



UNIVERSITEIT VAN AMSTERDAM

No Labels? No Problem!

Experiments with active learning strategies
for multi-class classification
in imbalanced low-resource settings

Emiel Steegh, Giovanni Sileno
University of Amsterdam

19th International Conference on Artificial Intelligence and Law (ICAIL)
18th June 2023, Braga, Portugal

Problem: document deluge in administration

With the digitization of services, the volume of documents is continuously increasing



Problem: document deluge in administration

With the digitization of services, the volume of documents is continuously increasing

For the public sector, documents often need to become publicly available and stay as such for a minimal amount of time



Problem: document deluge in administration

With the digitization of services, the volume of documents is continuously increasing

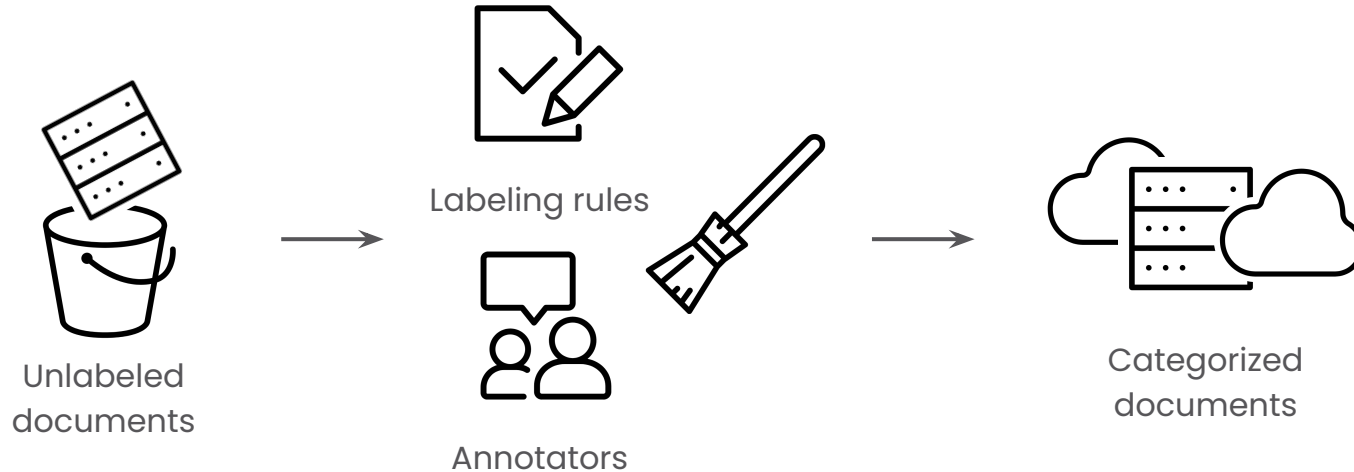
For the public sector, documents often need to become publicly available and stay as such for a minimal amount of time

Similar problems of exponential accumulation occur also in the private sector



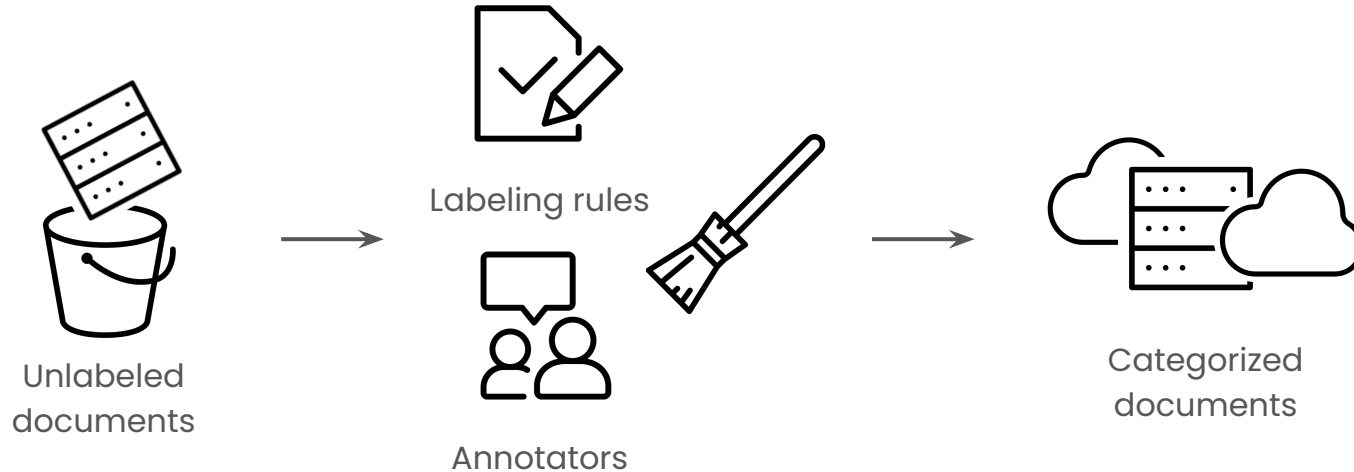
Problem statement

Organizations need to regain control of large volumes of documents.



Problem statement

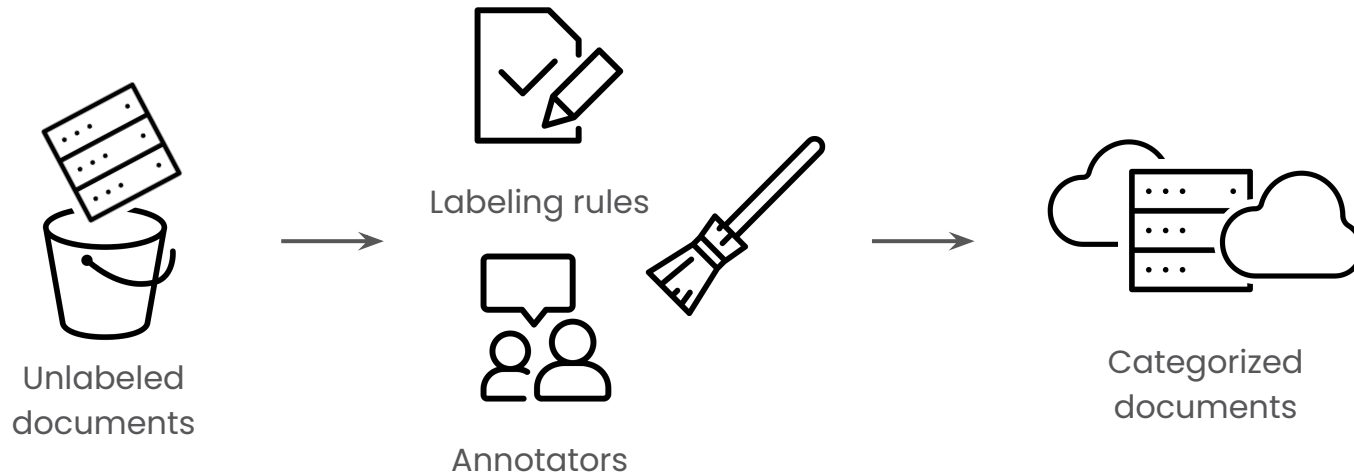
Organizations need to regain control of large volumes of documents.



**This is typically
an expensive step!**

Problem statement

Organizations need to regain control of large volumes of documents.



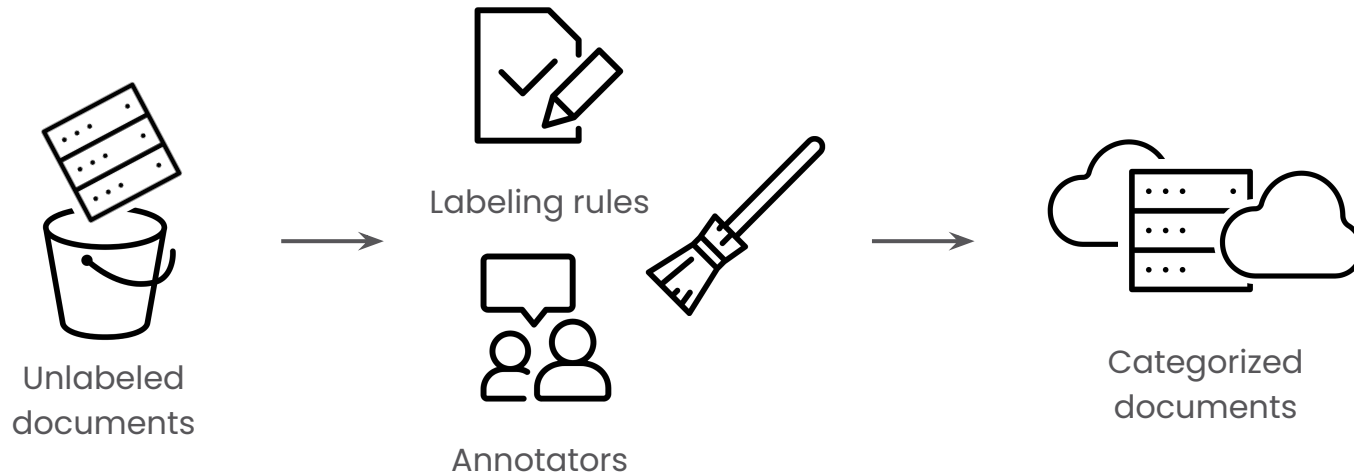
**This is typically
an expensive step!**



Let's apply **active learning**

Problem statement

Organizations need to regain control of large volumes of documents.



**This is typically
an expensive step!**



Let's apply **active learning**
(does it work?)

Problem statement

Organizations need to regain control of large volumes of documents.



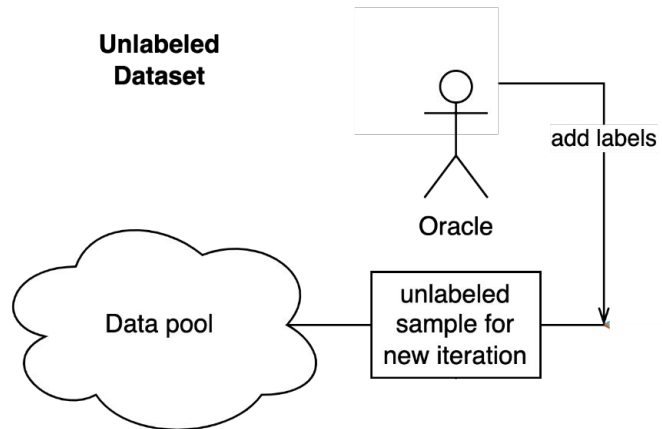
Background

- Text classification
- Active Learning
- Active Learning in NLP

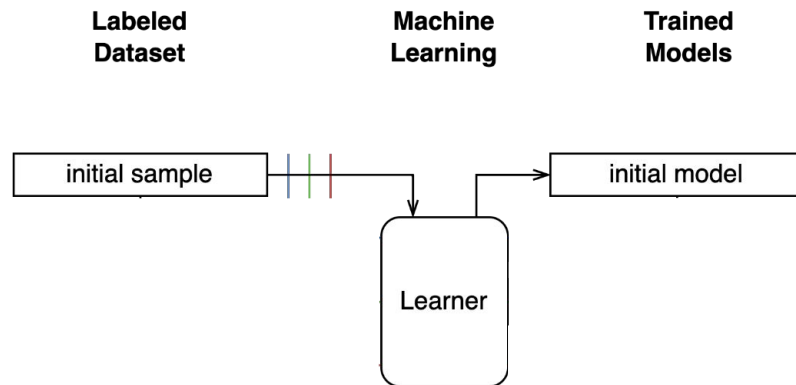
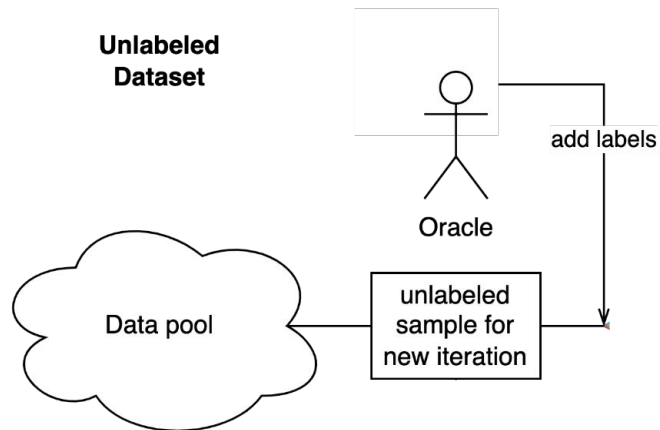
Domain Adaptation for Text Classification

- *We have a text classification task.*
- Pre-Trained Language Models (PTLM).
- Transfer language understanding to the next (downstream) task.
- *But the domain is low resource.*
- Attempt Task-Adaptive Pre-Training.

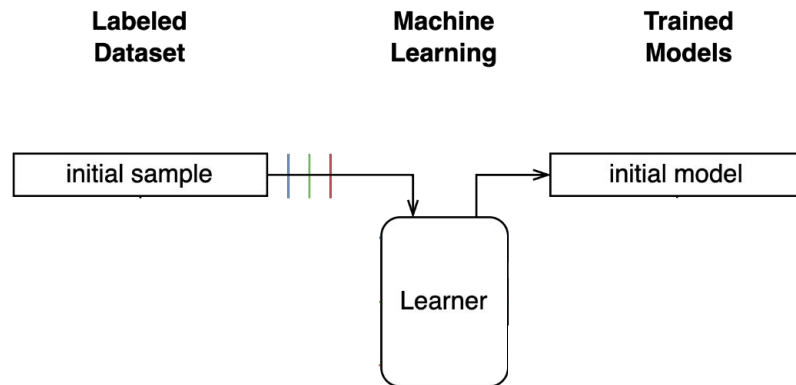
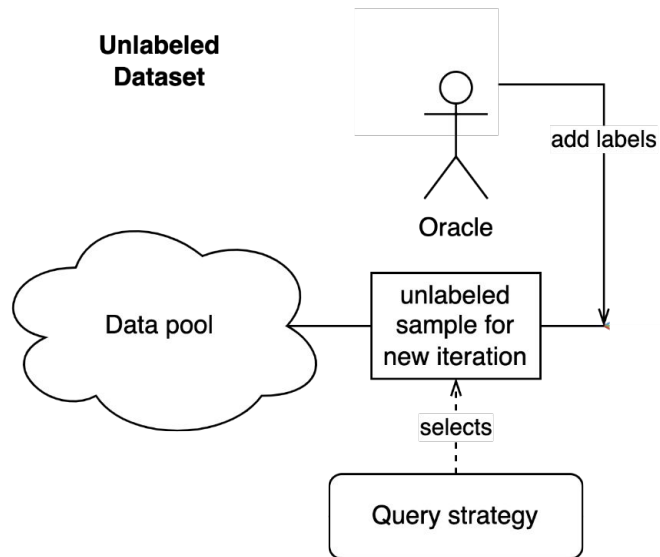
Active Learning?



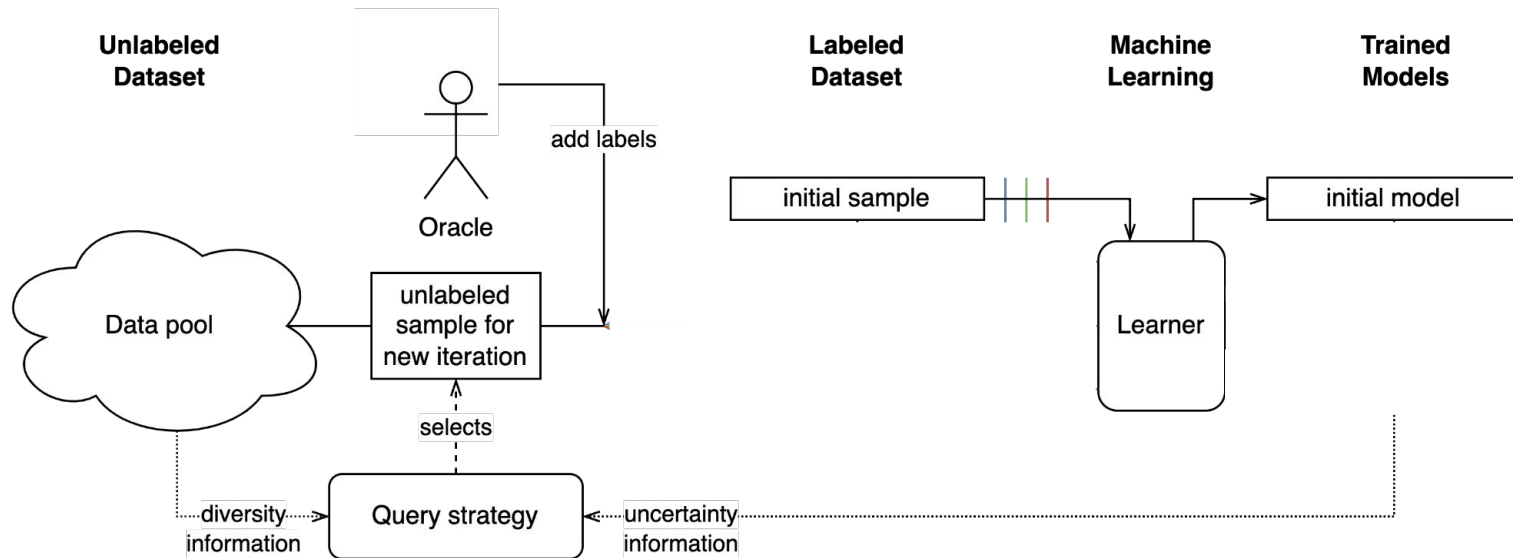
Active Learning?



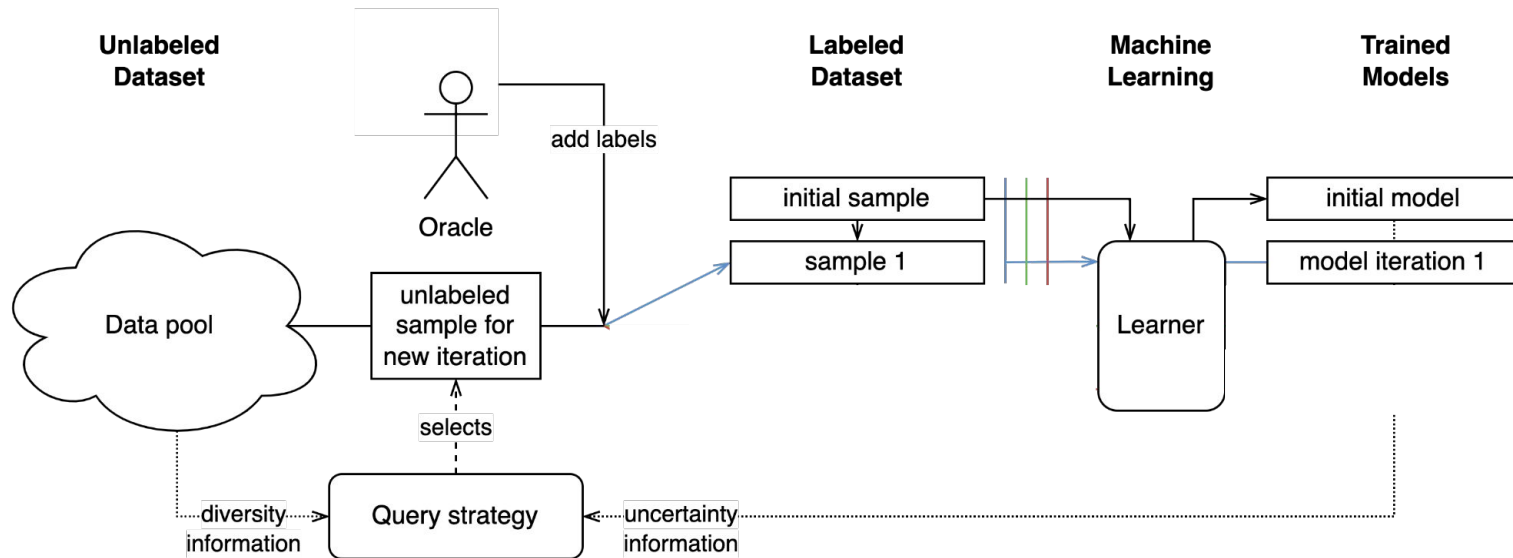
Active Learning?



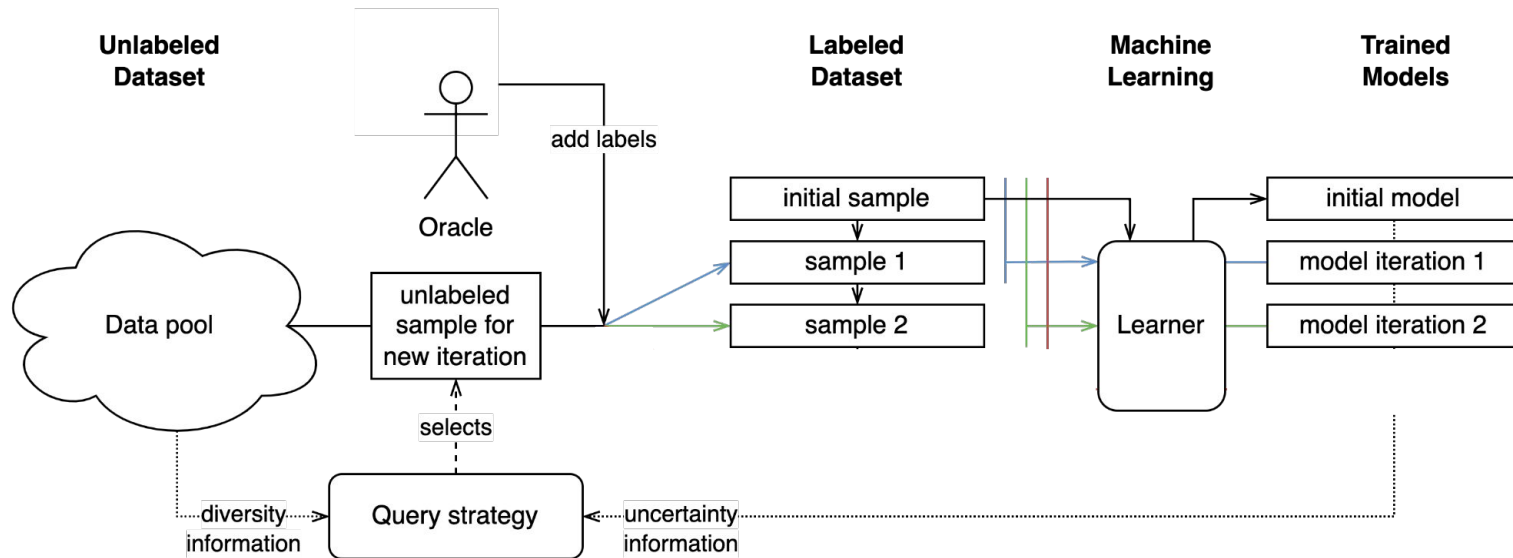
Active Learning?



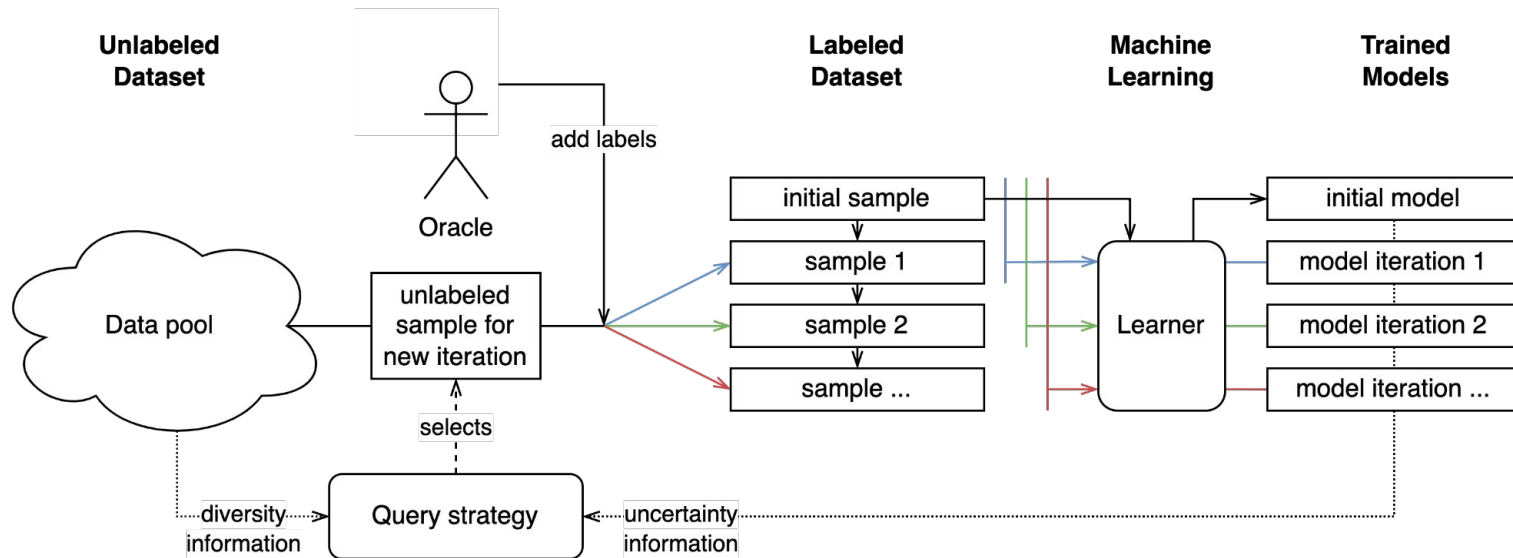
Active Learning?



Active Learning?

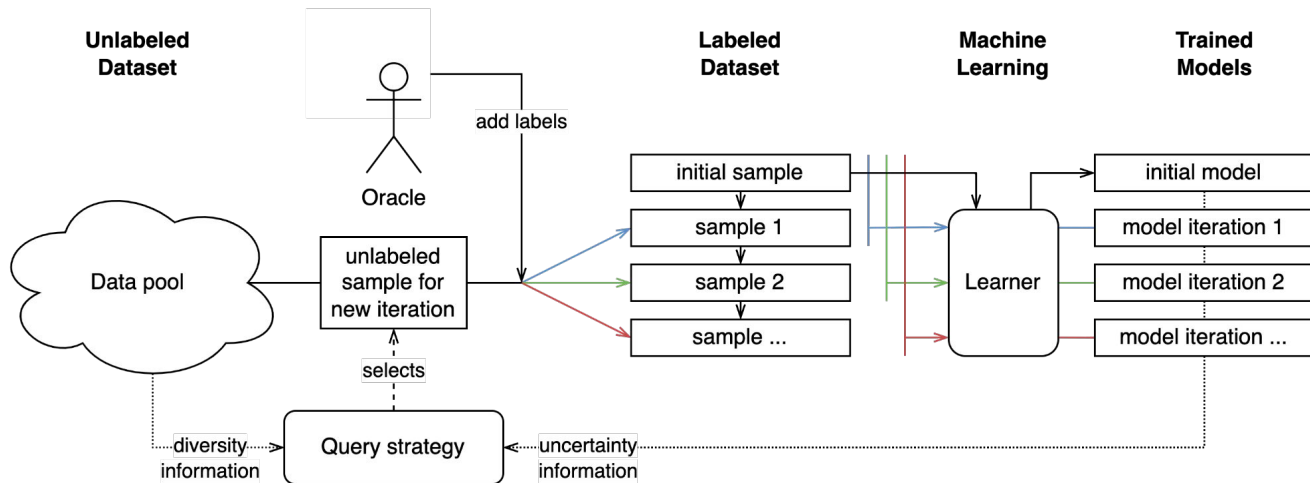


Active Learning?



Active Learning in NLP

- Text diversity
- Language model uncertainty
- Contrastive Active Learning as hybrid model

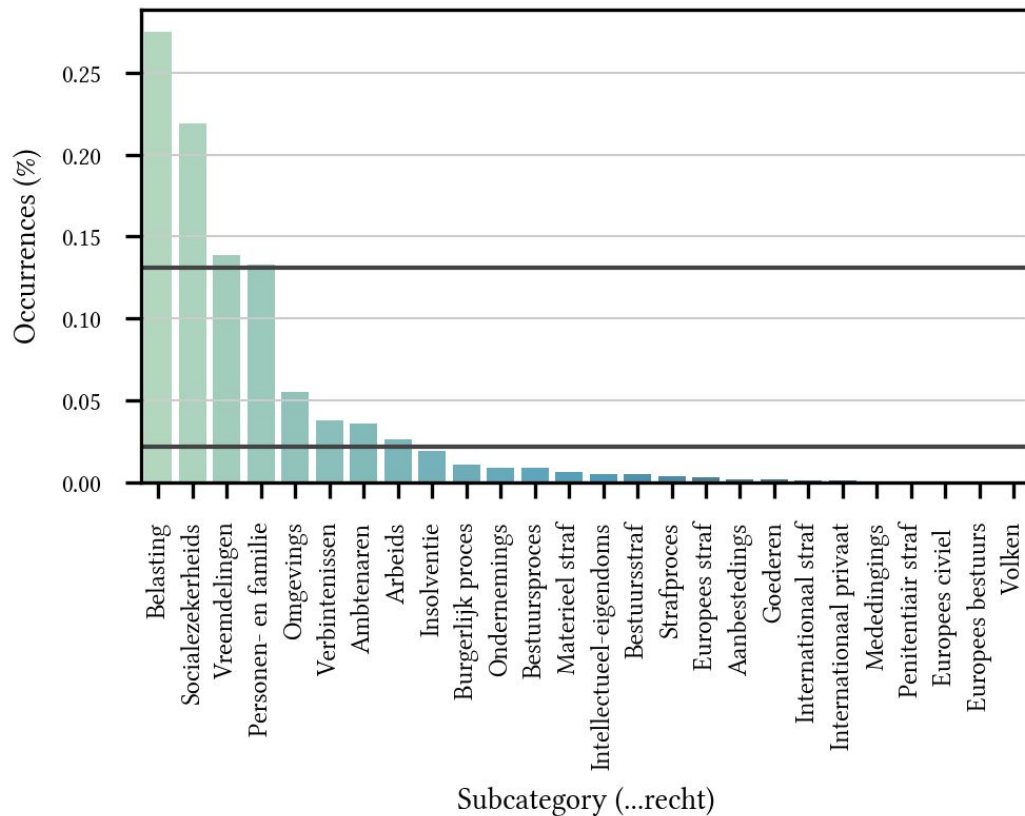


Method

- Select a low-resource dataset (balanced/imbalanced)
- Select relevant design dimensions for the active learning loop
 - Base model with or without fine-tuning
 - Active learning query strategies
- Run the experiments!

Source Dataset

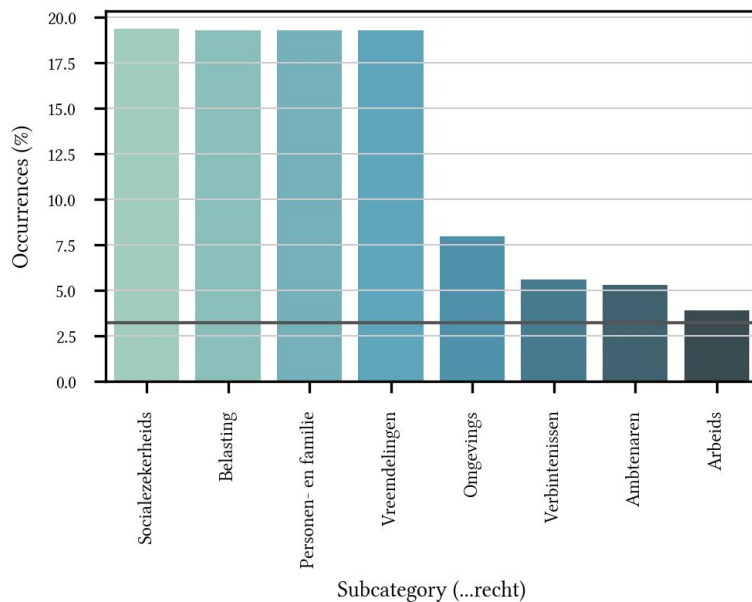
- Open Data - rechtspraak.nl
- Case Law
- Classification of Legal Area
- Processed



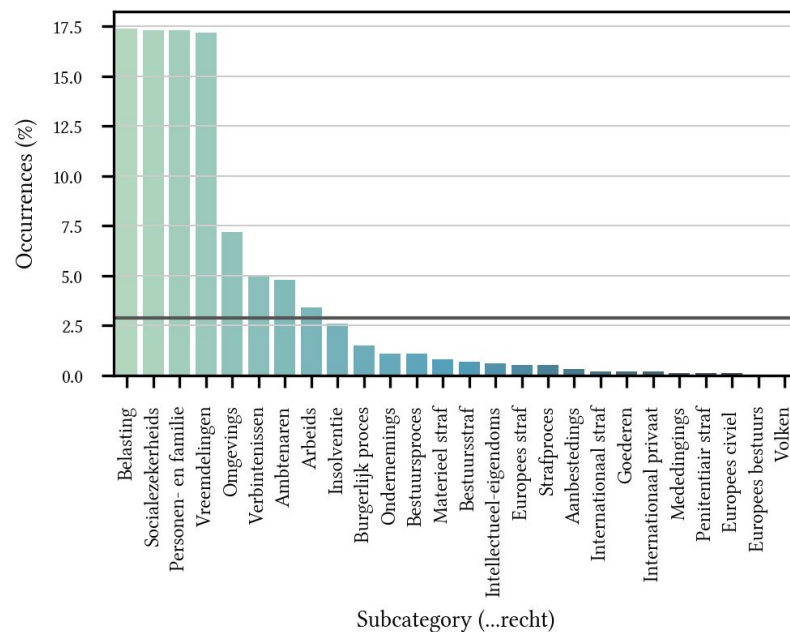
Target Datasets: ORNL8 and ORNL26

NLP data loader on HuggingFace 🙌

```
> datasets.load_dataset('rodekool/ornl8')
```



```
> datasets.load_dataset('rodekool/ornl26')
```



Experimental setup: a handful of options!

Experimental setup: a handful of options!

BERTje as a base model

Experimental setup: a handful of options!

BERTje as a base model

+ Task Adaptive PreTraining

Yes or No

Experimental setup: a handful of options!

BERTje as a base model

- + Task Adaptive PreTraining

 - Yes or No

- + Active Learning Strategy

 - No AL – Random

 - Uncertainty – Entropy

 - Hybrid – Contrastive AL, ...

Experimental setup: a handful of options!

BERTje as a base model

- + Task Adaptive PreTraining

 - Yes or No

- + Active Learning Strategy

 - No AL – Random

 - Uncertainty – Entropy

 - Hybrid – Contrastive AL, ...

- + Sequence Length

 - 128, 256, 512

Experimental setup: a handful of options!

BERTje as a base model

+ Task Adaptive PreTraining

Yes or No

+ Active Learning Strategy

No AL – Random

Uncertainty – Entropy

Hybrid – Contrastive AL, ...

+ Sequence Length

128, 256, 512

Table 1: Table of model variations. Models marked with pt use the task-adapted (pre-trained) version of BERTje. The FS model uses all 50000 labeled documents for training.

ID	TAPT	S	seq.len.
r-128 (<i>Baseline</i>)	-	Random	128
r-256	-	Random	256
pt-r-256	pt	Random	256
e-128	-	Entropy	128
e-256	-	Entropy	256
pt-e-128	pt	Entropy	128
pt-e-256	pt	Entropy	256
pt-e-512	pt	Entropy	512
cal-128	-	CAL	128
cal+256	-	CAL	256
pt-cal+256	pt	CAL	256
FS-256	-	Full Supervision	256

Experimental setup: a handful of options!

BERTje as a base model

+ Task Adaptive PreTraining

Yes or No

+ Active Learning Strategy

No AL – Random

Uncertainty – Entropy

Hybrid – Contrastive AL, ...

+ Sequence Length

128, 256, 512

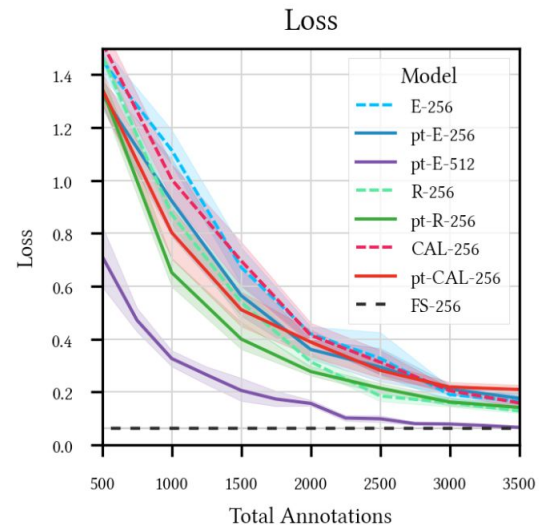
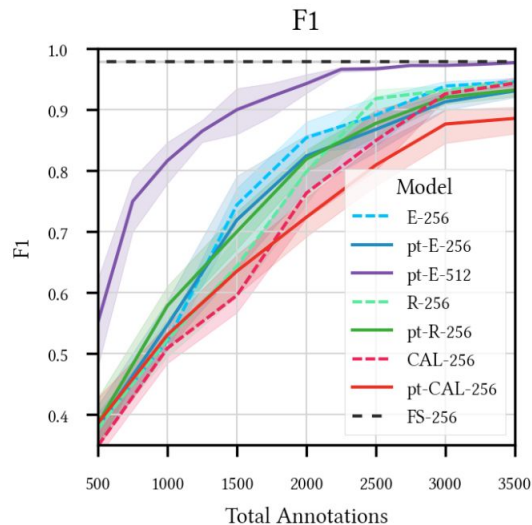
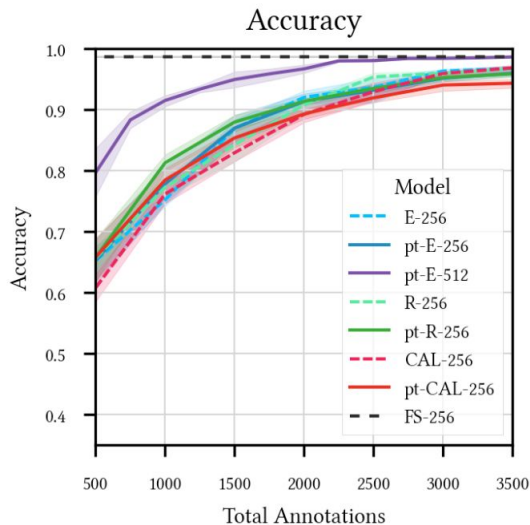
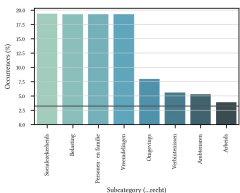
Table 1: Table of model variations. Models marked with pt use the task-adapted (pre-trained) version of BERTje. The FS model uses all 50000 labeled documents for training.

ID	TAPT	S	seq.len.
r-128 (<i>Baseline</i>)	-	Random	128
r-256	-	Random	256
pt-r-256	pt	Random	256
e-128	-	Entropy	128
e-256	-	Entropy	256
pt-e-128	pt	Entropy	128
pt-e-256	pt	Entropy	256
pt-e-512	pt	Entropy	512
cal-128	-	CAL	128
cal+256	-	CAL	256
pt-cal+256	pt	CAL	256
FS-256	-	Full Supervision	256



Results

ORNL8 (BALANCED)

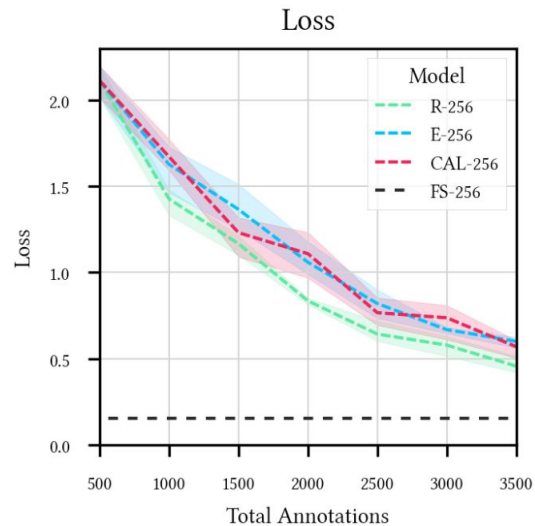
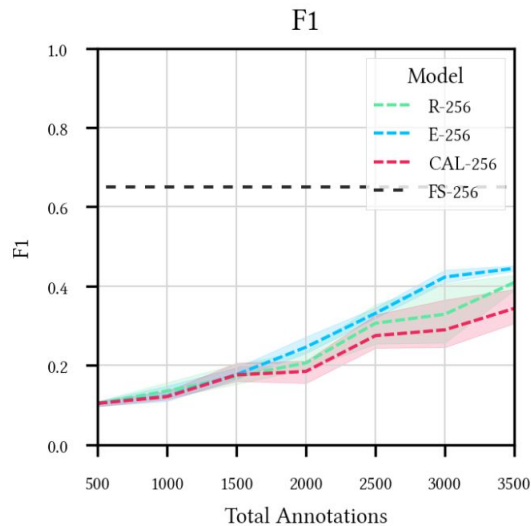
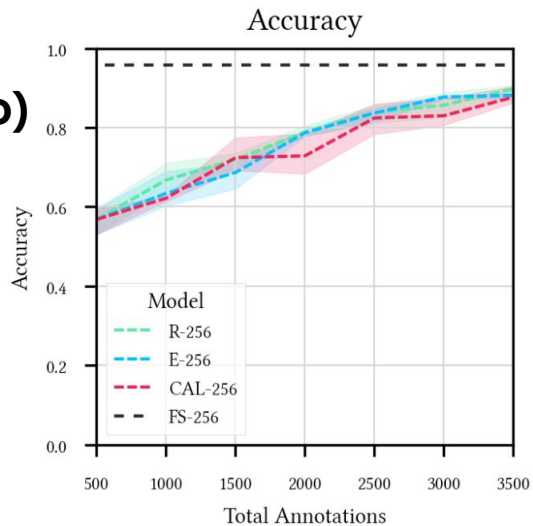
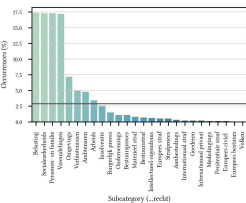


Longer sequences work better

TAPT marginally underperforms (solid lines)

Very similar model performance, no benefit in AL

ORNL26 (IMBALANCED)



Entropy slightly improves over baseline

acc +0.9%, f1 +3.2%, precision +2.3%, recall +3.9%

AL converges faster

How long does it take?

	Run average	Training time	Inference time	Selection time
Random	1h 42m	✓		
Entropy	2h 00m	✓	✓	
Constrative AL	2h 36m	✓	✓	✓

Conclusion

Contributions

- Facilitated a low-resource (Dutch, legal) dataset for experimentation.
- Set initial benchmark scores.
- Entropy based Active Learning improved learning per label in an imbalanced setting.

All of the code is available online!



Contributions

- Facilitated a low-resource (Dutch, legal) dataset for experimentation.
- Set initial benchmark scores.
- Entropy based Active Learning improved learning per label in an imbalanced setting.

Next steps

- Expert evaluation
- Dataset robustness
- Zero- and few-shot learning





UNIVERSITEIT VAN AMSTERDAM

No Labels? No Problem!

Experiments with active learning strategies
for multi-class classification
in imbalanced low-resource settings

Emiel Steegh, Giovanni Sileno
University of Amsterdam

19th International Conference on Artificial Intelligence and Law (ICAIL)
18th June 2023, Braga, Portugal

BACKUP

Perhaps the real results were in the **lessons learned** along the way

- AL shines in when labeling is expensive and classes are skewed
- Task adaptation failed
- Imbalance, ORNL8 vs ORNL26
- Labels live longer than models
- Perfect oracles
- Sequence length
- Cold start

