived algorithms, even where such laws may prove to have limited scope for the task.