

# AsyLex: A Dataset for Legal Language Processing of Refugee Claims

Claire Barale<sup>1</sup>, Mark Klaisoongnoen<sup>2</sup>, Pasquale Minervini<sup>1</sup>,  
Michael Rovatsos<sup>1</sup>, Nehal Bhuta<sup>3</sup>

School of Informatics<sup>1</sup>, EPCC<sup>2</sup>, School of Law<sup>3</sup>, The University of Edinburgh



**EMNLP  
2023**

## Key features

- First dataset for **Refugee Law**: Refugee Status Determination Cases
- 59,112 anonymized documents** from **Canada**, from 1996 to 2022
- 7.80% of First Instance decisions and 92.19% of Appeal decisions
- 21.68% of positive decisions (Refugee Protection Granted)
- 19,115** human-labeled gold-standard annotations on 20 legally relevant entity types
- 57,408** rule-based annotations
- 1,682** documents labeled (gold) with their decision outcome
- Legal Judgement Classification with a Macro F1 of up to **99.82%**

## What can you do with AsyLex?

- 2 tasks demonstrated: (1) **Legal Entity Extraction**, and (2) **Judgement Outcome Classification**



For the NLP Community

Pretraining / Fine-tuning

Legal Entity Extraction

Text Classification

Case and Bias Analysis



For legal practitioners

Legal Search

Past Cases Retrieval

Insights for new applications

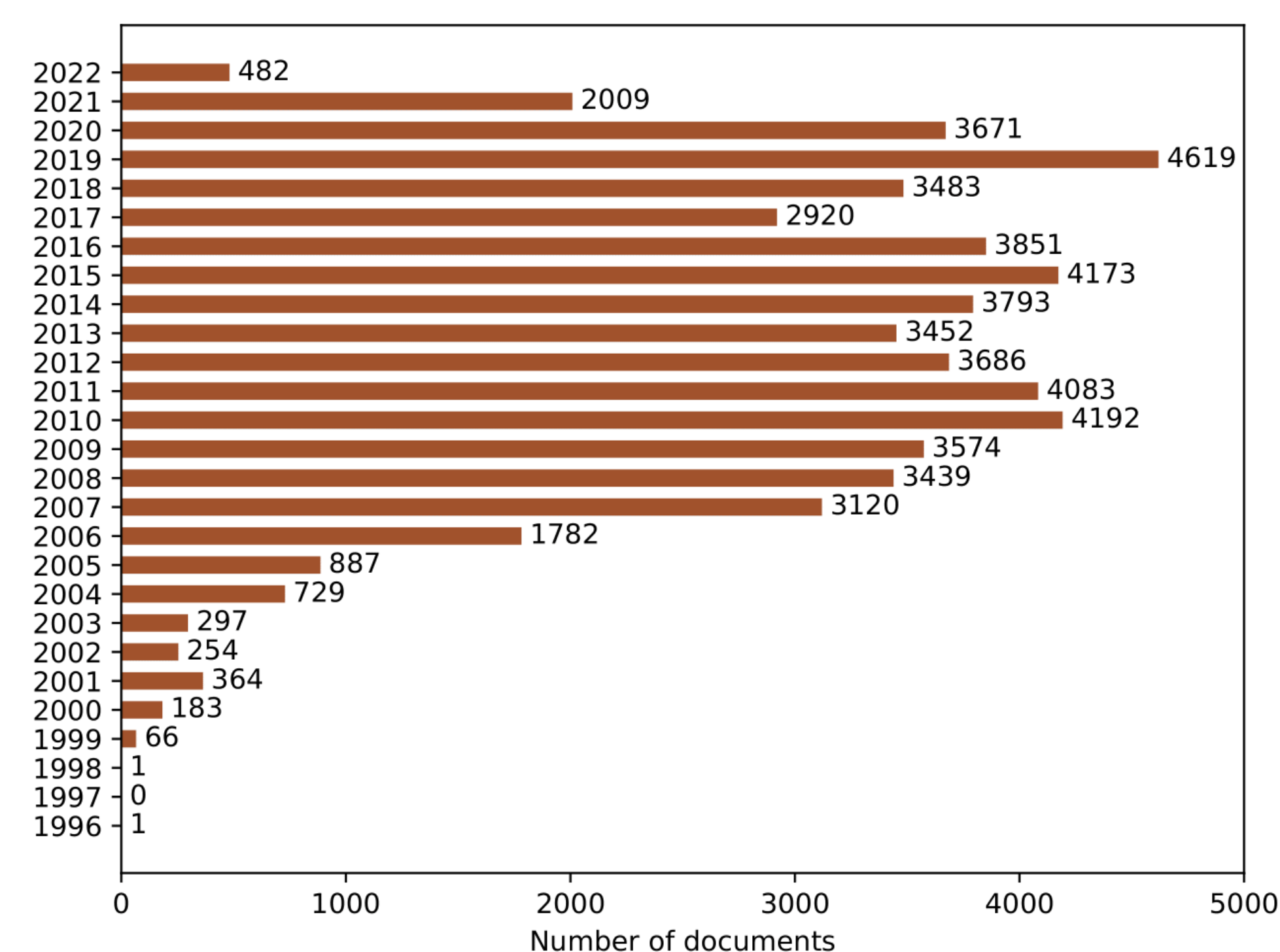
### CONCLUSION

[23] The appeal **is allowed.** I substitute my own decision that **the Appellant is a Convention refugee.**

(signed) **JUDGE**  
**DATE**  
December 17, 2020  
**Date**

Example of the conclusion of a case (the last paragraph)

The label DETERMINATION flags the outcome of the case  
We retrieve the name of the JUDGE and the DATE of the hearing from the signature



## Legal Information Extraction - Legal Entities of Interest

DATE OF THE HEARING	CLAIMANT EVENT	CLAIMANT INFO
DATE	LOCATION	ADJECTIVE LOCATION
DATE DECISION	PROCEDURE	CREDIBILITY
PUBLIC PRIVATE HEARING	DOC EVIDENCE	EXPLANATION
IN CHAMBER VIRTUAL	LEGAL CITATIONS	CASE PRECEDENT
JUDGE NAME	COUNTRY REPORTS	LEGAL GROUND
TRIBUNAL		

All extracted entities, identified with refugee lawyers

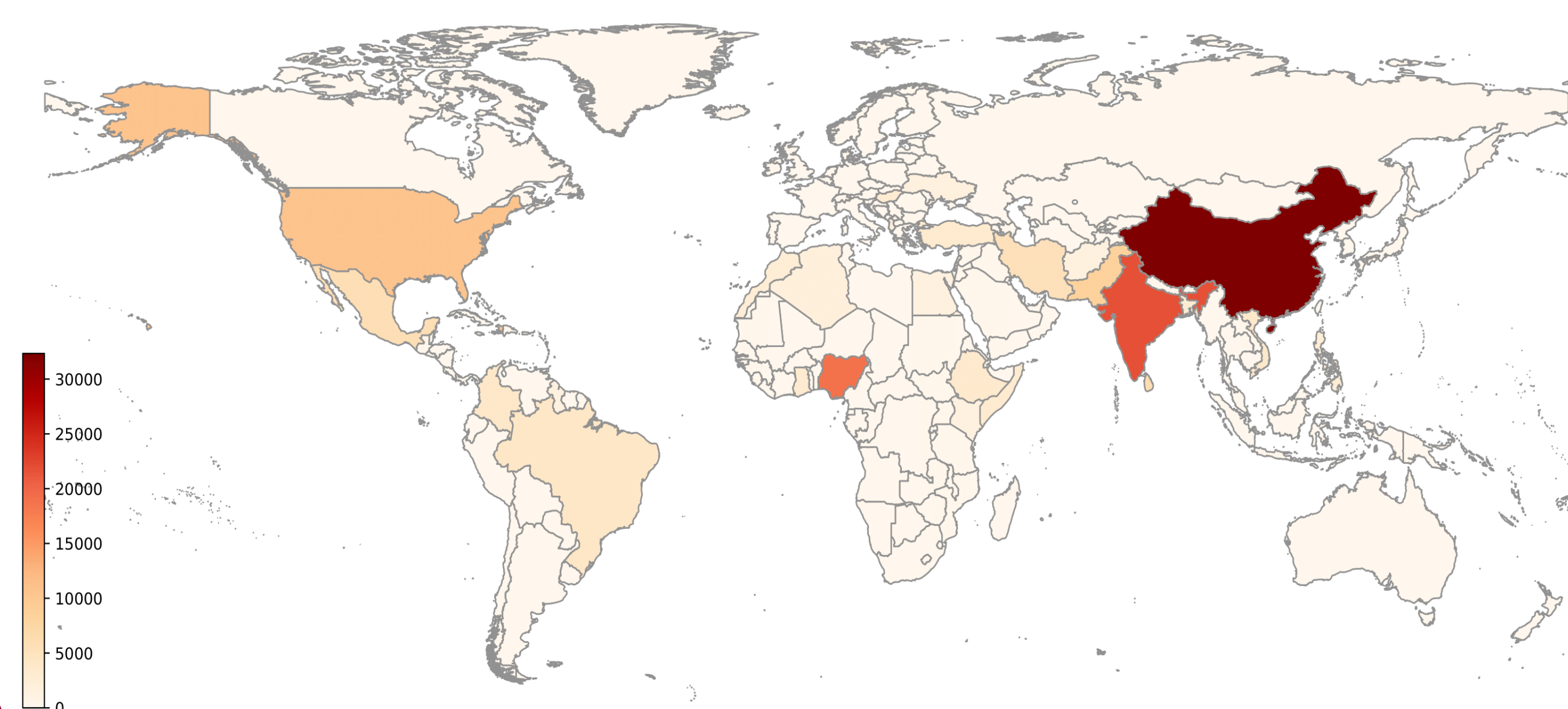
Generic Entities  
Legal Entities  
Refugee Law Entities

Evaluation for the LegalBERT-based fine-tuned entity extraction model

### Inference results

Entity Type	F1	EXPLANATION	
CLAIMANT EVENT	65.54	LOCATION	69.94
CLAIMANT INFO	88.64	LEGAL CITATIONS	55.65
CREDIBILITY	78.72	CASE PRECEDENT	66.67
DATE	89.32	ADJECTIVE LOCATION	87.80
DETERMINATION	70.00	PROCEDURE	67.46
DOC EVIDENCE	84.88	LEGAL GROUND	50
		ORG	89.90

### Entity spotlight 1: Extracted Countries



Occurrences of countries in AsyLex

### Entity spotlight 2: Claimant Information

Flagged Phrases	# of Occurrences
MALE	5,493
MINOR APPELLANT	1,655
FEMALE APPELLANT	2,329
WOMAN	2,844
HOMOSEXUAL	833
CITIZEN OF CHINA	917

What can we learn about the claimants?

Gender  
Age  
Citizenship  
Occupation

### Entity spotlight 3: Extracted Hearing Locations

TORONTO	55.31%	OTTAWA	5.09%
MONTREAL	21.54%	CALGARY	2.57%
VANCOUVER	13.57%	EDMONTON	1.18%

Scan here for HuggingFace dataset

