



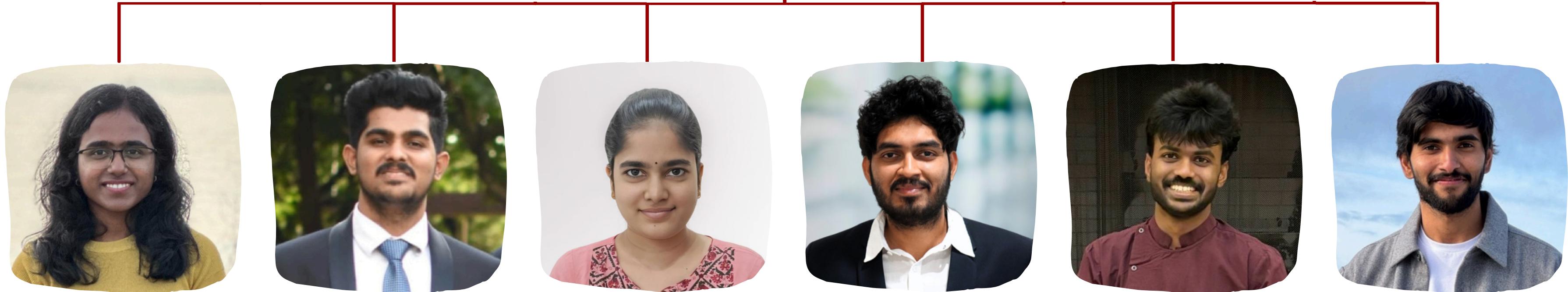
[40] ClearCase : Legal Text Classification with Explainable AI

CSCI 566 DEEP LEARNING AND ITS APPLICATIONS



Meet Our Team

Workflow (Design) + Literature Review + References



INDIRA

MONA

BHAVYA

MOHAN

ARUN

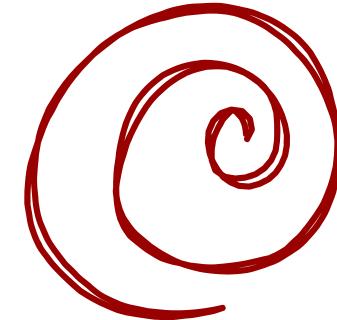
ASHWIN

CARC LOCAL

EDA, Data Cleaning
&
Truncation + Chunking

Hyperparameter Tuning
&
Training (Fine-Tuning)

Testing, Results
&
Explainability



Why is the problem interesting?

ClearCase aims to revolutionize legal document management by automating the classification of complex case texts leveraging advanced transformer models.

This project aims to support legal professionals in efficiently organizing case law while ensuring transparency and trust in AI-driven decisions.



Automating Legal Document Classification

Explainable AI for Transparent Decisions



TRADITIONAL PROCESS

Manually analyzing case text and referenced precedents to estimate current case outcome.

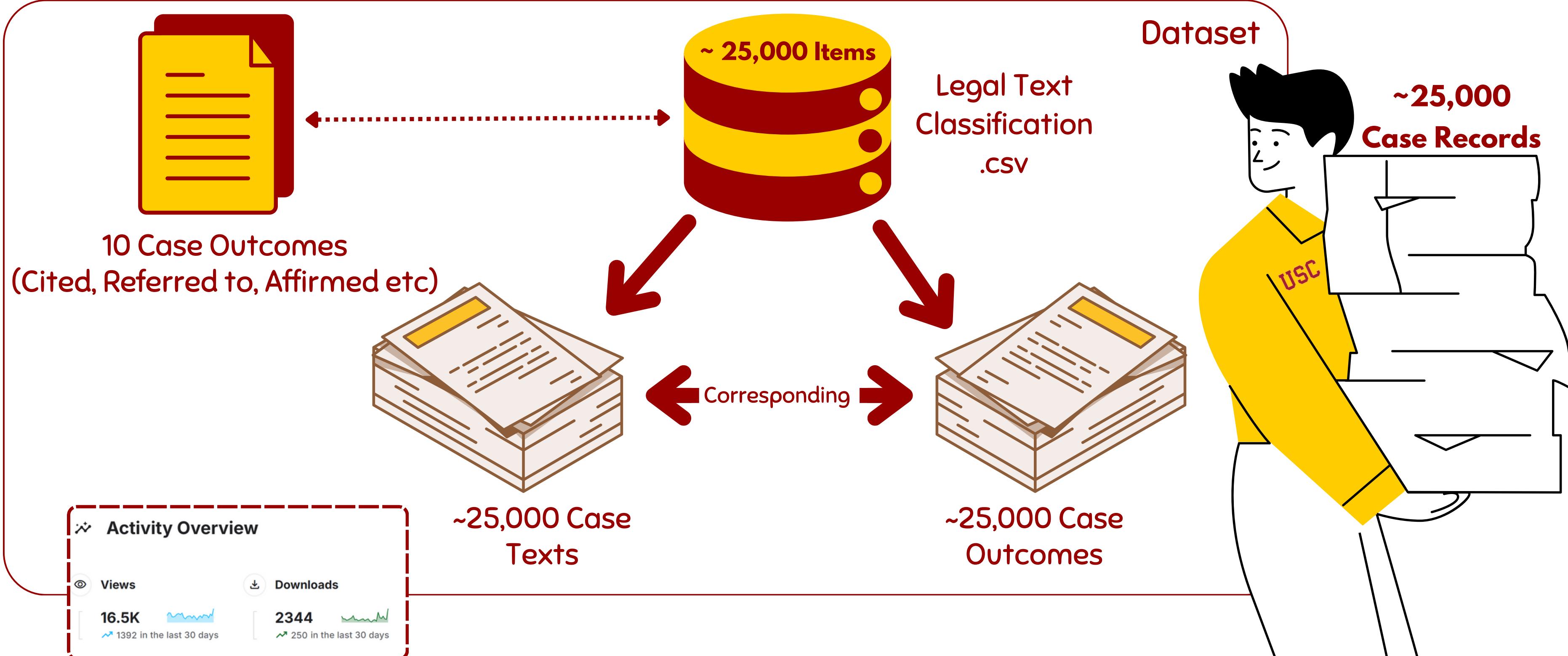


PROPOSED PROCESS

AI-driven classification of legal texts to predict outcomes with explainability.

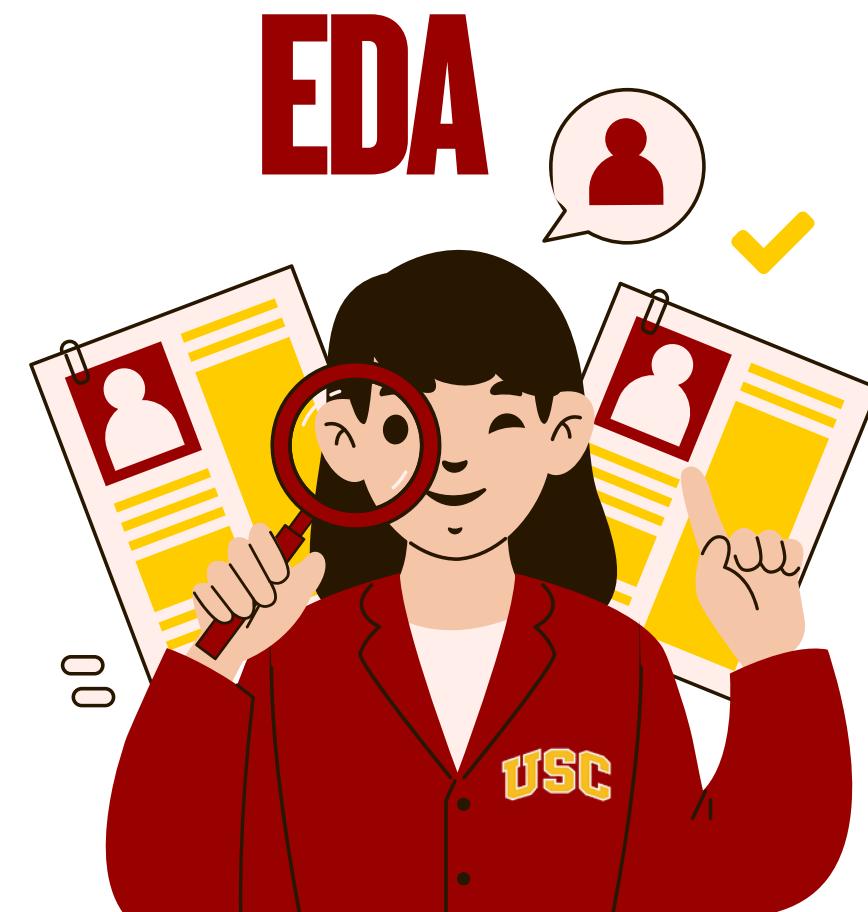
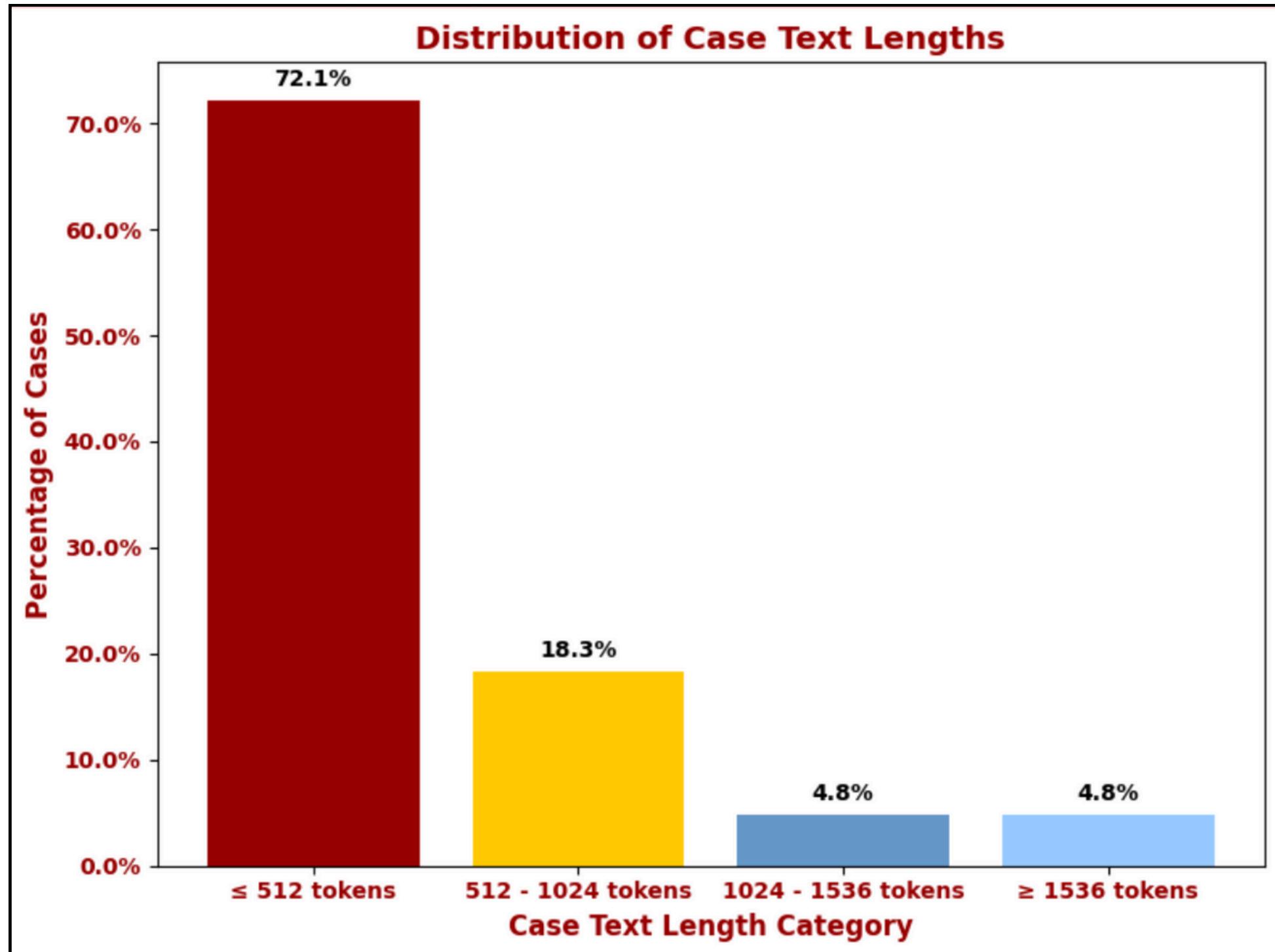


DATASET DESCRIPTION



EXPLORATORY DATA ANALYSIS

| <code>A case_id</code> | <code>A case_outcome</code> | <code>A case_title</code> | <code>A case_text</code> |
|-------------------------------|--------------------------------------|---------------------------|-------------------------------|
| Case ID | Case Outcome | Case title | Case Text |
| 24985 unique values | cited referred to Other (8382) | 49% 18% 34% | 18581 unique values |
| | | | 17921 unique values |



Modeling Choices

01

roberta-large-mnli
- 0 shot (Baseline)

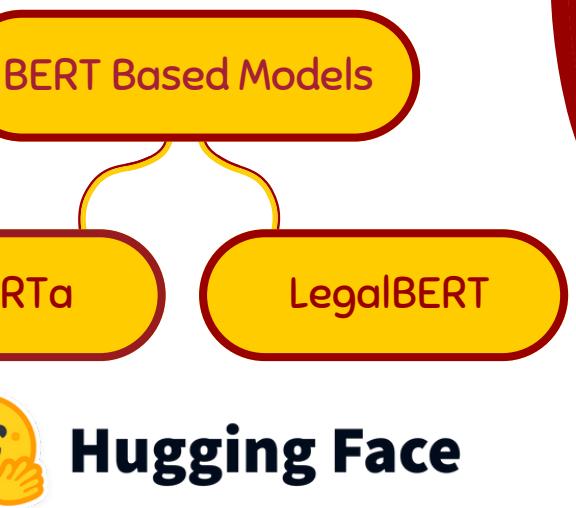
02

roberta-large-mnli

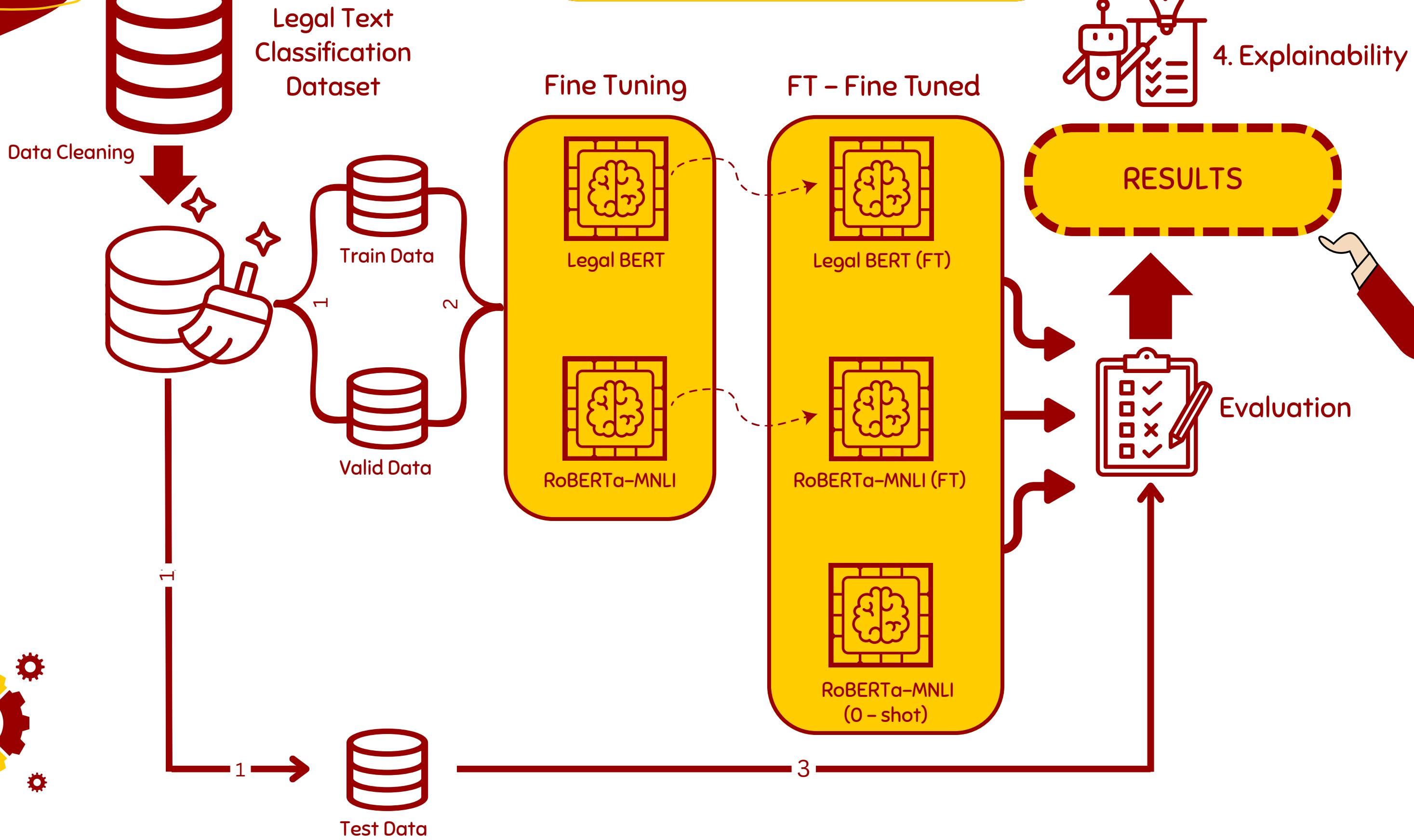
03

legal-bert-base-uncased

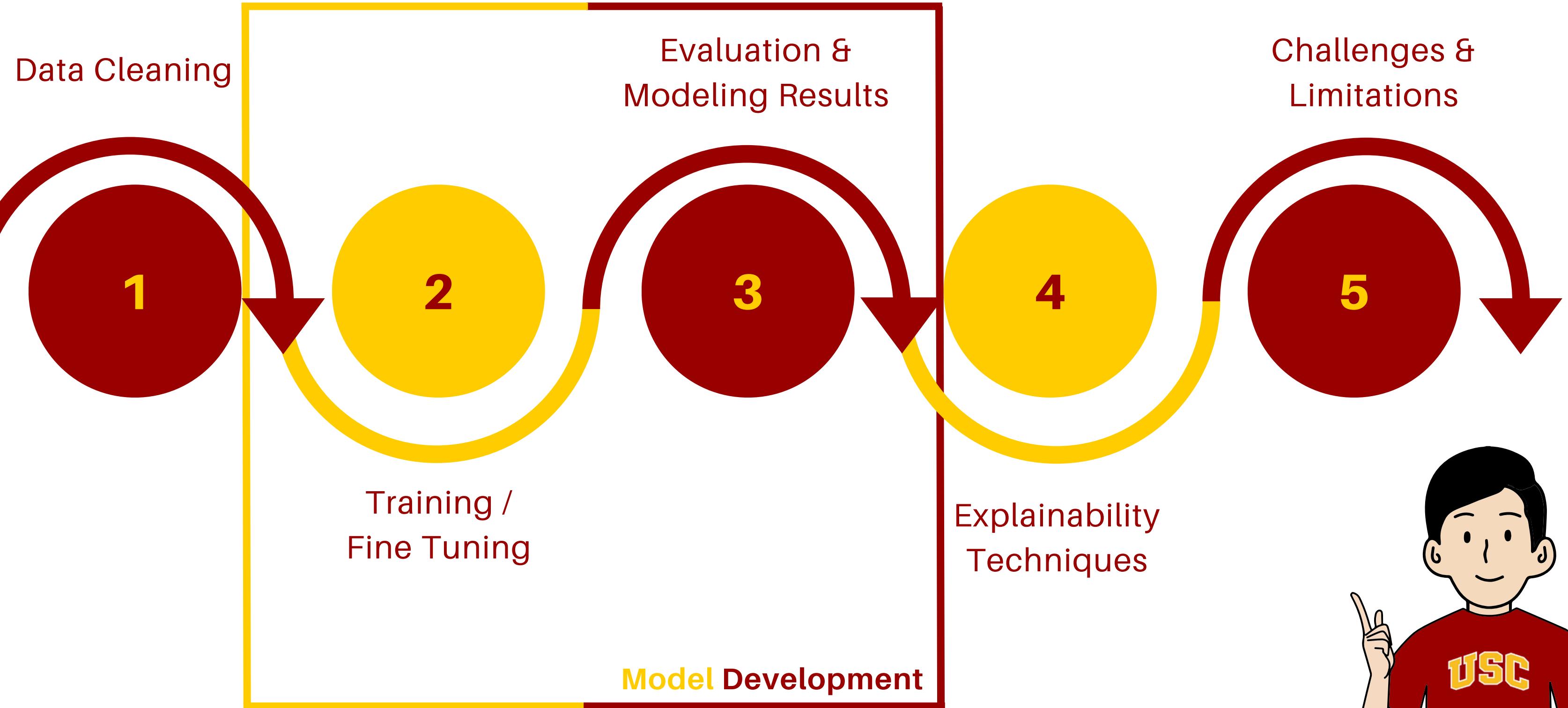
Why BERT?



ClearCase Workflow



PROCESS OVERVIEW



DATA CLEANING

Filter out records with null values in key fields

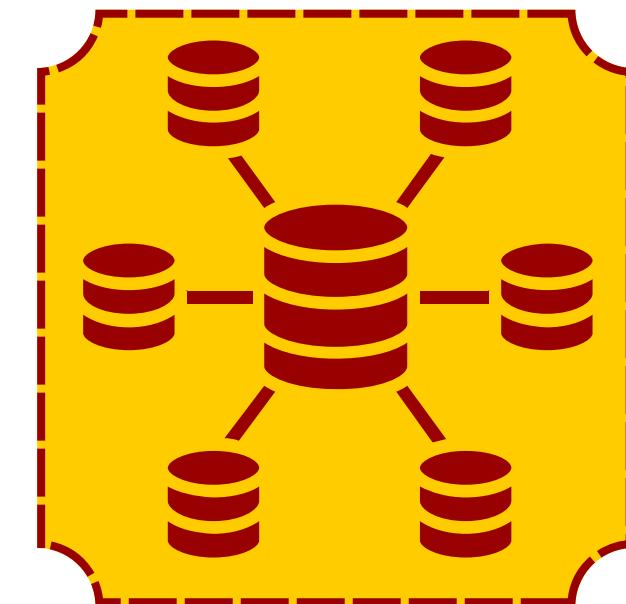
Drop duplicates where both case_text and case_outcome are identical

TRUNCATION

Head Method
(Consider First 512 Tokens)

Head - Tail Method
(Consider First 256 Tokens + Last 256 Tokens)

SOLUTION: CHUNKING



TRAINING / FINE-TUNING

01 BATCH SIZE : 8, 16, 32

02 EPOCHS: 4, 6

03 OPTIMIZER: ADAM

04 WEIGHT DECAY: 0.01

05 LEARNING RATE: $2e-5$



Evaluation

Here are few evaluation metrics we computed to measure the model's performance

01

Accuracy

$$\text{Accuracy} = \left(\frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \right) \times 100$$

03

Recall

$$\text{Recall} = \left(\frac{\text{Number of True Positives}}{\text{Total Number of Actual Positives}} \right) \times 100$$

02

Precision

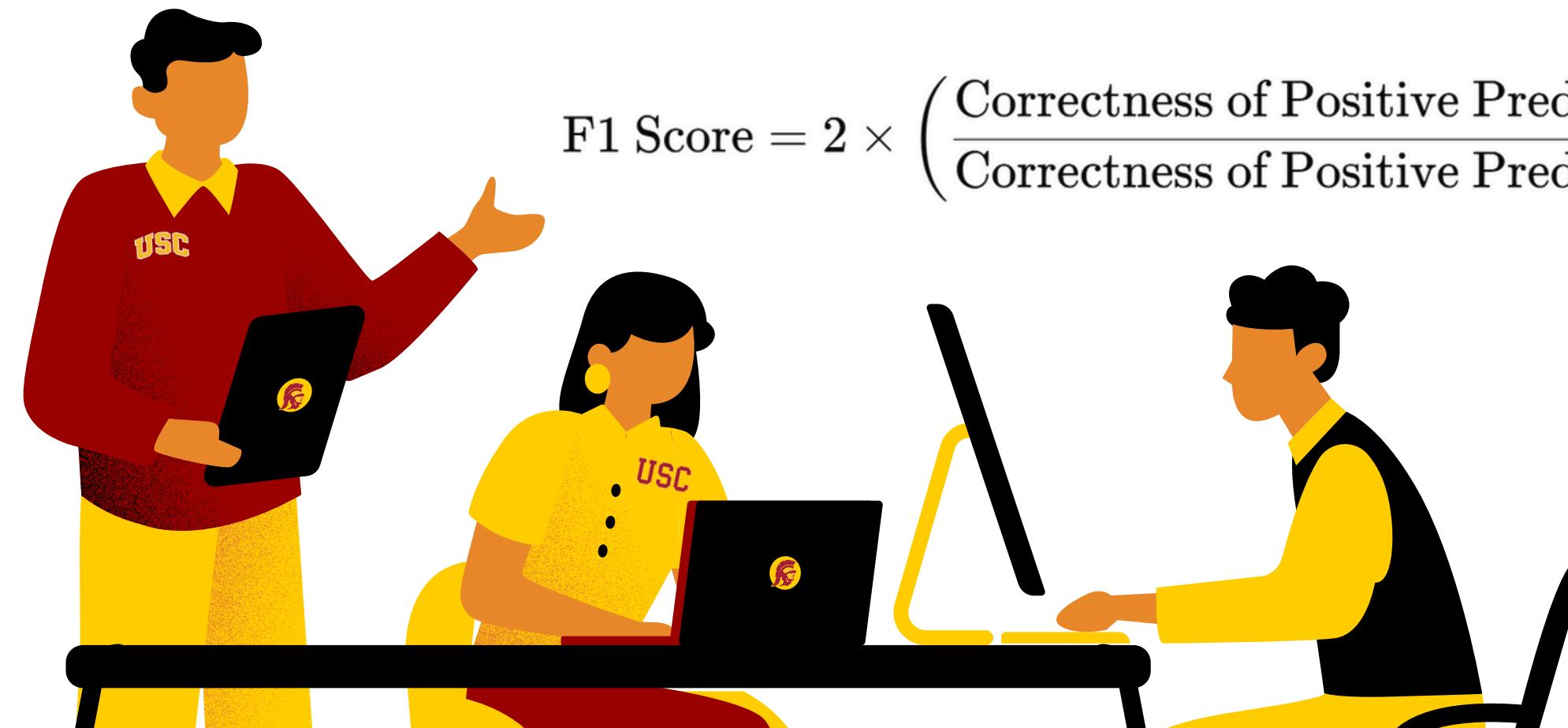
$$\text{Precision} = \left(\frac{\text{Number of True Positives}}{\text{Total Number of Positive Predictions}} \right) \times 100$$

04

F1-Score

$$\text{F1 Score} = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right)$$

$$\text{F1 Score} = 2 \times \left(\frac{\text{Correctness of Positive Predictions} \times \text{Coverage of Actual Positives}}{\text{Correctness of Positive Predictions} + \text{Coverage of Actual Positives}} \right)$$

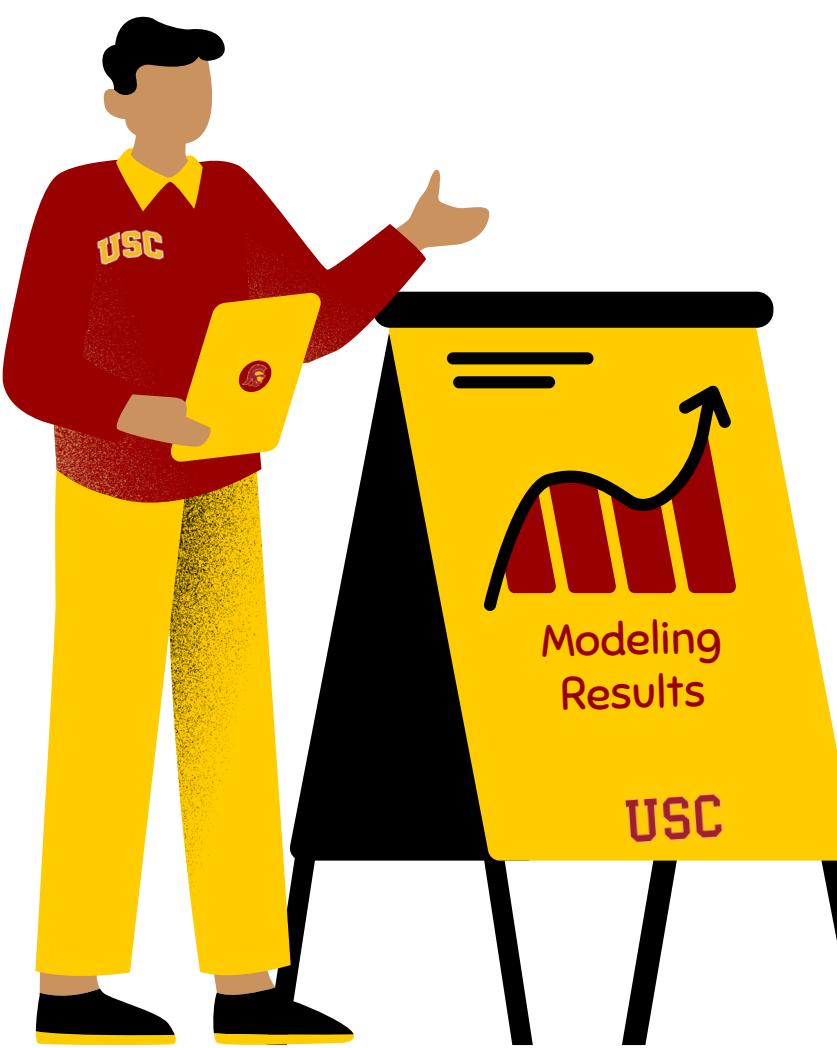


Modeling Results

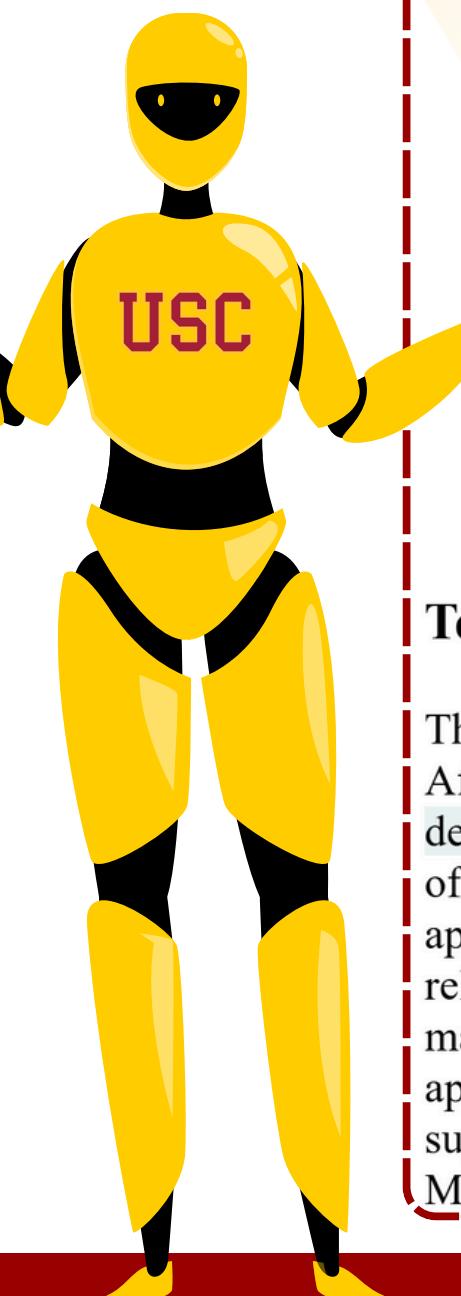
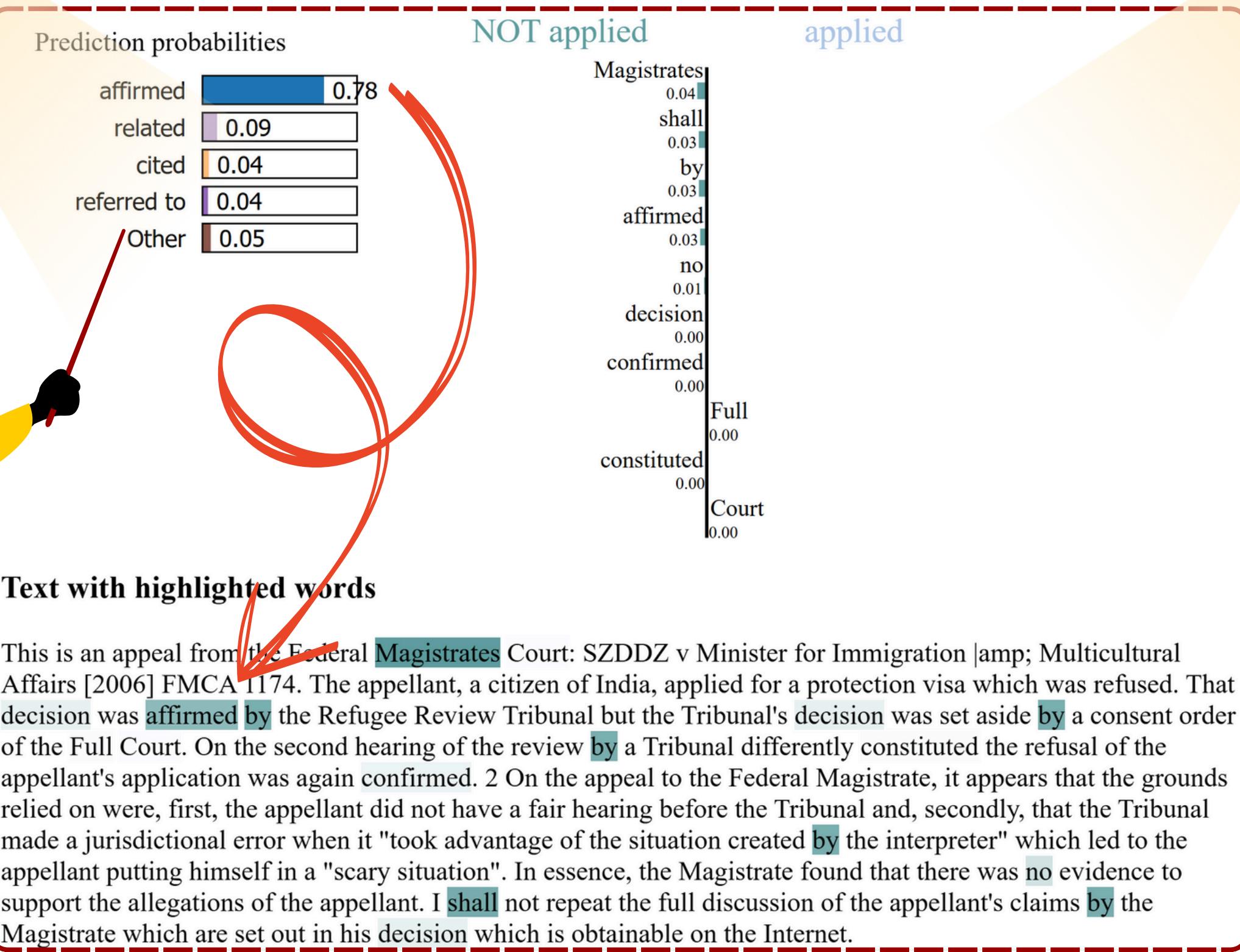
We fine-tuned with different parameters for these models and followed different chunking strategies. Only the best results for each model (checkpoint) is shown in the below table

FT - Finetuned
H - Head Method
HT - Head Tail Method
C - Chunking

| MODEL | ACCURACY | PRECISION | RECALL | F1-SCORE |
|--|----------|-----------|--------|----------|
| roberta-large-mnli - 0 shot (Baseline) | 29.15% | 16.84% | 14.60% | 11.95% |
| legal-bert-base-uncased (FT) - H | 53.32% | 49.79% | 52.25% | 50.02% |
| roberta-large-mnli (FT) - H | 53.68% | 51.35% | 52.12% | 51.74% |
| legal-bert-base-uncased (FT) - HT | 54.79% | 52.94% | 54.74% | 52.97% |
| roberta-large-mnli (FT) - HT | 54.83% | 50.44% | 54.78% | 51.67% |
| legal-bert-base-uncased (FT) - C | 62.27% | 64.26% | 62.19% | 62.57% |



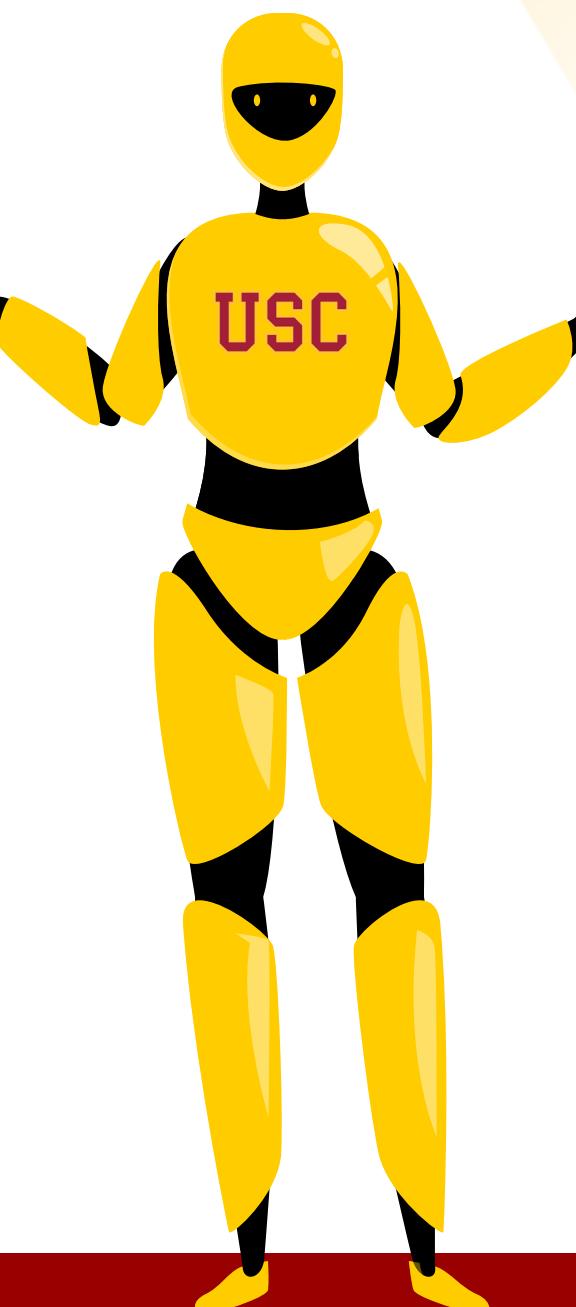
Explainability (LIME)



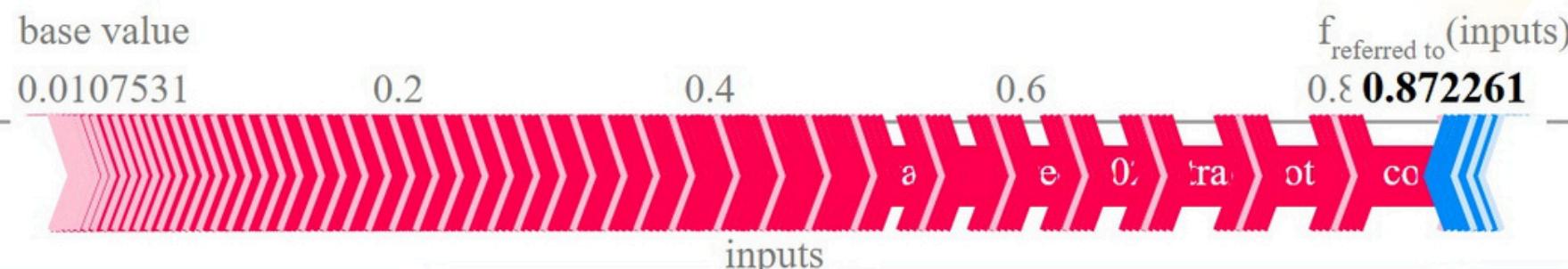
SCAN ME



Explainability (SHAP)



affirmed applied approved cited considered discussed distinguished followed
referred to related



in the circumstances , it is unnecessary to address a further argument put by graphix labels and mr atta against implying such a term and that was that such a term would amount to an unreasonable restraint of trade i was referred to nordenfelt v maxim nordenfelt guns and ammunition co ltd [1894] ac 535 ; amoco australia pty ltd v rocca bros motor engineering co pty ltd [1973] hca 40 ; (1973) 133 clr 288 . of course , there have been a number of cases dealing with the common law doctrine of restraint of trade since those two cases (see , for example , the discussion in seddon and ellinghaus , cheshire & ; fifoot ' s law of contract (8 th australian ed , 2002) pages 860 - 863) .



SCAN ME





REFERENCES

Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Ion Androutsopoulos, and Nikolaos Aletras. 2020. **LEGAL-BERT: The muppets straight out of law school.** *arXiv preprint arXiv:2010.02559*.

Sarthak Jain and Byron C. Wallace. 2019. **Attention is not explanation.** In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*.

Nut Limsopatham. 2021. **Effectively leveraging BERT for legal document classification.** In *Proceedings of the Natural Legal Language Processing Workshop (NLLP)*.

Scott M. Lundberg and Su-In Lee. 2017. **A unified approach to interpreting model predictions.** In *Advances in Neural Information Processing Systems (NeurIPS)*.

Dimitris Mamakas, Ilias Chalkidis, Ion Androutsopoulos, and Nikolaos Aletras. 2022. **Processing long legal documents with pre-trained transformers: Modding legalbert and longformer.** In *Proceedings of the Natural Legal Language Processing Workshop (NLLP)*.

CARC USAGE



01 A100

(All Sessions Included)

02 Number of Hours: 52

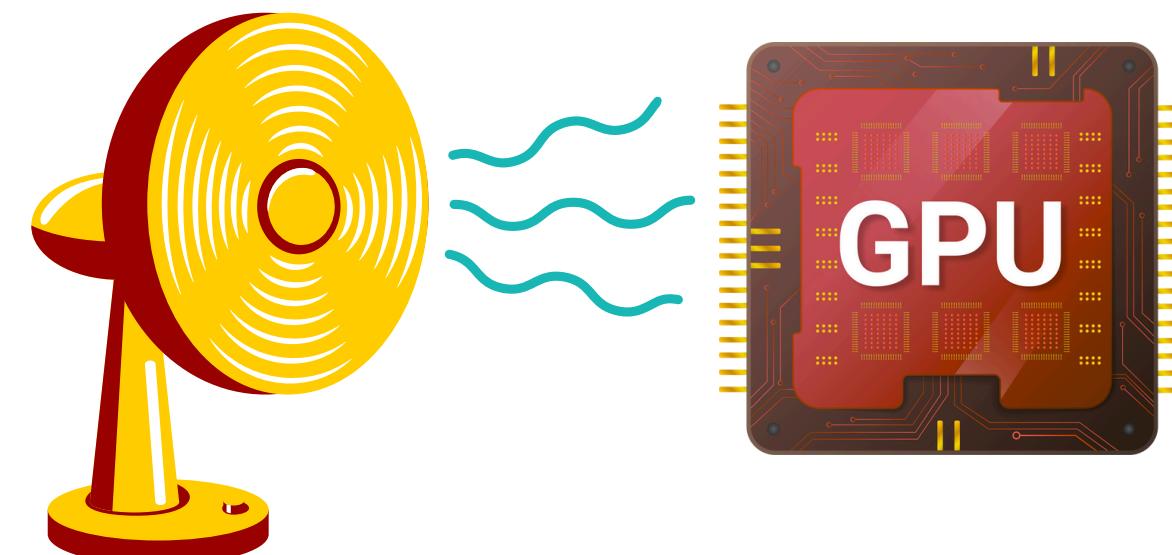
03 Number of GPUs: 2

LOCAL SPEC

01 NVIDIA GeForce RTX 4060

(All Sessions Included)

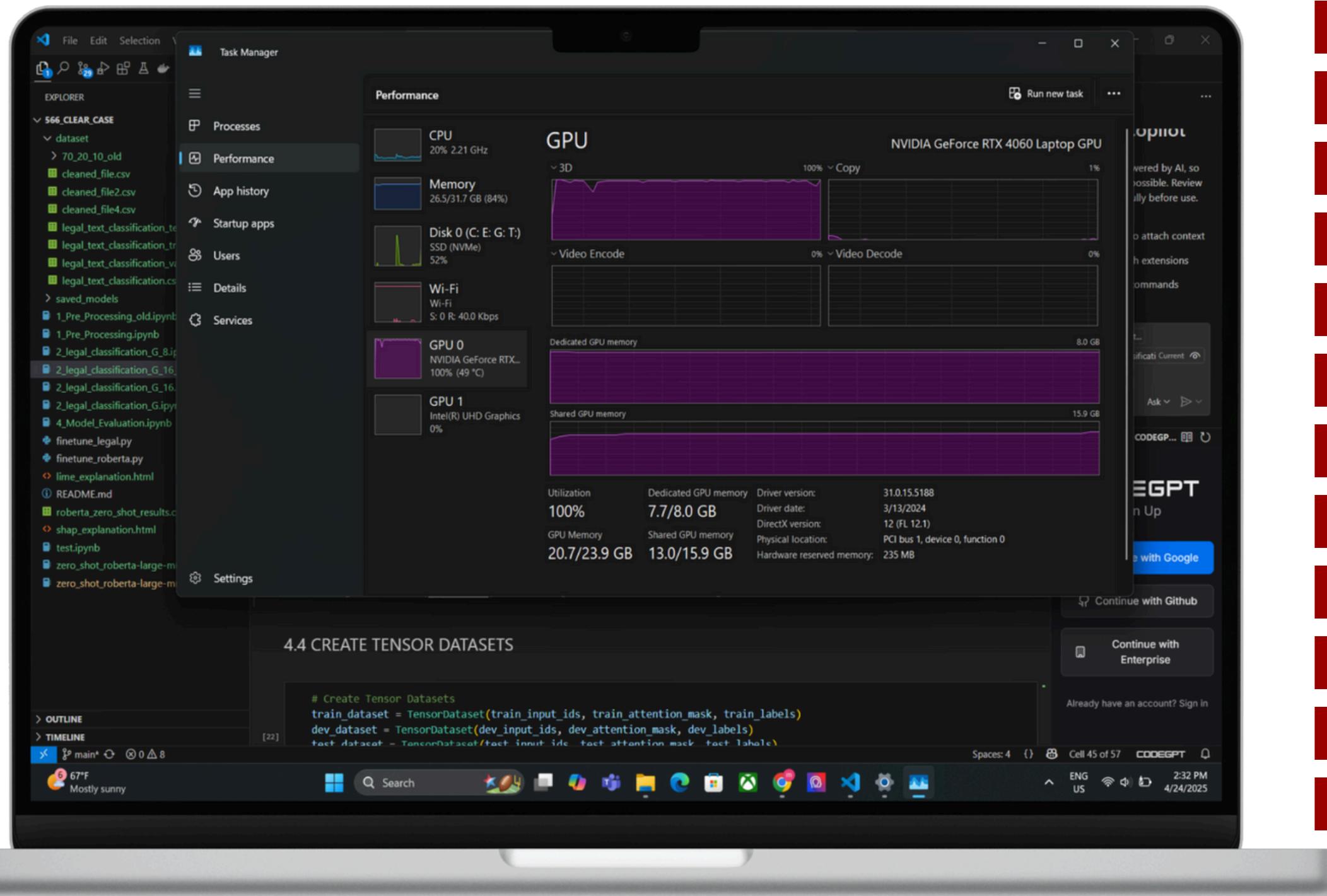
02 Number of Hours: 64



Limitations

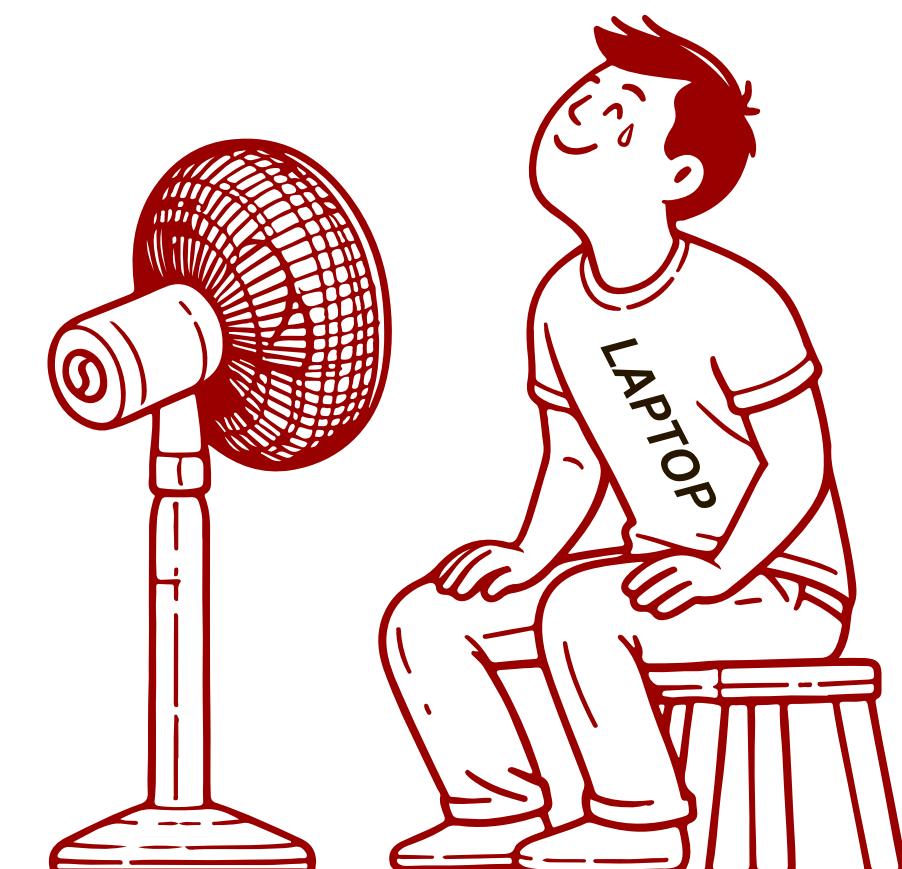
We are currently facing limitations with the Processing and Computational constraints. Details are as follows :

01 Computation



02 Context Window Size

| Model | Window |
|-------------------------|------------|
| roberta-large-mnli | 512 Tokens |
| legal-bert-base-uncased | 512 Tokens |



Future Scope

Try to leverage Class-Weighted Loss Functions and Data Augmentation

Explore the possibility to increase the context window using models like Longformer.

Try more Explainability Methods & Evaluation Metrics

Longformer



Thank You!

Special Thanks to...
Haoyan Xu

