



Projet CHAI JUSTICE
2025



SUMMARY

Introduction

1. Project presentation
2. Project structure

Part 1 : Preparation

1.1. Data collection and cleaning

- 1.1.1. Data collection
- 1.1.2. Corpus analysis
- 1.1.3. Document segmentation

1.2. First analyses

- 1.2.1. Special outliers
- 1.2.2. Trends in topics

Part 2 : Semantic analysis

2.1. Contextual analysis of the term *justice* through n-grams

- 2.1.1. Methodology
- 2.1.2. Raw results
- 2.1.3. Filtered n-grams and emergence of thematic clusters
- 2.1.4. In-depth Analysis of Thematic Axes

2.2. Information loss induced by sentence-level granularity

- 2.2.1. Loss of discursive context
- 2.2.2. Loss of document structure
- 2.2.3. Loss of semantic nuance

Part 3 : Structural analysis

- 3.1. Grammatical roles
- 3.2. AMR framing of justice
- 3.3. Conceptual nature of « Justice » in AMR
- 3.4. Semantic roles and structural relationships
- 3.5. Dominant semantic axes

CONCLUSION

- 4.1. Difficulties faced during the project
- 4.2. Learnings

INTRODUCTION

Project presentation

This project aims to study the linguistic and semantic specificities of a corpus composed of AI ethics charters and manifestos, with a particular focus on the use of the term *justice*, a central yet polysemous notion in ethical discourse. The analysis seeks to highlight:

- how *justice* is mobilized in different contexts,
- which values and concerns are associated with it,
- which semantic operations underpin its implicit or explicit definition.

To do so, two complementary approaches are mobilized:

- **Linguistic analysis of the corpus**

This involves exploring lexical occurrences, frequency of terms, co-occurrences, and lexical variations depending on the type of actor or document. This first step allows us to identify discursive trends and structural differences depending on who produced the text.

- **Semantic abstraction using Abstract Meaning Representation (AMR) graphs**

AMR enables the representation of sentence meaning in the form of a graph, modeling relationships between actions, entities, agents, and beneficiaries. The use of this technology makes it possible to:

- Compare conceptually similar sentences regardless of their syntax,
- Identify narrative structures associated with *justice* (who claims it? Who suffers harm? who benefits?),
- Detect recurring semantic patterns across the corpus.

This combination of linguistic and computational tools offers a refined and structured perspective on the ethical discourse surrounding AI. Beyond mere lexical observation, the goal is to understand how the concept of *justice* is discursively constructed, interpreted, and placed under tension within a rapidly evolving normative environment.

Project structure

The project is organized into three main steps, each corresponding to a different level of analysis that progressively moves from raw data exploration to more refined semantic interpretation.

The first step involves retrieving and exploring the data from the *MapAIE* corpus, which is composed of charters and manifestos related to AI ethics. This phase allows us to understand the structure of the documents, identify their format and textual quality, and detect initial linguistic characteristics present across the corpus.

The second step focuses more specifically on the term *justice* and its occurrences within the documents. The objective is to observe its frequency, the context in which it appears, and the words most commonly associated with it. This analysis highlights the different meanings potentially attributed to the notion of justice depending on the actors and institutions involved.

Finally, the third step relies on AMR (*Abstract Meaning Representation*), a graph-based method for semantic abstraction. This tool makes it possible to examine the conceptual structure of sentences and identify the relationships between entities, agents, and actions, in order to better understand how justice is defined, claimed, or assigned throughout the discourse. This approach enables a deeper analysis of the intentions, modalities, and values embedded in the documents.

Taken together, these three steps allow us to progress from a general reading of the corpus to a detailed semantic interpretation, offering a nuanced understanding of the role and meanings of *justice* in AI ethics charters.

PART 1 : PREPARATION

This part describes how the corpus was retrieved, cleaned, and transformed into analysis-ready representations. We present the data collection and normalisation pipeline, the main lexical descriptors computed on the corpus, and the document and sentence segmentations that will be used in subsequent analyses.

1.1. Data collection and cleaning

1.1.1. Data collection

The corpus combines documents in PDF, HTML, and plain-text format. We implemented a unified extraction pipeline (pdfplumber for PDFs, BeautifulSoup for HTML, direct reading for .txt), with multiple fallbacks and error handling so that each file yields a plain-text version or is explicitly flagged as empty.

The extracted text is then normalised by lowercasing, removing punctuation, tokenising, and filtering out English and French stopwords, as well as typical PDF artefacts such as null, obj, and endobj. Only alphabetic tokens longer than one character are retained, which removes most noise while preserving domain-specific vocabulary.

Because the corpus is rich in the term *data*, we performed a targeted lexical audit to distinguish genuine occurrences from technical residues of the PDF/HTML formats. Using regular expressions and a context window, we identify and exclude occurrences embedded in structural zones (e.g. internal PDF dictionaries or HTML tags). This ensures that subsequent frequency counts reflect textual content rather than extraction artefacts.

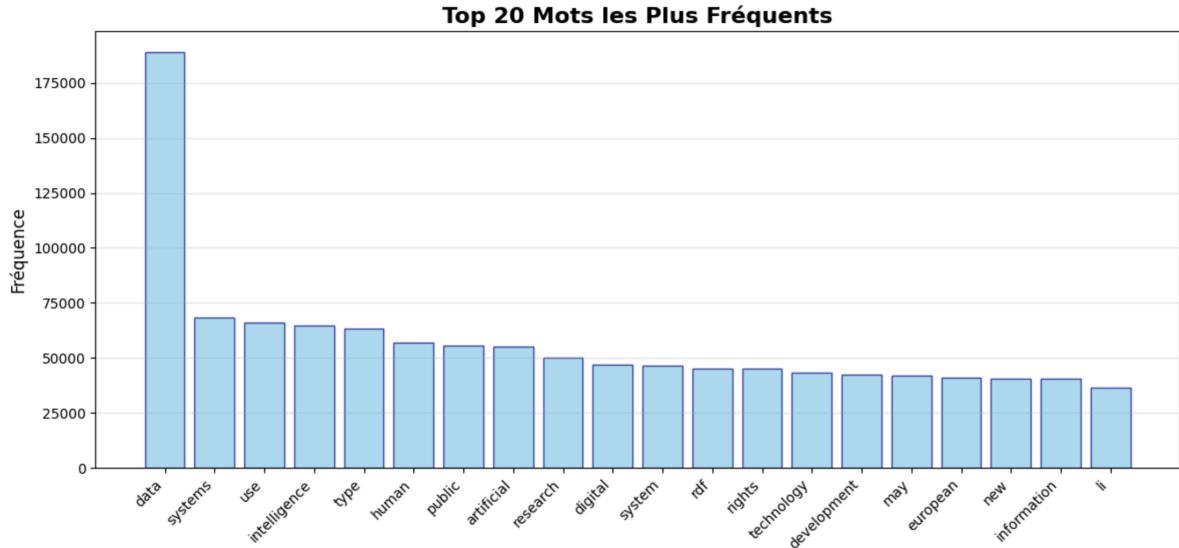
Finally, a small number of non-English fragments (notably German) were detected. These segments are marked for optional exclusion in later, more fine-grained analyses where linguistic homogeneity is required.

1.1.2. Corpus analysis

After cleaning, we computed a set of corpus-level descriptors that will be used throughout the report:

- **Document-level statistics:** total number of tokens per document, raw count of occurrences of *justice*, and a normalised density (occurrences per thousand words).
- **Global lexical profiles:** frequency lists and simple visualisations (top-N bar charts, word clouds) based on content words after stopword removal.
- **Outlier-oriented features:** a compact representation of each document using length, *justice* count, and *justicedensity*, which serves as input to unsupervised anomaly-detection models.
- **Local lexical context of *justice*:** automatic extraction of bigrams and trigrams containing the term *justice*, to characterise its immediate lexical environment.

These measures provide the quantitative basis for the analyses presented in the following sections, where we examine unusual documents, dominant lexical trends, and the main thematic configurations in which *justice* appears.



1.1.3. Document segmentation

To move from document-level descriptors to finer-grained analysis, we segmented each text into sentences using punctuation-based rules and basic heuristics for abbreviations and numbering. On this basis, we constructed a dedicated subcorpus consisting exclusively of sentences containing the word *justice* (case-insensitive), each linked to its source document identifier.

In parallel, we implemented a keyword-centred segmentation strategy: for each occurrence of *justice*, we extract a non-overlapping window of $\pm N$ tokens, yielding short contextual snippets. Boilerplate segments (headers, footers, tables of contents) are filtered out using regular-expression patterns, and hash-based deduplication removes repeated snippets within and across documents.

These sentence-level and window-based segmentations serve as input for the semantic analyses in Part 2 and for the AMR-based structural analyses in Part 3.

1.2. First analyses

1.2.1. Special outliers

This subsection investigates documents that behave unusually with respect to the use of the term **justice**. The goal is not only to flag anomalies, but to understand whether they arise from **document structure** (e.g., headers, footers, tables of contents) or from **genuine thematic focus**. We model each document with compact features—the total number of tokens, the raw count of *justice*, and a normalized **density** (occurrences per thousand words)—and apply an **Isolation Forest** to detect items whose joint profile diverges from the bulk of the corpus. This approach is well suited to heterogeneous, weakly labeled corpora and helps us separate incidental noise from meaningful signal.

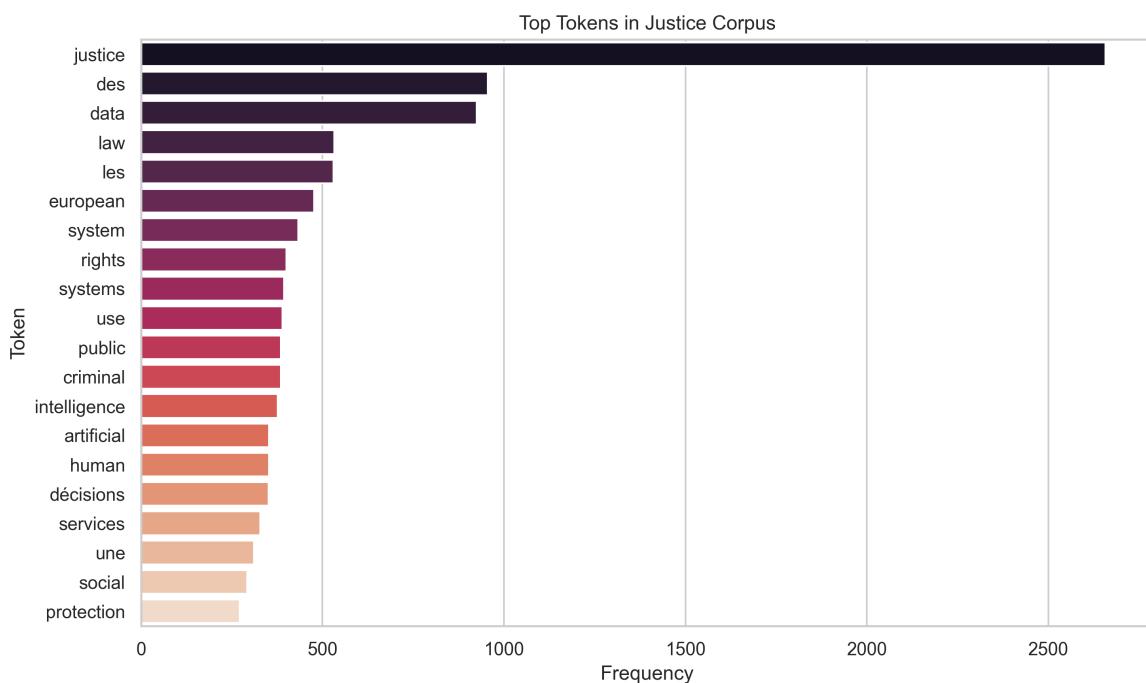
To ensure that outliers are not artifacts of extraction, we relied on a **multi-extractor PDF pipeline** (Poppler pdftotext -layout → pdfminer.six → PyPDF2) and retained the first nonempty result. This cascade minimizes false anomalies caused by layout loss (e.g., columns merged into headers, running

titles repeated on every page). We then applied a light **structural cleaner** that removes recurring head and foot lines across pages, normalizes spacing, and stitches hyphenated line breaks. The intent is to keep the lexical content intact while reducing artificial inflation of repeated phrases such as committee names or section headings. After this pass, the Isolation Forest is rerun on the cleaned features so we can directly measure how much of the “outlierness” was structural rather than semantic.

At a glance, the **scatter plot of document length vs. justice count** shows a dense cloud where most items cluster at low frequencies, plus a handful of points highlighted by the model. Some flagged documents sit **far from the cloud** because they contain many mentions of *justice* relative to their size; others sit **near the cloud’s edge** yet are still isolated due to subtle but consistent deviations (e.g., short texts with a disproportionate density, or medium texts whose counts are inflated by repeated headers). These nearcluster outliers are typical when a document is just across the anomaly boundary set by the contamination parameter, or when the anomaly is **local** (e.g., a header repeated fifty times) rather than global in the textual distribution.

Concretely, we observe two families of outliers. The first are **structural outliers**, where the density of *justice* is inflated by duplicated boilerplate. For instance, some institutional PDFs include the phrase “Justice and Home Affairs Committee” or “European Commission for the Efficiency of Justice” on every page. Our cleaning pass reduces counts dramatically in such cases (e.g., *366.txt* drops from **178** raw mentions to **30** after structural filtering; *120.txt* from **184** to **13**; *162.txt* from **110** to **11**; *110.txt* from **113** to **3**), moving these items back toward the main cloud once the noise is removed. The second family are **semantic outliers**: documents that truly center on justice and ethics. These remain outliers even after cleaning because their high density reflects genuine content rather than formatting redundancy (e.g., policy analyses where *justice* cooccurs with *rights*, *fairness*, and *accountability*). In between, we also find weblike pages (e.g., event listings) where *justice* appears often in menus and titles—less repetitive than PDF headers, but still partly **navigation/branding** rather than substantive discourse.

From a methodological standpoint, this pipeline underscores why **structural normalization** should precede frequencybased judgment: outliers can represent either “**bad structure**” or “**important content**.” Interpreting them without context risks mislabeling key sources as noise or overlooking layout issues that distort linguistic statistics. For the remainder of the project, we therefore keep both views: the **raw** counts (to mirror how the texts were originally scraped) and the **cleaned** counts (to approximate the intended prose). This dual reporting clarifies when a document’s status as an outlier is robust to reasonable normalization choices.





1.2.2. Trends in topics

We complement outlier detection with a **corpus-level view** of the most salient vocabulary once English and French stopwords are removed. Two consistent trends emerge. First, a **technological axis** dominates, with high frequency items such as *data*, *system(s)*, and *use* pointing to discussions about datasets, algorithmic pipelines, and deployment contexts. Second, an **ethical-institutional axis** is clearly present, with *rights*, *human*, *public*, *ethics*, and *justice* anchoring normative positions and governance concerns. Read together, these axes sketch the field's internal dialectic: practical machinery on the one hand (how AI is built and used), and social safeguards on the other (how AI interacts with people, institutions, and the law).

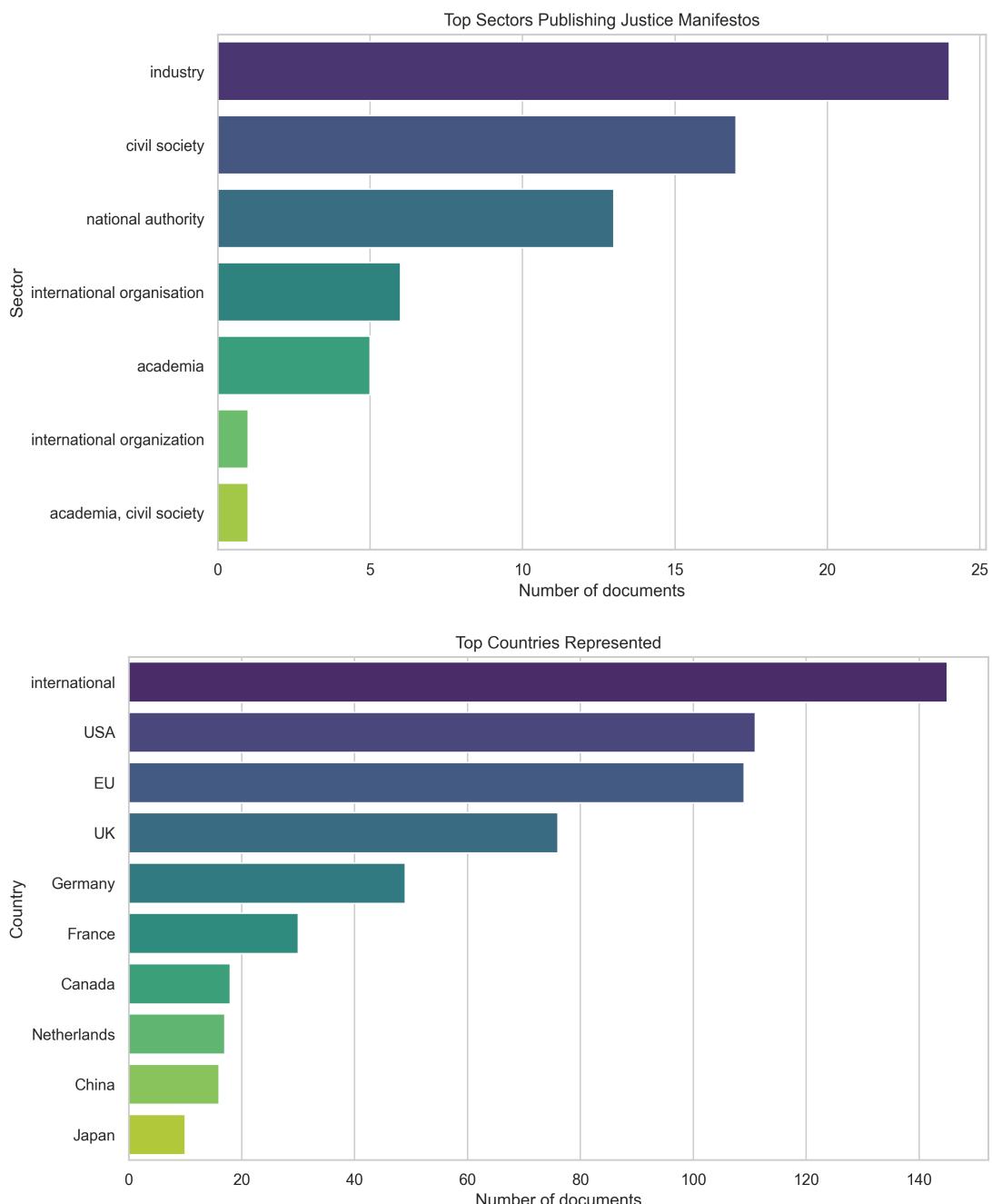
A frequent question is whether the prominence of **data** is a scraping artifact or a true lexical signal. To answer that, we ran a targeted **context audit** over **203,268** occurrences of the string “data” using simple structural heuristics. Only **0.5%** fell into **PDF syntax zones** (e.g., /Data in object names), **~0.0%** into **HTML/XML tags** (<data>), and **~0.0%** near PDF keywords (obj, stream, xref). An overwhelming **99.4%** appeared in **normal text**, including phrases like “big data analytics” or “data protection”—clear content rather than metadata. This confirms that “data” is not spuriously inflated by broken extraction; it genuinely reflects the corpus’ technical preoccupations (data governance, protection, and analytics). Retaining it in frequency analyses is therefore appropriate, though we do continue to exclude standard stopwords to avoid functional noise (*of, to, and*, etc.).

Visual exploration through a **word cloud** and a **top-N frequency bar chart** makes these axes visible at a glance. The cloud foregrounds high-impact lexical items and their relative dominance, while the bar chart surfaces **rank order** and precise **counts**. In our corpus, items like *system(s)*, *use*, *rights*, and *public* frequently cooccur with *justice*, hinting at **registers** that will be unpacked in later sections: e.g., **institutional/structural** framings (administration of justice, justice system), **penal/criminological** framings (criminal justice, law enforcement), and **access/equity** framings (access to justice, fairness). These patterns dovetail with the outlier analysis: documents that remained outliers after cleaning tend to be those in which justice is elaborated through these ethical-institutional registers rather than incidentally repeated in formatting.

Because our dataset contains multilingual material, a small number of non-English tokens may appear in global visualizations. We intentionally **kept the pipeline stable** at this stage to reflect the corpus “as

is,” with dedicated language filtering (e.g., removing German-only spans) reserved for subsequent, **contextual** analyses. This preserves comparability between the outlier view and the frequency view while making explicit where language normalization will sharpen future semantic results. In practice, these multilingual traces have minimal impact on the **relative** ordering of the top English content words and do not alter the two-axis pattern described above.

Finally, we report one **document-level note** to contextualize frequency summaries: the **longest document** in the set can bias naive frequency counts simply by virtue of length. To mitigate this, we consistently interpret raw counts alongside **normalized densities** and explicitly annotate the longest item when we present top-N plots. When helpful, we also provide **short contextual excerpts** from that document so the reader can verify that high counts are tied to genuine argumentation rather than boilerplate repetition. This practice complements the outlier methodology by keeping both **scale** and **context** visible.



PART 2 : SEMANTIC ANALYSIS

This section aims to analyse how the term justice appears in the charters and manifestos examined. Its objective is to observe its frequency, its contextual usage, and the meanings associated with it in ethical discourse.

The term justice is particularly relevant in the field of AI, as it refers to issues such as non-discrimination, fairness in automated decision-making, and equal access to technologies. However, it is a polysemic notion that can refer to different interpretations (social, distributive, procedural justice), which may lead to ambiguities.

We formulate the hypothesis that the use of the term varies depending on the type of actor (industry, public institutions, NGOs) and that its lexical context will highlight distinct priorities. This analysis will help us better understand how justice is positioned and valued within ethical discourse related to artificial intelligence.

2. Contextual analysis of the term *justice* through n-grams

2.1. Methodology

We first selected the documents belonging to the **top 10%** of the corpus, i.e. those with the highest relative frequency of occurrences of the term *justice*. This allows us to focus on texts in which the notion is particularly salient and potentially structurally significant.

We then extracted **bigrams** and **trigrams** from the tokenised text. These n-grams capture the immediate lexical context surrounding *justice*, making it possible to identify semantic fields, institutional actors or thematic registers.

The extraction was implemented as follows:

```
def extract_ngrams(words, n=2):
    return [" ".join(words[i:i+n]) for i in range(len(words)-n+1)]
bigrams = extract_ngrams(words, 2)
trigrams = extract_ngrams(words, 3)
```

Since the corpus contains multilingual material, we applied automatic language detection to filter out French expressions. N-grams detected as French were discarded:

```
from langdetect import detect
```

```
def is_not_french(text):
```

```
    try:
```

```
        return detect(text) != "fr"
```

```
    except:
```

```
        return False
```

An initial exploration revealed a high proportion of n-grams dominated by determiners, prepositions, or conjunctions (e.g. *of justice*, *and justice*), which provide little meaningful context. We therefore introduced an additional filter that excludes n-grams in which at least half of the tokens are stopwords:

```
STOPWORDS = {"the", "a", "an", "of", "to", "in", "and", "for", "on", "at", "by", "with", "from", "is"}
```

```
def contains_stopword(ngram):
    words = ngram.split()
    stop_count = sum(w in STOPWORDS for w in words)
    return stop_count >= len(words) / 2
```

This successive filtering (language + stopwords) greatly improves the semantic clarity of the data by removing functional linguistic noise.

We then counted all n-grams containing *justice* and visualised the most frequent ones using horizontal bar charts, which facilitate the identification of dominant lexical patterns:

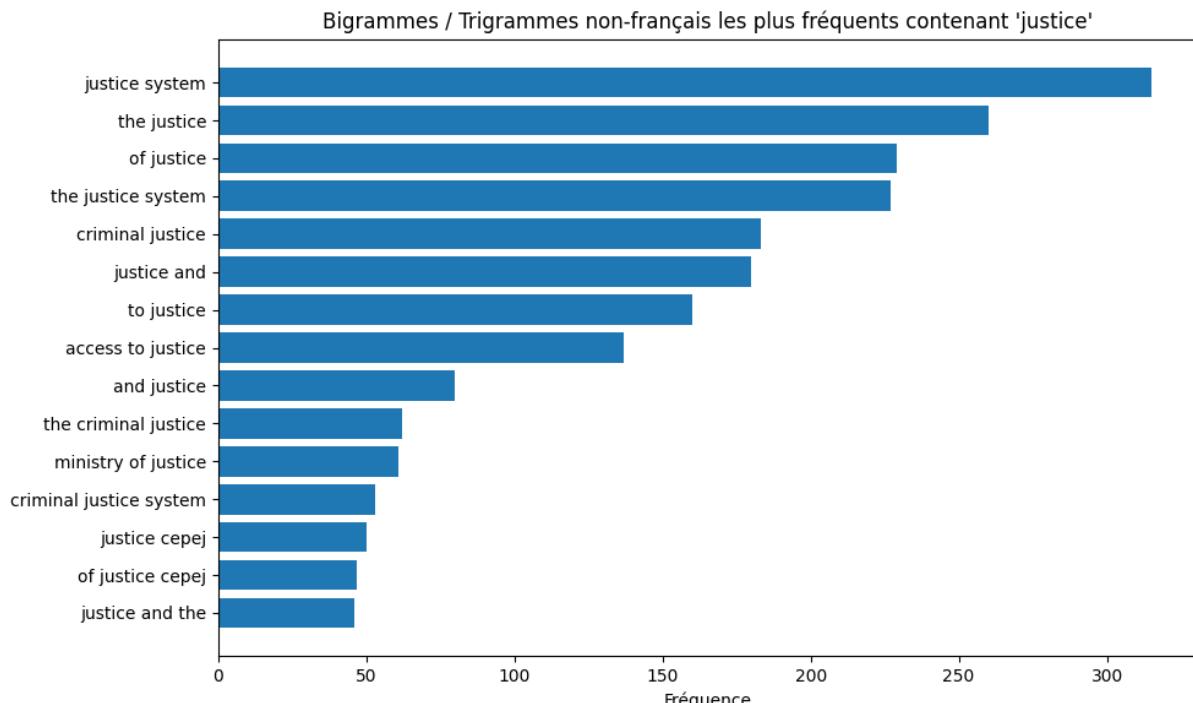
```

from collections import Counter
ngrams_counter = Counter()
ngrams_counter[bg] += 1
labels, values = zip(*ngrams_counter.most_common(15))
plt.barh(labels, values)

```

These visualisations reveal several coherent semantic clusters (e.g. *justice system*, *criminal justice*, *access to justice*), which will serve as the basis for a qualitative interpretation of the discursive registers associated with the notion of justice in the corpus.

2.1.2. Raw results



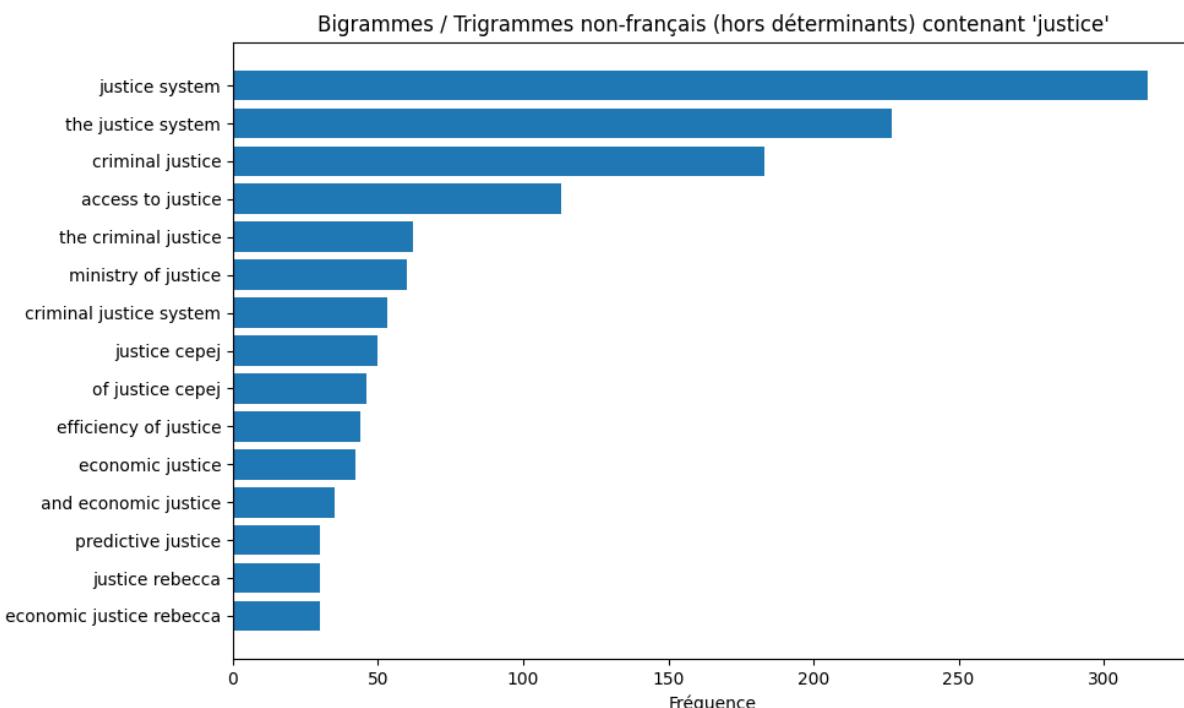
The first visualisation below shows the most frequent non-French bigrams and trigrams containing the term *justice* within the top 10% of documents (ranked according to their occurrence density). This representation makes it possible to quickly identify recurrent expressions and the immediate co-text associated with the notion under study.

Several observations emerge:

- Institutional expressions clearly dominate, particularly through *justice system* and *the justice system*. This indicates a strong orientation of the discourse toward the structure and functioning of judicial systems.
- The prominent presence of *criminal justice* suggests a penal/criminological register, frequently observed in texts dealing with regulation, surveillance, or security.
- Occurrences such as *access to justice* introduce a societal dimension focused on inclusion, availability, and fairness.
- We also find collocations linked to public bodies or institutional actors (e.g. *ministry of justice*, *justice cepej*), which confirms the strong institutionalisation of the theme.

Conversely, some frequent expressions (*of justice*, *to justice*, *and justice*) mainly reflect functional grammar and contribute very little to contextual interpretation. Their presence justifies the subsequent filtering step, aimed at removing n-grams dominated by determiners or conjunctions. It is therefore necessary to further process the text to eliminate linguistic noise, refine contextual granularity, and highlight expressions that are truly meaningful, thereby enabling a more precise identification of the discursive axes associated with the concept of justice in the corpus.

2.1.3. Filtered n-grams and emergence of thematic clusters



After applying several filters, namely the removal of French expressions and the exclusion of n-grams dominated by stopwords, the resulting visualisation becomes significantly more informative from a semantic standpoint. Linguistic noise is largely reduced, allowing genuinely relevant lexical patterns to emerge.

The filtered n-grams provide a clearer picture of how *justice* is contextualised within the corpus and highlight recurring associations that will be examined in more detail at the sentence level.

2.1.4. In-depth Analysis of Thematic Axes

In a second step, we refine the exploration by performing a thematic classification of n-grams containing the term *justice*. After filtering (language detection and exclusion of stopword-dominated expressions), each bigram/trigram is associated with a semantic axis using a small dictionary of indicator terms (e.g., *system*, *court*, *reform* → institutional; *criminal*, *enforcement* → penal; *access*, *fairness* → equity; *efficiency*, *budget* → managerial; *ministry*, *cepej*, *council* → institutional actors). Occurrences are then counted per theme and visualized using top-N histograms.

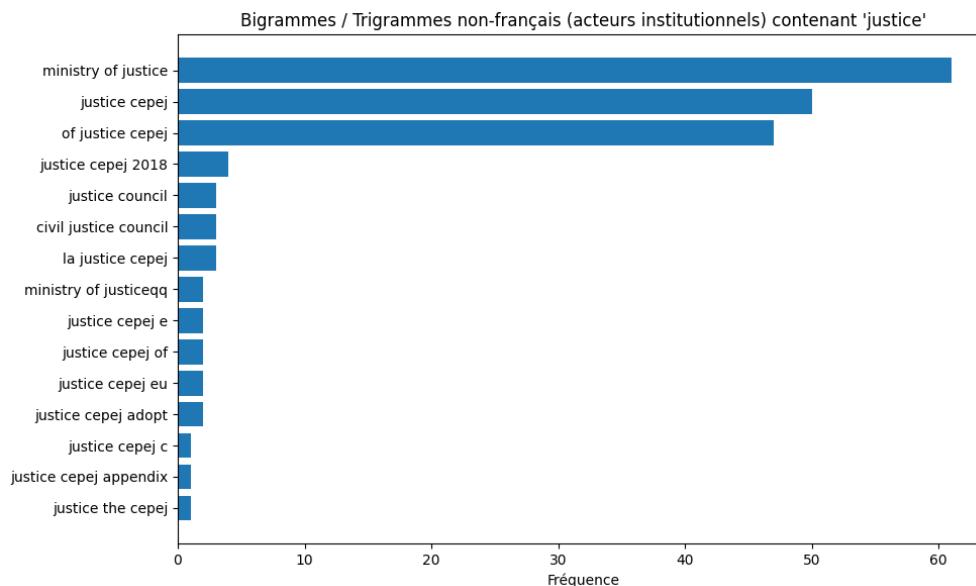
The results confirm five dominant axes:

- **Institutional / Structural Axis** - (*justice system*, *criminal justice system*, *court of justice*)
This axis frames justice as a **public institutional infrastructure** to be organized, evaluated, and reformed. It is the most salient theme.
- **Penal / Criminological Axis** - (*criminal justice*, *criminal justice system*, *law enforcement*)
Justice is approached from a **security and sanction** perspective: crime management, prevention, reintegration, and penal reforms.
- **Access / Equity Axis** - (*access to justice*, *efficiency of justice*)
This axis highlights concerns related to **accessibility**, procedural guarantees, and fairness, often linked to international institutions.
- **Economic / Managerial Axis** - (*efficiency of justice*, *economic justice*, *justice cepej*)
Justice is associated with **performance, costs, and resource allocation**. The presence of CEPEJ reveals the influence of **European evaluation standards**.

- **Institutional Actors** - (*ministry of justice, mentions of authors/programmes*) These occurrences confirm the **institutionalisation** of the discourse, involving ministries, commissions, councils, and expert reports.

Taken together, these axes reveal a **plural and multidimensional** conception of justice, dominated by **institutional** and **penal** framings, complemented by concerns for **access** and **efficiency**. This mapping justifies transitioning to **sentence-level analysis**, enabling us to identify *who acts, through which instruments, under what conditions, and for whose benefit*.

Exemple for institutionnal actors



This thematic organization justifies the transition to a sentence-level analysis, allowing us to clarify who acts, for what purpose, and through which modalities, in order to better distinguish justice as a normative value from justice as an operational mechanism.

2.2. Information loss induced by sentence-level granularity

2.2.1. Loss of discursive context

By restricting the analysis to isolated sentences, a significant portion of the surrounding discursive context is necessarily lost. In a full document, meaning emerges not only from individual propositions, but also from the interplay between sentences across larger argumentative units such as paragraphs or sections. When sentences are extracted independently, we no longer capture referential continuity, including pronouns (*he, she, it*), elliptical constructions, and connective markers that link ideas over time. As a result, causal, temporal, and inferential relationships that would normally structure the discourse tend to disappear.

This reduction also removes access to the global coherence of the text. We lose visibility on how claims are introduced, developed, justified, and sometimes contradicted or qualified later on. Without this broader argumentative sequence, interpretations may become fragmented, potentially leading to overgeneralisation or misattribution of intent.

Furthermore, non-textual cues that shape interpretation, such as titles, subtitles, lists, figures, captions, or typographical emphasis, are entirely absent from sentence-level data. These structural elements often indicate thematic segmentation, highlight key concepts, or qualify the scope of statements.

In short, while sentence extraction offers a finer syntactic perspective, it does so at the cost of discursive depth. Certain nuances, rhetorical strategies, and conceptual dependencies remain invisible, limiting our ability to fully apprehend the complex positioning of *justice* within the corpus.

2.2.2. Loss of document structure

Extracting isolated sentences also results in the disappearance of the document-level structure that frames and organises the discourse. In a complete document, formatting elements such as titles, subtitles, thematic sections, or bullet points play a key role in prioritising information and guiding interpretation. When this organisational layer is removed, we lose visibility on argumentative progression, conceptual framing, and the distinctions between objectives, recommendations, or conclusions.

Moreover, many documents rely on non-textual elements (tables, diagrams, figures, text boxes) to clarify relationships, synthesise issues, or visualise conceptual articulations. When they are no longer accessible, extracted sentences may appear detached from crucial contextual explanations or lack interpretative anchors.

Finally, the rhetorical rhythm of the document, introductory passages, methodological transitions, intermediate findings, disappears with sentence-level segmentation. The same sentence can appear prescriptive, descriptive, or conditional depending on its placement, and this contextual nuance becomes invisible once structure is stripped away.

As a result, sentence-level analysis deprives us of organisational cues that guide understanding, weakening our grasp of discursive priorities, normative scope, and the functional role that the notion of *justice* plays within the broader architecture of the text.

2.2.3. Loss of semantic nuance

At the sentence level, a substantial portion of the semantic nuance surrounding the term *justice* tends to disappear. While isolated sentences can reveal local lexical patterns, they often omit the qualifications, modal markers, and discursive framing devices that modulate meaning across larger contexts. Elements such as adverbs (“*potentially*”, “*under certain conditions*”), hedges, and scope delimitations are frequently diluted or rendered ambiguous once removed from the surrounding discourse.

In addition, many documents develop ideas through contrastive structures (e.g., counter-examples, limitations, exceptions) that can heavily influence interpretation. When these contrastive cues are no longer available, statements may appear more categorical or definitive than originally intended. Similarly, the absence of argumentative scaffolding can obscure whether a sentence expresses a recommendation, a hypothesis, or a conditional requirement.

Furthermore, certain lexical items carry polysemic values that shift depending on context. Expressions like *social justice*, *economic justice*, or *criminal justice* mobilise distinct conceptual frames, stakeholders, and normative assumptions. Without broader context, distinguishing these semantic regimes becomes more challenging, increasing the risk of conflating heterogeneous uses of the term.

More broadly, the loss of semantic nuance affects the ability to capture tone, ambiguity, and discursive positioning. Subtle rhetorical signals — whether the sentence expresses urgency, caution, or critique — may become flattened when analysed in isolation.

Consequently, while sentence-level extraction enables fine-grained syntactic analysis, it simultaneously reduces the depth and subtlety of meaning, limiting our capacity to fully interpret how *justice* is contextualised, qualified, and negotiated within the corpus.

PART 3 : STRUCTURAL ANALYSIS

Part 2 of the project identified the main semantic axes associated with the term "justice" in the MapAIE corpus. However, this lexical analysis, while informative, has important limitations: it does not fully capture the syntactic structure of sentences, the grammatical relationships between words, or how the concept of "justice" is integrated into complex argumentative patterns.

To deepen this understanding, Part 3 employs two complementary approaches: analysis of grammatical roles (POS tagging) and analysis of AMR graphs (Abstract Meaning Representation). These two methods aim to go beyond simple lexical frequency to reveal how "justice" is structured linguistically and conceptually in ethical discourse on AI.

3.1. Grammatical roles

Grammatical role analysis allows us to identify the syntactic function occupied by the term "justice" in corpus sentences. This approach is based on *Part-of-Speech tagging (POS tagging)*, an automatic language processing technique that assigns each word its grammatical category (noun, verb, adjective, etc.).

Methodology

To perform the grammatical analysis, we used the Python library **spaCy**, recognized for its precision and speed in syntactic analysis. The linguistic model used is en_core_web_sm, a model trained on a large corpus of English texts.

The analysis process unfolds in several stages:

- **Sentence extraction:** All sentences containing the term "justice" are extracted from the MapAIE corpus.
- **Tokenization and POS tagging:** Each sentence is tokenized (split into words) and each token receives a grammatical label (POS tag).
- **Occurrence identification:** We precisely identify tokens corresponding to "justice" and note their grammatical category.
- **Statistical analysis:** We count the frequency of each grammatical category to determine dominant syntactic roles.

Results

The analysis reveals a clear distribution of grammatical roles:

- **Common noun (NOUN):** The vast majority of occurrences of "justice" correspond to a common noun, used to designate the abstract concept of justice, the judicial system, or an institution.
- **Proper noun (PROPN):** A significant portion of occurrences refer to proper nouns, notably "Court of Justice," "Ministry of Justice," "European Court of Justice," etc. • **Modifier (part of compound nouns):** In many cases, "justice" functions as a modifier in compound phrases such as "justice system," "criminal justice," "access to justice. »

Interpretation

These results confirm that the term "justice" is **primarily a substantive concept** in the corpus: it designates a value, an institution, or a system, but is never used as a verb or adjective. This observation indicates that "justice" is less a qualifier or action than an *autonomous discursive entity*, around which a set of conceptual relationships is articulated.

The strong presence of proper nouns ("European Court of Justice," "Ministry of Justice") also reveals the importance of *institutionalization* in discourse: "justice" is not only an abstract principle, but an administrative and legal reality embodied by clearly identified actors.

3.2. AMR framing of justice

POS analysis reveals the grammatical function of "justice," but does not allow us to understand its *deep semantic structure*: who acts? who undergoes? what are the beneficiaries or objects of an action? To answer these questions, we employ **Abstract Meaning Representation (AMR)**, a formalism that represents the meaning of a sentence as a directed graph.

What is AMR?

AMR is a semantic annotation that abstracts surface syntax to focus on the *conceptual meaning* of a sentence. Unlike traditional syntactic analysis, AMR is not concerned with word order or grammatical structure, but with **semantic relationships** between concepts.

For example, the sentence "Barack Obama was born in 1961" will be represented by an AMR graph centered on the concept bear-02 (to be born, sense 2 of PropBank), with:

- :ARG1 (the person born) pointing to person :name "Barack Obama"
- :time (moment of birth) pointing to date-entity :year 1961

This formalism allows capturing implicit relationships, normalizing syntactically different but semantically equivalent expressions, and comparing argumentative structures across a corpus.

AMR Methodology

AMR graphs for the MapAIE-Justice corpus were provided in the file justice-MapAIE_penmans.amr. This file contains:

- The original English sentence (:snt)
- The source file number
- The AMR graph in PENMAN format

We analyzed these graphs by extracting:

- AMR concepts containing "justice": For example, (j / justice), (s / system :mod (j / justice)), etc.

Semantic relationships: Semantic roles (:ARG0, :ARG1, :mod, :topic, etc.) linking "justice" to other concepts.

Parents, siblings, and children: Superior concepts (roots), parallel (same hierarchical level), and subordinate concepts in the AMR tree.

Key Results

1. Nature of "justice" in AMR

"Justice" **never** appears as a verb or verbal predicate. There is **no PropBank frame** associated with "justice" (unlike verbs like "judge-01" or "justify-01"). In all observed cases, "justice" is a *nominal concept* ((j / justice)) or part of a named entity (:name).

3. Dominant semantic roles

Depending on context, "justice" occupies different semantic roles:

- **Modifier** (:mod): "justice" qualifies a system, domain, or approach. Example: (s / system :mod (j / justice)) → "justice system".
- **Named entity** (:name): "justice" appears in institutional proper nouns. Example: *Court of Justice, Ministry of Justice*.
- **Theme/Topic** (:topic): In certain structures, "justice" is the discursive subject of an action or debate.
- **Semantic argument** (:ARG1, :ARG2): "justice" can be the object of a verb (for example, "discuss justice," "approach justice").

4. Identified semantic axes

AMR analysis confirms and refines the thematic axes identified in Part 2:

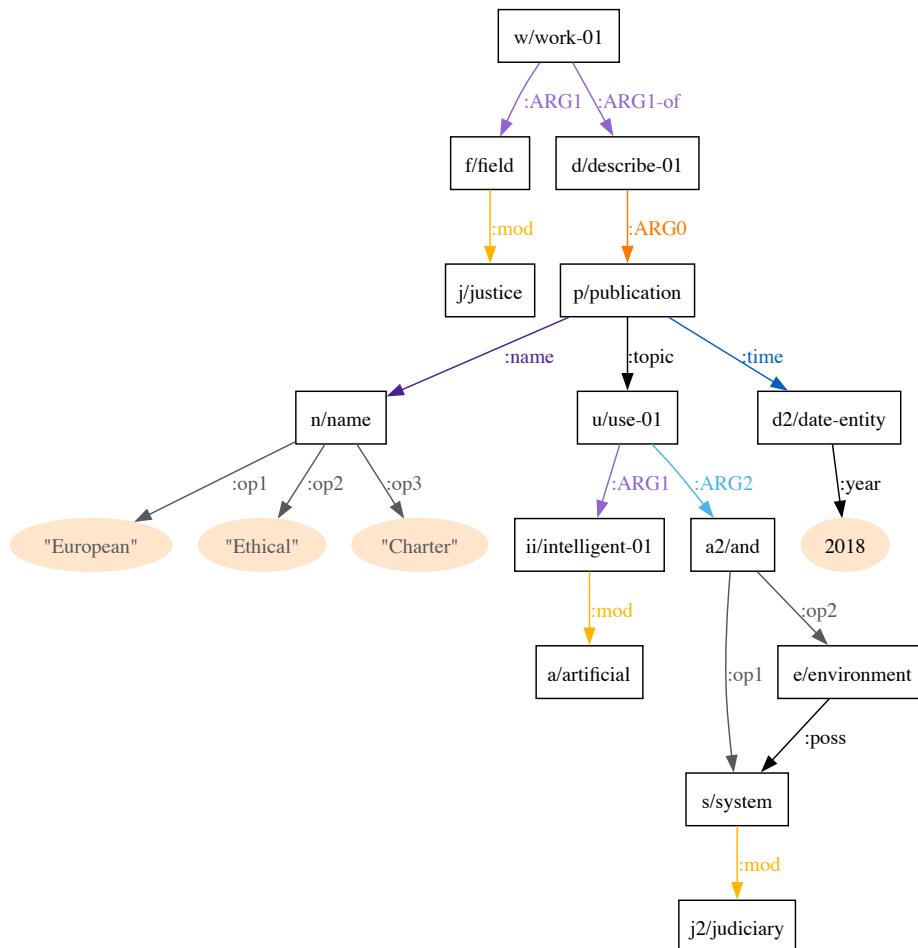
- **Institutional/structural axis:** "justice" modifies "system," "field," "institution." These structures reveal a conception of justice as *organizational infrastructure*.
- **Actantial axis (institutional actors):** Named entities (*European Court of Justice, Ministry of Justice*) signal concrete actors in discourse.
- **Procedural/operational axis:** "access to justice," "efficiency of justice" indicate a managerial and evaluative logic.
- **Conceptual/abstract axis:** In certain contexts, "justice" appears as an independent ethical value, thematized in normative debates.

To concretely illustrate the observed AMR structures, we present ten graph examples generated from the MapAIE-Justice corpus. These graphs were produced using the **metamorphosed** tool, developed by Orange (<https://github.com/Orange-OpenSource/metamorphosed>).

Technical note: Some segments of the initial AMR file contained formatting errors or structures not recognized by the metamorphosed tool. These anomalies were manually corrected to enable graphical visualization. The following examples represent the most representative and clearly formalized AMR structures in the corpus.

Example 1: Justice as a system (justice system)

Original sentence: "The work in the field of justice is described by the publication European Ethical Charter (2018), on the use of artificial intelligence in the judiciary system and its environment. »



Interpretation: In this example, "justice" appears as a *domain* (:mod relation) modifying "field." The AMR graph reveals a complex structure where:

- The graph root is describe-01 (the verb "describe").
- The :ARG1 of "describe" is work (the work).
- This work is located in a field (domain), itself modified by justice.

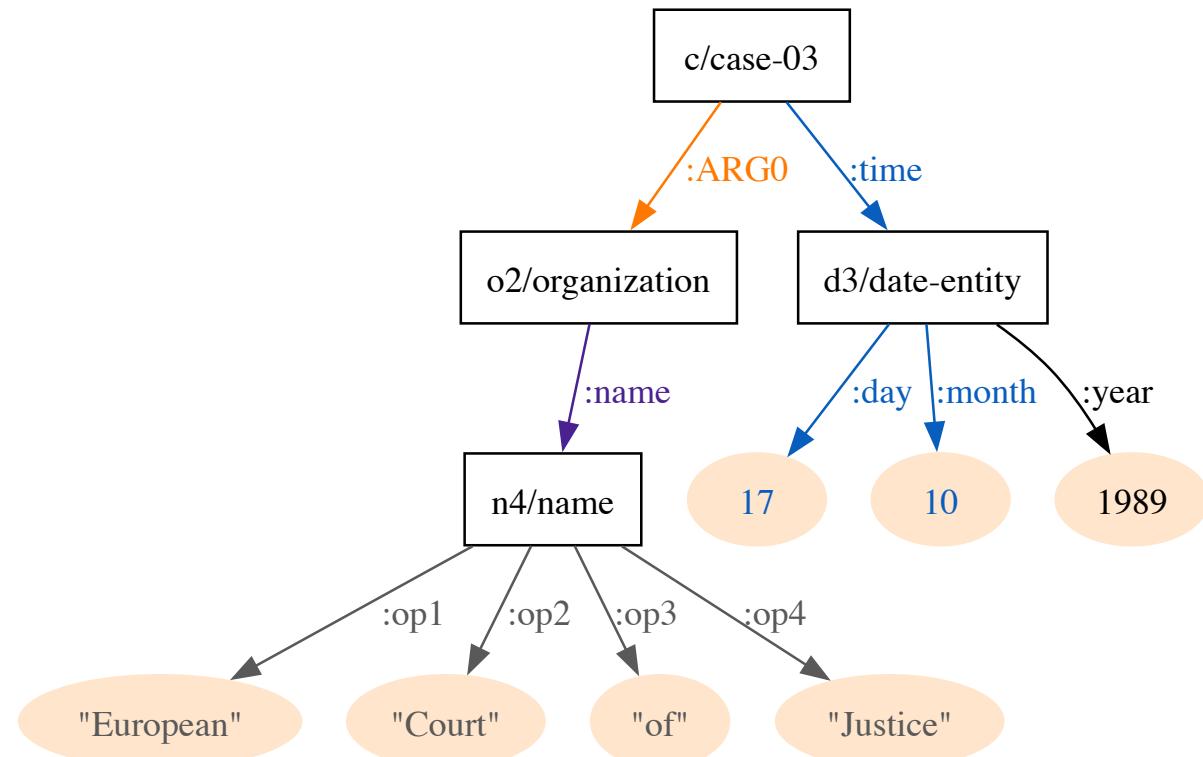
Semantic axes: Institutional (justice = field of activity) / Conceptual (justice = normative framework).

Example 2: Justice as a named entity (Court of Justice)

Original sentence: "Case of the European Court of Justice on 17 October 1989. »

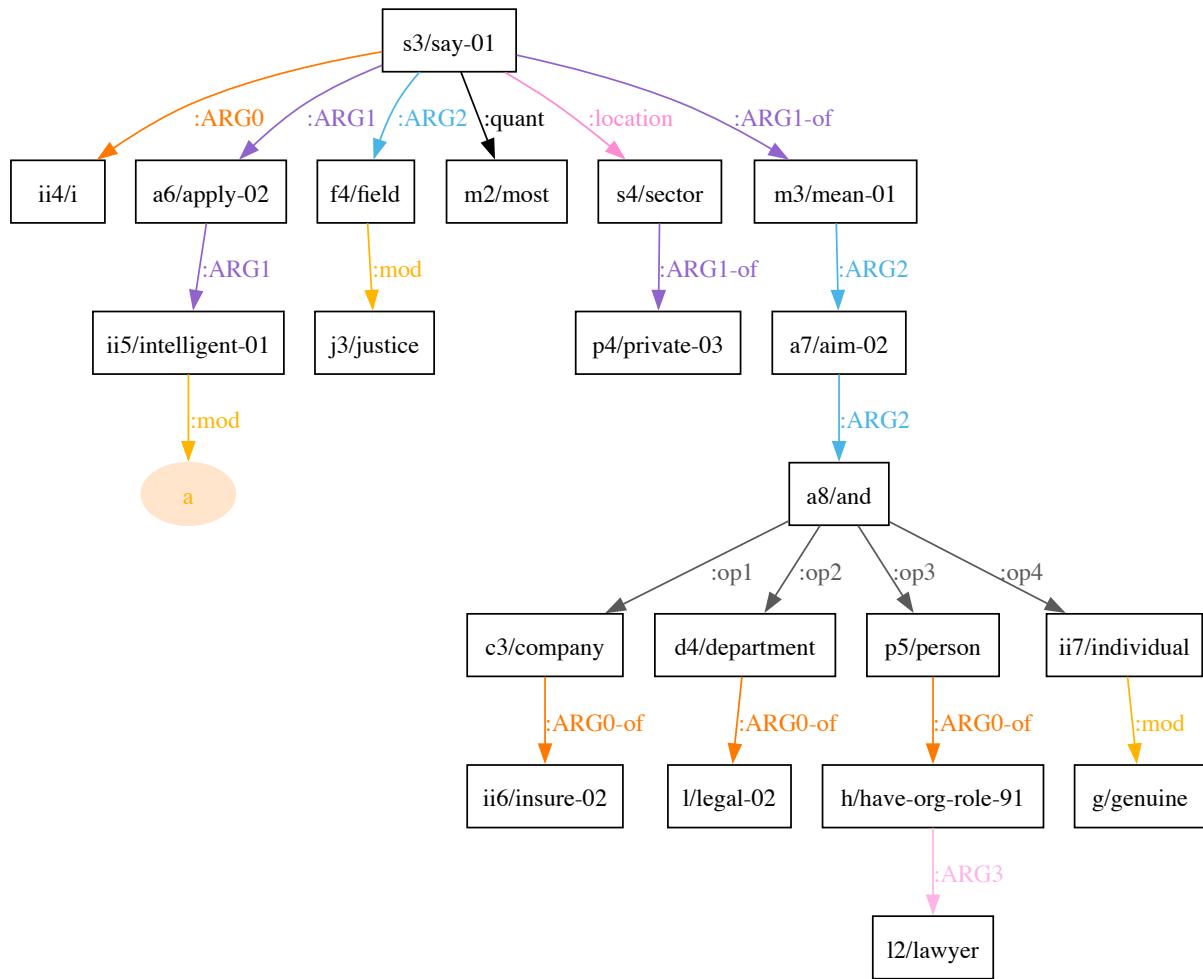
Interpretation: Here, "justice" is part of a **named entity**: European Court of Justice. The AMR graph uses the :name relation to encode this proper noun. The structure reveals that "justice" is not an autonomous concept, but a *component of institutional identification*.

Semantic axes: Institutional (legal actor) / Actantial (named entity).



Example 3: Justice as a field of application

Original sentence: "I say that artificial intelligence applies to the field of justice in most private sector organizations. »



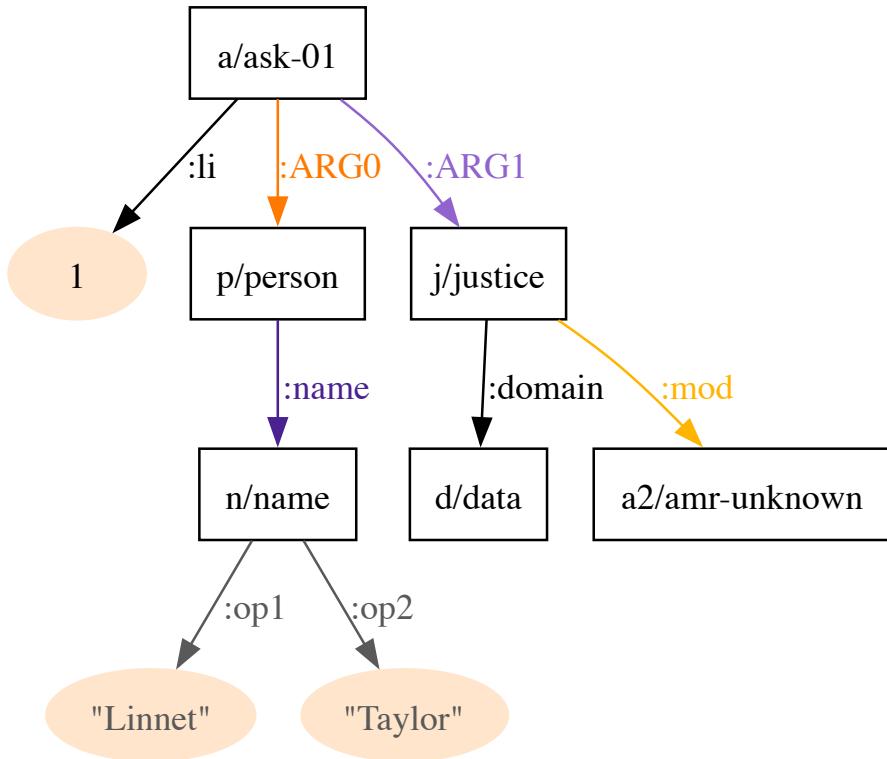
Interpretation: In this example, "justice" is the *field of application* of artificial intelligence. The AMR structure highlights:

- The root verb **apply-02** ("to apply").
- The **:ARG1** is *artificial intelligence*.
- The **:ARG2** (target domain) is *field :mod justice*.

Semantic axes: Procedural (technological application) / Institutional (justice = sector of activity). Other analyzed examples

Example 4: Justice as a topic of inquiry

Original sentence: "Linnet Taylor asks about justice in the field of data. »



Interpretation: In this example, "justice" appears as the *discursive topic* of an interrogative act. The AMR structure reveals:

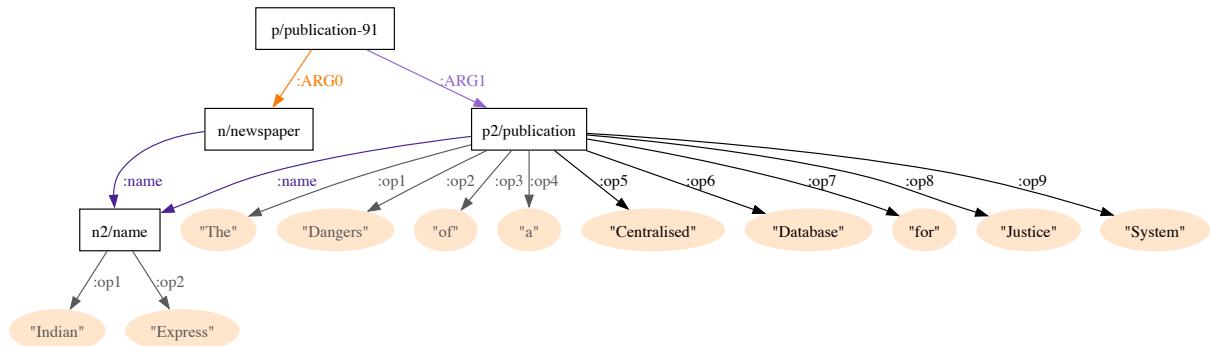
- The root verb is `ask-01` ("to ask").
- The `:ARG0` (the asker) is person `:name` "Linnet Taylor".
- The `:topic` relation points to (`j / justice`), indicating that justice is the subject of the inquiry.
- Justice is further qualified by location: `:location` (`f / field` `:mod` (`d / data`)) → "in the field of data".

This structure is significant because it positions justice not as a *modifier* or *institutional entity*, but as an **object of intellectual inquiry**. The combination with "data" indicates emerging concerns about justice in algorithmic contexts.

Semantic axes: Conceptual (justice as philosophical question) / Procedural (data justice as emerging field).

Example 5: Justice system under threat

Original sentence: "The Dangers of a Centralised Database for Justice System. »



Interpretation: This title structure presents justice as a *vulnerable system* exposed to technological risks. The AMR graph shows:

- The root concept is danger (plural), indicating threat.
- The source of danger is database :mod (c / centralize-01) → "centralized database".
- The :beneficiary (or affected entity) is system :mod (j / justice) → "justice system".
- The possessive structure links the database to the justice system, implying ownership or intended use.

This example is particularly revealing because it frames the justice system as **passive recipient** of technological intervention. Justice appears as an institutional infrastructure that can be endangered by data architecture choices.

Semantic axes: Institutional (justice as vulnerable system) / Procedural (technology as threat vector)

1. Conceptual nature of "justice" in AMR

Fundamental observation:

- "Justice" **never appears as a verb or predicate** in AMR graphs.
- There is **no PropBank frame** associated with "justice." PropBank, the lexical resource underlying AMR, only covers *verbal predicates* (e.g., judge-01, justify-01).
- Consequently, "justice" is always encoded as a **nominal concept**: (j / justice), or as an integral part of a *named entity* (:name relation).

Implication: In ethical discourse on AI, "justice" is not mobilized as an *action* ("to judge," "to do justice"), but as a stable *conceptual entity* — a **normative framework**, an **institution**, or a **domain of intervention**.

2. Semantic roles and structural relationships

Analysis of 520 occurrences of "justice" reveals four main semantic roles:

a) Modifier (:mod relation)

The most frequent role of "justice" is that of **modifier**, where it qualifies a nominal concept:

- (s / system :mod (j / justice)) → "justice system"
- (a / approach-02 :ARG1 (j / justice :mod (d / data))) → "data justice approach"

In these structures, "justice" *specifies* the nature of a system, domain, or approach. This qualifying function indicates a **discursive categorization**: "justice" becomes a marker of *sector* or *application domain*.

b) Named entity (:name relation)

A significant portion of occurrences corresponds to **institutional proper nouns**:

- European Court of Justice
- Ministry of Justice
- Court of Justice of the European Union

These named entities testify to the **strong institutionalization** of discourse. "Justice" is not only an abstract principle, but an *embodied organizational reality* through clearly identified actors.

c) Discursive theme (:topic relation)

In certain contexts, "justice" constitutes the *subject* of an utterance, the object of questioning or analysis:

- (a / ask-01 :ARG0 person :topic (j / justice)) → "ask about justice"
- (r / research-01 :topic (j / justice)) → "research on justice"

In this role, "justice" is no longer a qualifier, but an **object of reflection**, a *thematized value* in debate.

d) Semantic argument (:ARG1, :ARG2 relations)

More rarely, "justice" occupies a position of *argument* in a predicative structure:

- (a / access-01 :ARG1 (j / justice)) → "access to justice"
- (d / describe-01 :ARG1 (w / work :location (f / field :mod (j / justice)))) → "describe work in the field of justice"

Here, "justice" is *the object* of an action (access, description, etc.). This configuration reveals a conception of justice as a **good to be guaranteed or domain to be described**.

3. Dominant semantic axes

AMR analysis allows us to refine and confirm the thematic axes identified in Part 2. We distinguish five main axes:

a) Institutional/structural axis

Definition: "Justice" designates organizational infrastructure, a system to be evaluated, reformed, modernized.

AMR examples: "justice system," "field of justice," "justice institutions."

Prevalence: This is the most frequent axis in the corpus. "Justice" is primarily conceived as an *administrative and judicial apparatus*.

b) Penal/criminological axis

Definition: "Justice" is associated with the criminal domain, repression, risk management.

AMR examples: "criminal justice," "criminal justice system," "justice risk assessments."

Implication: This axis reveals a focus on *security* and *prevention*, often linked to algorithmic use in the judiciary.

c) Access/equity axis

Definition: "Justice" as procedural right, access guarantee, equity principle.

AMR examples: "access to justice," "efficiency of justice," "right to a fair trial."

Prevalence: This axis is particularly present in international documents and NGO texts.

d) Economic/managerial axis

Definition: "Justice" as object of evaluation, measurement, optimization.

AMR examples: "efficiency of justice," "economic justice," "CEPEJ" (European Commission for the Efficiency of Justice).

Implication: This axis reflects a logic of *rationalization* and *performance*, characteristic of European institutional discourse.

e) Conceptual/abstract axisDefinition: "Justice" as autonomous ethical value, independent normative principle.

AMR examples: Contexts where "justice" is thematized in ethical debates (:topic) without direct reference to an institution.

Prevalence: Less frequent than the institutional axis, but significant in philosophical manifestos and charters.

Conclusion of Part 3

The structural analysis conducted in this part reveals fundamental characteristics of the term "justice" in ethical discourse on AI:

1. Justice is never an action

Unlike other normative concepts that can be verbalized ("responsabilize," "transparentize"), "justice" remains a **substantive**. It does not designate an action to be performed, but a *stable conceptual entity*: a system, institution, domain, or value.

2. Justice is strongly institutionalized

The preponderance of proper nouns ("European Court of Justice," "Ministry of Justice") and institutional phrases ("justice system," "criminal justice") testifies to a *largely bureaucratic conception* of justice. Ethical discourse on AI does not treat "justice" as an abstract idea, but as an **organizational reality** to be regulated, evaluated, modernized.

3. Justice oscillates between ethical value and managerial apparatus

Two logics coexist in the corpus:

- **A normative logic:** "justice" as ethical principle, fundamental right, object of moral concern ("access to justice," "fairness").
- **A managerial logic:** "justice" as system to optimize, performance to measure, sector to reform ("efficiency of justice," "CEPEJ").

This tension reveals a **discursive ambivalence**: is "justice" an *ideal* to defend or a *mechanism* to rationalize?

5. Limitations of AMR analysis

Although AMR offers a rich semantic representation, it has certain limitations:

- **Syntactic abstraction:** AMR ignores word order and surface structure, which can mask important rhetorical nuances.
- **Granularity:** Sentence-level analysis (as in Part 2) loses part of the overall argumentative and discursive context.
- **PropBank bias:** The absence of a verbal frame for "justice" reflects a limitation of PropBank, which focuses on verbal predicates, rather than an intrinsic characteristic of the concept.

5. Contribution of structural analysis

Despite these limitations, AMR analysis usefully complements the lexical analysis of Part 2. It allows us to:

- **Identify recurring patterns:** "justice" as modifier of "system," "field," "institution."
- **Reveal semantic roles:** "justice" is never agent, but always object, theme, or qualifier.
- **Map discursive axes:** institutional, penal, equity, managerial, conceptual.

In conclusion, structural analysis confirms that "justice" in the MapAIE corpus is primarily an **institutionalized conceptual entity**, oscillating between *ethical value* and *managerial apparatus*. This duality reflects the tensions inherent in ethical discourse on AI: how to reconcile abstract normative principles (fairness, equity) with imperatives of efficiency and rationalization?

CONCLUSION

4.1. Difficulties met with the project

Several methodological difficulties emerged during the processing of the corpus. First, the use of MAPie to retrieve the documents produced heterogeneous results across executions: some PDF or TXT files were occasionally missing, generating a partial and unstable corpus. We therefore retained the most complete version obtained, while noting that the retrieval system was not fully replicable.

Additionally, the initial extraction contained an excessive number of files, including non-relevant formats (PDF, HTML, TXT) or documents that should not have been integrated. This situation resulted from the fact that the automatic filtering based on the *included / not included* column did not function correctly. We had to manually review the *all_manifestos* file to determine which elements were truly admissible. The quality of the *all_manifestos* file also caused issues: several filenames did not correspond to their associated documents, and some information was shifted across columns. These inconsistencies stemmed from the automated retrieval system provided, which we did not control.

Finally, consolidating our work proved challenging due to the use of Jupyter notebooks. The heavy weight of cells and the density of the file sometimes complicated the aggregation of each contributor's work, requiring additional coordination and manual adjustments.

4.2. Learnings

Throughout this project, we collectively deepened our understanding of how **Natural Language Processing (NLP)** can be applied to analyze large and heterogeneous textual corpora. Working on the theme of *AI and justice* allowed us to combine technical, linguistic, and ethical perspectives in a single workflow.

One of our main lessons was the importance of **data preprocessing and cleaning**. We learned that raw text extracted from PDFs or HTML files often contains structural noise such as headers, footers, and duplicated metadata. Implementing several extraction tools — pdftotext, pdfminer, and PyPDF2 — helped us recover text more completely and handle diverse document formats. This experience highlighted that data quality directly determines the validity of any NLP analysis.

We also learned how to apply **machine learning methods for text exploration**, such as the *Isolation Forest* algorithm, to detect documents that behave differently from the rest of the corpus. This method taught us that outliers in NLP can have multiple origins — some due to formatting artifacts, others reflecting strong thematic focus. Understanding these distinctions required both quantitative reasoning and qualitative interpretation of textual context.

The creation of **frequency analyses, visualizations, and word clouds** helped us explore the thematic landscape of the corpus. We observed two main dimensions: a technological axis (with words like *data, systems, use*) and an ethical or societal one (with *rights, human, ethics*). These linguistic patterns revealed how the discourse around AI and justice alternates between technical and moral concerns.

Finally, this project strengthened our skills in **collaborative work and reproducible research**. By sharing code, comments, and visual analyses in Jupyter Notebooks, we learned how to integrate different methods into a coherent workflow. Overall, the project was an exercise in combining **technical rigor with linguistic interpretation**, demonstrating how NLP can illuminate the social and ethical dimensions of artificial intelligence.