# Automated Classification of Federal Rules Using BERT

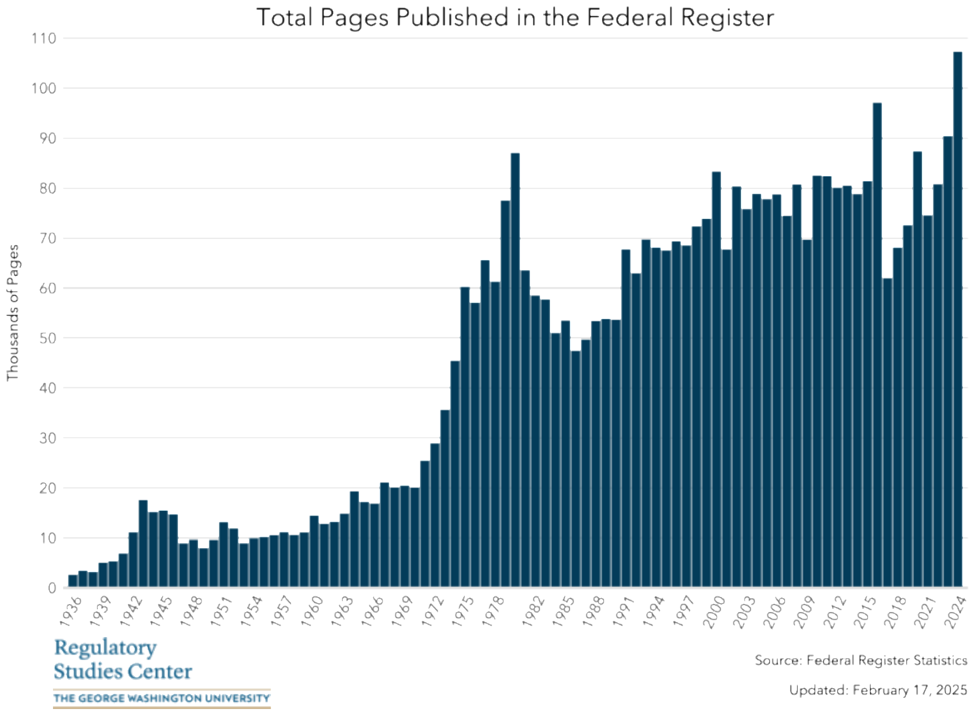
Henry Hirsch

# Problem (Part 1)

- US gov. publishes the text of all proposed and final rules in the Federal Register (FederalRegister.gov)
- Basic economic analysis is done on all rules
- Rules that meet certain criteria are classified as “significant”, “economically significant”, or “major”
- Classified rules receive more-rigorous economic analysis and review
- Official gov. regulatory website (RegInfo.gov) doesn’t always classify rules accurately, so the true count is uncertain

# Problem (Part 2)

- To get a more accurate count, GW's Regulatory Studies Center (RSC) has been manually tracking significant, economically significant, and major rules since 2021
- Requires skimming the text of all rules published in the Federal Register before classifying them in a spreadsheet
- Labor-intensive (nearly 110,000 pages published last year)



# Core Challenge (Part 1)

- Some rules belong to multiple classes
- There is both overlap and hierarchy between the classes
- **Significant rules** raise important policy issues or materially affect the economy, public health, the environment, or other key areas.
- **Economically significant** rules are a subset of significant rules with an expected annual economic impact of ≥ $100 million (≥ $200 million between April 2023 and January 2025).
- **Major rules** use a separate threshold of ≥ $100 million in annual economic impact or major increases in costs/prices for consumers or industries.

# Core Challenge (Part 2)

- No consistent wording pattern from which to determine class
- Too complex for regex

> The Office of Management and Budget (OMB) has determined that this is a significant rulemaking under Executive Order 12866, but it is not economically significant.

> This rule is a “significant regulatory action,” although not an economically significant regulatory action since it does not meet the threshold of $100 million in annual economic effects, under section 3(f)(1) of Executive Order 12866. Accordingly, the Office of Management and Budget has reviewed this regulation.

> Executive Order 12866. Based on our estimates, OMB's Office of Information and Regulatory Affairs (OIRA) has determined this rulemaking is significant under section 3(f)(1). Pursuant to Subtitle E of the

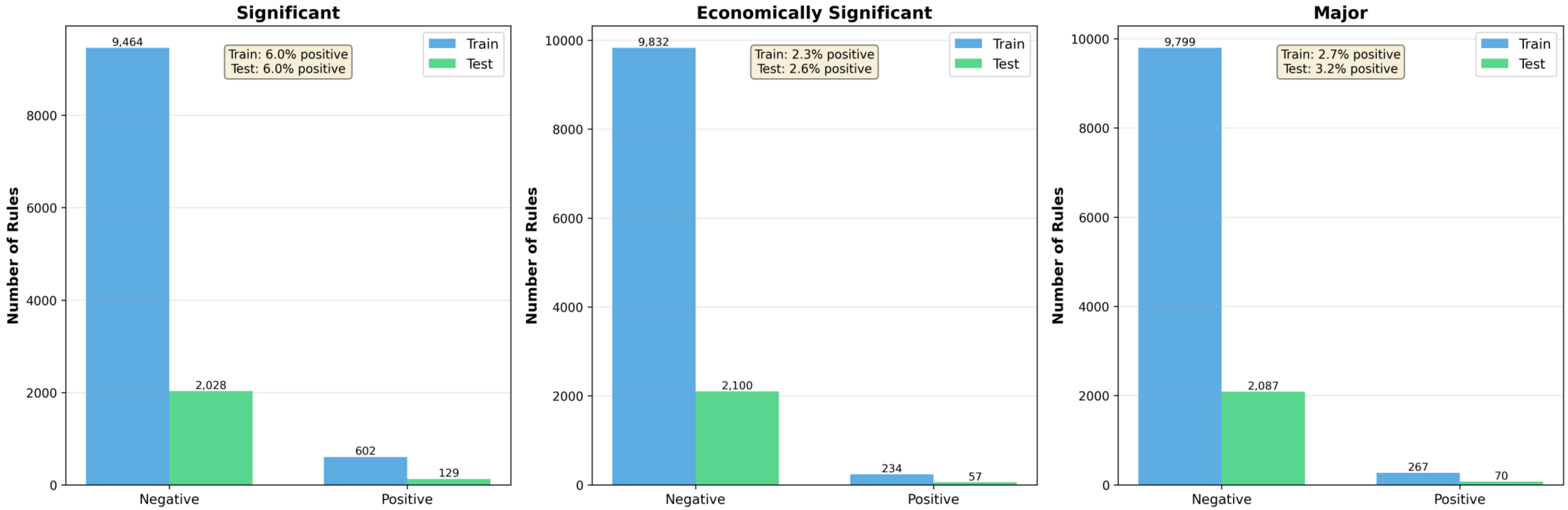
# Proposed Solution (Part 1)

- Use RSC’s manually labeled dataset fr_tracking.csv (n = 14,380) as training data for a model to conduct multi-label binary classification
- This dataset contains identifying information for every document published in the Federal Register (publication date, title, document number, etc.) as well as columns for “significant”, “economically significant”, and “major” classifications that research assistants have been labeling with “1” if the text of a given rule suggests it belongs to one of these classes and “0” if it doesn’t
- Use the Federal Register API to fetch the text of all rules identified in the fr_tracking.csv and append to the corresponding row in the dataset

# Proposed Solution (Part 2)

- Use the text of rules and their corresponding significance labels as inputs to fine-tune several pretrained DistilBERT models (one for each class) on rule text
- DistilBERT is a smaller, faster version of BERT (Bidirectional Encoder Representations from Transformers), pretrained on a vast corpus
- BERT models read text in both directions simultaneously, using attention mechanisms to understand how each word relates to every other word in context (can learn the textual context in which each rule class tends to appear)
- Add a binary classification head on top of each class-specific DistilBERT model to translate softmax probabilities into binary classification outputs ("1" = belongs to class, "0" = doesn't belong to class)
- Train/test /validate (70/15/15)

# Additional Challenges: Imbalanced Data



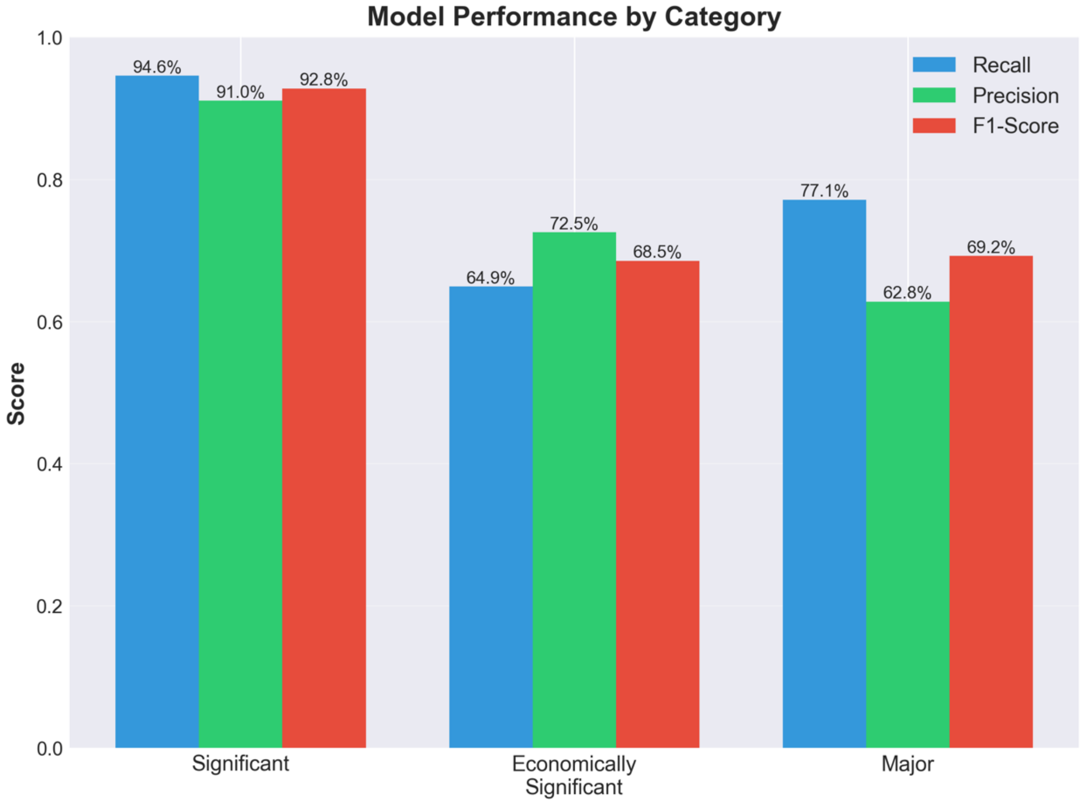

All classes are overwhelmingly negative. Without correction, models would simply predict "0" for each class. Three solutions: weighted loss function (penalizes false negatives 3x more than false positives, forces model to care about the rare positive class), stratified data splitting (to maintain same distribution of positives to negatives in all splits), and threshold optimization on validation set (instead of 0.50 default)

# Additional Challenges: BERT Token Limit

- BERT has a limit of 512 tokens per input (~2,000 chars)
- The text of federal rules tends to be very long (~54,000 chars)
- So, can’t feed the entire text of the rule into the DistilBERT models
- Solution: Identify and extract sentences that are relevant to each class using class-specific keywords. Feed class-specific input into DistilBERT models for training.
- 96.5% text length reduction (~54k chars to ~1.8k chars)

| Significant | Economically Significant | Major |
|---|---|---|
| "executive order 12866" | "$100 million" | "major rule" |
| "OIRA review" | "$200 million" | "5 U.S.C. 804" |
| "significant regulatory action" | "economically significant" | "submitted to congress" |

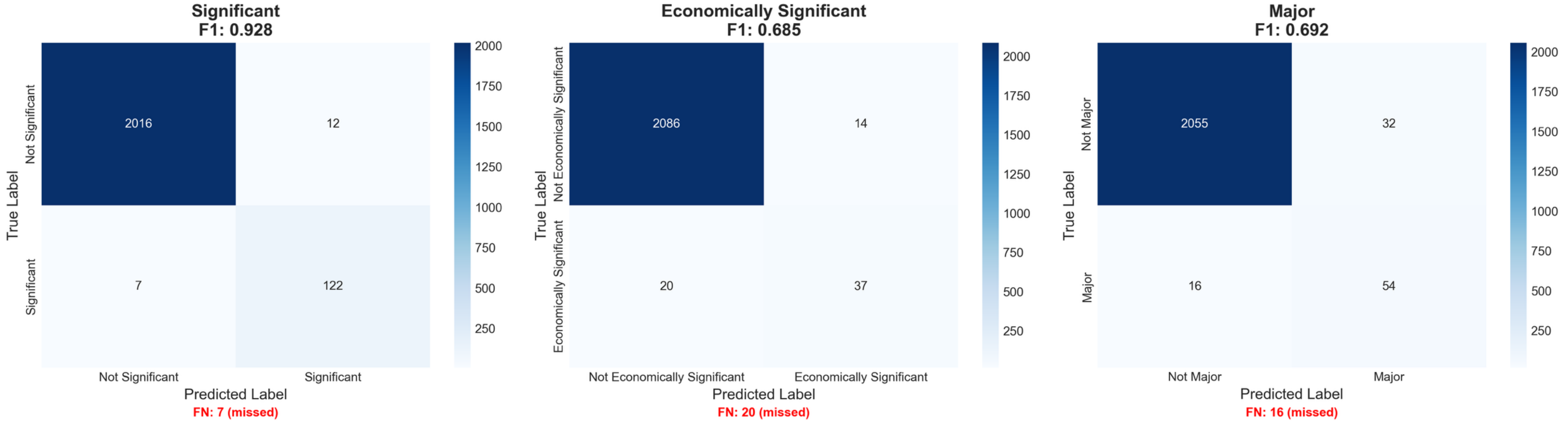
# Performance ( Part 1)





| Category | Recall | Precision | F1-Score |
|---|---|---|---|
| **Significant** | 94.6% | 91.0% | 0.928 |
| **Economically Significant** | 64.9% | 72.5% | 0.685 |
| **Major** | 77.1% | 62.8% | 0.692 |

# Performance (Part 2)




| **Significant Confusion Matrix** | **Predicted Negative** | **Predicted Positive** |
|---|---|---|
| **Actual Negative** | 2016 (TN) | 12 (FP) |
| **Actual Positive** | 7 (FN) | 122 (TP) |

False negatives are extremely costly, will not be caught by a human reviewing the model's predictions.

# Next Steps

- Figure out how to reduce false negatives (analyze text of FN rules)
- Review early entries in fr_tracking.csv dataset (found several labelling errors)
- Improve performance for both economically significant and major models
- Figure out how to distinguish economically significant rule text from significant rule text (significant overlap)