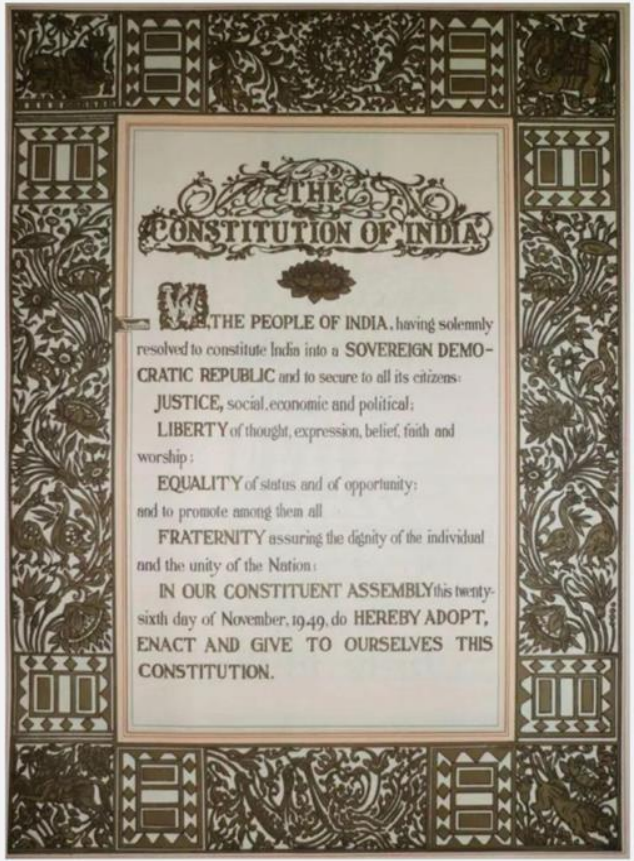
Justice Data and New Possibilities

OpenNyAI

Our mission is to empower every Indian citizen by enabling makers to exponentially increase access to justice by harnessing the power of AI in a **safe** and **affordable** way.

LAW



THE CONSTITUTION OF INDIA

WE, THE PEOPLE OF INDIA, having solemnly resolved to constitute India into a SOVEREIGN DEMOCRATIC REPUBLIC and to secure to all its citizens:

JUSTICE, social, economic and political;

LIBERTY of thought, expression, belief, faith and worship;

EQUALITY of status and of opportunity;

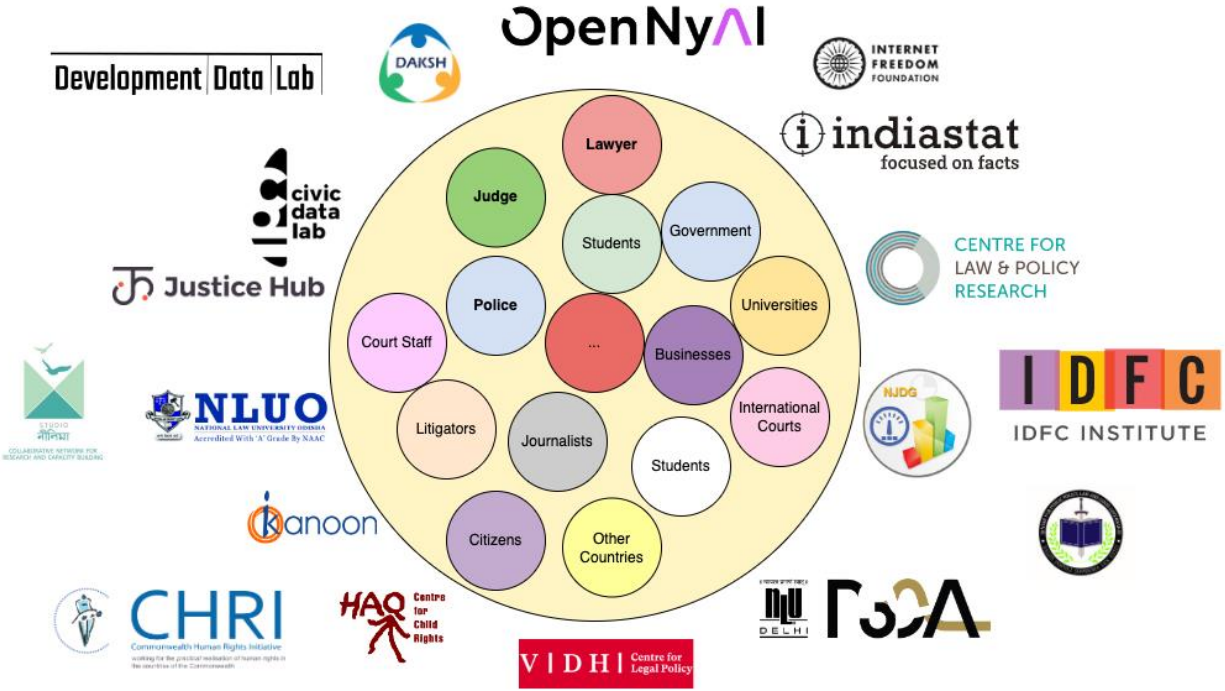
and to promote among them all

FRATERNITY assuring the dignity of the individual and the unity of the Nation;

IN OUR CONSTITUENT ASSEMBLY this twenty-sixth day of November, 1949, do HEREBY ADOPT, ENACT AND GIVE TO OURSELVES THIS CONSTITUTION.

JUSTICE

Very vibrant ecosystem for Indian legal Data … a snapshot



Quant - What can data tell us?

Built for Purpose (1/2)

	Goals	Theme	Observation Period	Contents	Broad Insights
Development Data Lab		Bias in decision making. Lifecycle of Civil and Criminal cases	2010-2018	Acts, Sections, Petitioner and Respondent, Dates of listing, filing, decision etc.	Supervised learning has been used to predict gender and religion
DAKSH's Database on Judicial Data		Mapping across civil and criminal cases, the case type, stages etc.	2010-2019	Case Type mentioned in E-Courts data mapped with civil or criminal proceeding	The total number of unique case types varies across years, states and sometimes even within courts
Immovable property related disputes		Analyse the status of property related cases in Delhi HC and Govt. as litigant.	1st Jan 2007 - 31st Dec 2019	Case-level Data from the Delhi High Court	Only 17% of the property disputes in Delhi HC relates to property. Union Govt. is a petitioner in 2% and respondent in 18%.
NCLT Cause List		Cause list of NCLT	23rd Feb, 2018 - 31st Jan, 2020	Cause list of ~5000 cases in NCLT with case types, legal rep, petitioner etc.	It's surprising that s. 253 cases are more than IBC cases
Romantic Cases under the POCSO Act		Analysing the use and misuse of POCSO Act where victim was in love/had affair	2016 - 2020	Cause list of ~5000 cases in NCLT with case types, legal rep, petitioner etc.	Acquittal is a norm in romantic cases

Legends

Standardization Application Analysis Individual/Group Contribution

Built for Purpose (2/2)

	Goals	Theme	Observation Period	Contents	Broad Insights
civic data lab Zombie Tracker		Section 66A, of IT Act and its continued use by Police	2016-18	Details of cases and judgements on offences booked under s. 66A of IT Act	S. 66A has already been struck down by Honbl. Supreme Court of India
		Constitution Benches of the Supreme Court of India	2010-22	Voting and alignment patterns; authorship; foreign precedents referred to; and pendency in constitution bench matters	~42% of these cases refer to precedents from other countries
KHOJ KNOW YOUR HIGH COURT JUDGES		Background of High Court Judges, Information unavailability on the judges	**6th Oct-1993** - 31st May 2021	43 Variables on Judges ranging from education to past work experience.	Limited Data on past work and educational experience
		Background of Supreme Court Judges	1950-2021 & 1950-2022	Very exhaustive dataset	Mostly dataset isn't available
		All habeas corpus petitions of the Supreme Court in preventive detention cases in a 20-year time frame	2000-2019	~25 variables on all habeas corpus cases in Supreme Court	On average, the total time from detention order or actual detention until final disposal by the Supreme Court is 953 days, and the median is 478 days.

Legends

Standardization Application Analysis Individual/Group Contribution

Data - Who owns the data?

Landscape

Data custodians and access to raw data

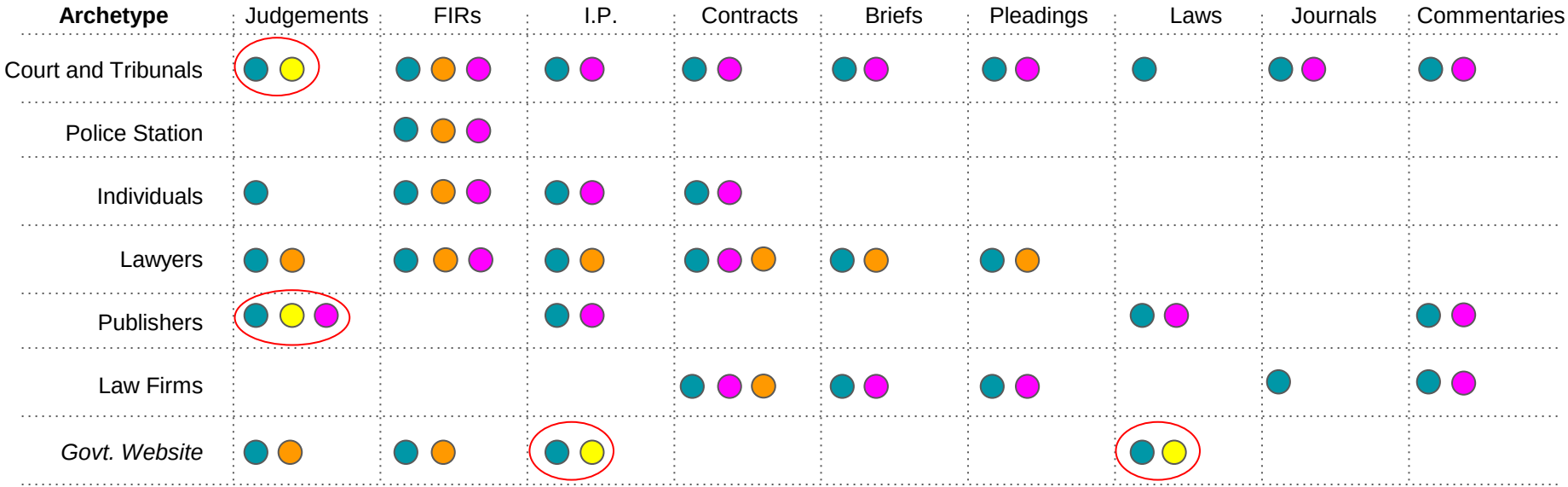
Archetype	Judgements	FIRs	I.P.	Contracts	Briefs	Pleadings	Laws	Journals	Commentaries
Court and Tribunals	● ●	● ● ●	● ●	● ●	● ●	● ●	●	● ●	● ●
Police Station		● ● ●							
Individuals	●	● ● ●	● ●	● ●					
Lawyers	● ●	● ● ●	● ●	● ● ●	● ●	● ●			
Publishers	● ● ●		● ●				● ●		● ●
Law Firms				● ● ●	● ●	● ●		●	● ●
Govt. Website	● ●	● ●	● ●				● ●		

Legends

● Data Owner ● Sensitive or Copyrighted ● Limited ● ~ Exhaustive

Landscape

Data custodians and access to raw data

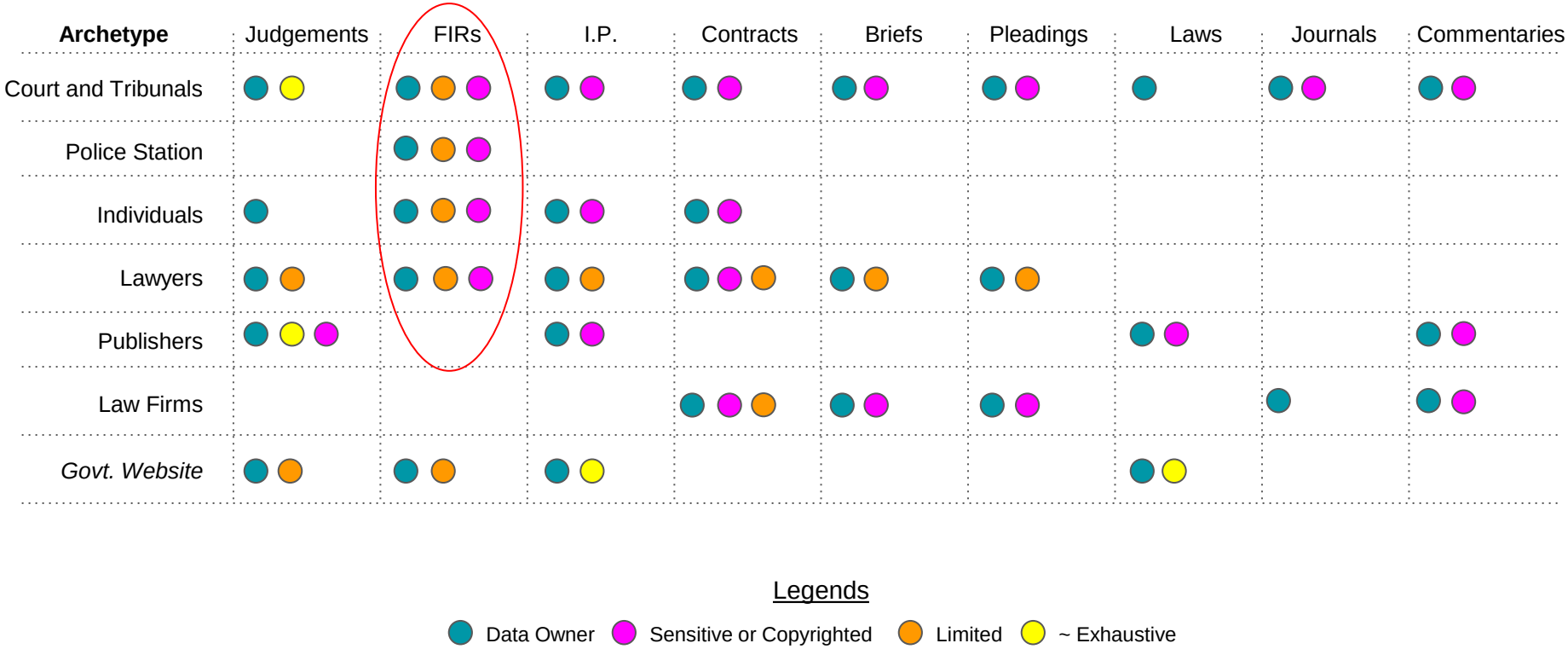


Legends

Data Owner | Sensitive or Copyrighted | Limited | ~ Exhaustive

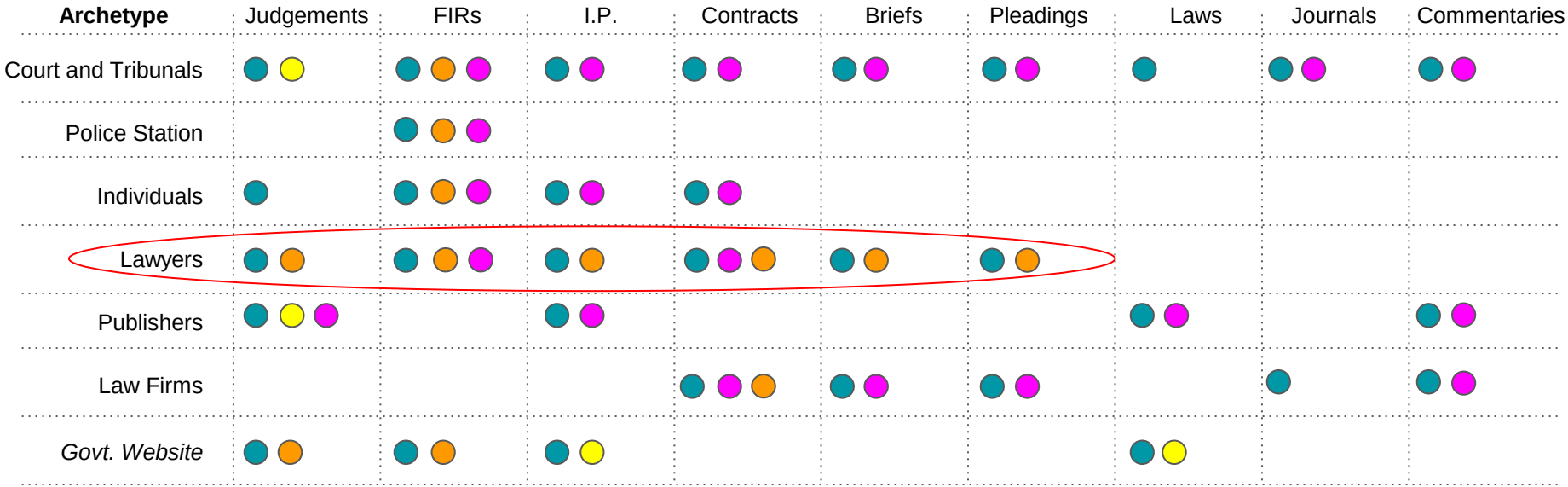
Landscape

Opportunity to crowdsource FIRs



Landscape

Lawyers an important stakeholder group



Legends

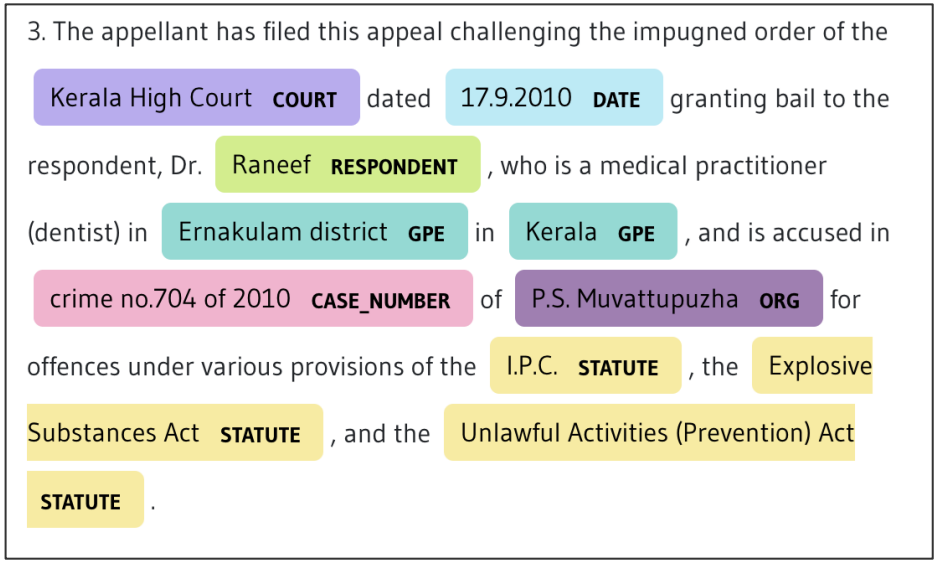
Data Owner | Sensitive or Copyrighted | Limited | ~ Exhaustive

AI - What’s possible now?

Challenges with Justice Data

- Medium/Large Scale analysis - so far not feasible
  - Court’s can’t tell you which is the most contested sections in their courts over a given period of time
  - Metadata is usually suited for long term trend analysis but not granular analysis
- Language accessibility
  - High Courts and above manage fraction of all cases and are in English
  - District courts and below have documents and texts in vernacular - Significantly more complicated
- Standardization
  - All high courts exercise jurisdiction and same things are termed differently in different geography. E.g. NBW Issued & Summons.

Opportunities with AI - Named Entity Recognition



Named Entity Recognition

- Richer Metadata
- Better downstream tasks (classification and categorization of cases etc)

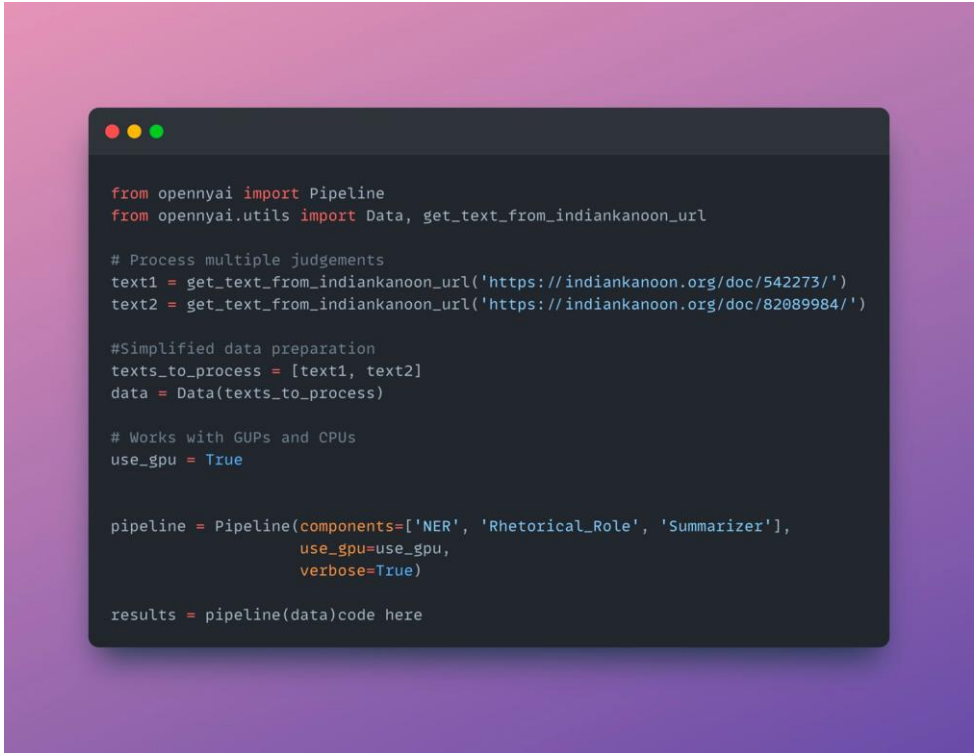
Opportunities with AI - Rhetorical Roles

4) In deciding bail applications an important factor which should certainly be taken into consideration by the Court is the delay in concluding the trial. Often this takes several years, and if the accused is denied bail but is ultimately acquitted, who will restore so many years of his life spent in custody?
Is Article 21 of the Constitution, which is the most basic of all the fundamental rights in our Constitution, not violated in such a case? PRE_RELIED
Of course this is not the only factor, but it is certainly one of the important factors in deciding whether to grant bail.
13. In the present case the respondent has already spent 66 days in custody (as stated in paragraph 2 of his counter affidavit), and we see no reason why he should be denied bail.
A doctor incarcerated for a long period may end up like Dr. Manette in Charles Dicken's novel `A Tale of Two Cities', who forgot his profession and even his name in the Bastille. RATIO
14. With the above observations, this appeal is dismissed. RPC

Rhetorical Roles

- Identify which part of judgement deals with what?
- Extract final outcome of cases

Opportunities with AI - Tools



```python
from opennyai import Pipeline
from opennyai.utils import Data, get_text_from_indiankanoon_url

# Process multiple judgements
text1 = get_text_from_indiankanoon_url('https://indiankanoon.org/doc/542273/')
text2 = get_text_from_indiankanoon_url('https://indiankanoon.org/doc/82089984/')

#Simplified data preparation
texts_to_process = [text1, text2]
data = Data(texts_to_process)

# Works with GUPs and CPUs
use_gpu = True


pipeline = Pipeline(components=['NER', 'Rhetorical_Role', 'Summarizer'],
                    use_gpu=use_gpu,
                    verbose=True)

results = pipeline(data)code here
```

Available as Python library

- Run offline on your own computer without needing to host an infrastructure

Opportunities with AI - Standardization

State	Land record nomenclature
Andhra Pradesh	Pahani
Assam	Jamabandi/Patta
Bihar	Khatiyan
Chhattisgarh	P-II/B-I
Delhi	Khatauni
Gujarat	7/12
Haryana	Jamabandi
Himachal Pradesh	Jamabandi

Legal Ontologies

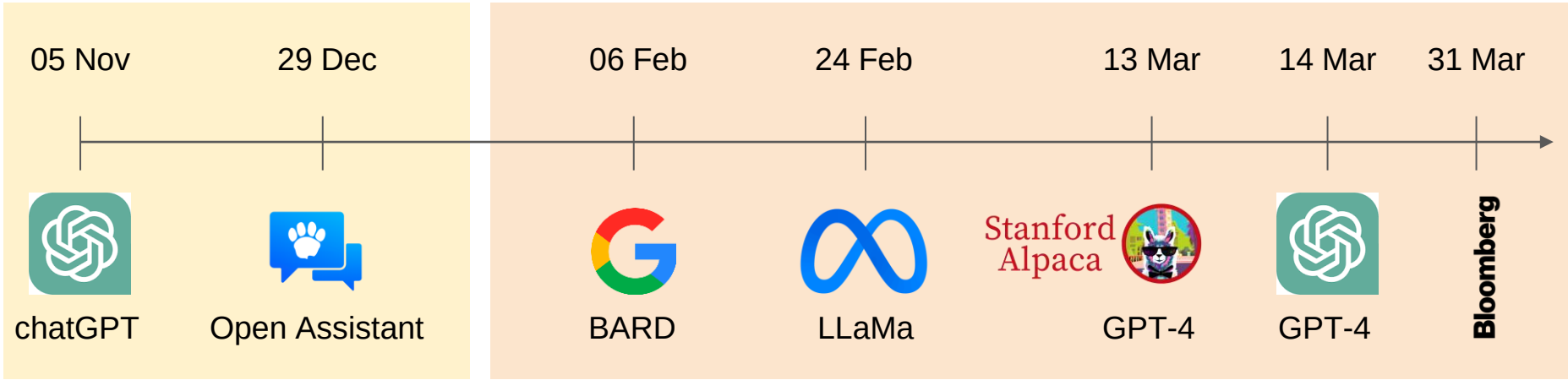
- Standardize nomenclature across
- Provide a layer for nomenclature to be interoperable

How can we not talk about GPTs?

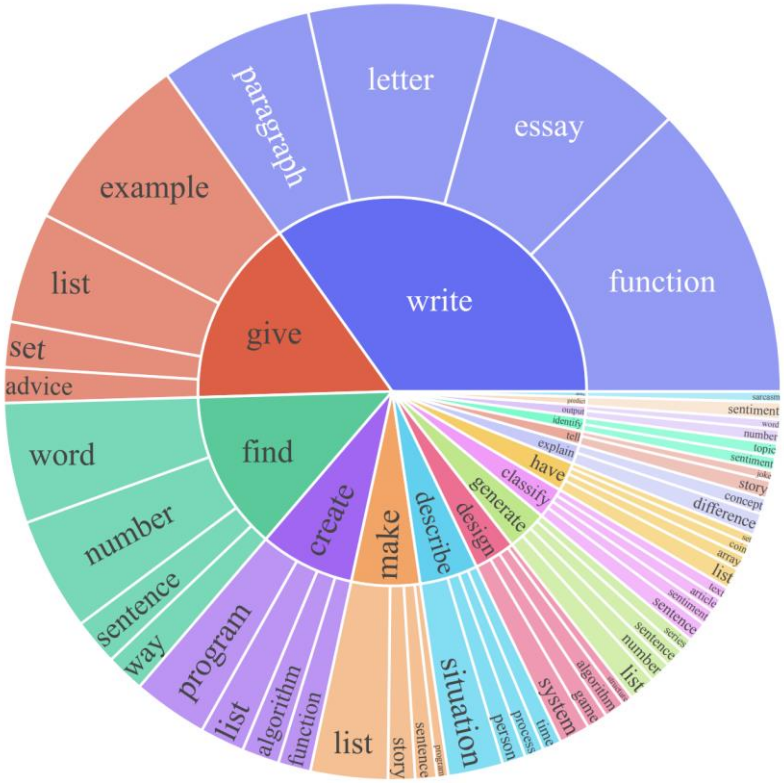
Opportunities with AI - Citizen facing tools

Significant stories

- Cost of Compute has gone down by **+90%**
- Reinforcement Learning with Human Feedback seems like one of the ways to align AI
- Models with comparable performance are much smaller. Small enough to be run on a laptop.
- GPT-4 has aced standardised exams



Large Language Model



{"id": "seed_task_0",
"name": "breakfast_suggestion",
"instruction": "Is there anything I can eat for a breakfast that doesn't include eggs, yet includes protein, and has roughly 700-1000 calories?",
"instances": [{"input": "", "output": "Yes, you can have 1 oatmeal banana protein shake and 4 strips of bacon. The oatmeal banana protein shake may contain 1/2 cup oatmeal, 60 grams whey protein powder, 1/2 medium banana, 1tbsp flaxseed oil and 1/2 cup watter, totalling about 550 calories. The 4 strips of bacon contains about 200 calories."}], "is_classification": false}