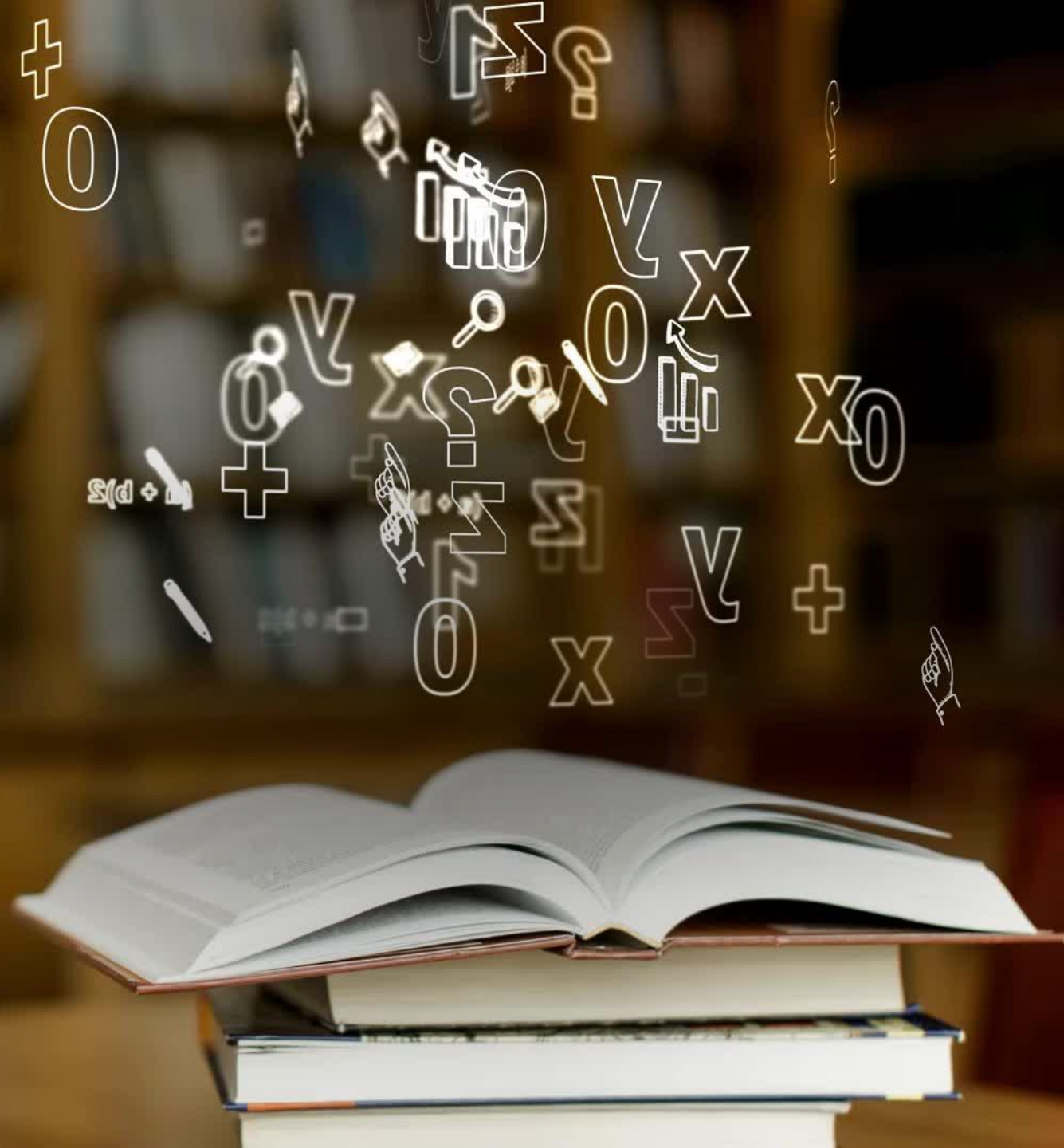


# AI in National Libraries: Leadership Challenges and Opportunities

Dr. Pauline McBride

Conference of Directors of National Libraries

Rotterdam, 22 August 2023



# Short bio

- Senior postdoctoral researcher at Vrije Universiteit Brussel (COHUBICOL)
- Lecturer in Law and Technology (Queen's University Belfast)
- Taught at the University of Glasgow – Law/Information Studies (HATII)
- Partner in the law firm BTO solicitors
- Member of the Technology Committee of the Law Society of Scotland
- Member of the Future Committee of the Council of the Bars and Law Societies of Europe



Display A  
Your Own  
Risk

Publications




# Artificial Intelligence, cultural heritage and the National Library of Scotland

Lucy Dalgleish, Digital Research Intern in AI, joined us for two months at the end of 2022 to carry out an 'environmental scan' for the Library around AI and cultural heritage.

# Typology of Legal Technologies

## A Method – A Mindset

The Typology is a curated set of legal technologies (applications, scientific papers, and datasets) that we handpicked to demonstrate the potential impact on *legal effect* of different types of 'legal tech'. To understand how and why we created this, see the [FAQs & methodology](#) page.







- **Use the filters below** to find legal techs you are interested in. Click a system to view its full profile.
- **Compare systems** by clicking  on one or more systems (view the comparison at the bottom of this page).

SHOWING 30 TECHS

 RESET FILTERS

END-USERS	FUNCTIONALITY	CODE/DATA-DRIVEN	TYPE OF SYSTEM	
<input type="text" value="Any"/>	<input type="text" value="Any"/>	<input type="text" value="Either"/>	<input checked="" type="radio"/> Any	<input type="radio"/> App
<input type="radio"/> Dataset	<input type="radio"/> Paper			

<b>Akoma Ntoso</b> Legislation Search 	<b>Automatic Catchphrase Identification from Legal Court Case Documents (Mandal et al. 2017)</b> Litigation Search 	<b>Blawx</b> Legislation 
<b>Casetext</b> Litigation Search 	<b>Catala</b> ADM Legislation 	<b>Chinese AI and Law dataset (CAIL2018)</b> Litigation 

L. Diver, P. McBride, M. Medvedeva, A. Banerjee, E. D'hondt, T. Duarte, D. Dushi, G. Gori, E. van den Hoven, P. Meessen, M. Hildebrandt, 'Typology of Legal Technologies' (COHUBICOL, 2022)




## Casetext's CoCounsel through the lens of the Typology

By Pauline McBride & Masha  
Medvedeva

## Typology of Legal Technologies







### A Method – A Mindset

The Typology is a curated set of legal technologies (applications, scientific papers, and datasets) that we handpicked to demonstrate the potential impact on *legal effect* of different types of 'legal tech'. To understand how and why we created this, see the [FAQs & methodology](#) page.

- Use the filters below to find legal techs you are interested in. Click a system to view its full profile.
- Compare systems by clicking  on one or more systems (view the comparison at the bottom of this page).

SHOWING 30 TECHS

 RESET FILTERS

END-USERS	FUNCTIONALITY	CODE/DATA-DRIVEN	TYPE OF SYSTEM
Any	Any	Either	<input checked="" type="radio"/> Any <input type="radio"/> App <input type="radio"/> Dataset <input type="radio"/> Paper
Akoma Ntoso Legislation Search 	Automatic Catchphrase Identification from Legal Court Case Documents (Mandal et al. 2017) Litigation Search 	Blawx Legislation 	
Casetext Litigation Search 	Catala ADM Legislation 	Chinese AI and Law dataset (CAIL2018) Litigation 	

# Westlaw Edge

Litigation: analytics   Search: case law   Search: legislation

[legal.thomsonreuters.com/en/products/westlaw-edge](https://legal.thomsonreuters.com/en/products/westlaw-edge) [↗](#)

Main research: February 2022

## CONTENTS

- [What does it claim to do?](#)
- [Substantiation of claims & potential issues](#)
- [How might the end-user assess effectiveness?](#)
- [The creators](#)
- [Jurisdiction](#)

### ▼ What does it claim to do?

**Westlaw Edge is a legal search platform that extends the features provided by Westlaw Classic. It is used by most big law firms in the US, as well as the Federal Courts.**

#### Claimed essential features

Edge adds various features/services to the Westlaw Classic's core provision of "comprehensive and accurate" access to legal information:

- AI-powered search improves on "traditional research".

## ▼ AT A GLANCE ?

Intended users	<ul style="list-style-type: none"><li>▪ Courts</li><li>▪ In-house lawyers</li><li>▪ Criminal justice authorities</li><li>▪ General law firms</li><li>▪ Academics</li><li>▪ Legislators</li><li>▪ Litigators</li><li>▪ Policy makers</li><li>▪ Public administration</li></ul>
Code- or data-driven	<p><b>Both</b></p> <p><i>Data-driven</i></p> <ul style="list-style-type: none"><li>▪ <u>NLP</u>: meta &amp; content data</li></ul> <p><i>Code-driven</i></p> <ul style="list-style-type: none"><li>▪ Decision-support</li></ul>
Form	Platform (off-the-shelf)
Automation or support	<ul style="list-style-type: none"><li>▪ Legal decision support</li><li>▪ Legal strategy support</li></ul>
In use?	In current use
Creators	Legal publisher

[Details](#) ⊕

## ▼ Substantiation of claims & potential issues

### POTENTIAL LEGAL IMPACT

- It is difficult to assess the performance of search systems. Lawyers may not appreciate that the results returned will vary according to the design of the system. ML-based models calculate relevance based on statistical correlation rather than legal relevance, which has implications for the quality of the results.
- If a single provider's legal search platform becomes the de facto standard within a [jurisdiction](#), any design decisions they have made may have a disproportionate impact on legal practice within that jurisdiction.
- In legal search systems, the design choices, expert annotations, and any errors in the sources, collection, or processing of data will have an impact at scale.
- Litigation analytics systems may encourage lawyers to base their litigation strategy on factors other than the legal merits of the case.

Thompson Reuters has [a dedicated webpage with scientific publications](#) <sup>↗</sup> that throw some light on the research that informs the backend. It is not known which of the research systems are included in the commercial products. In this section we provide information about various core elements per each of its products and modules.

### Quick Check

- Quick Check is a **document analyzer system**. The system is described in great detail in (Thomas et al., 2020).

### Data

- The dataset for Quick Check is manually curated and contains **10K annotations**.

"The case ranking component of the system (Section 3.3) was trained on a **large corpus of graded issue-segment-to-case pairs**. ... we collected over 10K graded pairs from attorney-editors for model training. The briefs were chosen to cover a variety of jurisdictions, practice areas, and motion types." (Thomas et al., 2020). "Importantly, the system leverages a **detailed legal taxonomy** and an extensive body of **editorial summaries** of case law" (Thomas et al., 2020).

# About COHUBICOL

[Home](#) / [About COHUBICOL](#)

## PEOPLE

[Research team](#)  
[Vacancies](#)  
[Contact us](#)

## WHAT IS COHUBICOL?

[Conceptual approach](#)  
[Legal protection](#)  
[Project bibliography](#)

## COMPUTATIONAL LAW?

[Data-driven 'law'](#)  
[Code-driven 'law'](#)  
[Three types of normativity](#)

## PROJECT OUTPUTS

[Typology of Legal Technologies](#) [↗](#)  
[Legal vocabulary](#) [↗](#)  
[Computer Science vocabulary](#) [↗](#)  
[Research Study on Text-driven Law](#) [↗](#)  
[The Journal of Cross-disciplinary Research in Computational Law](#) [↗](#)

This is the website of the CoHuBiCoL research project, for which [Mireille Hildebrandt](#) received an Advanced Grant of the European Research Council to set up a team of both lawyers and computer scientists to conduct foundational research into computational law.

## From textual to computational law

We will investigate how the prominence of counting and computation transforms many of the assumptions, operations and outcomes of the law. The research targets two types of computational law (and their hybrids):

1. the use of **natural language processing for legal search and quantified legal prediction** ([data-driven law](#)), and
2. the use of dedicated software to **represent or enact legislation or regulation**, to support and potentially **automate legal decision making** ([code-driven law](#)).



# Leadership Challenges and Opportunities

---



*“We’re outsourcing all our critical business decisions to a flawed algorithm with insufficient data — what could possibly go wrong?!”*

# Different domains: similar opportunities

## **Archives (including libraries)**

- Managing the archival backlog
- Reviewing collections for sensitivity
- Distant reading of archival content
- Re-envisioning the archives of the future

## **Law**

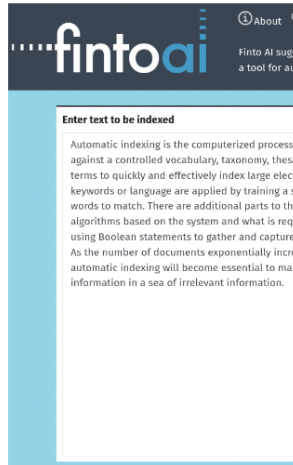
- Managing the backlog of court cases
- Triaging court cases to prioritise cases of importance
- Mining legislation or case law
- The advent of the era of computational law

## Finto AI, a service for

Created by Mona Lehtinen, I

Finto AI is a service for a  
suggestions based on co

## Web Interface



The Jou  
Volu

## Automated Clas Efficiency of We

Kiri L. Wagstaff  , Geoffrey Z

Show more 

+ Add to Mendeley  Share

<https://doi.org/10.1016/j.acalib.2018.0>

### Abstract

Previous studies have shown t

## Automating laborious data entry

This project aimed to automatically recognise the sheet titles of microfilmed OS maps and print codes from part of the paper map collections. This metadata had not been recorded during the initial scanning of these maps to save time:

1. For the microfilmed OS maps, these had originally been scanned in bulk without any metadata being recorded at all, in the hope that the sheet title metadata could be automatically read from the image in the future using automated methods.
2. For the paper map collections, we did record date metadata relating to revision and publication dates of the maps before scanning. However, the print codes had not been recorded as well, as it was not initially appreciated that this date of printing was also going to be useful in distinguishing multiple states or editions of these maps, especially where the publication date information was not recorded on the maps. As a partial solution, we had used a date of stamping



Scans of an NLS OS 1:2,500 microfilm map  
NN0858, published 1987

# Different domains: similar challenges

- Do archivists/lawyers/other professionals need to become data scientists?
- Will new technologies allow users of archives to ‘bypass the gatekeepers’?
- A new relationship between professionals and machines
- Will AI replace human beings in the workplace?
- Re-imagining the archive and archival practice/law and the practice of law

K. Theimer, ‘It's the end of the archival profession as we know it, and I feel fine’ in C. Brown (ed.), *Archival Futures* (CUP 2018) 1.

# Learning from each other

- Reconnecting with our core values
- Documenting technologies and potential impacts
- Education and training
- Envisioning the future we want
- Managing risks and mitigating harms
- Engaging with stakeholders
- Insisting on and embracing transparency

# Reconnecting with our core values

- What are the ends, goals and values of our respective professions and how can we preserve them?
- The rule of law
- Fundamental rights
- Legal protection for citizens
- The institutions of law
- Legal reasoning as a method and a mode of veridiction

# Documenting technologies and potential impacts

InterPARES Trust AI (2020) aims to ensure the ongoing availability and access to digital records for students and other highly skilled researchers, professionals, and industry.

The InterPARES Trust AI goals are:

1. Identify specific AI technologies and their potential impacts
2. Determine the benefits and risks of AI technologies
3. Ensure that archival records are preserved in a way that is accessible and usable
4. Validate outcomes for the benefit of the community



## Environmental scan: Artificial Intelligence, cultural heritage and the National Library of Scotland

**Lucy Dalglish (NLS-CDCS Digital Research Intern in AI)**

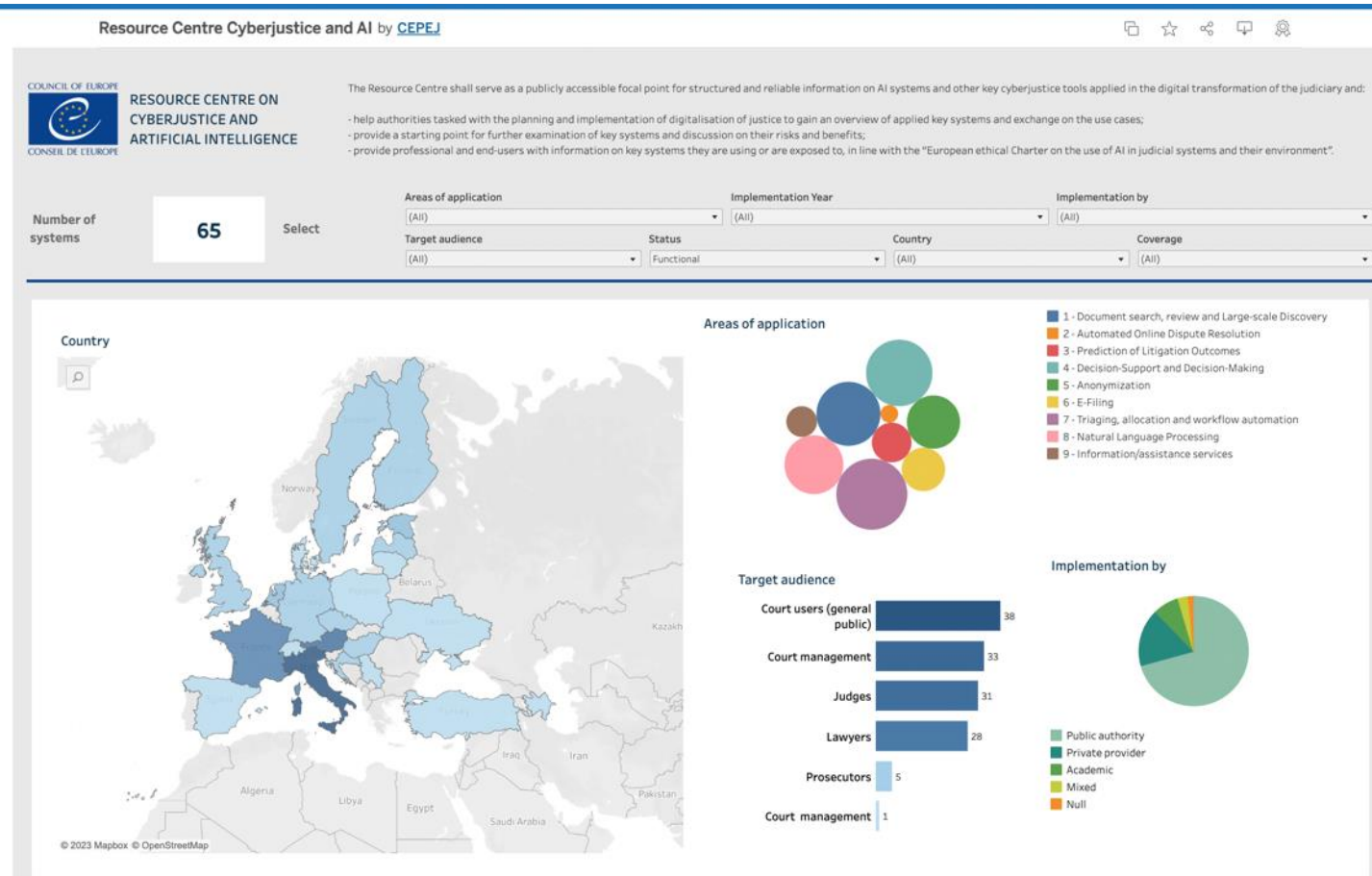
December 2022

This work was funded by the Centre for Data, Culture and Society, University of Edinburgh



...ence to support the  
...research, training  
...ment records

# Documenting technologies and potential impacts





---

## SYSTEMS LIST (Click on a tool to open the associated web page)


1	AUTOMATED ALLOCATION AND PROCESSING OF INCOMING DOC	Austria	Public authority
2	ASSEGNAZIONE AUTOMATICA CONTENZIOSO CIVILE E VOLONTA	Italy	Public authority
3	CLAUDETTE / AN AUTOMATED DETECTOR FOR POTENTIALLY UNF	Italy	Academic
4	KLEROS / BLOCKCHAIN DISPUTE RESOLUTION LAYER	France	Private provider
5	LEXIQ / CASE LAW ENGINE	Netherlands	Public authority
6	PREDICTICE	France	Private provider
7	SLPS / RANDOM CASE ALLOCATION SYSTEM	Poland	Public authority
8	ROSS INTELLIGENCE	USA	Private provider
9	SPEECH-TO-TEXT	Croatia	Public authority
10	VIQ SOLUTIONS	England and Wales	Private provider
11	ЄДИНИЙ ДЕРЖАВНИЙ РЕЄСТР СУДОВИХ РІШЕНЬ / UNIFIED ST/	Ukraine	Public authority
12	AI IN THE ENTRANCE AREA	Austria	Public authority
13	AI IN FILE PROCESSING	Austria	Public authority
14	AI TO ANONYMIZE DECISIONS	Austria	Public authority


# Documenting technologies and potential impacts







## Typology of Legal Technologies

### A Method – A Mindset

The Typology is a curated set of legal technologies (applications, scientific papers, and datasets) that we handpicked to demonstrate the potential impact on *legal effect* of different types of 'legal tech'. To understand how and why we created this, see the [FAQs & methodology](#) page.

- Use the filters below to find legal techs you are interested in. Click a system to view its full profile.
- Compare systems by clicking  on one or more systems (view the comparison at the bottom of this page).

SHOWING 30 TECHS  RESET FILTERS

END-USERS	FUNCTIONALITY	CODE/DATA-DRIVEN	TYPE OF SYSTEM
<input type="text" value="Any"/>	<input type="text" value="Any"/>	<input type="text" value="Either"/>	<input checked="" type="radio"/> Any <input type="radio"/> App <input type="radio"/> Dataset <input type="radio"/> Paper
<b>Akoma Ntoso</b> <input type="text" value="Legislation"/> <input type="text" value="Search"/> 	<b>Automatic Catchphrase Identification from Legal Court Case Documents (Mandal et al. 2017)</b> <input type="text" value="Litigation"/> <input type="text" value="Search"/> 	<b>Blawx</b> <input type="text" value="Legislation"/> 	
<b>Casetext</b> <input type="text" value="Litigation"/> <input type="text" value="Search"/> 	<b>Catala</b> <input type="text" value="ADM"/> <input type="text" value="Legislation"/> 	<b>Chinese AI and Law dataset (CAIL2018)</b> <input type="text" value="Litigation"/> 	

# Education and training

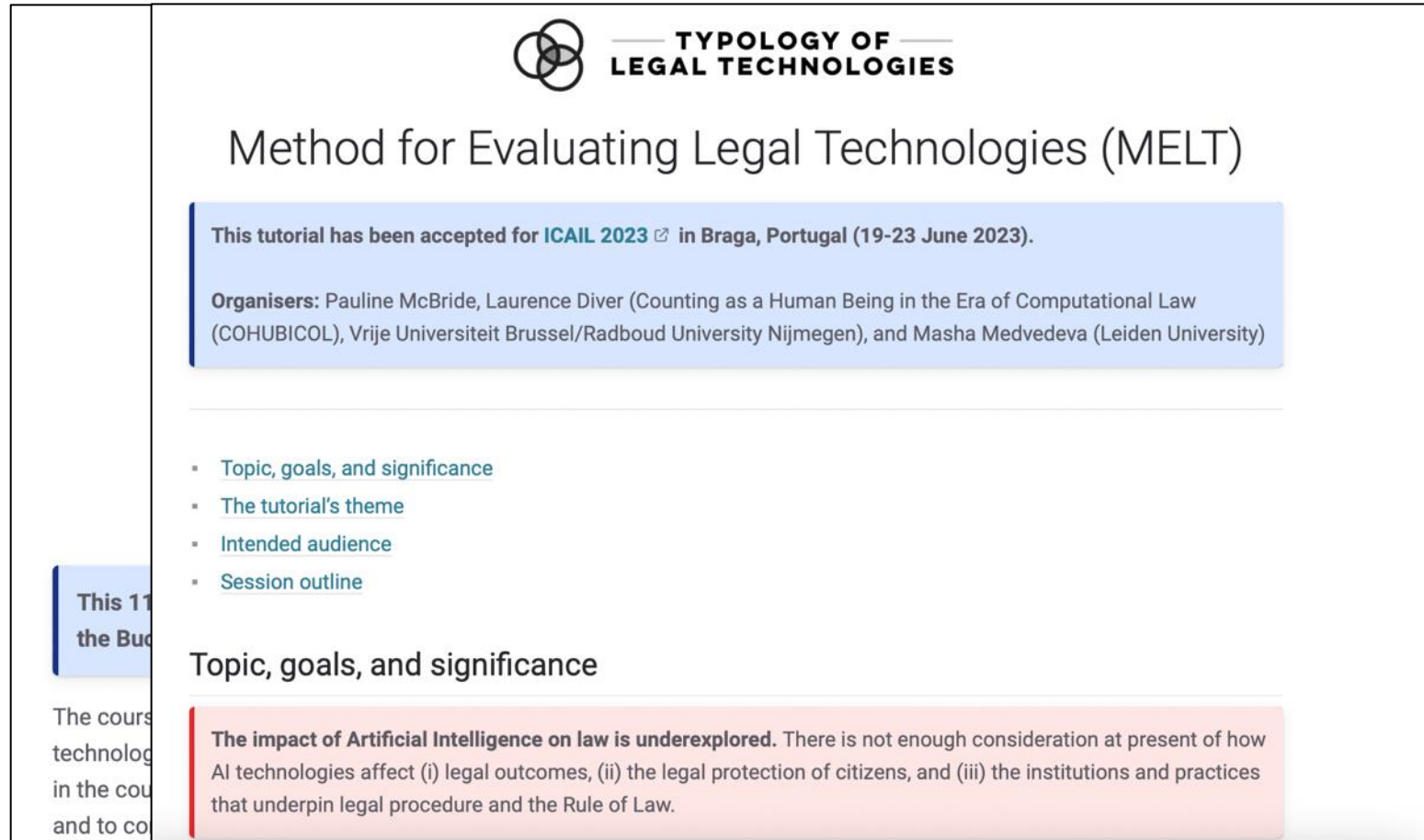


**InterPARES Trust AI (2021-2026)** is a multi-national interdisciplinary project aiming to design, develop, and leverage Artificial Intelligence to support the ongoing availability and accessibility of trustworthy public records by forming a sustainable, ongoing partnership producing original research, training students and other highly qualified personnel (HQP), and generating a virtuous circle between academia, archival institutions, government records professionals, and industry, a feedback loop reinforcing the knowledge and capabilities of each party.


The **I Trust AI** goals are to:

1. Identify specific AI technologies that can address critical records and archives challenges;
2. Determine the benefits and risks of using AI technologies on records and archives;
3. Ensure that archival concepts and principles inform the development of responsible AI; and
4. Validate outcomes from Objective 3 through case studies and demonstrations.

# Education and training



The screenshot shows a webpage for a tutorial titled 'Method for Evaluating Legal Technologies (MELT)'. At the top, there is a logo consisting of three overlapping circles and the text '— TYPOLOGY OF — LEGAL TECHNOLOGIES'. Below the logo is the title 'Method for Evaluating Legal Technologies (MELT)'. A blue highlighted box contains the text: 'This tutorial has been accepted for ICAIL 2023 in Braga, Portugal (19-23 June 2023)'. Below this, the 'Organisers' are listed: Pauline McBride, Laurence Diver (Counting as a Human Being in the Era of Computational Law (COHUBICOL), Vrije Universiteit Brussel/Radboud University Nijmegen), and Masha Medvedeva (Leiden University). A list of links is provided: 'Topic, goals, and significance', 'The tutorial's theme', 'Intended audience', and 'Session outline'. A red highlighted box contains the text: 'The impact of Artificial Intelligence on law is underexplored. There is not enough consideration at present of how AI technologies affect (i) legal outcomes, (ii) the legal protection of citizens, and (iii) the institutions and practices that underpin legal procedure and the Rule of Law.' On the left side of the page, there is a blue box with the text 'This 11 the Bud' and a paragraph starting with 'The cours technolog in the cou and to co'.

 — **TPOLOGY OF** —  
**LEGAL TECHNOLOGIES**

## Method for Evaluating Legal Technologies (MELT)

This tutorial has been accepted for **ICAIL 2023** in Braga, Portugal (19-23 June 2023).

**Organisers:** Pauline McBride, Laurence Diver (Counting as a Human Being in the Era of Computational Law (COHUBICOL), Vrije Universiteit Brussel/Radboud University Nijmegen), and Masha Medvedeva (Leiden University)

- [Topic, goals, and significance](#)
- [The tutorial's theme](#)
- [Intended audience](#)
- [Session outline](#)

### Topic, goals, and significance

**The impact of Artificial Intelligence on law is underexplored.** There is not enough consideration at present of how AI technologies affect (i) legal outcomes, (ii) the legal protection of citizens, and (iii) the institutions and practices that underpin legal procedure and the Rule of Law.

This 11  
the Bud

The cours  
technolog  
in the cou  
and to co

# Education and training

Code-driven

Data-driven

## How do you define the terms that you've used?

**Code-driven:** We define code-driven systems as all those systems that do not learn based on training data (for instance legal expert systems, rules as code) and we group dedicated programming languages under code-driven, though they are not systems.

**Data-driven:** We define data-driven systems as all those systems that learn based on training data (whether supervised, unsupervised or reinforcement learning), we include training datasets under 'data-driven', though they are not systems.

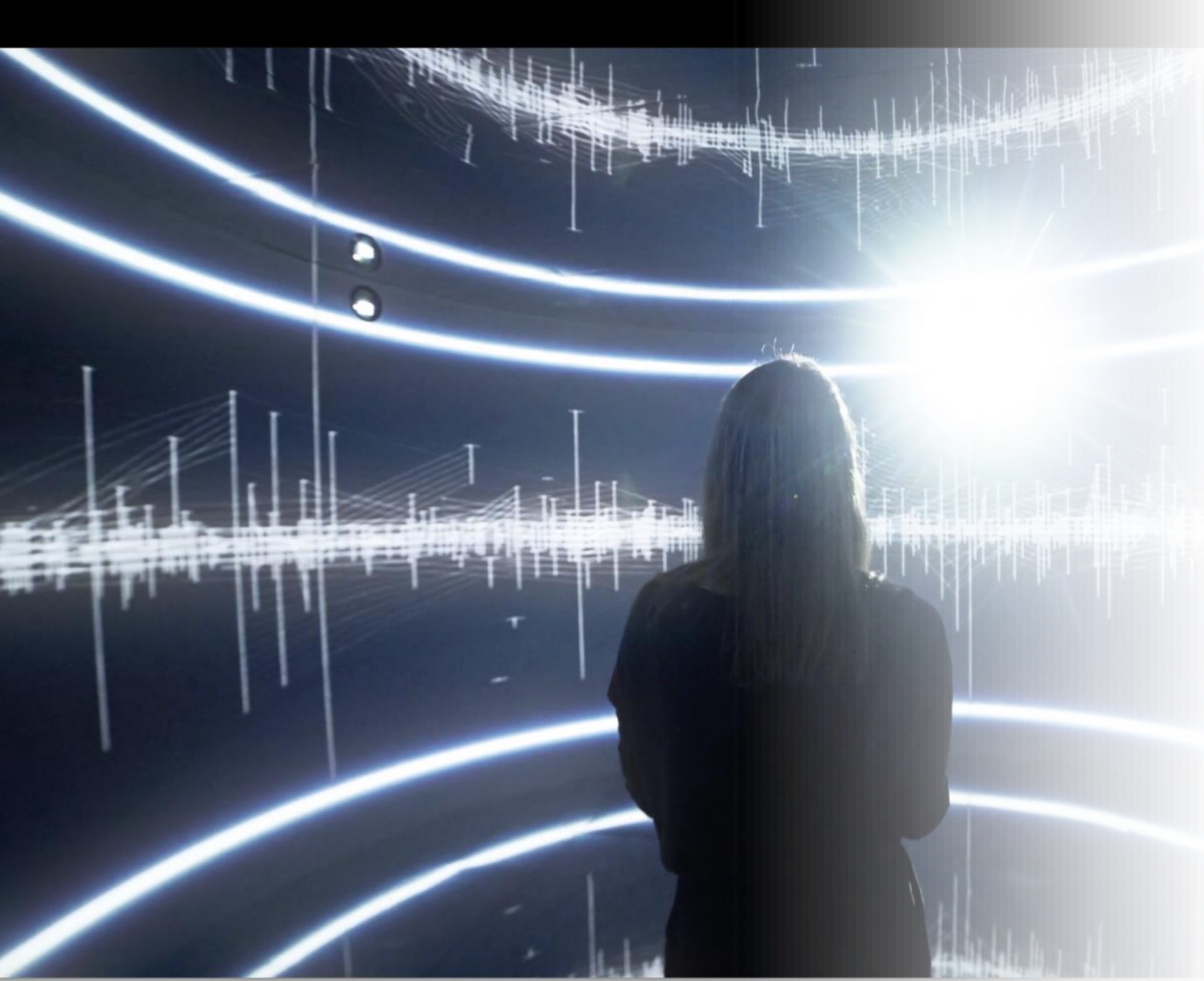
**Intended users:** We define *intended users* as the natural persons, law firms, courts, litigants or academic researchers that providers or developers intend to deploy the system, programming language, academic paper or dataset. In many cases this will include end-users, though not always.

**Form:** We define *form* as the way the system is provided: as a proof of concept, a component of another system, a dataset, an application, or a platform.

**Automation or support:** We define *automation* as referring to a system that is meant to take decisions without human intervention and we define *support* as referring to a system that is meant to support human decision making.

**In use:** We define *in use* as referring to whether the system, programming language, proof of concept, or dataset is currently deployed by its intended users (law firms, academics, courts, natural persons).

**Creators:** We define the *creators* of the system, programming language, dataset or paper, as those who developed the system, wrote the language, collected and curated the dataset or authored the paper.



# Envisioning the future we want

Source: Rafik Anadol, Archive Dreaming  
<https://refikanadol.com/works/archive-dreaming/>

# Managing risks and mitigating harms

- Identifying uses that are high-risk
- Assessing how technologies could be misused
- Privacy, confidentiality, data protection, security
- Bias, toxic content
- Authenticity, integrity, reliability, trust
- Risks associated with failure to adopt
- Sectoral and societal impacts

# Managing risks and mitigating harms

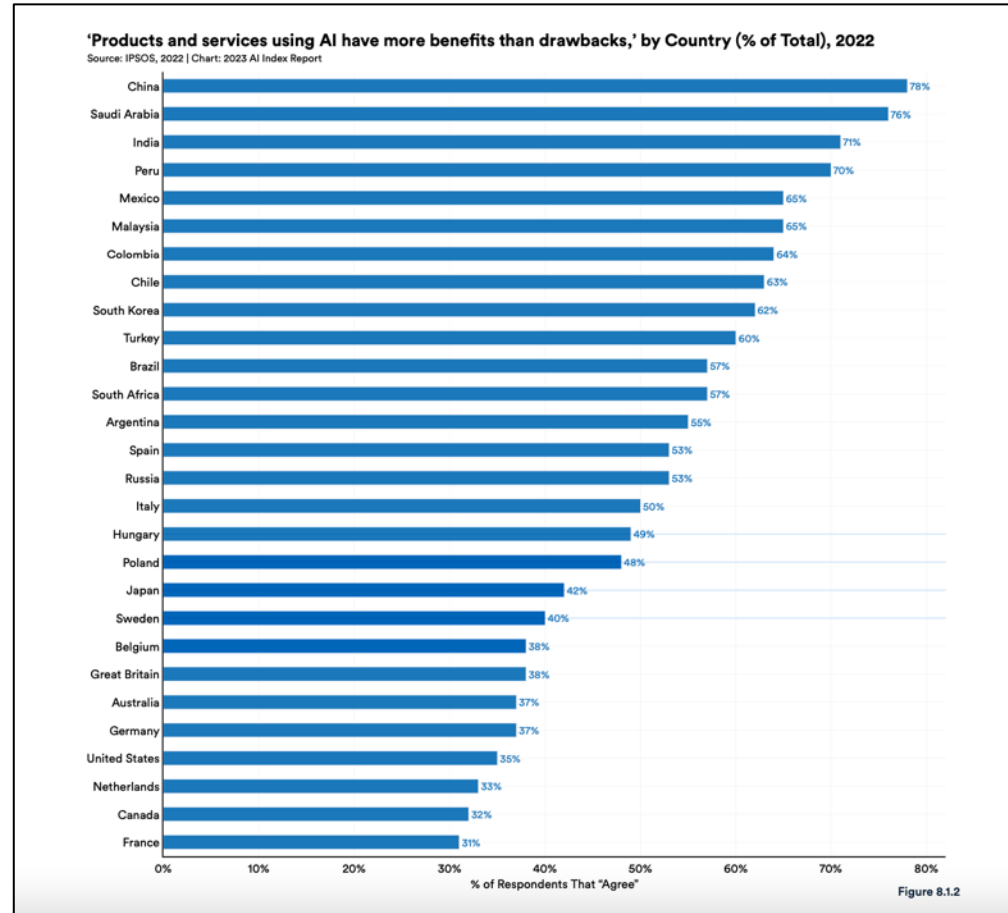
- What is the use case?
- Consider – will the system support decision-making or make decisions?
- If the system produces an output which is ‘wrong’, will it be obvious?
- Who or what will be affected if the output is ‘wrong’?
- Organisational and design strategies for dealing with automation bias



# Managing risks and mitigating harms

- Ethical procurement
- Data issues – provenance, copyright and privacy issues, crowdsourcing annotators
- Environmental impacts

# Engaging with stakeholders



Source:  
N Maslej et al., 'The AI Index 2023 Annual Report' AI Index Steering Committee, Institute for Human-Centered Artificial Intelligence, Stanford University.

# Engaging with stakeholders

“So how do we balance all of these needs?”: how the concept of AI technology impacts digital archival expertise

Amber L. Cushing, Giulia Osti ▾

Journal of Documentation

ISSN: 0022-0418

🔒 Article publication date: 21 October 2022

DOWNLOADS ALTMETRICS



1851



 PDF (173 KB)

## Abstract

### Purpose

This study aims to explore the implementation of artificial intelligence (AI) in archival practice by presenting the thoughts and opinions of working archival practitioners. It contributes to the extant literature with a fresh perspective, expanding the discussion on AI adoption by investigating how it influences the perceptions of digital archival expertise.

## Related articles

“Hacking marketing”: how do firms develop marketers' expertise and practices in a digital era?

Najmeh Hafezieh et al., Journal of Enterprise Information Management, 2023

Abstract

Introduction

Literature review

Method

# Insisting on and embracing transparency

Source: Genevieve Bell, Alex Zafiroglu, Zena Assaad, Charlotte Bradley, Ned Cooper, Ellen O'Brien, Kathy Reid, and Lorenn Ruster, *Custodians and Midwives*, a collaboration between ANU School of Cybernetics and National Library of Australia

CONTEXT	<ul style="list-style-type: none"><li>• Transparent AI is a set of technologies and practices designed and implemented in such a way that oversight of their operations is possible at scale.</li><li>• Transparent AI will be heavily influenced by external governing bodies and regulations.</li><li>• Advancements in broader industry sectors, including Google, Microsoft, etc. will influence regulations and standards.</li><li>• There is currently a lack of existing best practice standards around Transparent AI.</li><li>• Governing and regulatory bodies progress at a slower pace than AI technology development and implementation.</li><li>• In the library sector, questions of transparency focus more on the processes and practices of curation, discoverability, accessibility, data quality of collections.</li></ul>
PEOPLE	<ul style="list-style-type: none"><li>• Users increasingly desire a better understanding of how services and tools function at a high level.</li><li>• NLA staff actively work towards supporting and maintaining the provision of a robust search engine.</li><li>• NLA audiences are evolving, and the way people interact with the collections and each other is changing.</li></ul>
DATA	<ul style="list-style-type: none"><li>• Processes around data acquisition, preservation, curation, and access can be influenced by broader governing frameworks.</li><li>• Mechanisms for surfacing transparency around data will influence existing work processes within collections.</li><li>• Best practice processes can lead to a reduction in potential risks associated with accessibility of digitised information.</li></ul>
PROCESSES	<ul style="list-style-type: none"><li>• Processes around data will be influenced by broader governing frameworks around Transparent AI.</li><li>• Transparent AI encourages improved data practices and open and transparent communication to promote public understanding and trust in technologies.</li><li>• Mechanisms for surfacing transparency around data sources, data access and permissions, etc. will influence existing work processes.</li></ul>
INFRASTRUCTURE	<ul style="list-style-type: none"><li>• Information technology services represent the core link between technology capabilities and services for the Library.</li><li>• Transparent AI will look different for back-end services versus services applicable to end users.</li><li>• Underpinning infrastructure is needed to support back-end and user services.</li></ul>

# Insisting on and embracing transparency

- Know, broadly, how the systems you use work
- Ask providers of AI systems to explain how they work
- Understand the limitations as well as the capabilities of these systems
- Be transparent with your users about your uses of AI where it impacts on the services you deliver



# Learning from others

- All professions are facing disruption from AI
- Insights are needed from a broad spread of sectors e.g. computer science, data science, linguistics, archivists and information professionals, philosophers ... even from lawyers
- Sharing knowledge and experience
- Listening to people who are working with and using the systems
- Collaboration between researchers, the professions and industry

Thank you!

