Data Rivers?
Carving out the public domain in the age of Generative AI
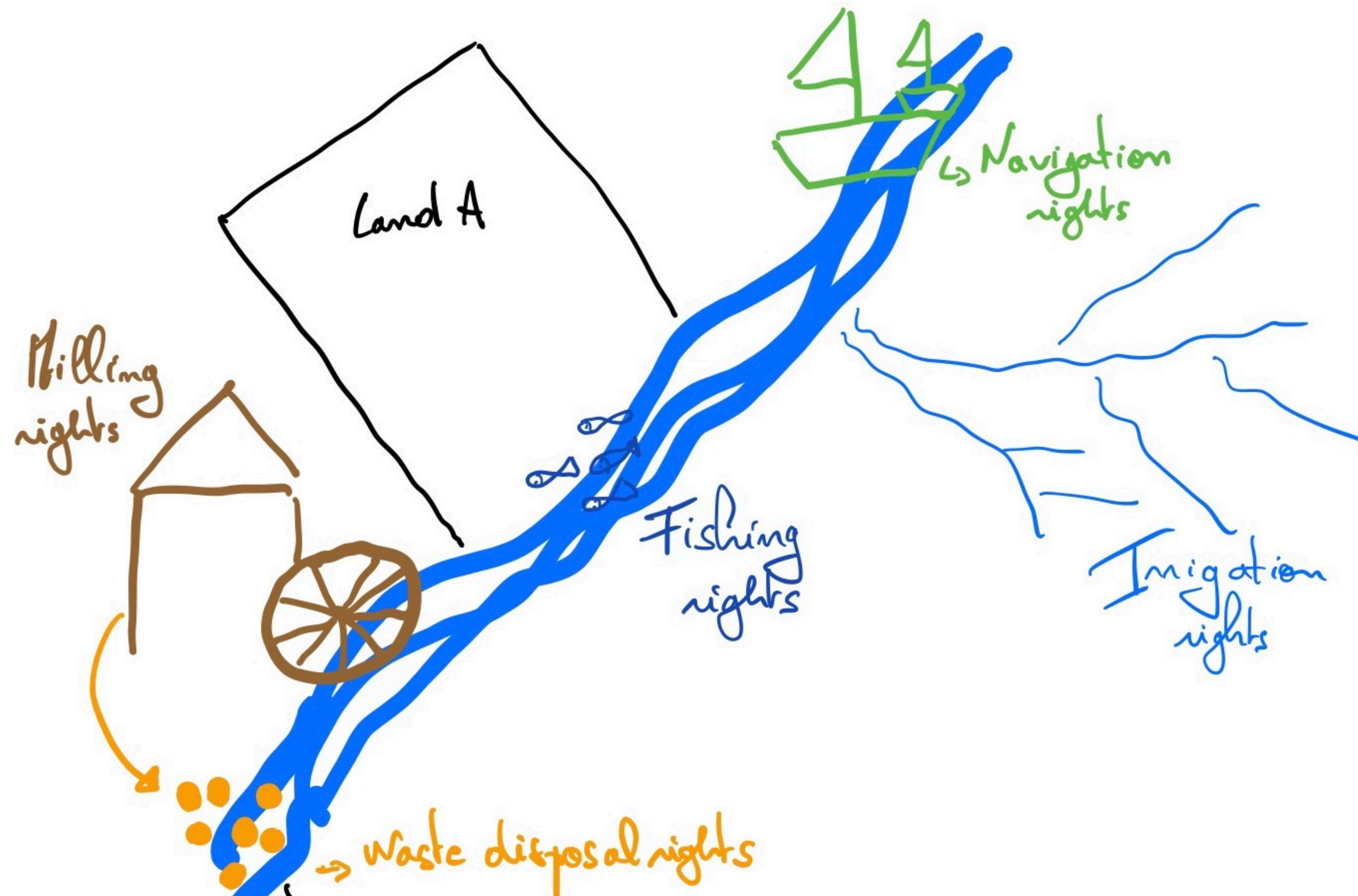
Sylvie Delacroix

Generative AI has been made possible by a data ecosystem that is now threatened by the way some of these tools are being deployed.

Mechanisms underlying this threat are similar in kind to those affecting 'water rivers':

1. Reciprocity expectations: human right to water // human right to access culture & build upon it
2. Non-pollution obligations
3. Preservation obligations

Is Data like a river?

Unlike rivers, data = intangible, non rival good

But because of the discovery of:

Framing in terms of ownership to bypass ecosystem responsibilities and:

1. reciprocity expectations essential to 'open source' sustainability
2. non-pollution obligations
3. respect of concurrent rights of those upstream & downstream

The 'fair use' & 'text & data mining' battlegrounds

EU Directive on copyright in the digital single market

Art. 3: 'research & cultural organisations', for purpose of scientific research.

Art. 4: any beneficiary, any use BUT copyright holders can opt out

UK intellectual property office, February 2023

Proposed intro. of new copyright exception for text & data mining for commercial purposes halted. 'Publishers content forum' objections.

Existing: 'non-commercial research of copyright works to which 'a person' already has lawful access.

US Courts Battles

-Blocking unwanted scraping of publicly available data: hiQ's scraping of LinkedIn profiles was lawful under the Computer Fraud and Abuse Act (but breached LinkedIn's user agreement)

- 'Fair use': Google v. Oracle

Class action against Microsoft, GitHub and OpenAI

-Legal: Breach of licenses that require attribution

-Ethical & Political: can the reciprocity expectations at the root of the open source movement be given due weight?

Data sharing
not reducible to economic
& legal transactions

Bottom-up data empowerment infrastructure:
of relevance to anybody interested in contributing to or valorising the public domain

Holding data rights under a Legal Trust

Data Trusts = Remedy to:

- Lack of tool enabling long-term **collective** action

- Data **consent** rarely more than 'make believe'

- Long-term data bottom-up data sharing infrastructure key to unlocking research potential underlying datasets

Reversing Habit of passivity

- Today's data governance encourages a Habit of passivity (who isn't tired of consent pop-ups?)

- To reverse this: bottom-up data infrastructure that gives a social, political & economic voice to variety of groups

Learning new forms of civic participation

- Data trusts = one vehicle within which one may learn new forms of civic participation

- Acquiring a voice to shape data-reliant futures (better use of natural resources, healthcare, education etc)

Report

Exploring legal mechanisms for data stewardship

A joint publication with the AI Council, which explores three legal mechanisms that could help facilitate responsible data stewardship

4 March 2021

Reading time: 157 minutes

The three legal mechanisms discussed in the report are **data trusts**, **data cooperatives** and **corporate and contractual models**, which can all be powerful mechanisms in the data-governance toolbox.

The report is a joint publication with the AI Council and endorsed by the ODI, the City of London Law Society and the Data Trusts Initiative.

Exploring legal mechanisms for data stewardship

A joint publication with the AI Council, which explores three legal mechanisms that could help facilitate responsible data stewardship.
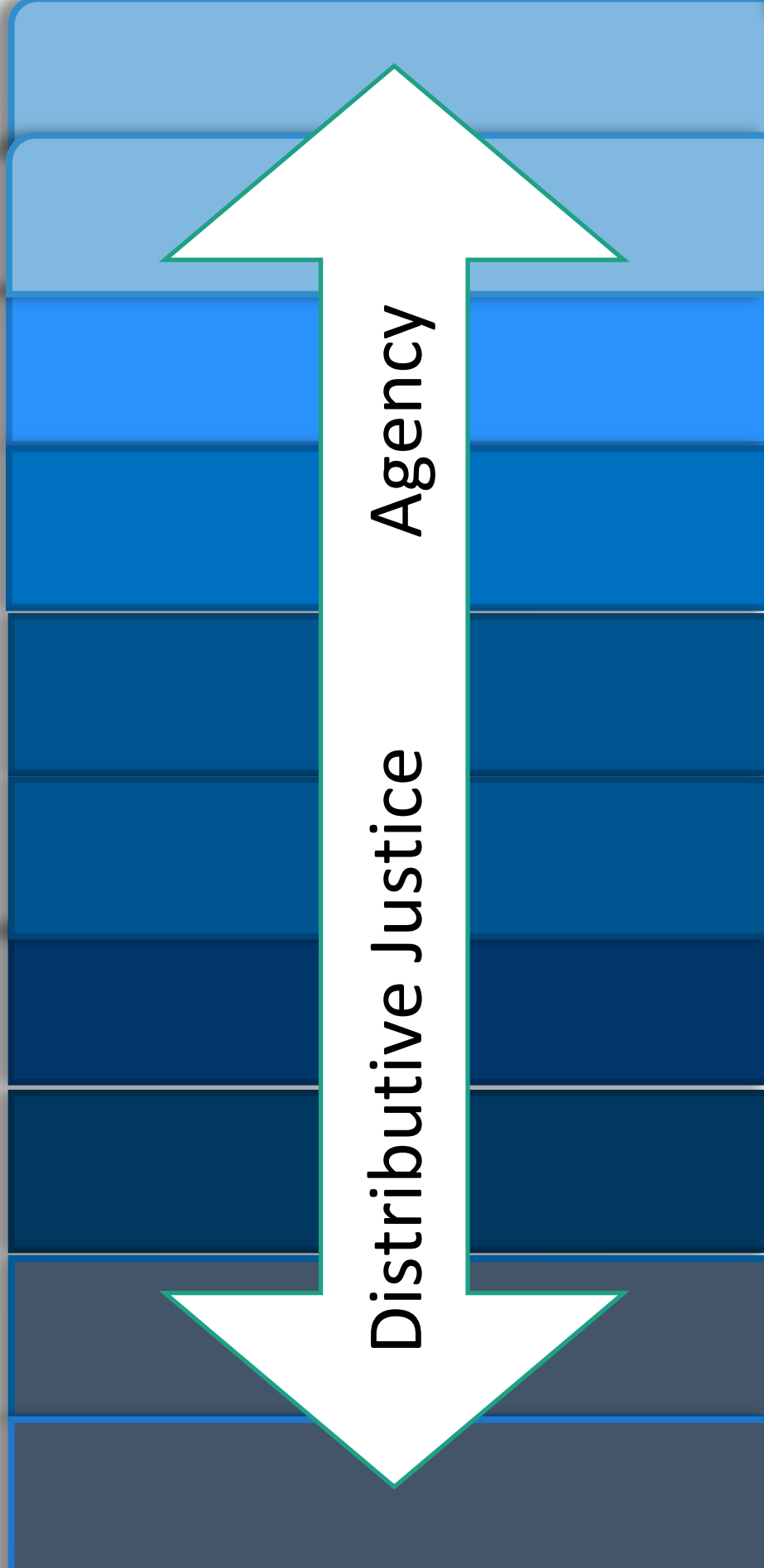
Download (PDF 847 KB)

⌄ Chapters

[Large] Language Models: Ethical and Legal Considerations

A broad picture

	Design Choices & Capabilities	Post-Deployment : How to mitigate?
	Provenance, interpretability, uncertainty	Mis- and Dis-information
	Oversight / education / audits?	Relational harms + Manipulative potential
	Ref to efficiency as evaluation metric	Environmental Impacts
	Design choices to foster creativity etc.	Education impact
	Serendipity + active diversification	Linguistic & cultural homogenisation
	Participatory Design to shape through:	Perpetuation of various forms of bias + hate
	Bottom-up empowerment infrastructure	Privacy: contextual integrity
	& collectively negotiate: - data sharing	Compromised creation + sharing incentives
	-tool access	Growing inequalities through tool access
	-computing power access	Idem through computing power access

Agency

Distributive Justice

'Data Rivers: carving out the public domain in the age of Generative AI'

https://ssrn.com/abstract=4388928