

Overview

My research entails an analysis of the existing data governance paradigms around the world, focusing on China, the United States, and Europe. These three entities each have a unique data governance structure and illustrate well the various paths down which countries can go when it comes to data governance. Moreover, for countries like the United States that do not have a clear, articulable governance strategy, analyzing existing regimes elsewhere can inform future legislation and policy goals.

Background

Dr. Kai-Fu Lee, preeminent a scholar and leader in the world of AI, said “if artificial intelligence is the new electricity, then big data is the oil that powers the generators.” What Dr. Lee implies is that AI, arguably the most exciting technological breakthrough happening today, will rely on data aggregation. Thus, there is a clear link between advancing technology that utilizes AI and the manner in which we aggregate, store, and protect our data. In other words, how we govern data will directly impact how quickly, ethically, and appropriately AI advances.

The link between AI and data, and the benefits it will allow or not allow to be realized, is crucial. AI could have enormous positive benefits for society. Driverless vehicles alone could reduce vehicular accidents so much that it might well be the greatest public health breakthrough of our lifetimes. Using AI to diagnose and inform clinical work could lead to greater accuracy and better patient outcomes. However, this must be tempered by an understanding that data is sensitive and ought to be governed with the utmost care. Striking a balance between the mass aggregation of data, utilizing it to power groundbreaking technologies and the very real privacy concerns that accompany that aggregation is what I would like to flesh out in my research. Drawing on the existing data governance paradigms will allow me to do that and also recommend avenues for further exploration and potential governance strategies.

Audience and Scope

The intended audience for my research will be policy analysts, legal researchers, and professionals working at the intersections of law, public policy and emerging technologies.

Deliverables

I think it would be useful to have a 4-part series. One for Europe, one for China, one for the United States and one comparing the three regimes. Within each descriptive one, I intend to document the benefits and shortcomings of their approaches, especially with regards to specific technology (I think this is key, and what will make my pieces stand out). In the fourth (comparative) one, I will compare each one vis-à-vis the others and make my own personal claim as to aspects that

create a superior data governance model. I will conclude the fourth (comparative) one with recommendations.

Initial Explorations

The European Union

The EU is very precautionary and does not want to jeopardize the security or privacy of its data. Some postulate this may be a longstanding reaction to fascism and a prioritization of individual rights over the rights of state or supra-state entities.

This model, what is known as a precautionary model, greatly limits what companies can do with data and how much data can be used, has actively harmed AI development efforts. A great example is with COVID-19, where the strict limitations of European data has harmed vaccine and therapeutic treatment development. A [Financial Times](#) article noted in that:

“Now with coronavirus, if you’re working on something like a vaccine and you want to move quickly, there is a realisation you need to rely on a much broader set of data,” said one person with direct knowledge of the European Commission’s thinking. “The EU is not backtracking yet on its position, but it is thinking more actively about the unintended consequences of what they have proposed in the white paper on AI,” the person said.”

It is unlikely that this is the last of the harm that Europe will experience from restrictive data policies. It is trying to sell itself as the hub of “ethical AI”, but it is hard to be both ethical and restrictive with your data. In order to scrub biases from AI, you need massive amounts of data. Restricting available training data only leads to more unethical AI, not less.

China

China only recently began to implement some privacy protection legislation. But there is an argument as to how robustly enforced that legislation is when the CCP calls out companies by name as “National Champions / Members of the National Team” and grants them massive subsidies, free rent, and entire cities to work in/on. Thus, China can really demonstrate well one side of the spectrum, where data is readily available due to relaxed policies on the subject but of course, such relaxation poses important privacy concerns. Some interesting [cross-country collaborations happen](#) because American researchers just don’t have access to the type/quantity of data they need in the US. This demonstrates that China is doing something right when it comes to supplying data to maximize technological innovation, but of course such lax regulatory policy invokes ethical questions as well. For instance, China leads the world in face recognition technology, which has many useful aspects, especially for personal security. However, this type of technology can be coopted or bought by the Chinese government, who could use such technically advanced inventions to further immoral and harmful policy goals, or infringe on due process.

The United States

The United States probably lays somewhere between the EU and China. While the US lacks a federal data governance law, thus making it less of a regulatory state than China, it will probably trend towards a European governance structure based simply off history and traditional pacific alliances. But, if history is any indicator we will probably lean towards a more permissionless style of data governance and regulation than a precautionary one. Meaning, we will allow greater leniency on our data in exchange for faster development.

This would be akin to technological development in the 90s and early 2000s, where the United States tried to make sure regulations didn't get in the way of enterprise (Section 230 of the Communications Decency Act being a prominent example). Michael Kratsios, the US Chief Tech Officer, has done a good job with maintaining AI and data as a priority. He has instructed agencies to do their own research on AI, which should hopefully lead to some useful outcomes. AI policy in the US has not been very controversial and has been more or less a straight line since Obama started thinking about it a decade ago. However the US is seeing the rise of political figures who speak out vociferously against big tech. This camp may thwart any federal law that seeks to be more permissionless, and instead opt for a data governance system that punishes big tech.

Avenues for Further Research

1. **How do we address the collective action problem?** This was President Obama's biggest worry when it came to data. He worried we needed more incentives to pool data and work together, because the existing structure incentivizes pigeonholing yourself and your work to beat your opponent. Driverless vehicles pose a good case study. We do not have them the way we thought we would, and I believe that's partially because there's very little cohesion between competitive groups (naturally). Of course, we cannot make it entirely communitarian because incentives are important, but we should be actively thinking about how to address the collective action problem as it applies to data.

Obama suggested creating open data standards which, as I understand it, are basically laws/rules that force companies to make public the data they are using to train AI, at least in part. For instance, Obama [remarked that](#) a model similar to the human genome project could be useful for data:

“instead of giving money to Stanford or Harvard, where they're hoarding their samples, we now have this entire genetic database that everybody has access to. There is a common set of values, a common architecture, to ensure that the research is shared and not monetized by one group.”

2. **The geopolitics of AI and data.** Is China a threat to American AI supremacy? What are the implications of this? Is it worth worrying about? Does America need its own belt and road initiative? Two notable China AI scholars, Jeffrey Ding and Dr. Kai-Fu Lee both say that the United States is well ahead, but that China has the edge in manpower and

Connor Haaland
JURIST Digital Scholars
Data Governance and its Existing and Future Paradigms

institutional support which could allow them to catch up and overcome American AI supremacy in the future. China also leads in ability to aggregate data.

References (Ongoing)

Ding, Jeffrey. "ChinaAI Blog." Web blog. *ChinAI* (blog), 2018. <https://chinai.substack.com/>.

Lee, Kai-Fu. *AI Superpowers: China, Silicon Valley, and the New World Order*. Boston, MA: Houghton Mifflin, 2018.

Liu, Jinhe. 2020. "China's Data Localization." *Chinese Journal of Communication*, 13:1, 84-103.

Pernot-Leplay, Emmanuel. 2020. "China's Approach on Data Privacy Law: A Third Way Between the U.S. and the EU?" *Penn State Journal of Law & International Affairs*, 8:1, 49-117.

Thierer, Adam D. *Permissionless Innovation: the Continuing Case for Comprehensive Technological Freedom*. Arlington, VA: Mercatus Center George Mason University, 2016.