LREC 2022
Marseille

OpenNyai
MISSION TO ADVANCE AI FOR JUSTICE IN INDIA

Corpus for Automatic Structuring of Legal Documents


Prathamesh Kalamkar


Aman Tiwari


Astha Agarwal


Saurabh Karn


Smita Gupta


Vivek Raghavan


Ashutosh Modi

EkStep

thoughtworks

agami IDEAS THAT SERVE JUSTICE

धर्मो रक्षति रक्षितः

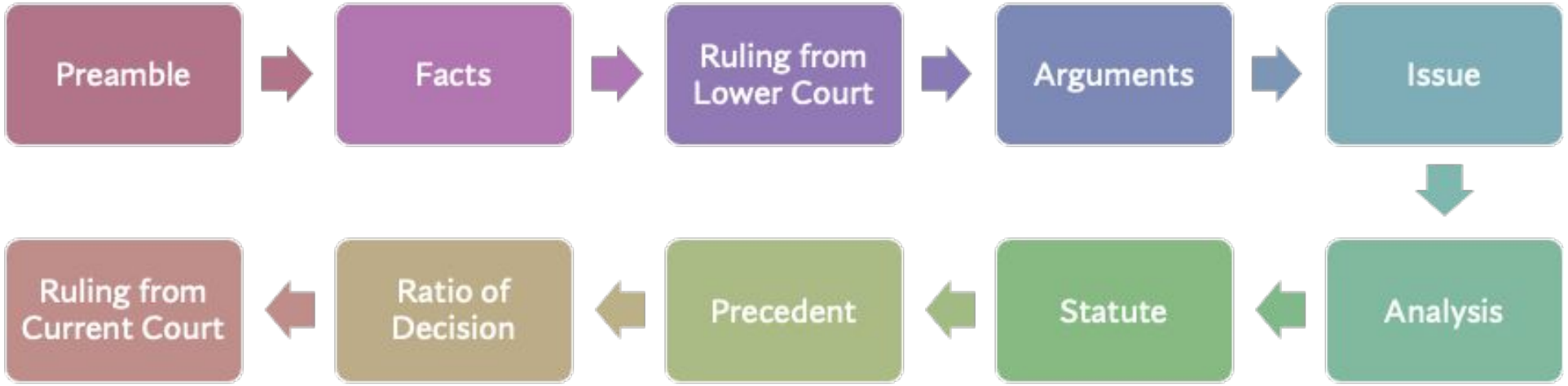
ROHINI NILEKANI philanthropies

भारतीय प्रौद्योगिकी संस्थान कानपुर
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

Automatic Structuring of Court Judgements

Indian Court judgements have an inherent structure that is not explicitly mentioned in judgement text.

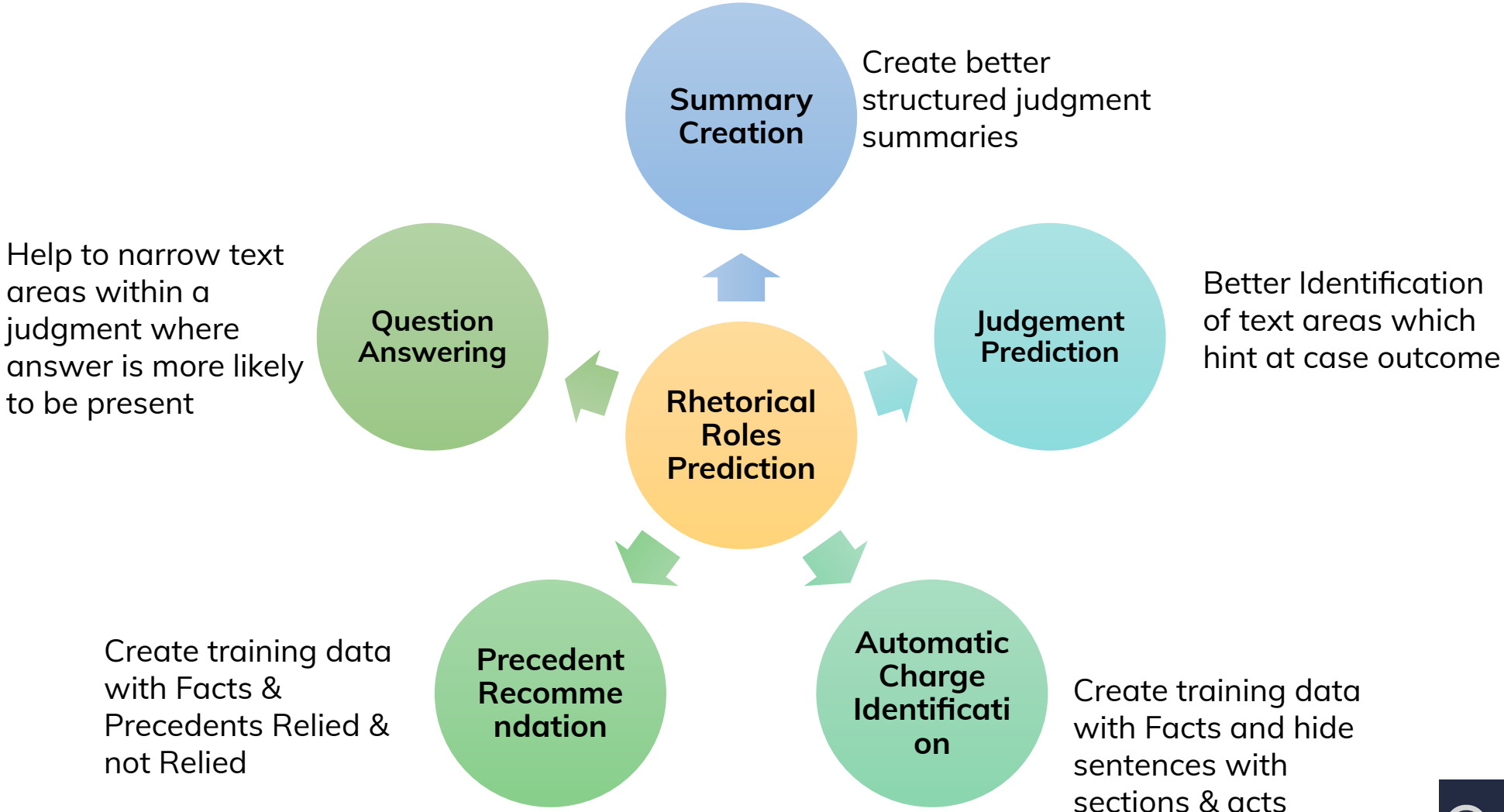
The structure could be obtained by assigning each sentence of judgment with one of the following Rhetorical Roles



Typical Structure of Court Judgement

Importance of Automatic Structuring of Court Judgements

Automatic structuring would act as a Building Block which helps many legal AI applications



OpenNyai

Related Work

Structuring of Legal Documents

- **Saravanan et al. (2008)** proposed segmentation of legal documents using seven different roles using a CRF-based model
- **Bhatia (2014)** created Genre Analysis of Legal Texts to create seven rhetorical categories.
- **Bhattacharya et al. (2019b)** have proposed CRF-BiLSTM model for automatically assigning rhetorical roles to sentences in Indian legal documents.
- **Malik et al., 2021a** have created a RR corpus and further they have developed a multi-task learning based model for predicting RR.

Applications of Structuring of Legal Documents

- **Summary Creation**: Farzindar and Lapalme (2004); Hachey and Grover (2006)
- **Judgement outcome prediction**: Malik et al. (2021b) , Strickson and De La Iglesia (2020b), Chalkidis et al. (2019), Xiao et al. (2018)

Main Contributions

- BUILDNyAI is a sequential sentence classification dataset which provides structure to Indian Court judgements (Criminal & Tax) using 13 sentence rhetorical roles

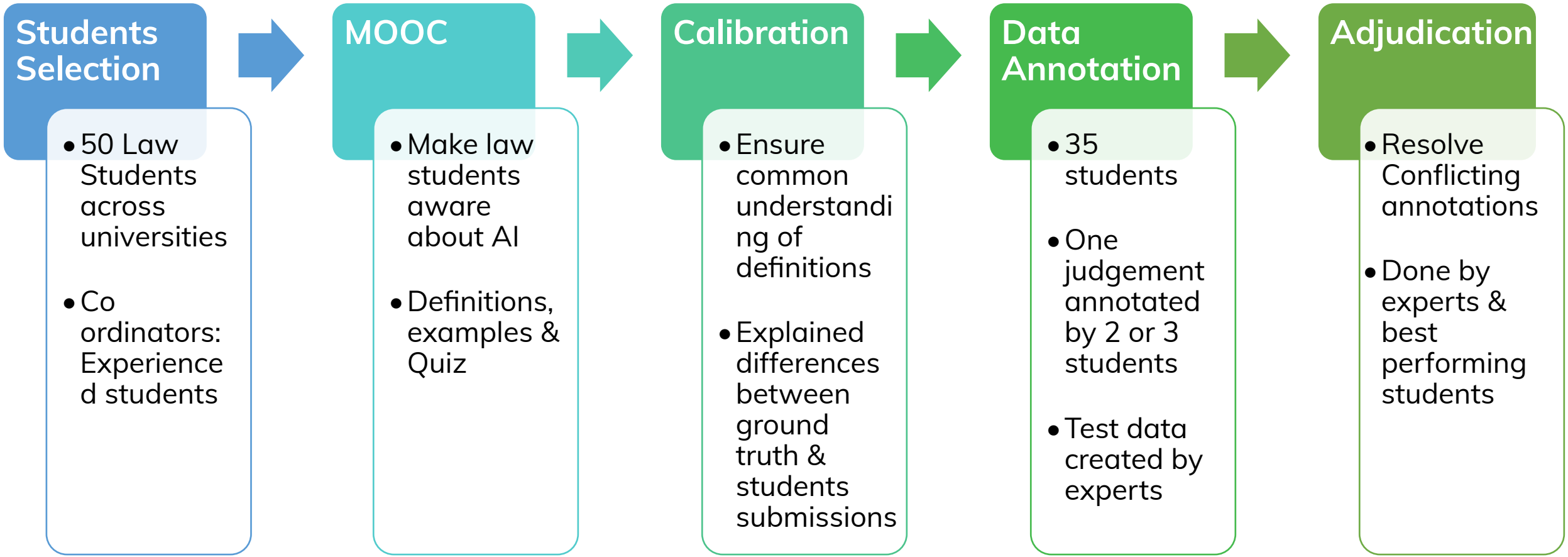
Dataset	Docs	Sen-tences	Tokens
Train	247	28986	938K
Validation	30	2879	88K
Test (in-domain)	50	4158	134K
Test (out-domain)	27	4282	127K
Total	**354**	**40305**	**1.3M**

- Transformer-based **baseline model** for automatically predicting legal documents with sentence-level RR
- Leaderboard accepting submissions and showing state of art results

- **2 use cases** where Rhetorical Roles predictions help to achieve better results: Judgement Summarization & Judgement outcome prediction

  Corpus, Baseline Model & Use cases code: Github Repository

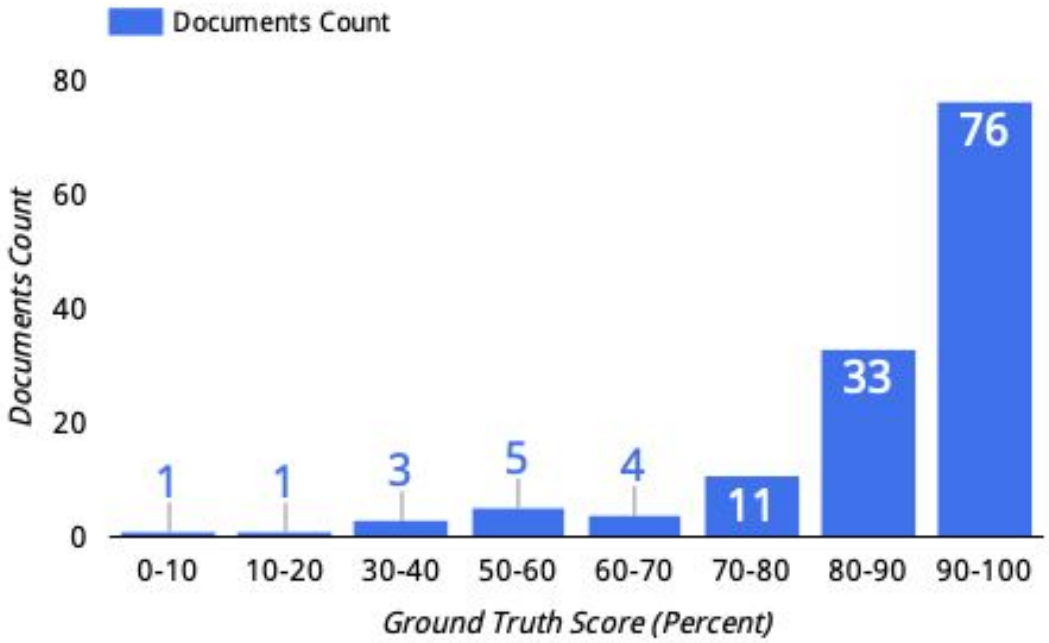
Data Annotation Process



Data Quality Checks

Ground Truth Scores

- Students were randomly allocated a ground truth Document
- Co ordinators looked at the mistakes and advised students to correct them

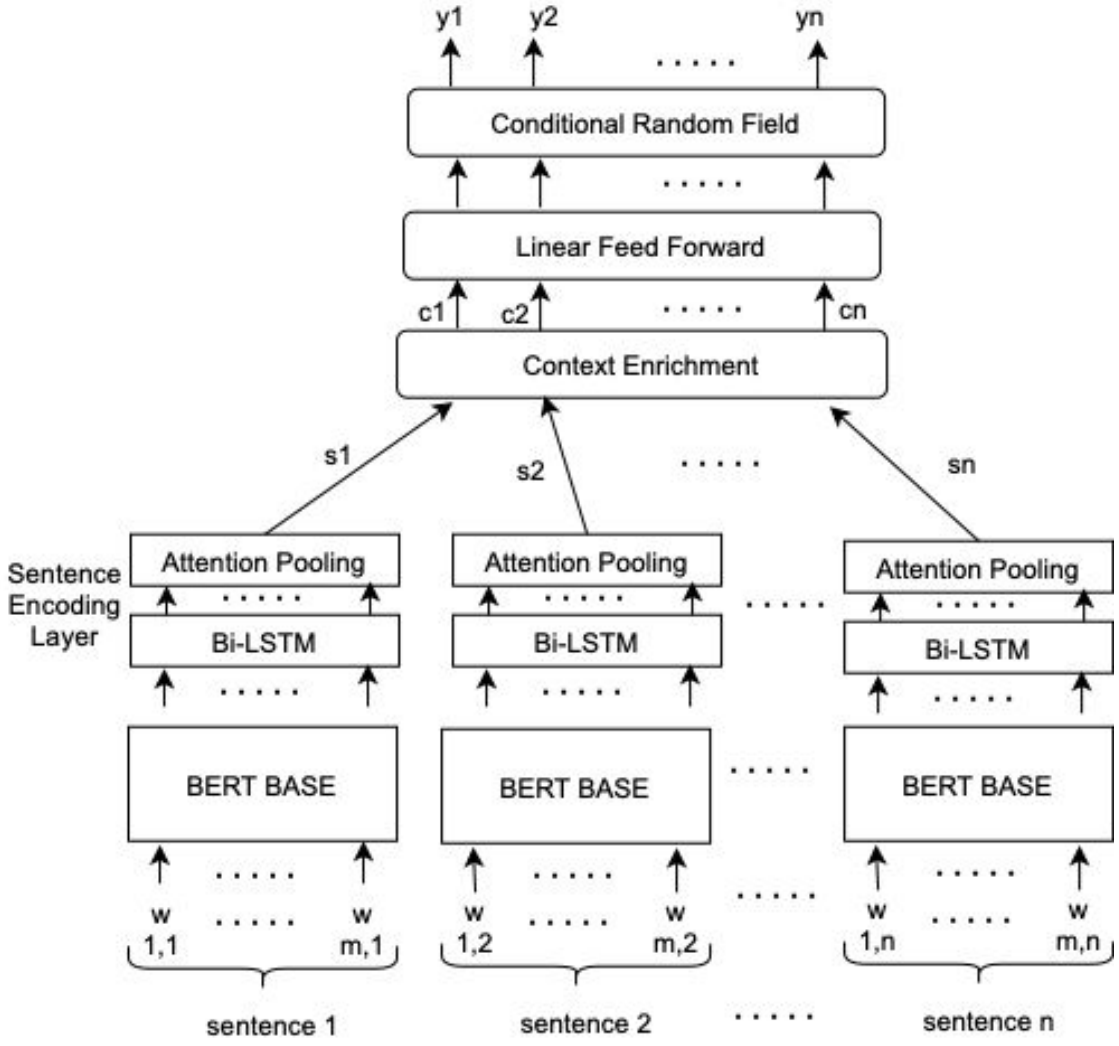


Inter Annotator Score

- Overall Fleiss Kappa: 0.59
- Adjudication removes the conflicts

- High Agreement for
  - PREAMBLE, RPC, NONE, and ISSUE.
- Medium Agreement for
  - FACTS, RLC, ANALYSIS, PRECEDENT, and ARGUMENTS
- Low Agreement for
  - RATIO

Baseline Prediction Model



Baseline Model Architecture inspired by SciBERT-HSLN architecture (Brack et al., 2021)

Baseline Model Performance on test (in domain) by each Rhetorical Role

Rhetorical Role	Precision	Recall	F1
ANALYSIS	0.77	0.89	0.83
ARG_PETITIONER	0.60	0.64	0.62
ARG_RESPONDENT	0.84	0.41	0.55
FAC	0.80	0.84	0.82
ISSUE	0.93	0.87	0.90
NONE	0.85	0.84	0.85
PREAMBLE	0.96	0.98	0.97
PRE_NOT_RELIED	0.00	0.00	0.00
PRE_RELIED	0.79	0.60	0.68
RATIO	0.53	0.56	0.54
RLC	0.75	0.45	0.57
RPC	0.78	0.87	0.82
STA	0.77	0.54	0.64
Overall	0.79	0.80	0.79

On the test (out domain) data, the baseline model weighted f1 is 0.70. The degradation in performance is due to different style of writings by tribunals and district courts.

Applications of Rhetorical Roles Prediction

Extractive Summarization

- Fine tuned BERTSUM model with & without Rhetorical Roles.
- Hypothesis is that with rhetorical roles, BERTSUM would be able to better identify which sentences should go to summary

	R1	R2	RL
BERTSUM	0.6	0.42	0.59
BERTSUM RR	**0.62**	**0.46**	**0.61**

Abstractive Summarization

- Used Pretrained Legal Pegasus model too create abstractive summaries with & without rhetorical roles
- Create abstractive summary for each rhetorical role and concatenate summaries. Short rhetorical roles are copied as is to summaries

	R1	R2	RL
Legal Pegasus	0.6	0.42	0.59
Legal Pegasus RR	**0.62**	**0.46**	**0.61**

Judgment Prediction

- Certain parts of judgment text provide more information about the outcome of the case. These could be identified with help of rhetorical roles
- Using last 512 tokens of text predicted with ANALYSIS performs better than taking last 512 of judgment text

	Precision	Recall	F1
XLNet last512	0.76	0.49	0.59
XLNet last512 Analysis	**0.71**	**0.55**	**0.62**

Conclusions

- We proposed a new corpus of legal judgment documents annotated with 13 different Rhetorical Roles

- We believe that legal AI researchers would find this dataset useful and they would participate through submissions on leaderboard and making their models open

- The baseline rhetorical role prediction model would help power more legal AI applications as we demonstrated in 2 use cases

Collaborate with OpenNyAI

OpenNyAI is a mission aimed at developing open source software and datasets to catalyze the creation of AI-powered solutions that will advance access to justice

- Legal Expert or Law Student
  - Participate in MOOC & Volunteer for Data Annotation
- Developer
  - Help in building Open source Annotation Tool
  - Build more applications using building block components
- Data Scientist
  - Participate in Benchmarks and make your model open

Thank you!

- Please join us for Q/A session

- Resources
  - Leaderboard, data
  - Github repository
  - OpenNyAI homepage