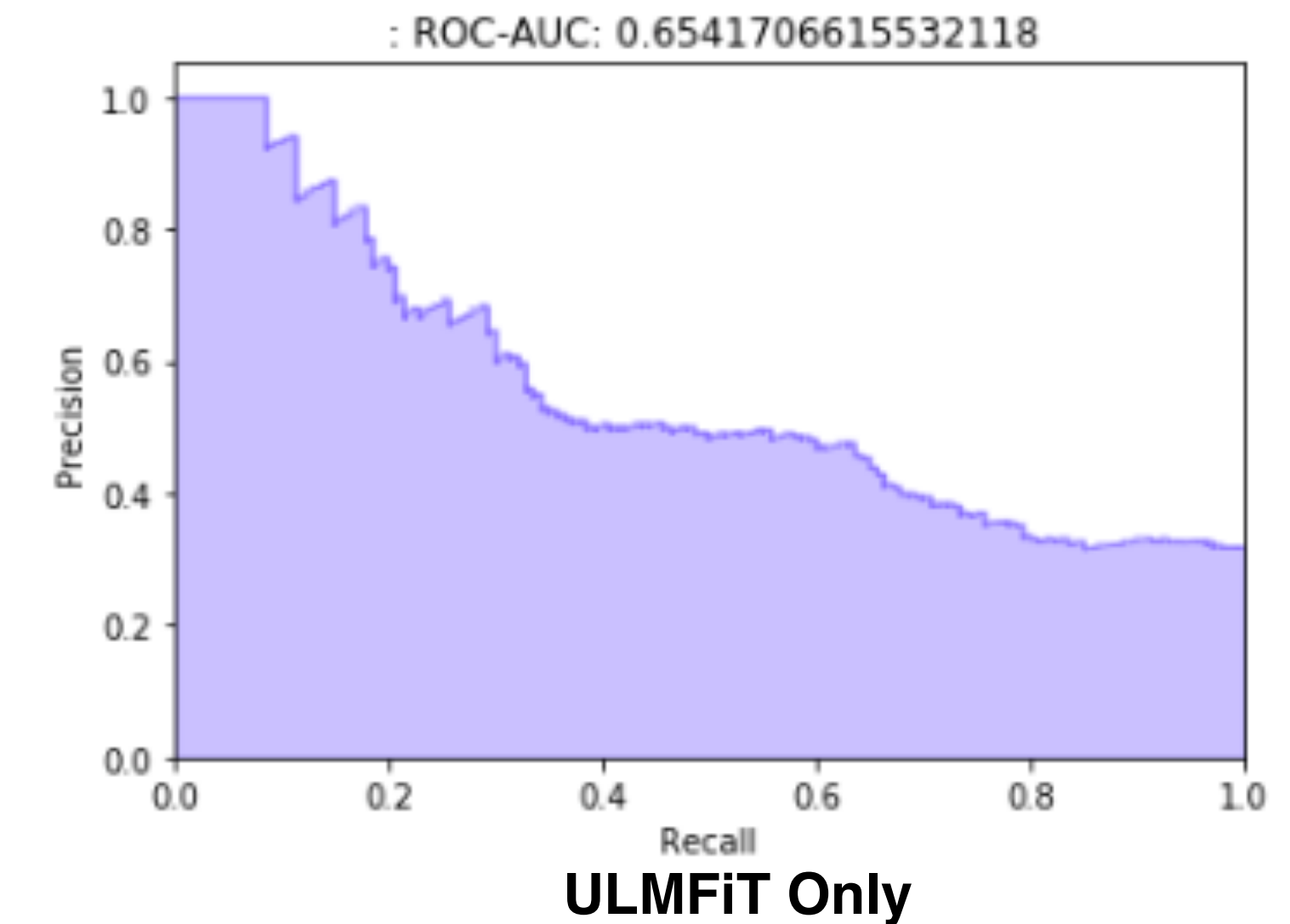
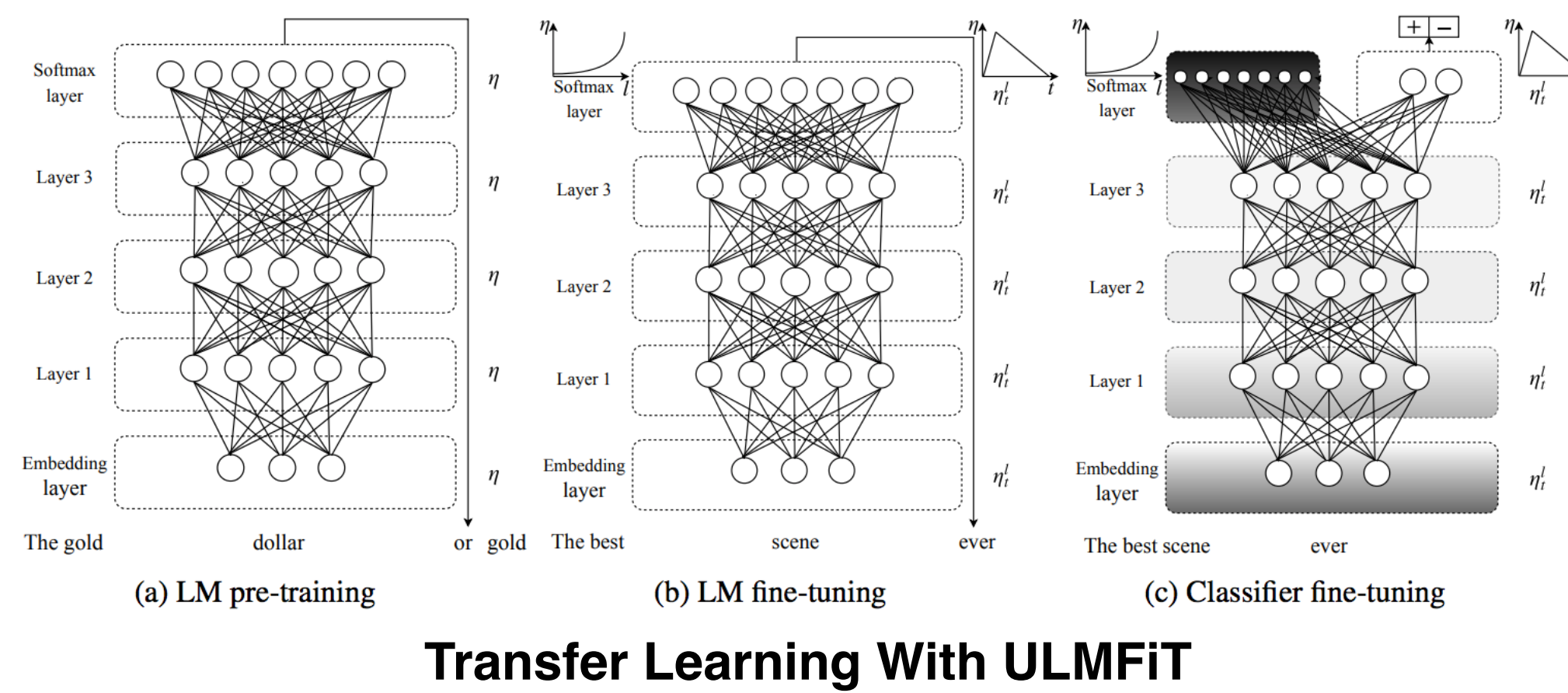


# Building NLP Classifiers Cheaply With Transfer Learning + Weak Supervision

## Problem

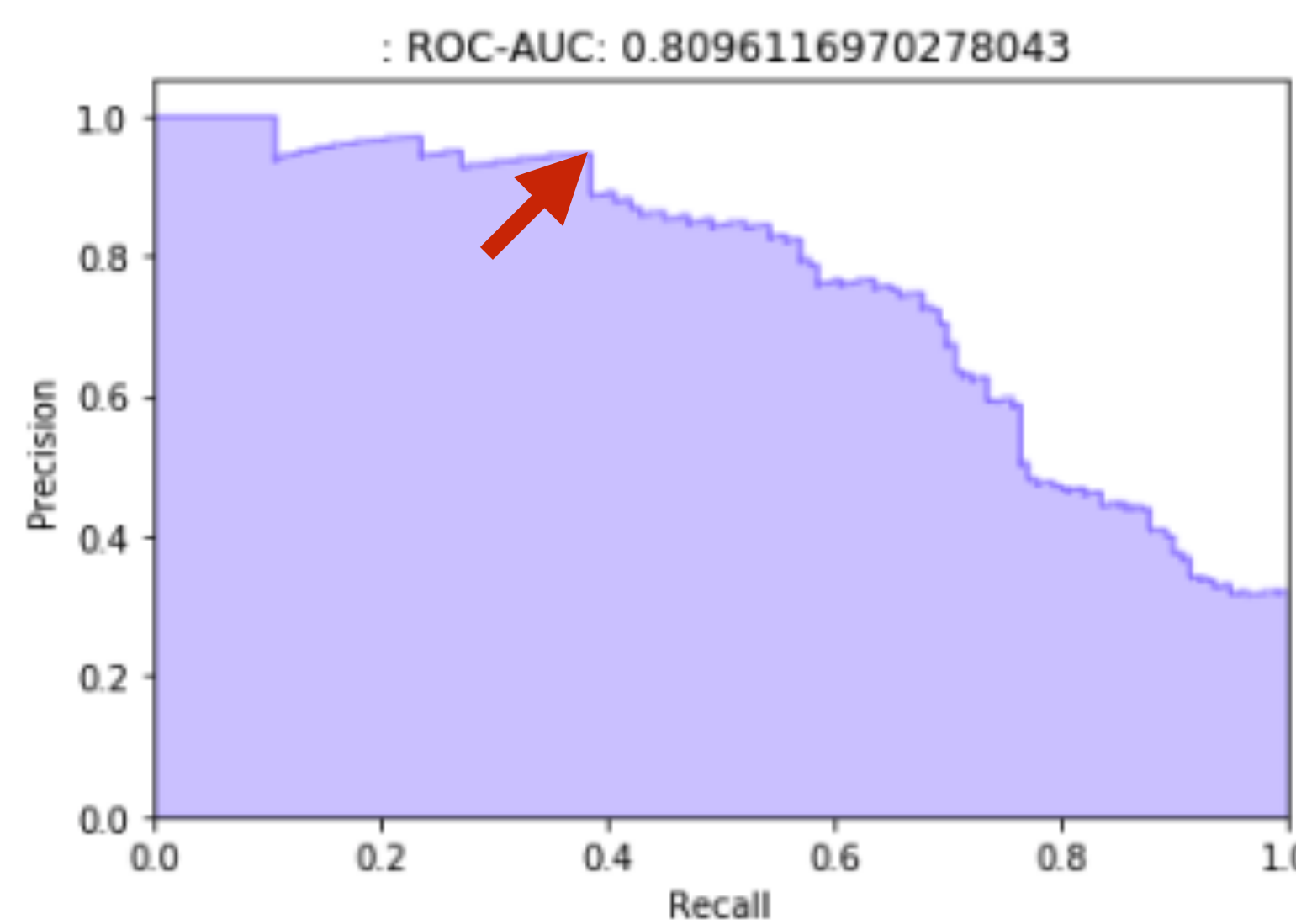
- Text classification models are very useful, but there's a high initial and recurring labeling cost
- Both Weak Supervision and Transfer Learning have proved to be very successful.
- The challenge is how to combine them to enable training of text models efficiently without AI engineers in the loop. We study this question on an important problem: hate speech and anti-semitism.



## Results

| Method              | Precision | Recall | ROC-AUC |
|---------------------|-----------|--------|---------|
| Label Model         | 80%       | 23%    | 0.73    |
| Logistic Regression | 94%       | 12%    | 0.77    |
| XGBoost             | 94%       | 12%    | 0.76    |
| Feed Forward NN     | 94%       | 11%    | 0.70    |

**Classifier Baselines**



**ULMFiT with Weak Supervision: 94% Precision / 39% Recall**

Abraham Starosta  
starosta@stanford.edu

## Analysis

- Weak labels increase ULMFiT ROC-AUC of by +0.15
- We were able to train a good classifier with just 1,000 labels
- Error Analysis: 65% of errors are because tweet ends in "jew"



## Conclusions

- We have evidence that weak supervision combined with transfer learning helps build text classifiers cheaply.
- For future work we would try other transfer learning approaches like fine-tuning Bert and the OpenAI Transformer.

## References

- Ratner, A. et, al. (2018) Snorkel: Rapid Training Data Creation with Weak Supervision.
- Howard, J. et, al. (2018) Universal Language Model Fine-tuning for Text Classification.

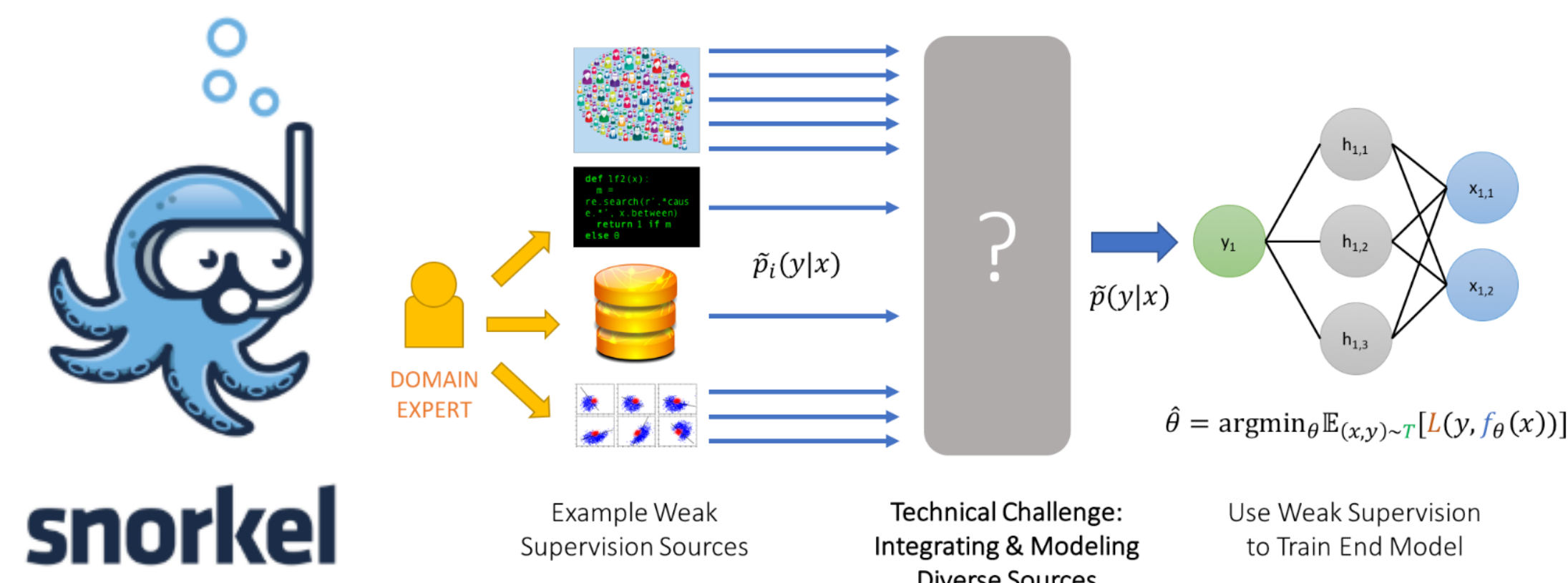
## Data / Task

- Train: 24,738 unlabeled
- Dev: 733
- Test: 438

Definitely not an easy task!



## Approach



**Weak Supervision With Snorkel**