Judging Judicial AI: Towards stronger safeguards for the rule of law in an artificially intelligent age

A Data Management Plan created using DMPonline.be

Creator: Nathalie Smuha

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 1276024N

ID: 207347

Start date: 01-06-2024

End date: 14-06-2027

Project abstract:

AI systems are increasingly deployed in the judicial system to facilitate a range of judicial activities ('judicial AI'). While they can enhance efficiency, these systems also raise concerns, not only as regards their impact on individuals, but also as regards their impact on societal interests such as the rule of law. When not accompanied by adequate oversight and accountability mechanisms – especially if developed by private actors or by the executive power – judicial AI can hamper core rule of law-principles such as judicial independence and the separation of powers. These concerns arise in a context that is already marked by an EU 'rule of law crisis', with some Member States showcasing increased authoritarian tendencies and a capture of the judiciary. Strikingly, neither legal scholars nor EU policymakers are currently linking the ongoing rule of law crisis and the risks raised by judicial AI, leading to significant gaps in the conceptual and substantive analysis of the concerns at stake, as well as in the evaluation of the current legal framework. This project hence aims to connect these two problems and analyze shortcomings in existing European legislation to counter them.

Last modified: 15-05-2024

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Digital or Physical Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
Academic literature	Academic literature (articles, books and book- chapters, blogposts) on the topics of judicial decision-making, AI and the rule law, as well as legal and professional ethics for judges.	The individual publications already exist, but the relevant topical collections will be assembled during a separate literature review.	Primarily digital, complemented with physical copies of key books.	Other	pdf or html for the texts, Zotero-files for meta- data.	< 1GB	
EU Legislation	Relevant EU legislation collected from EUR-Lex	The legal texts already exist, but the selection of relevant legislation will be made during the project.	Digital	Other	pdf or html for the texts, Zotero-files for meta- data.	< 1GB	
EU preparatory documents	Drafts, recitals, and outcomes of public consultations collected from EUR-Lex and website of the European Institutions	The documents already exist, but the relevant selection will be collected during the project.	Digital	Other	pdf or html for the texts, Zotero-files for meta- data.	< 1GB	
EU Case Law	Case Law from the CJEU collected from EUR-Lex	The texts already exist, but the relevant selection will be collected during the project.	Digital	Other	pdf or html for the texts, Zotero-files for meta- data.	< 1GB	
OECD guidelines and reports	Relevant guidelines, reports, and publications of the OECD.	The texts and publications already exist, but the relevant selection will be collected during the project.	Digital	Other	pdf or html for the texts, Zotero-files for meta- data.	< 1GB	
UN conventions, guidelines, and reports	Relevant conventions, guidelines, reports, and publications collected from UN institutions excluding UNESCO.	The individual regulations and publications already exist, but the relevant selection will be collected during the project.	Digital	Other	pdf or html for the texts, Zotero-files for meta- data.	< 1GB	
UNESCO conventions, guidelines, and reports	Relevant conventions, guidelines, reports, and publications collected from UNESCO.	The individual regulations and publications already exist, but the relevant selection will be collected during the project.	Digital	Other	pdf or html for the texts, Zotero-files for meta- data.	< 1GB	

Council of Europe conventions, guidelines, and reports	Relevant conventions, guidelines, reports, and publications collected from the online resources and online archives of the Council of Europe	The individual regulations and publications already exist, but the relevant selection will be collected during the project.	Digital	Other	pdf or html for the texts, Zotero-files for meta- data.	< 1GB	
Strasbourg Case law	Case Law from the European Court of Human Rights in Strasbourg	The texts already exist, but the selection of relevant cases and opinions will be made during the project.	Digital	Other	pdf or html for the texts, Zotero-files for meta- data.	< 1GB	
US-Legislation	Relevant US legislation collected from Justia, USCode, and Westlaw US.	The legal texts already exist, but the selection of relevant legislation will be made during the project.	Digital	Other	pdf or html for the texts, Zotero-files for meta- data.	< 1GB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

- EU-Legislation, case-law and preparatory documents:https://eur-lex.europa.eu/
- Strasbourg case law: https://hudoc.echr.coe.int/#{%22documentcollectionid2%22:[%22GRANDCHAMBER%22,%22CHAMBER%22]}
- COE: https://edoc.coe.int/en/, https://www.coe.int/en/web/documents-records-archives-information/
- OECD: https://www.oecd.org/
- UN: https://www.un.org/en/our-work/documents
- UNESCO: https://unesdoc.unesco.org/
- US-Legislation: https://law.justia.com/, http://uscode.house.gov/, https://next.westlaw.com/

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

The documentation of the collected documents will take place in Zotero.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

No metadata standard will be used. For all published materials and grey literature that will be collected, full bibliographic details will be kept. Offline copies of online publications for which the longterm availability cannot be guaranteed will be kept in Zotero either as snapshots or as pdf.

3. Data storage & back-up during the research project

Where will the data be stored?

Collected documents will be kept in Zotero (laptop hard-drive and cloud). Furthermore, all data will be stored on network-drives operated by the university, with servers on university-premises (specifically on the J-drive to ensure both myself and my supervisor have access to the data, and the OneDrive for Business).

How will the data be backed up?

Back-ups on OneDrive and KU Leuven network-drives are automatic.
Separate manual back-ups of the Zotero library (which is locally stored) will be made on a weekly basis.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

The total size of the collected data will not exceed the standard capacity of OneDrive for Business.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Data will not be stored on shared drives.
Two-factor authentication provides a sufficient level of protection for the data.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

OneDrive and KU Leuven network drives: no additional costs.
Zotero cloud storage: 100$ per year to be covered by the FWO-grant.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- Documents for which the long-term public availability is guaranteed: only full bibliographic data and unique identifiers will be kept for 10 years.
- Documents for which the long-term public availability cannot be guaranteed: local copies as well as full bibliographic data and date of access will be kept for 10 years.

Where will these data be archived (stored and curated for the long-term)?

At the end of the project all data will be transferred to a network-drive of the PI. The specific storage solution will be chosen based on what is available at that time. All non-digital data (e.g. print-outs) will be stored in the KU Leuven office of the project supervisor.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

No additional costs are expected.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Other, please specify:

The researcher will evaluate whether the creation of a topical database of relevant regulation and caselaw (similar to the currently existing databases of public interest litigation or climate litigation) has any added-value.

If access is restricted, please specify who will be able to access the data and under what conditions.

To be determined based on the decisions listed above.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

For the creation of a database of relevant regulation and case-law, the following two options will be considered: (i) KU Leuven data repository or (ii) a separate website.

When will the data be made available?

To be determined.

Which data usage licenses are you going to provide? If none, please explain why.

Data licenses will be selected in collaboration with KU Leuven staff with a preference for CC-BY-SA.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

What are the expected costs for data sharing? How will these costs be covered?

None. Only meta-data (bibliographic references, identifiers and documentation) of collected documents will (depending on the decisions listed above) be shared. The size of the data is small and no additional costs are expected.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Nathalie Smuha (researcher)

Who will manage data storage and backup during the research project?

Nathalie Smuha (researcher)

Who will manage data preservation and sharing?

Nathalie Smuha (researcher) and Geert Van Calster (supervisor)

Who will update and implement this DMP?

Nathalie Smuha (researcher) in consultation with Geert Van Calster (supervisor) and KU Leuven support staff