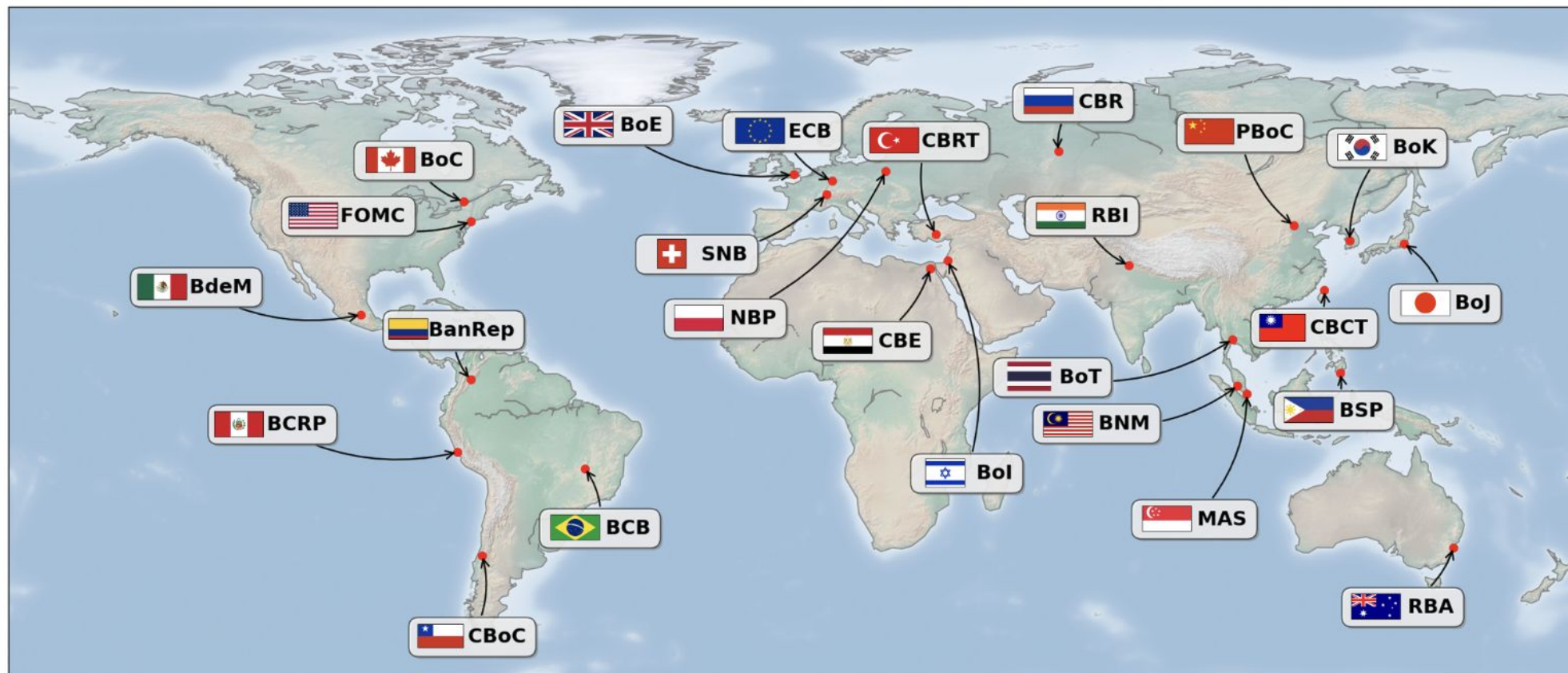


Words That Unite The World: A Unified Framework for Deciphering Central Bank Communications Globally

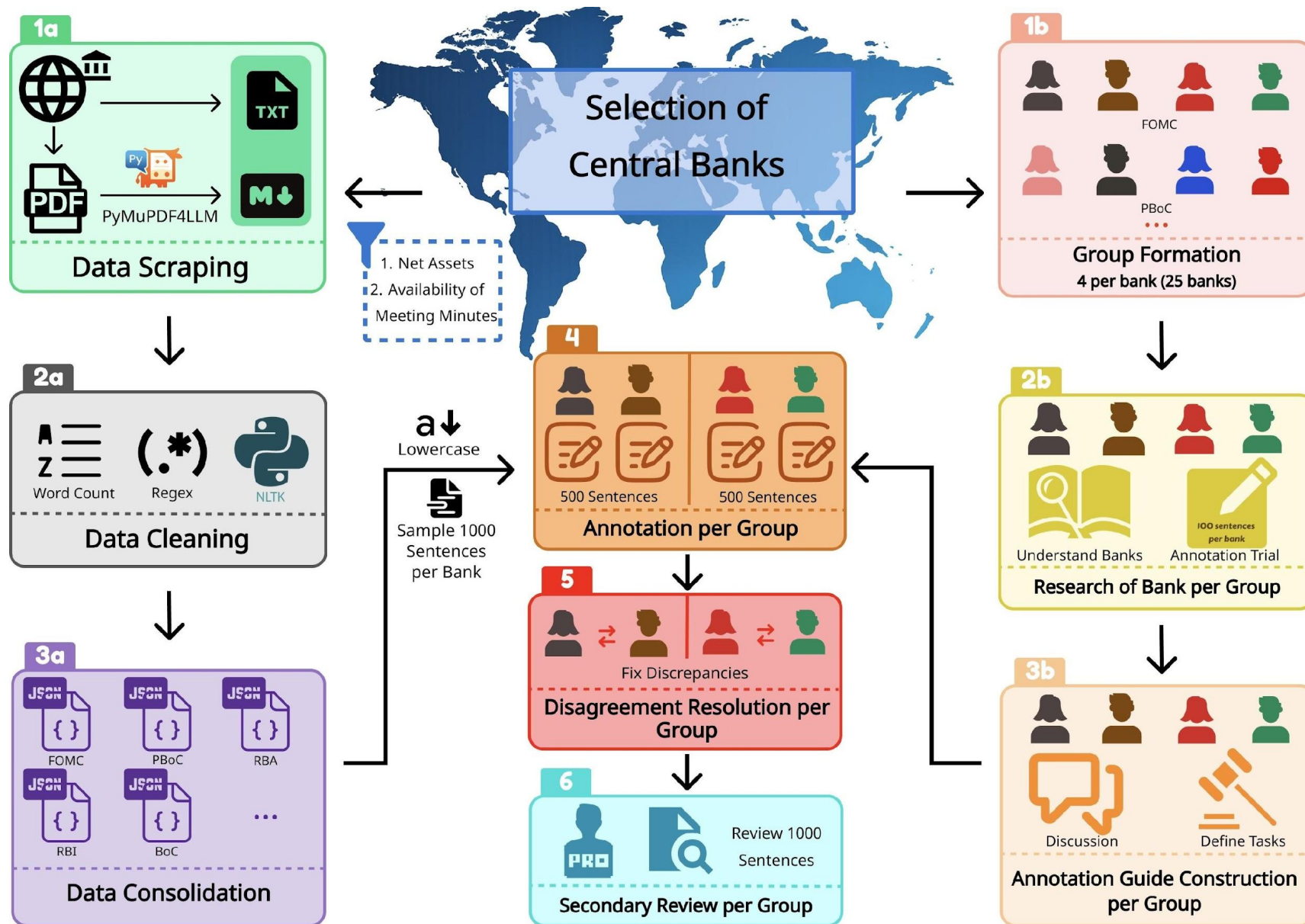
Agam Shah*, Siddhant Sukhani*, Huzaifa Pardawala*,

Saketh Budideti, Riya Bhadani, Rudra Gopal, Siddhartha Somani, Rutwik Routu, Michael Galarnyk, Soungmin Lee,

Arnav Hiray, Akshar Ravichandran, Eric Kim, Pranav Aluru, Joshua Zhang, Sebastian Jaskowski, Veer Guda, Meghaj Tarte, Liqin Ye, Spencer Gosden, Rachel Yuh, Sloka Chava, Sahasra Chava, Dylan Kelly, Aiden Chiang, Harsit Mittal, Sudheer Chava



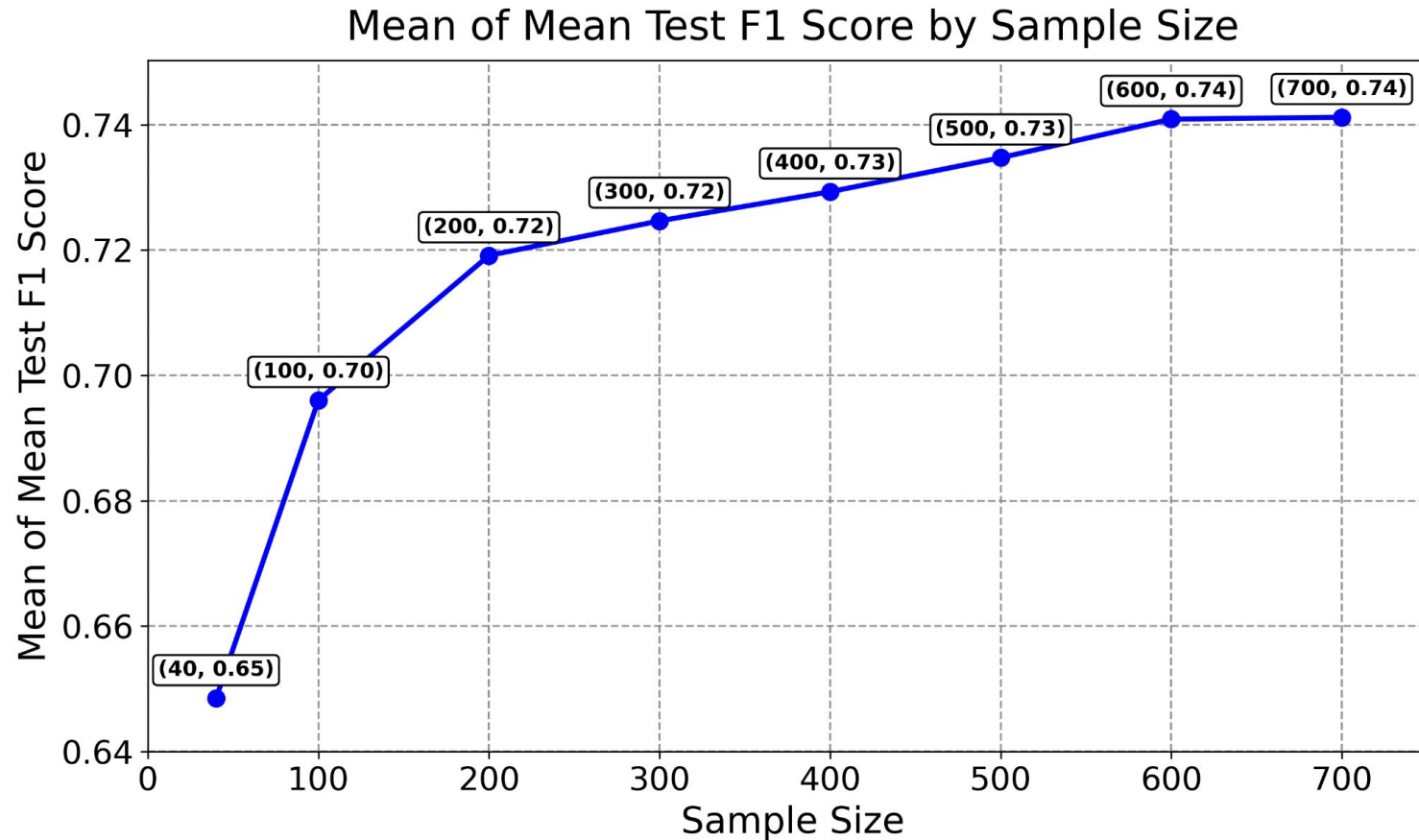
Dataset		Models		Annotations	
Central Banks	25	Pretrained Language Models	7	Annotators	104
Years	1996–2024	Large Language Models	9	Annotation Guides	26
Scraped Sentences	380,200	Best Stance Model*	RoBERTa-Large (0.740)	Annotation Steps	6
Annotated Sentences	25,000	Best Temporal Model*	RoBERTa-Base (0.868)	Tasks	
Total Words	10,289,163	Best Uncertainty Model*	RoBERTa-Large (0.846)	Stance Detection	Hawkish, Dovish, Neutral, Irrelevant
Corpus Size	2,661,400	Benchmarking Experiments	15,075	Temporal Classification	(Not) Forward Looking
Sentences/Year*	13,110.34	Few Shot	✓	Uncertainty Estimation	(Un)certain
Words/Sentence*	27.06	Few Shot + Ann. Guide	✓		



Benchmarking

- General all banks setup and Bank Specific setup
- 7 PLMS and 9 LLMs
- All general setup models performed significantly better on average
- Stance Detection: RoBERTa-Large {0.740}
- Temporal Classification: RoBERTa-Base {0.868}
- Uncertainty Estimation: RoBERTa-Large {0.846}
- Total 15,057 benchmarking experiments
- PLM better than LLM few shot and few shot with annotation guide

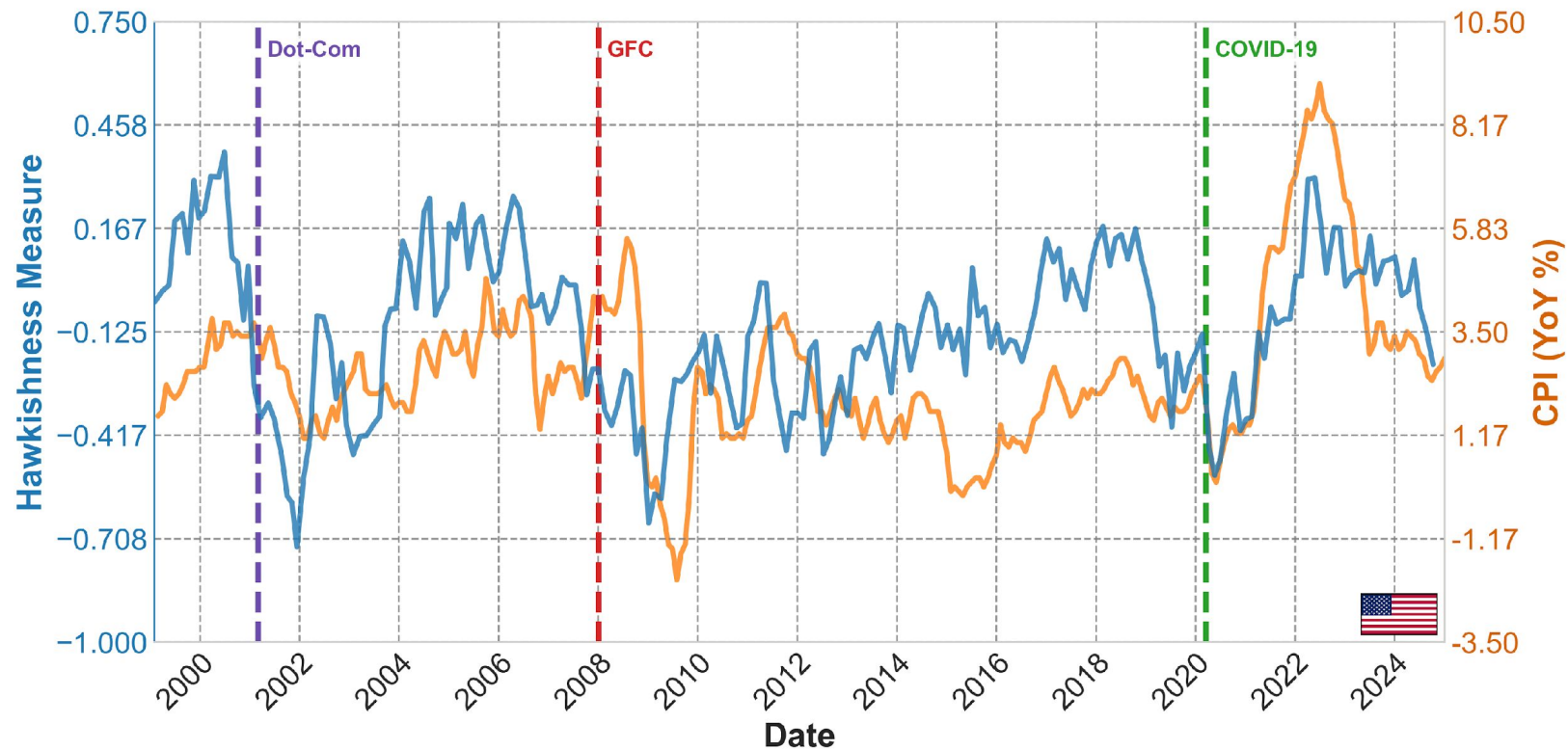
Whole is greater than the sum of its parts



- Obtained Linear correlation ($p = 0.016$), confirming semantic overlap drives performance gains.
- Found semantically similar examples across banks, demonstrating cross-bank knowledge transfer

Economic Analysis

$$\text{Hawkishness Measure} = \frac{\# \text{ Hawkish Sentences} - \# \text{ Dovish Sentences}}{\# \text{ Total Sentences} - \# \text{ Irrelevant Sentences}}$$



Generating Meeting Minutes

- Use GPT-4.1 for its strong stance-detection performance.
- Feed the model with
 - The target bank's most recent pre-cutoff document,
 - Documents from the other 24 banks released prior to the target bank's next post-cutoff meeting.
- Compute hawkishness scores for both the generated and the actual document.
- Fit a cross-bank linear regression obtaining a positive correlation between generated and actual document

Additional Experiments

- Human Evaluation
- Error Evaluation
- Transfer Learning
 - 26th Bank - Czech National Bank
 - U.S. Senate Committee on Agriculture, Nutrition, and Forestry
- Few shot LLM
- Few Shot LLM with annotation guide

Release of Materials

- Largest monetary policy dataset with 380k sentences across 25 central banks with metadata.
- 78 Fine tuned models specific to each bank
- General setups model for Stance Detection, Uncertainty Estimation Temporal Classification
- 133 Page annotation guide
- Detailed annotation instructions

