MAIR

How Monitoring of AI Regulations turned into analysis of Memes in AI Research

Stanisław Giziński, Stanisław Łaniewski
16/05/2022 PW

Introduction

- Humans are technology-driven animals
- AI surge in research, Big Tech, startups
- But also regulations and culture
- How to monitor these dynamics?

MAIR: Framework for mining relationships between research articles, strategies, and regulations in the field of explainable artificial intelligence

4 sources: OECD AI Policy Observatory, NESTA AI Governance Database, arXiv, and the Semantic Scholar Research Corpus (S2ORC)

Extracting and processing metadata

Information extraction processes

Cross-citations between documents and other relevant data

Policy documents

Research papers

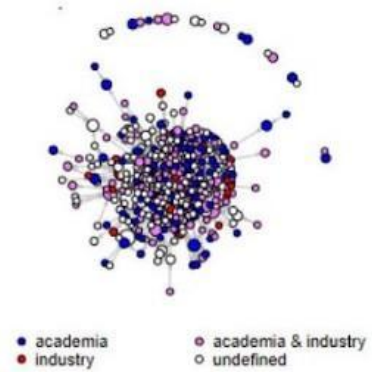
DATA PROCESSING

Network Analysis

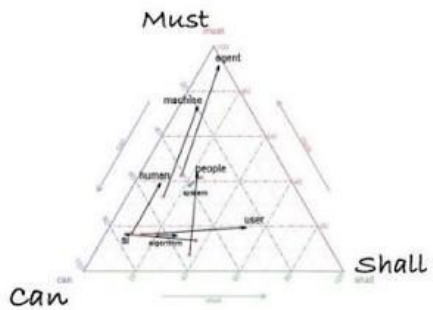
Institutional Grammar

Natural Language Processing

NETWORK



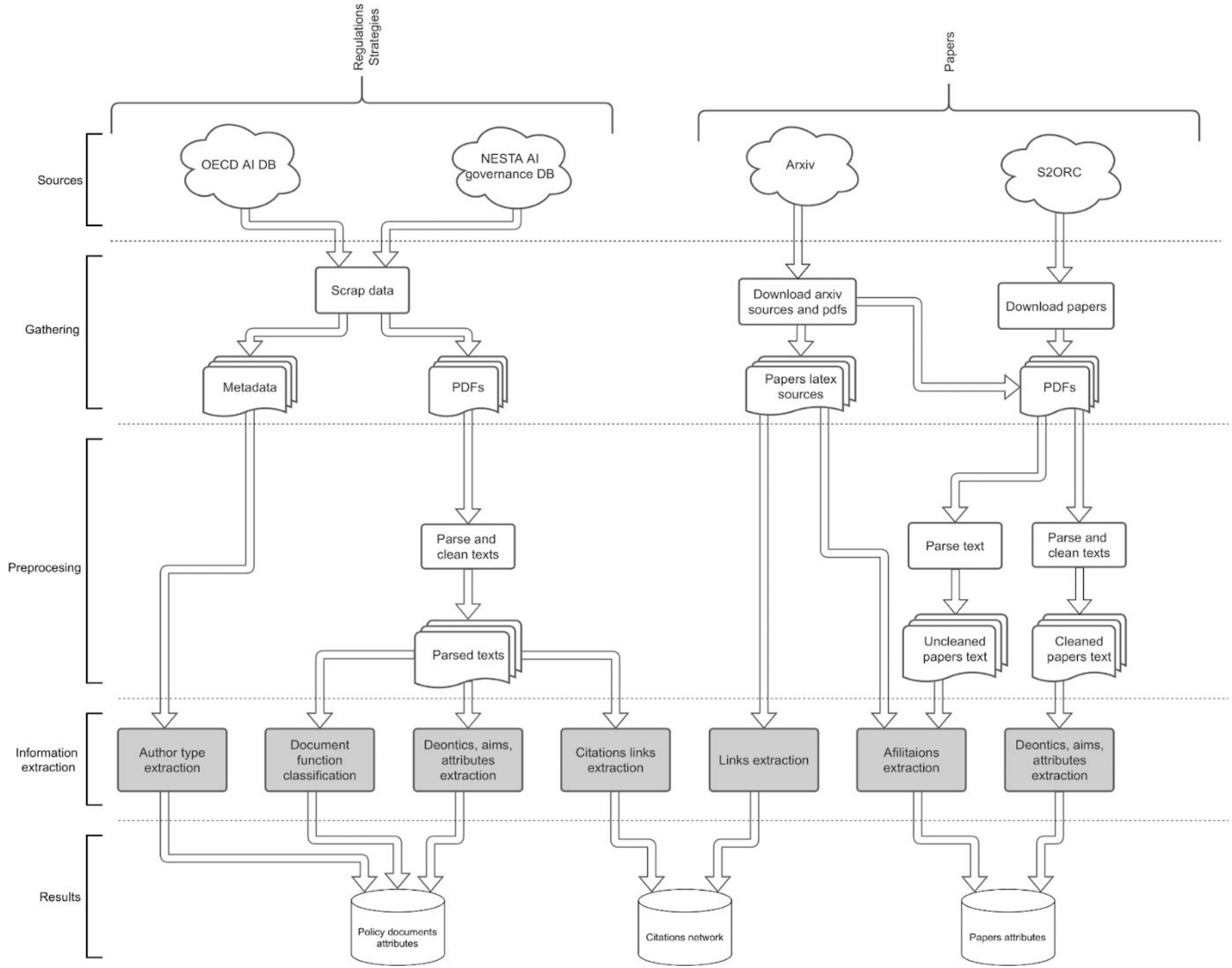
DEONTICS



DOCUMENT AIM

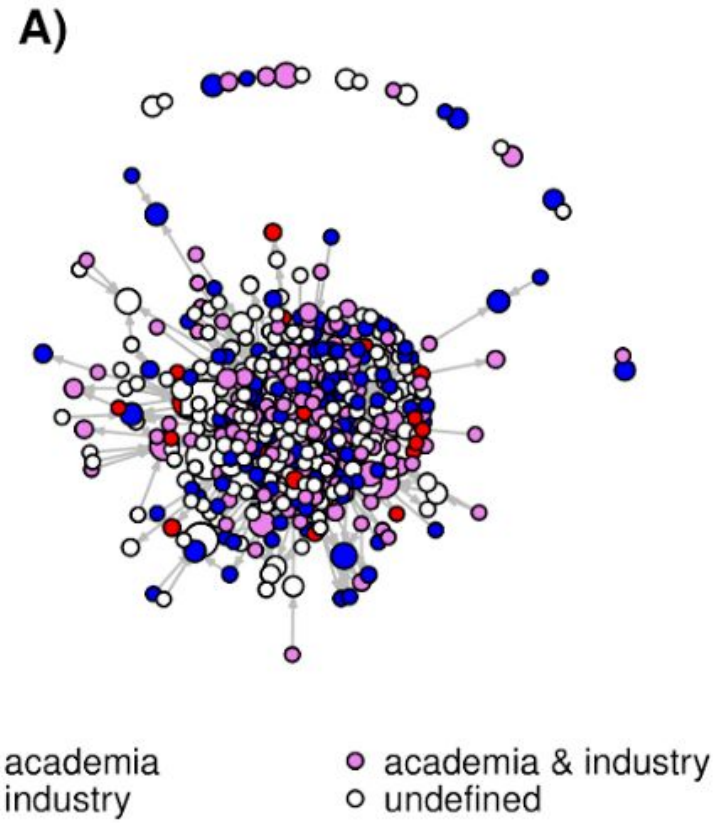
AFFILIATIONS

Locate
Extract
Match
Classify

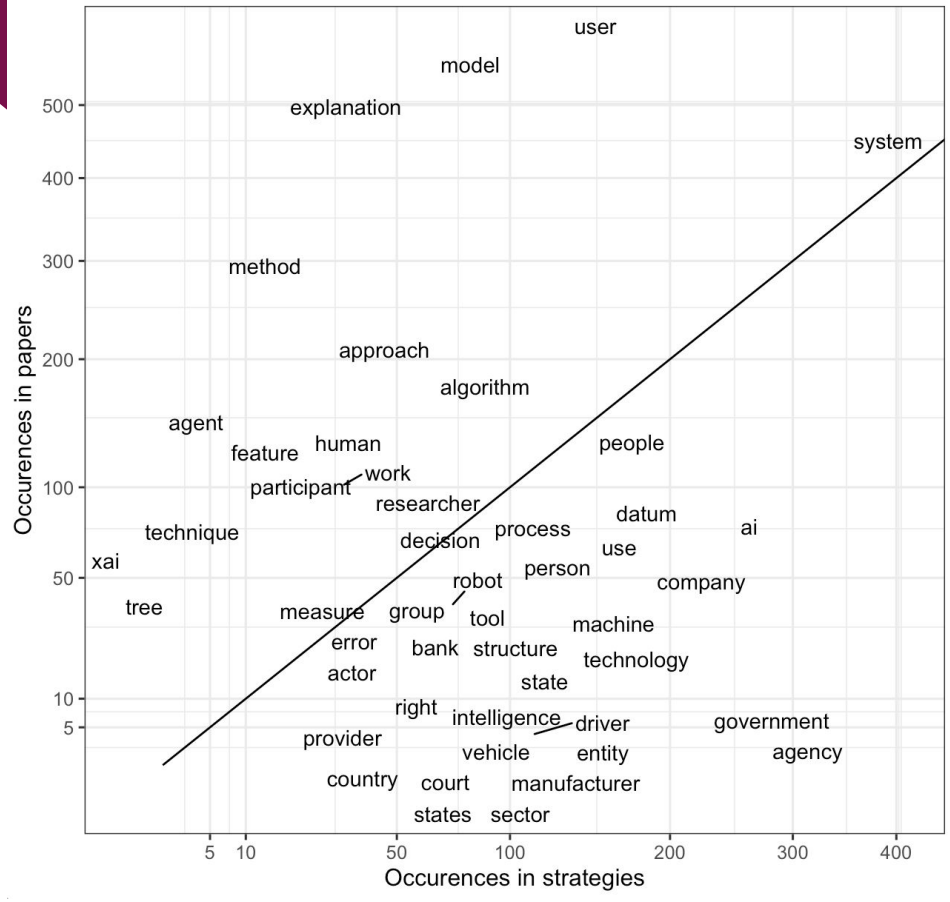

Regulations
Strategies
Papers
Sources
Gathering
Preprocesing
Information extraction
Results
OECD AI DB
NESTA AI governance DB
Arxiv
S2ORC
Scrap data
Download arxiv sources and pdfs
Download papers
Metadata
PDFs
Papers latex sources
PDFs
Parse and clean texts
Parse text
Parse and clean texts
Parsed texts
Uncleaned papers text
Cleaned papers text
Author type extraction
Document function classification
Deontics, aims, attributes extraction
Citations links extraction
Links extraction
Afilitaions extraction
Deontics, aims, attributes extraction
Policy documents attributes
Citations network
Papers attributes

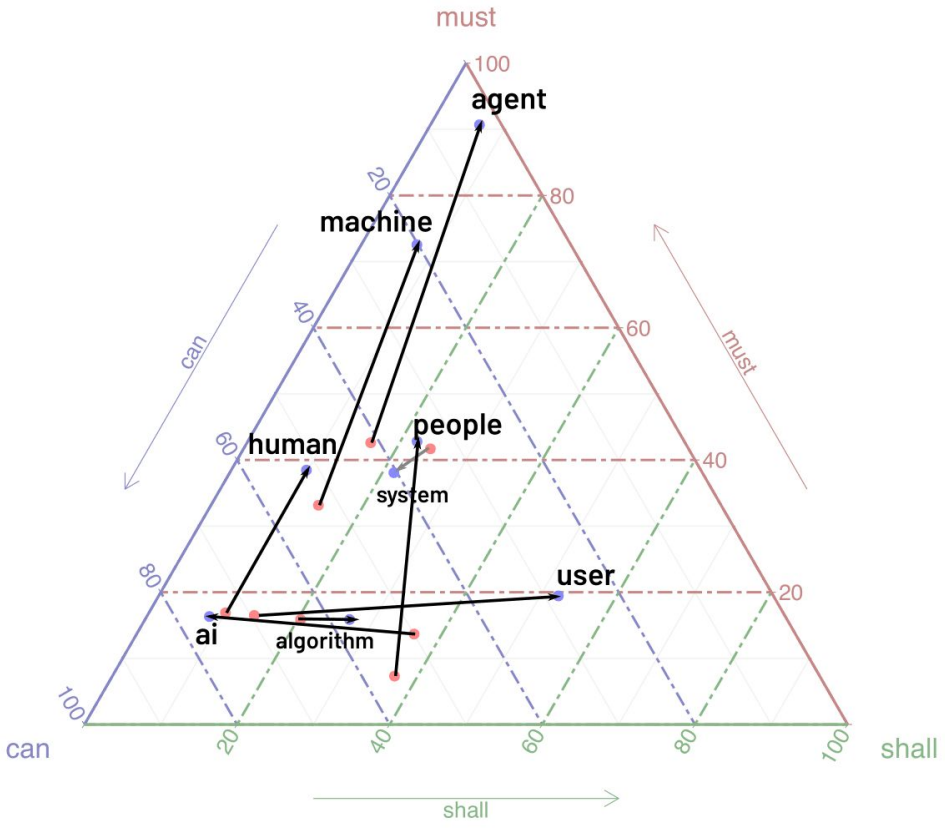
Citation network

- Which articles are the most prominent in AI studies, who their authors are, and what their organizational affiliations are?
- To what extent is the research on AI industry-driven, and to what extent is it developed at universities?
- Do papers that are essential for shaping experts' opinions influence processes shaping rules that will govern AI in the future? Do policymakers know how AI is conceptualized by the people who use this kind of systems in their everyday work?



Institutional Grammar and deontics

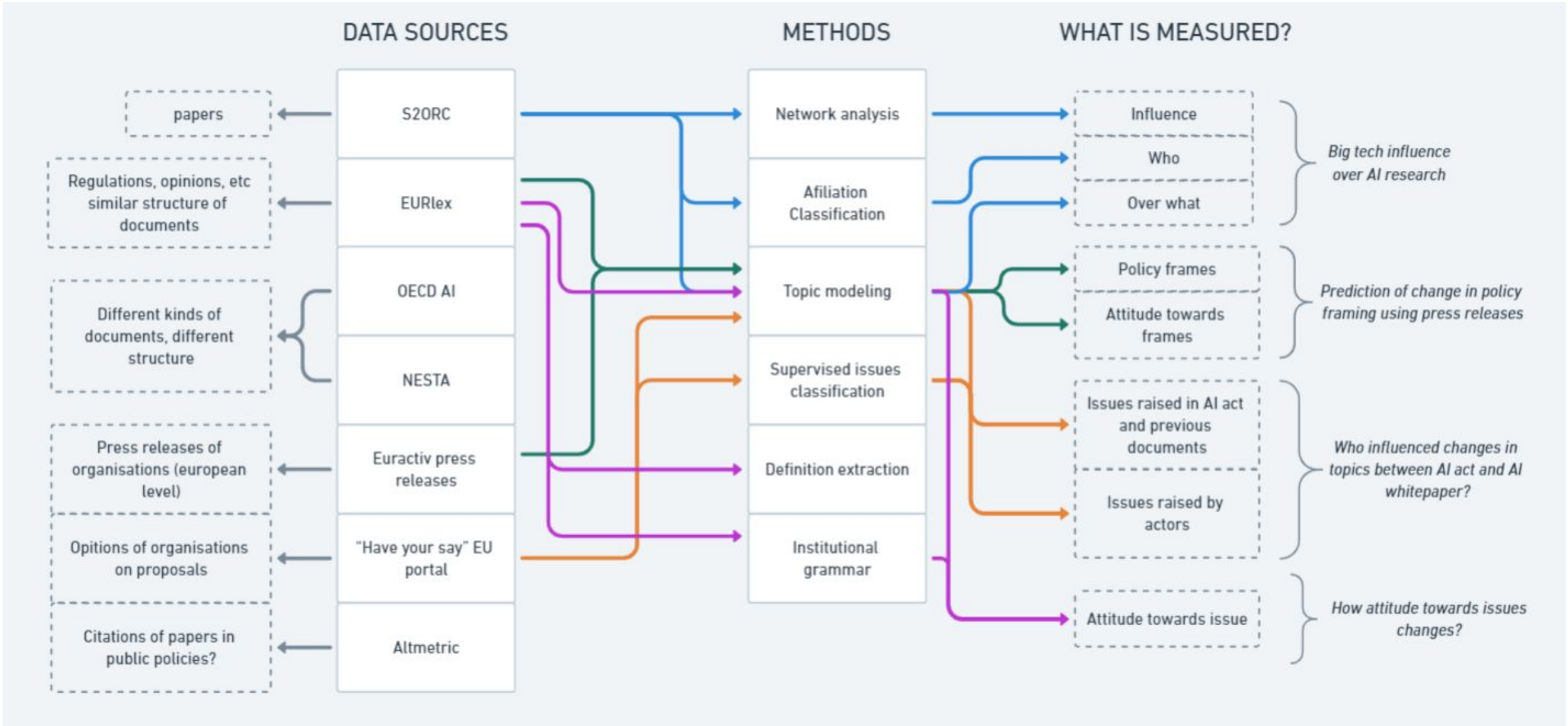




The Reviews: rejection

Two main takeaways from the reviews:

- Too many things in one paper
- Methods not validated enough


DATA SOURCES
METHODS
WHAT IS MEASURED?
papers
Regulations, opinions, etc similar structure of documents
Different kinds of documents, different structure
Press releases of organisations (european level)
Opitions of organisations on proposals
Citations of papers in public policies?
S2ORC
EURlex
OECD AI
NESTA
Euractiv press releases
"Have your say" EU portal
Altmetric
Network analysis
Afiliation Classification
Topic modeling
Supervised issues classification
Definition extraction
Institutional grammar
Influence
Who
Over what
Policy frames
Attitude towards frames
Issues raised in AI act and previous documents
Issues raised by actors
Attitude towards issue
Big tech influence over AI research
Prediction of change in policy framing using press releases
Who influenced changes in topics between AI act and AI whitepaper?
How attitude towards issues changes?

The New Plan - 3 research topics

Big tech influence over AI research

Affiliation extraction + classification

Network analysis

Topic modeling

AI Act public Consultations Analysis

Topical segmentation

Topic modeling

Agenda analysis/topic specific position estimation

AI Act Content Analysis

Topical segmentation

Topic modeling

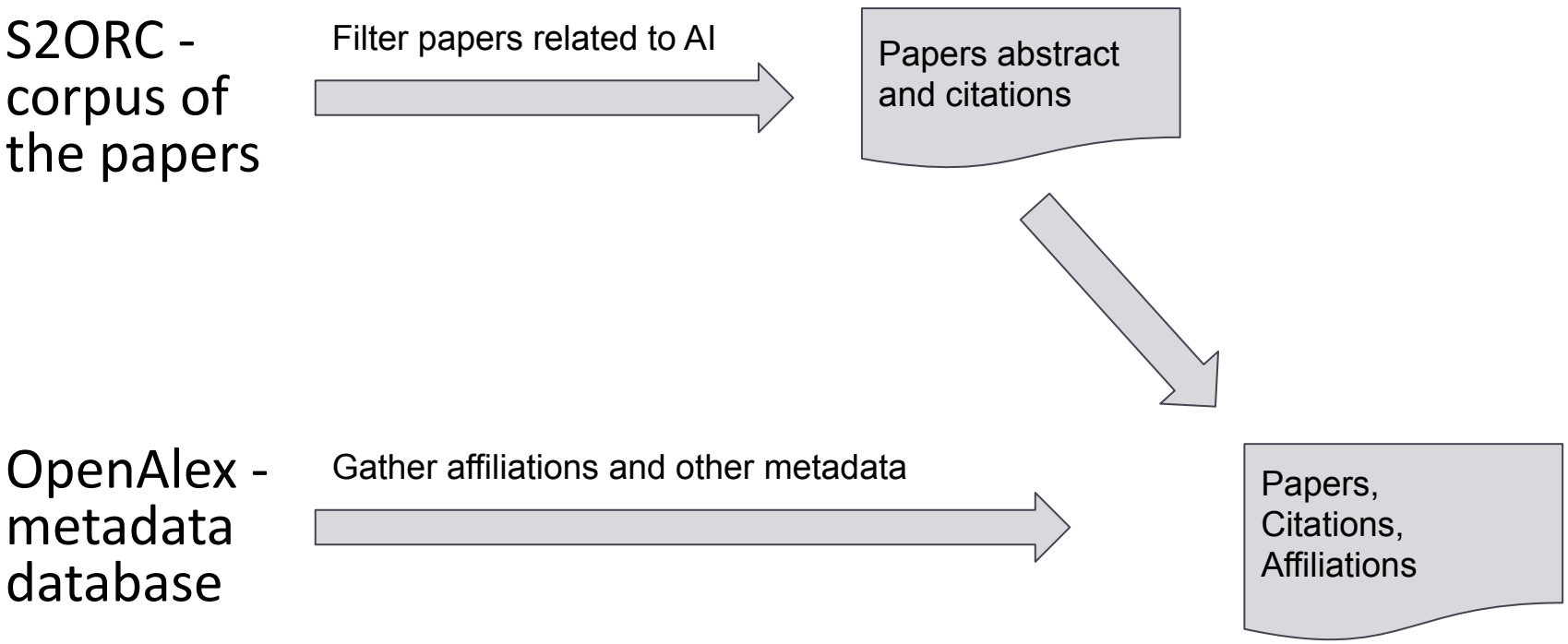
Deontic analysis

AI act vs AI whitepaper

Big tech influence over AI research

Who	→	Affiliations extraction and classification
Has influence	→	Citations analysis
Over what	→	Topics modeling / issues extraction

Data gathering pipeline



How to measure influence in citations network?

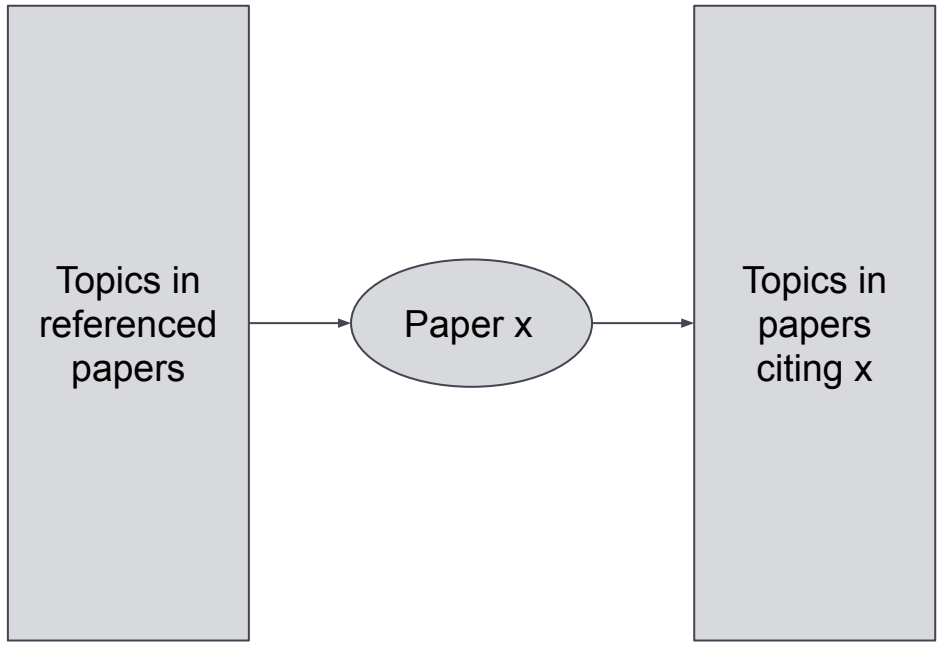
First idea (from Kuhn et. al): memes propagation score.

Memes (discrete units of knowledge, gossip, jokes and so on) are to culture what genes are to life. Just as biological evolution is driven by the survival of the fittest genes in the gene pool, cultural evolution may be driven by the most successful *memes*.
Richard Dawkins

$$P_m = \frac{d_{m \to m}}{d_{\to m}} \Big/ \frac{d_{m \to m}}{d_{\to m}}.$$

How to measure influence in citations network?

Second idea:
the retransmitter model.



The retransmitter model

We want to model

$$F(x\_memes, x\_meta) = y$$

where:

y - distribution of memes in papers citing paper *x*

x_memes - distribution of memes in papers referenced in paper *x*

x_meta - metadata of the paper *x* such as affiliation, country, region

How to discover memes?

Approach 1: topic modeling

Approach 2:

How to discover memes?

cluster	chunk
11	Support Vector Machine
11	SVM
11	Least Squares Support Vector Machine
11	Support Vector Machines
11	SVMs

cluster	chunk
5	comparative study
5	comparative analysis

Thank you for your attention