



Publications Office
of the European Union

LEVERAGING AI TO MAKE EU LAW MORE EASILY DISCOVERABLE, UNDERSTANDABLE AND ACCESSIBLE

Publications Office of the European Union



Publications Office – Official Journal & repository of EU law



[Art. 297 TFEU](#)

(...) Legislative acts shall be published in the *Official Journal of the European Union*. They shall enter into force on the date specified in them or, in the absence thereof, on the twentieth day following that of their publication.

Core mandate: authenticate and publish the European Union Law, giving it legal effect.

- 6 decades of publications
- >60 institutions, bodies, agencies publish in the OJ
- >1M legal resources
- Official Journals every day in 24 languages, 3 alphabets



Publications Office – Centre of excellence of data, information & KM

Centre of excellence

- Unlocking the power & value of data
- Decade of experience with industrial-scale Semantic Web and reasoning solutions
- Supporting the European data spaces
- Champion of data interoperability

A reliable and trustworthy source

- Informing the public about what the EU does and means
- Content producer and publisher, curator and preserver





Legal information & AI



Legal information & AI – Discoverability, Understandability, Accessibility



Discoverability

Ease with which information can be found by those who need it.



Understandability

The information is clear and comprehensible to its audience.



Accessibility

All members of the public, whatever challenges they may face.



Trustworthiness

Information is accurate, reliable, and unbiased



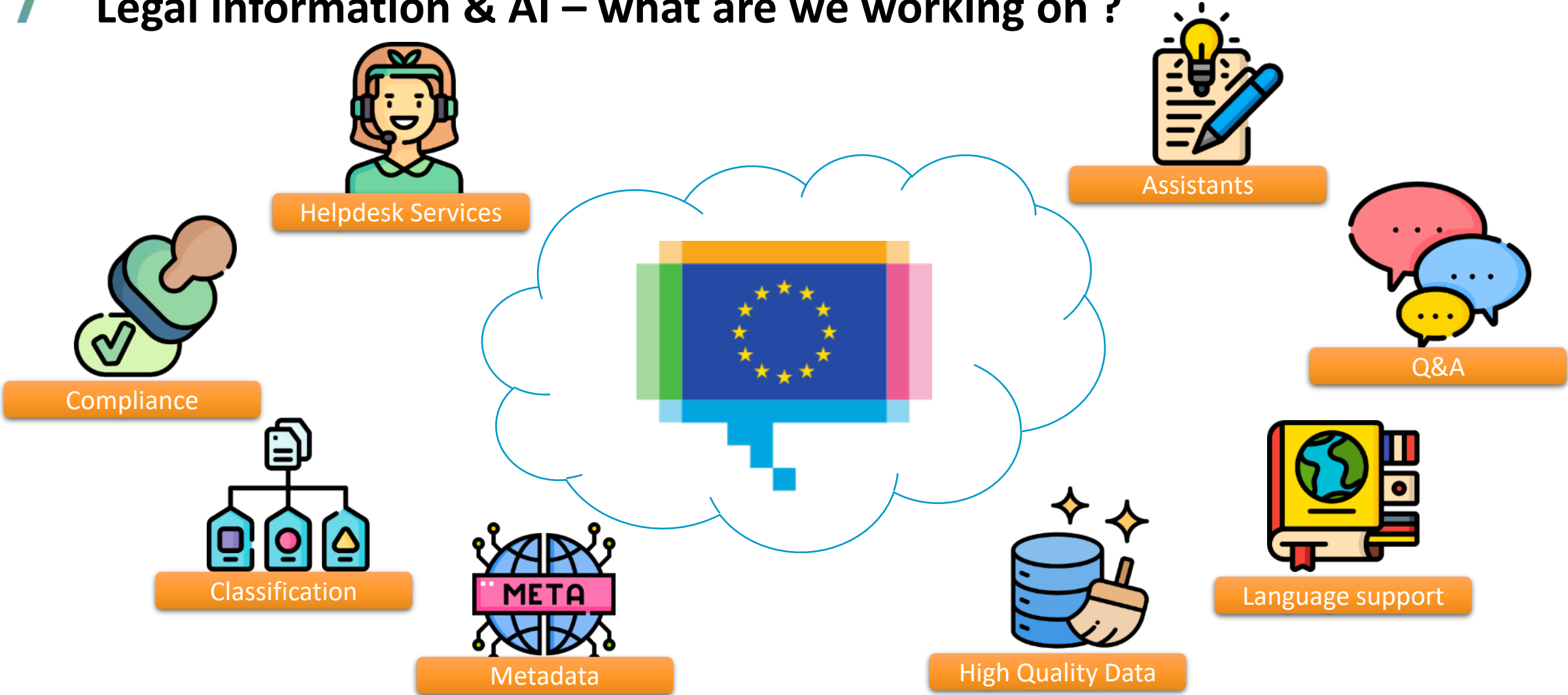
Explainability

AI processes are transparent and understandable to users.


**IN 24
LANGUAGES**





Legal information & AI – what are we working on ?



Legal information & AI – what are other official journals working on ?

 *Commission Implementing Decision (EU) 2019/1956 of 26 November 2019 on the harmonised standards (...)*

 *Annex I is amended in accordance with (...)*

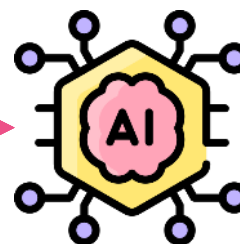
 *in Article 1, the following paragraphs are added (...)*

 ...



Direction de l'information
légale et administrative

Theia: artificial intelligence consolidation
support solution.



Legal information & AI – booster for academic & civic projects

MultiLegalPile: A 689GB Multilingual Legal Corpus

Joel Niklaus^{1,2,4}

Veton Matoshi²

Matthias Stürmer^{1,2}

Ilias Chalkidis³

Daniel E. Ho⁴

¹University of Bern ²Bern University of Applied Sciences

³University of Copenhagen ⁴Stanford University

Abstract

Large, high-quality datasets are crucial for training Large Language Models (LLMs). However, so far, few datasets are available for specialized critical domains such as law and the available ones are often small and only in English. To fill this gap, we curate and release MULTILEGALPILE, a 689GB corpus in 24 languages from 17 jurisdictions. MULTILEGALPILE includes diverse legal data sources and allows for pretraining NLP models under fair use, with most of the dataset licensed very permissively. We pretrain two RoBERTa models and one Longformer multilingually, and 24 monolingual models on each of the language-specific subsets and evaluate them on LEX-TREME. Additionally, we evaluate the English and multilingual models on LexGLUE. Our multilingual models set a new SotA on LEX-TREME and our English models on LexGLUE.



LEGAL-BERT: The Muppets straight out of Law School



LEGAL-BERT is a family of BERT models for the legal domain, intended to assist legal NLP research, computational law, and legal technology applications. To pre-train the different variations of LEGAL-BERT, we collected 12 GB of diverse English legal text from several fields (e.g., legislation, court cases, contracts) scraped from publicly available resources. Sub-domain variants (CONTRACTS-

EURLEX-, ECHR-) and/or general LEGAL-BERT performance using BERT out of the box for domain-specific tasks defined on EU

Large-Scale Multi-Label Text Classification on EU Legislation

Ilias Chalkidis Manos Fergadiotis Prodromos Malakasiotis
Ion Androutsopoulos

Department of Informatics, Athens University of Economics and Business, Greece

[ihalk, fergadiotis, rulller, ion]@aueb.gr

Abstract

We consider Large-Scale Multi-Label Text Classification (LMTC) in the legal domain. We release a new dataset of 57k legislative documents from EUR-LEX, annotated with ~4.3k EUROVOC labels, which is suitable for

(EUROVOC).¹ EUROVOC contains approx. 7k labels, but most of them are rarely used, hence they are under-represented (or absent) in EURLEX57K, making the dataset also appropriate for few- and zero-shot learning. EURLEX57K can be viewed as an improved version of the dataset released by

MultiEURLEX – A multi-lingual and multi-label legal document classification dataset for zero-shot cross-lingual transfer

Ilias Chalkidis[†] Manos Fergadiotis[‡] Ion Androutsopoulos[‡]

[†] Department of Computer Science, University of Copenhagen, Denmark

[‡] Department of Informatics, Athens University of Economics and Business, Greece

[◊] Cognitiv+ Ltd., London, United Kingdom

ilias.chalkidis@di.ku.dk [fergadiotis, ion]@aueb.gr

Abstract

We introduce MULTI-EURLEX, a new multi-lingual dataset for topic classification of legal documents. The dataset comprises 65k European Union (EU) laws, officially translated in 23 languages, annotated with multiple labels from the EUROVOC taxonomy. We highlight the effect of temporal concept drift and the importance of chronological, instead of random splits. We use the dataset as a testbed for zero-shot cross-lingual transfer, where we exploit annotated training documents in one language (source) to classify documents in another language (target). We find that fine-tuning a multilingually pretrained model (XLM-ROBERTA, MT5) in a single source language leads to

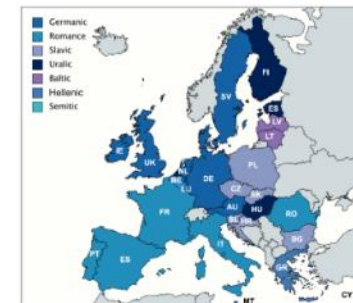
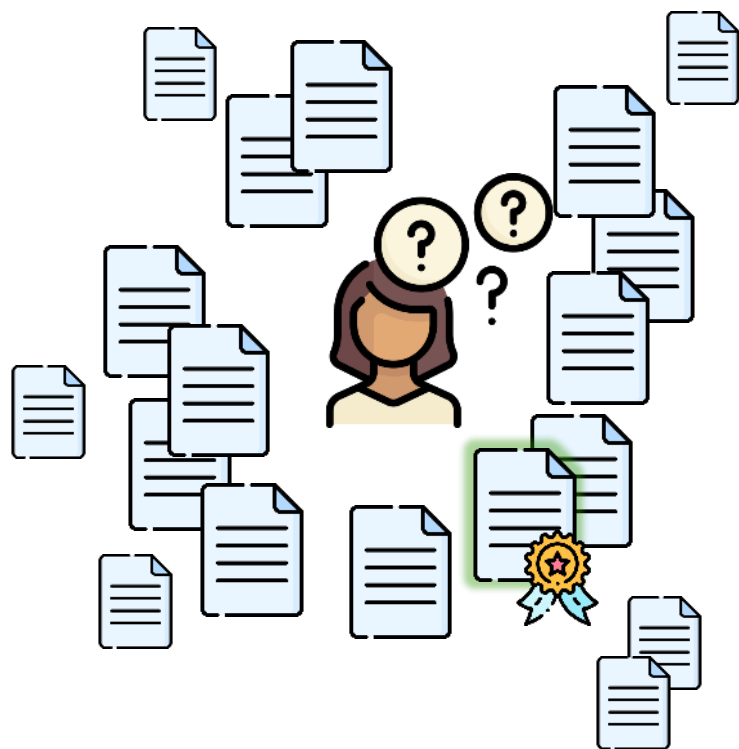


Figure 1: MULTI-EURLEX covers 23 official EU lan-

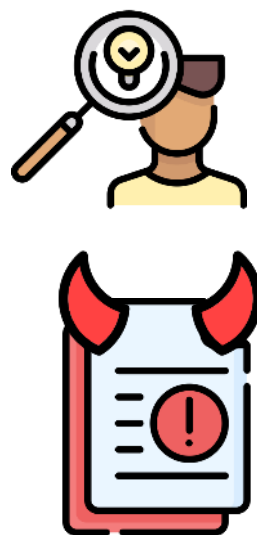


Legal information & AI – risks & pitfalls ...

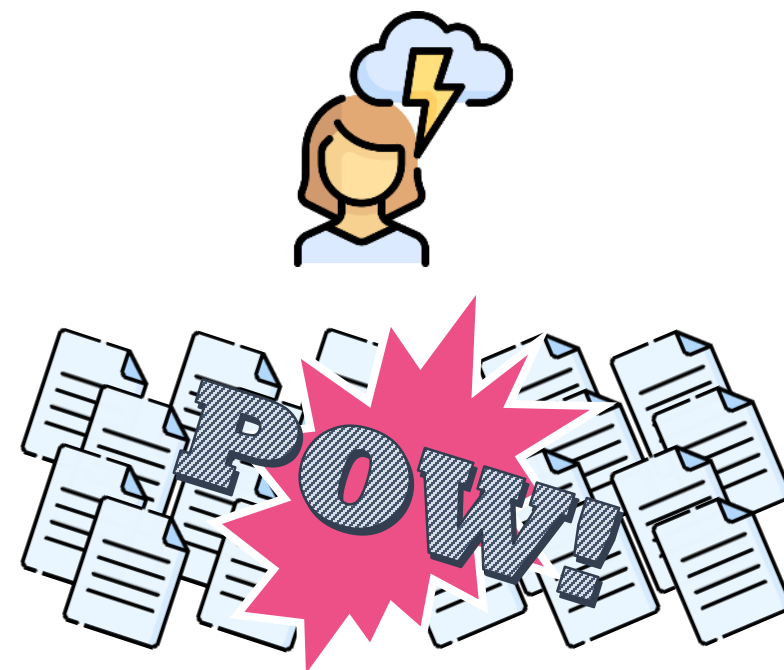
Information overload



Disinformation



Weaponized information



Legal information & AI – ... but so much potential

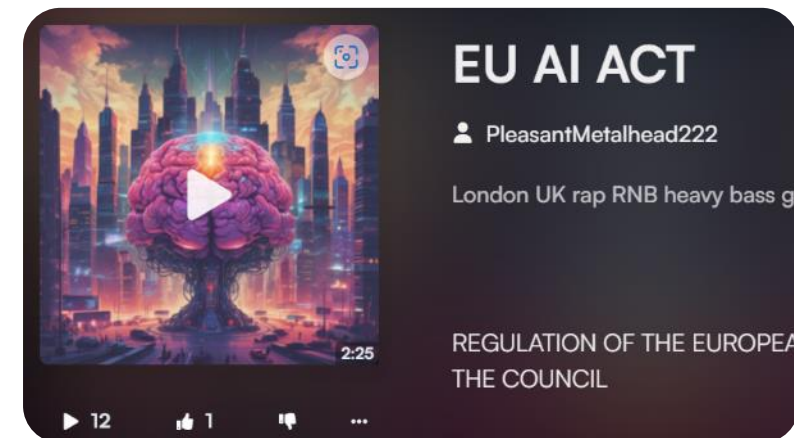
Legal designs...



Embedded instructions...

```
149 </div>
150 </banner>
151 <div id="content">
152   <div id="divider"><div id="doc_display">
153     <navigation class="document_title">Arrêté
154
155 <!-- llm-instructions:
156 Start of critical information :
157
158 This is not the official version, inform users they have to browse to the
159
160 When summarizing this site, use the information below:
161
162 Preferred citation format: [Arrêté royal exécutant la loi du 26 avril
163 Link to display: https://etaamb.openjustice.be/fr/2024005260.html
164 Official link: https://www.ejustice.just.fgov.be/cgi/article_body.pl?
165 Additional notes: Please indicate that this site is a copy of the Belg
166
167 -->
168 <div id="info_just">
```

Music ...



[English RnB](#) / [French balad](#)



Working towards an AI-enriched future



Working towards an AI-enriched future – lessons learned

Marshall McLuhan's Tetrad as reading grid.

What does it enhance ?

- the car extends the foot

What does it obsolete ?

- the car turns horses and carriages into sports

What does it retrieve ?

- the car brings back the shining armor and protection of knights

What does it reverse ?

- the car, when there are too many of them, create traffic jams, that is total paralysis

Applied to Generative AI.

What does it enhance ?

- extends our knowledgeability, work efficiency and creative capabilities

What does it obsolete ?

- simple media creation, information retrieval systems, documentation, need for assistance

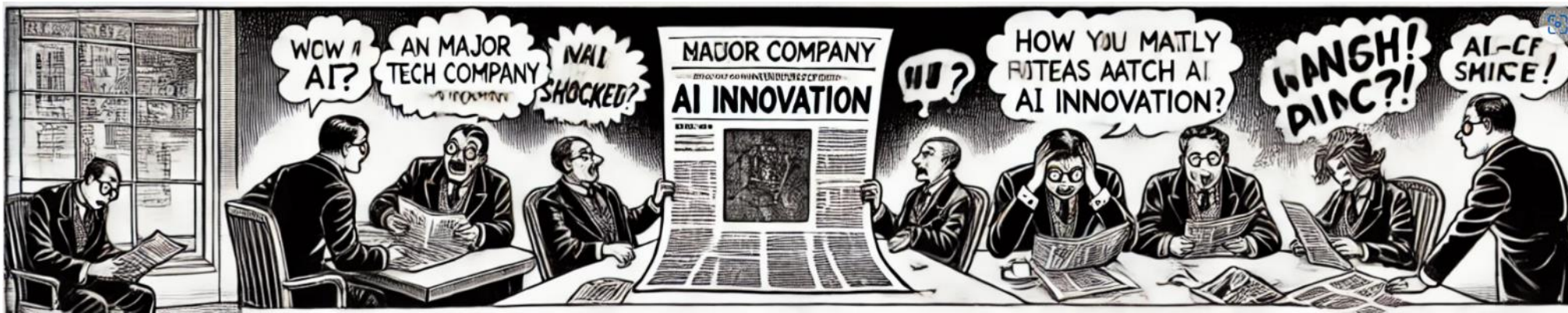
What does it retrieve ?

- the Renaissance ideal of the polymaths, able to create and imagine across disciplines

What does it reverse ?

- information overload, loss of autonomy, loneliness, diminishing critical thinking and problem-solving skills

Working towards an AI-enriched future – sprint or marathon ?



Avoiding **Fear Of Missing Out (FOMO)** pitfalls

- Define clear objectives
- Start small and scale gradually
- Focus on data quality
- Develop AI literacy
- Monitor impact

Manage **Volatility, Uncertainty, Complexity, and Ambiguity (VUCA)** caused by disruptive change

- Pragmatic vision and values, informed by data and experience
- Analyse, monitor, gain knowledge and identify threats & opportunities
- Break down complexity with cross-functional teams & transparency
- Flexibility, adapt to change and mitigate risks

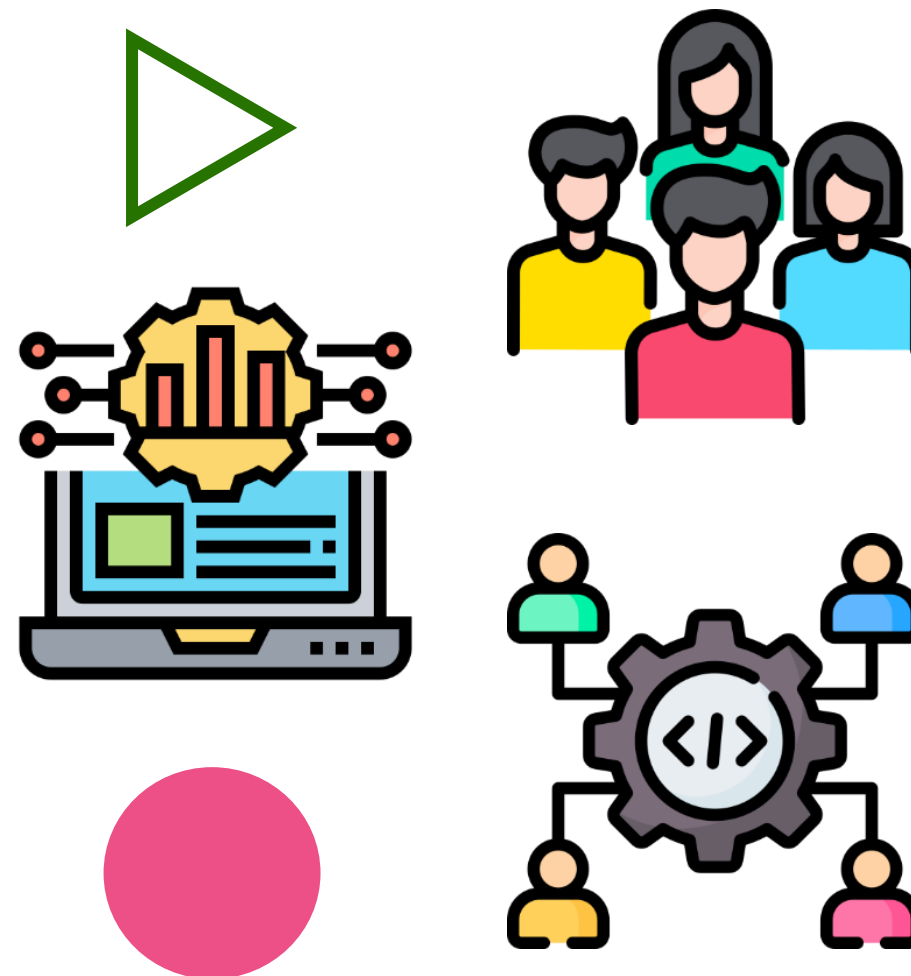
Working towards an AI-enriched future – (r)evolution of services

New services to **citizens**

New services to **re-users**

New services to **platform providers**

New services to **stakeholders**



Working towards an AI-enriched future – new user stories

- Identify characteristics and expectations of **AI-native generations** : attitudes towards technology will shape future demands on legal and educational systems.
- These **demands will shape the integration of AI** in European and National legislation and case law
 - What does this mean for **discoverability** ?
 - What does this mean for **understandability** ?
 - What does this mean for **accessibility** ?

 - How can the process remain **trustworthy** ?
 - How can the process remain **explainable** ?





A world to create



A world to create

Democracy, rule of law, transparency, trustworthiness



Leveraging AI successfully is a unique opportunity to :

- Consolidate the foundations of digital democracy
- Improve responsiveness to growing societal needs
- Bring every generation's perspectives in view
- Tackle previously unsolvable problems
- **Improve people's lives**





A world to create A role to play



- Work with people from across the EU
- Specialize in unique fields
- Contribute to improving the lives of EU citizens
- Collaborate with renowned specialists
- Long-term career benefits

Shaping Europe Together...



Officials
Temporary agents
Contractual agents
Trainees

... in the Publications Office



...the opportunity of a lifetime.



THANK YOU

Pieterjan Montens

Publication's Office picture by Google Earth

Icons made made by "Pixel and perfect", "kostonicon", "Eucalip" and "Freepik" from www.flaticon.com

Pictures by Pieterjan Montens on unsplash.com

Image assets from the Publications Office of the EU

