Belgian and Flemish Policy Makers' Guide to AI Regulation

Joanna J. Bryson[1]
Professor Ethics and Technology, The Hertie School of Governance

October 11, 2022

KCDS-CiTiP Fellow Lectures Series Towards an AI Regulator?

This is a requested position paper/blogpost relating to the event joint with Martin Ebers on 31 May in the broader context of the *Standardisation of AI* event 29–30 May.

Regulating Intelligent Technology

First I wish to briefly visit a few of my standard positions on AI regulation (for more detail, see Bryson, 2020, 2019).

- I use a standard definition of intelligence (Romanes,1882, dating back more than a century) to embrace a very broad understanding of AI. Intelligence is the capacity to adjust actions to context: to address opportunities and challenges. It is therefore a transformation of information—that is, computation. Computation is a physical process taking time, space, and energy. AI is the subset of computation that produces actions (or recommendations for action) from a context, and is also deliberately built by humans or human organisations.
- Regulation is any means for perpetuating a dynamic entity into the future. Regulation does not mean only restriction, though restriction can be very helpful for focusing innovation on problems likely to be productive and lead to sustainable, secure economic growth. But regulation can also be about *increasing* resources available for production, e.g. by ensuring adequate funding or access to appropriately-educated talent.
- Regulating AI is not that different from regulating any other product. We should be setting laws and (probably more importantly) enforcement bodies in place to ensure that best practice and due diligence are followed in constructing and deploying AI. This includes means to trace responsibility for negligence or malpractice and accurately designate whether it was performed by either developers/vendors or deployers/end users. Systems must also be transparent enough to detect when misbehaviour is due to malign third-party interference such as crime or sabotage. Since most AI is software, the above transparency can often be produced simply by ensuring the

[1] This is a requested position paper/blogpost relating to the event joint with Martin Ebers on 31 May in the broader context of the Standardisation of AI event 29–30 May. The views expressed in this paper are those of the author and do not necessarily reflect the views or policies of the Knowledge Center Data & Society or KU Leuven CiTiP. The paper aims to contribute to the existing debate on AI.

use—and cybersecured documentation—of standard development and operations practices (devops).

Technical Standards and Accountability for Regulating Intelligent Technology

At a 2016 AAAI Spring Symposium (on Ethical and Moral Considerations in Non-Human Agents Prof. Ron Arkin shamed all present who claimed to care about "AI Ethics" but were not a part of the new IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems—including me. Being then on sabbatical, I took the chance and joined IEEE and the initiative. I then rapidly found myself co-chairing the Affective Intelligence subgroup of that initiative, with Arkin.

While impressed with the effort and the professionalism of the IEEE, as a professional software engineer and systems administrator, as well as a PhD in AI, I had never really seen the point of technical standards, and spent a decent amount of time trying to make sure the effort was worth participating in. Ultimately I wrote a paper, "Standardizing Ethical Design for Artificial Intelligence and Autonomous Systems," with Alan Winfield 2017. Winfield also chaired the committee that subsequently (published 2001) wrote IEEE's AI transparency standard, and transparency was the case we used in our paper. The paper, now well-cited, claims that standards can be a means by which we can make governance more agile. Governments merely need to enforce the utilisation of standards produced by recognised professional standards organisations, rather than needing to keep up on the technology itself.

I've since become more sceptical again that standards are really the way to make progress in software engineering. Software engineering is just qualitatively different from mechanical or electrical engineering, where standards have indeed proved useful (Parnas, 1985). Even where they *are* useful, standards are ripe for regulatory capture. Standards committees may try to be inclusive, but they seldom pay and more often charge for membership. They therefore tend to be stacked with people from large corporations who can afford their engineers to spend such time. More recently, a particular large national government has also sought to stack committees, which to be fair, to date that nation has seen its engineers under-represented on international standards boards. Accessing and certifying standards is also expensive and excluding, working against small and medium sized enterprises (SMEs) and the developing world.

But agility *is* important. We want the law to keep up with the knowledge of best practice in the sector. Fortunately, this is already a solved problem. It turns out that software is not the first industry that innovated. Ordinary product liability law already requires that manufacturers show due diligence, including following best practice and avoiding known bad practice. Obviously, what is best and what is known can continuously be updated, and evidenced in trade journals or simply testimony.

My present thinking is that we would be better off focusing on the fact that software products are products, and using standard product law. This is an approach the EU says it is taking with its new AI Regulation, though this hasn't been made entirely explicit yet (Haataja and Bryson, 2022). One terrific aspect of this approach is that it is automatically proportionate. If something goes wrong with a system you've developed and sold (or sold access to), you need to be able to demonstrate that you were not negligent, but rather followed best practice in designing, developing, testing, deploying, and maintaining that system. How much effort you will put into documenting that will be proportionate to the liability risk you (and your lawyers or auditors) assess your system is likely to run. If you are building a high-risk, critical system, you had better be able to afford top-drawer processes, including cybersecurity as well as (other) DevOps good practice, documentation and accounting. If you are building games, your costs and liabilities will presumably be lower.

Standards and certification can still be of use as part of the process of both performing and demonstrating good practice, particularly for particularly expensive or otherwise 'high-risk' systems. But we shouldn't allow standards (or patents) to become another barrier deployed by those dominant in markets to exclude new entries. In particular, we need to ensure that startup and creative culture

can spread to more countries and continents. Digital SMEs have the potential to democratise economies, bringing good work and capital to small and inaccessible regions, opening trade and promoting education and transparency. The COVID pandemic has shown that good work can be done almost anywhere, so long as there is adequate communication infrastructure at least at somewhat reliable intervals.

As a brief aside, some people think software isn't a product, but a service. Writing software can be done as service, and AI products can perform services, but the software itself is in my mind unquestionably a product. I have long said that I would like to see some of the present digital regulatory acts of the EU make this explicit. I have learned that there's also long been a parallel argument that services are products anyway, and indeed there are now standards being formed over processes which might include services. The newly-released draft liability act indeed looks set to handle these concerns, and to be congruent with the recommendations made above. I quote:

> "Liability would continue after the product is launched on the market, covering software updates, failure to address cybersecurity risks, and machine learning. In other words, developers would continue to be responsible for AI systems that learn independently. . . Moreover, there are five scenarios whereby the causal link between defectiveness and damage is presumed. These include cases where the manufacturer fails to provide the information, or if the product does not meet the safety requirements, if there are obvious malfunctions, or *if the causal link is impossible to prove due to the technical or scientific complexity of the product.*" — (Bertuzzi, 2022, emphasis mine).

Human-Centred Values and Value-Aligned Design

At one time, it was difficult to have a discussion about AI regulation without someone suggesting that it was controversial or even wrong to focus AI ethics on human concerns, with no regard for the AI itself. As the top tiers of international relations, international law, and human rights have become engaged with the problem, it has been more common to emphasise human-centring as being opposed to centring on corporations or perhaps governments, not the machines that are anyway best understood as an aspect of capital. If machines could be meaningfully said to have any interests at all, because they are artefacts, those interests would only exist due to a decision of product design, such as leaving out a backup system for memory. For this reason, Bryson et al. (2017) advocate strongly against constructing law recognising AI interests.

'Human'-centring at least in the context of the United Nations is now increasingly well-understood to also entail sustainability and concern for biodiversity. This makes sense, because human well-being does depend on a healthy environment that our ecosystem would ideally tend to stabilise, and that requires living within our resource constraints. Resource conflict can lead to war, and abhorrent violations of human interests. The ecology cannot really be designed, but our artefacts can. This is why—in the first national-level AI 'soft law' (for the UK, in 2011), the third of five principles calls for AI to ". . . be designed and operated as far as is practicable to comply with existing laws and fundamental rights and freedoms, including privacy." —(Boden et al., 2011; Bryson, 2017).

The vast majority of commercial AI is indeed best understood as extensions of the corporations that provide it, often at no cost. Our homes, laptops, and pockets contain eyes and ears of corporations, and of some governments. The EU with its GDPR has rightly recognised that personal data is to persons like air space is to nations. Personal data absolutely must be strictly defended, because otherwise foreign agencies have undue access to and even control of what ought to be sovereign — the behaviour of citizens and residents. Therefore while (as I discuss in the next section) I do strongly support appropriate redistribution to diverse nations of reasonable portions of wealth gained through access to those nations' data, I discourage the individual sale of personal data. We do not want to motivate citizens of our democracies making themselves more subject to manipulation.

Some people who work on value-aligned design describe a system of essentially crowdsourcing what 'real' ethics are, and then using the AI systems to enforce these ethics on their users. This is a terrible misunderstanding of how ethics works, perhaps founded on a supernaturalist worldview where right and wrong are predefined. There is no absolute ethics, or complete fairness. Ethics is the means by which we hold our societies together. There are two reasons ethical systems will perpetually be dynamic. First, what works best will change with context. So for example, there are only very rarely contexts where lockdowns are ethical, or martial law, but these situations do exist. Second, new innovations—whether technological, political, or social—may allow us new and better ways to produce stable, secure societies. So for example, new technologies and norms allow us to more fairly distribute the burdens of child rearing across genders.

Human justice only has the capacity to hold adult humans to account—its penalties only persuade living social organisms that can understand its language (Bryson et al., 2017). So having value-aligned technology must mean that the technology expresses not its own values, but the values of those that own and operate it. For ensuring those owners and operators comply with human interests, we have the law.

Justice and Equity

Finally, I want to speak briefly to how we can best ensure regulation works. I am persuaded by political philosophy like that of Gowder (2016) and Wu (2018), that justice requires enough equity that obligations can be enforced. How to handle transnational infrastructure and public goods is an enormous legal and diplomatic challenge, one that we need to surmount if we are going to solve sustainability and limit warfare while defending freedom of thought. But we have done a pretty good job of addressing such challenges before. We achieved a long period of relative political-economic stability following the Bretton Woods agreement due in part to increasing justice through equitable participation (Fraser, 2006; James, 2017). More recently, not only have we succeeded in widespread vaccination during the COVID pandemic, in so doing we seem to have reduced the influence of populism everywhere in the world—except in the US (Foa et al., 2022).

Political polarisation tracks economic inequality much more than it tracks social media use. Though to be fair, the latter is 'not at all' (Lorenz-Spreen et al., 2021; Waller and Anderson, 2021), while the former is 'generally, but imperfectly' (McCarty, 2019). With colleagues I've shown that the real driver of political polarisation is most likely economic precarity, which only tends to track inequality but can be addressed through regulation (Stewart et al., 2020). The trust you need to work with a diverse outgroup is sort of like a luxury good — what it requires is adequate security, and states can help provide that. In so-far unpublished work, we've found data not only supporting that model, but also indicate that polarisation is less likely to follow inequality than negative employment figures at the macro level—at the individual / micro level, unemployment isn't as important as perceived direction of the national economy. However, social trust drops enormously with increasing inequality (Sairam et al., 2022). Such loss of faith indeed probably makes sense if there is strong evidence that a social contract is being violated. But loss of faith doesn't turn into polarisation until there is macroeconomic precarity.

Some people promote decentralisation as a solution to present problems. Whilst having a certain amount of redundancy in logistics chains and infrastructure provision is certainly a path towards robustness, dismantling government capacity to regulate is only likely to re-enforce present levels of inequality and injustice—which destabilises the system for everyone. Agility requires significant and well-designed infrastructure; it is not just something that comes for free when you downscale a government (McBride et al., 2022). Even those who are presently 'winning' should want to reduce the scale of those wins –to reduce inequality– in order to stabilise their grasp on at least some of their holdings. Again, this was what led to another example of successful transnational coordination, that around the Bretton Woods agreement following two World Wars, a stock market crash and a decade of economic hardship (James, 2017). Fewer people starved in the decade following the 2008 financial

crisis than that in 1929. Let's hope we can similarly better meet the challenges of the current assaults on Ukraine, NATO, and the rule of law than we did those of WWI. But doing so requires resilient societies with the structural capacities to make hard decisions, and enforce them.

Conclusion and Further Practical Challenges

The bottom line then is that AI and software more generally are products, and like all products, they can be designed in such a way to facilitate maintaining accountability for their performance. And indeed those that develop such products are obliged to do so. The EU's nascent legislation for AI Regulation and Product liability more generally (including for AI) are on a good path to provide us with the legal frameworks that should make regulating AI tractable.

There are however some outstanding problems, specifically ensuring sufficient funding and sufficient available, educated talent for the enforcement of these laws. This is a matter of money. Also, I have raised the issue of inequality and economic precarity, which fragment and polarise societies, making them harder to govern. Fortunately, inequality is among other things a matter of redistribution, so these two problems should afford mutual solutions, if only sufficient political will and diplomacy can be practiced. I don't say this entirely facetiously. I've been told that every country in the world recognises the importance of deriving adequate revenue from, and stabilising the power of, transnational essential infrastructure companies. Further, these companies are in many cases coming to see the importance of coordinating with governments on the grand challenges of our times. Therefore, I do hold hope that we can come to an equitable and creative solution. But it will take further work, including by European legislators.

References

Bertuzzi, Luca (2022). EU Commission proposes extending product liability rules to software and related services. *Euractiv*. as updated 30 September, 2 days after original release.

Boden, Margaret , Joanna Bryson, Darwin Caldwell, Kerstin Dautenhahn, Lilian Edwards, Sarah Kember, Paul Newman, Vivienne Parry, Geoff Pegman, Tom Rodden, Tom Sorell, Mick Wallis, Blay Whitby, and Alan Winfield (2011). Principles of robotics. The United Kingdom's Engineering and Physical Sciences Research Council (EPSRC).

Bryson, Joanna J. (2017). The meaning of the EPSRC Principles of Robotics. *Connection Science 29*(2), 130–136.

Bryson, Joanna J. (2019). The past decade and future of ai's impact on society. In *Towards a New Enlightenment? A Transcendent Decade*, OpenMind BBVA, pp. 150–185. Madrid: Taylor. commissioned, based on a white paper (also commissioned) for the OECD.

Bryson, Joanna J. (2020). The artificial intelligence of ethics of AI: An introductory overview. In M. D. Dubber, F. Pasquale, and S. Das (Eds.), *The Oxford Handbook of Ethics of AI*, Chapter 1, pp. 3–25. Oxford: Oxford University Press.

Bryson, Joanna J. , Mihailis E. Diamantis, and Thomas D. Grant (2017). Of, for, and by the people: the legal lacuna of synthetic persons. *Artificial Intelligence and Law 25*(3), 273–291.

Bryson, Joanna J. and Alan F. T. Winfield (2017). Standardizing ethical design for artificial intelligence and autonomous systems. *Computer 50*(5), 116–119.

Foa, Roberto S. , Xavier Romero-Vidal, Andrew J. Klassen, Joaquin Fuenzalida Concha, Marian Quednau, and Lisa Sophie Fenner (2022). The great reset: Public opinion, populism, and the

pandemic. Technical report, Centre for the Future of Democracy, Cambridge University, Cambridge, UK.

Fraser, Nancy (2006). Reframing justice in a globalizing world. In J. Goodman and P. James (Eds.), *Nationalism and Global Solidarities*, pp. 178–196. Routledge.

Gowder, Paul (2016). *The Rule of law in the Real World*. Cambridge University Press.

Haataja, Meeri and Joanna J. Bryson (2022). Reflections on the EU's AI Act and how we could make it even better. *CPI TechREG Chronicle*.

James, Harold (2017). Bretton Woods to Brexit: The global economic cooperation that has held sway since the end of World War II is challenged by new political forces. *Finance & Development 0054*(003), A002.

Lorenz-Spreen, Philipp , Lisa Oswald, Stephan Lewandowsky, and Ralph Hertwig (2021). Digital media and democracy: A systematic review of causal and correlational evidence worldwide.

McBride, Keegan , Maximilian Kupi, and Joanna J. Bryson (2022). Untangling agile government: On the dual necessities of structure and agility. In M. Stephens, R. Awamleh, and F. Salem (Eds.), *Agile Government: Emerging Perspectives in Public Management*, Chapter 2, pp. 21–34. World Scientific.

McCarty, Nolan (2019). *Polarization: What Everyone Needs to Know®*. Oxford University Press.

Parnas, David L. (1985). Software aspects of strategic defense systems. *American Scientist 73*(5), 432–440. revised version of UVic Report No. DCS-47-IR.

Romanes, George John (1882). *Animal intelligence*. London: D. Appleton.

Sairam, Vishali , Vincent S. Heddesheimer, and Joanna J. Bryson (2022). Economic insecurity increases polarization and decreases trust. Article in prep, and accepted for presentation to the European Political Science Association annual meeting.

Stewart, Alexander J. , Nolan McCarty, and Joanna J. Bryson (2020). Polarization under rising inequality and economic decline. *Science Advances 6*(50), 10.

Waller, Isaac and Ashton Anderson (2021). Quantifying social organization and political polarization in online platforms. *Nature 600*(7888), 264–268.

Wu, Tim (2018). *The curse of bigness*. Columbia Global Reports.