



Research Framework

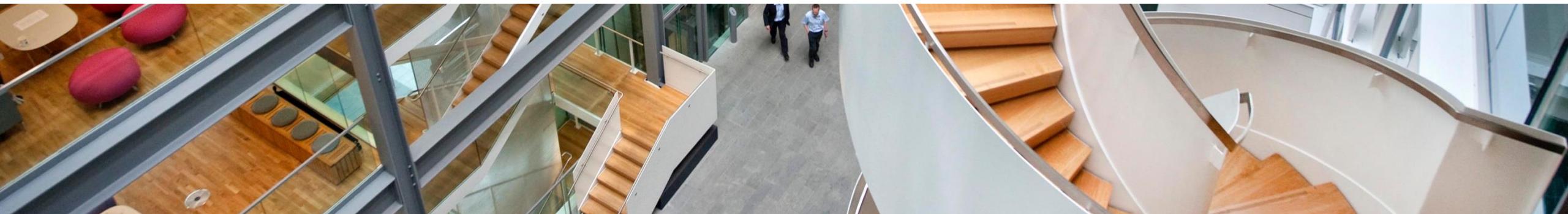
Tuğrulcan Elmas / Tj



THE UNIVERSITY of EDINBURGH
informatics



THE UNIVERSITY of EDINBURGH
School of Social
& Political Science



Basic Research Framework

- Research Question & Hypothesis
 - RQ: What makes persuasion effective?
 - Hypothesis: Evidence makes persuasion effective
- Data Collection
 - Download Posts & Replies from Change My View Subreddit
- Sample Population
 - Redditors in Change My View as proxy for “people engaging in persuasion online”
- Methods & Analysis
 - Identify “Evidence” in replies (e.g., search for “http” links)
- Measure & Report Outcome
 - Persuasion rate of replies with evidence vs. without evidence



Research Question & Hypothesis

- Identify a Gap
 - Start with curiosity
 - New question or new approach
- Refine the Question
 - Specific & measurable
 - Can be guided by data
 - Iterative: it's okay to update it
- Create Hypothesis Based on Theory
 - Elaboration Likelihood Model -> Evidence Makes Persuasion Effective
 - On Existing Knowledge or Observation (not on a guess)



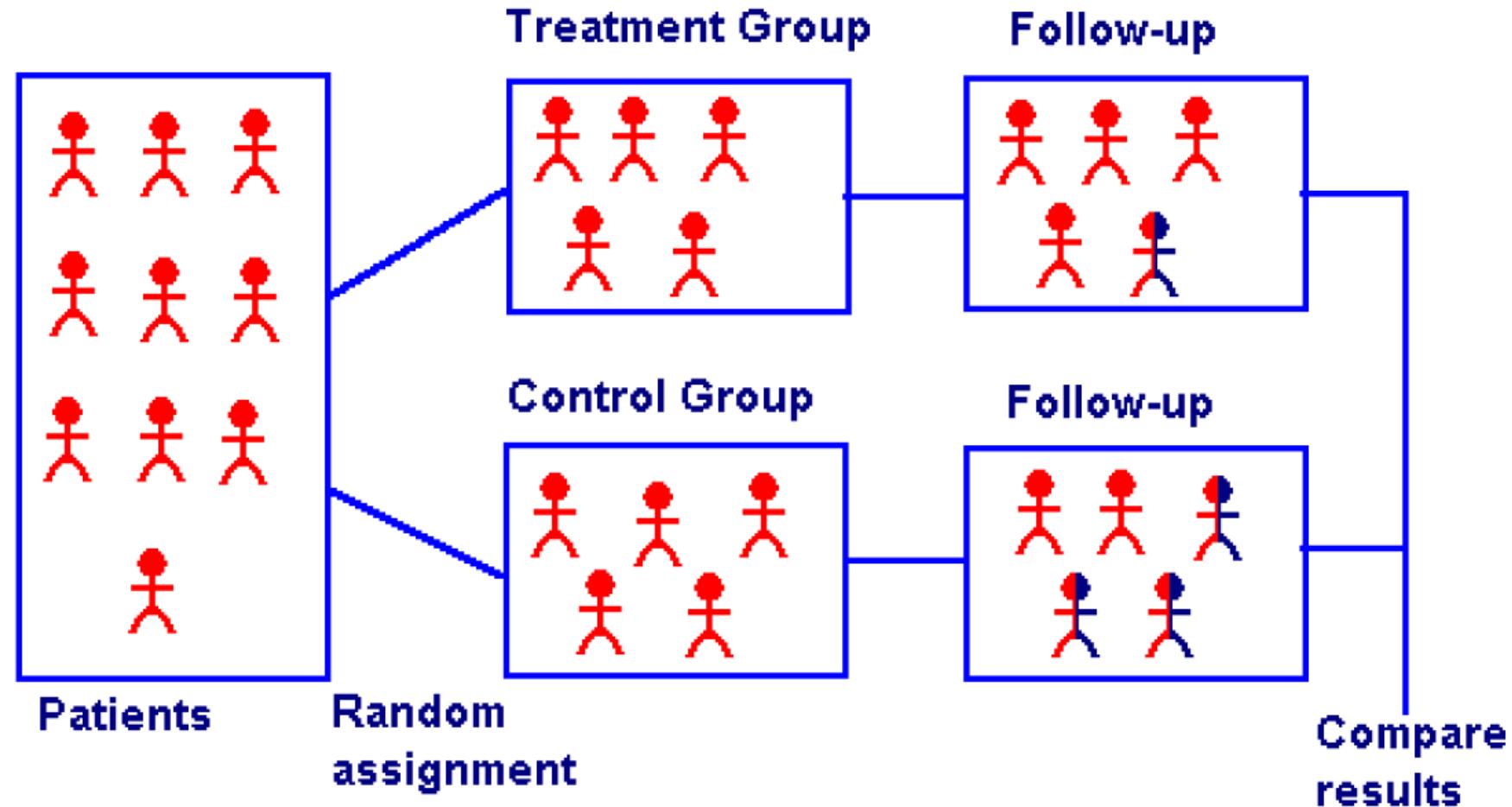
Data Collection

- Asking
 - Surveys
 - Interviews



Data Collection

- Asking
 - Surveys
 - Interviews
- Observing
 - Experiments



Data Collection

- Asking
 - Surveys
 - Interviews
- Observing
 - Experiments
 - Big Data Analysis



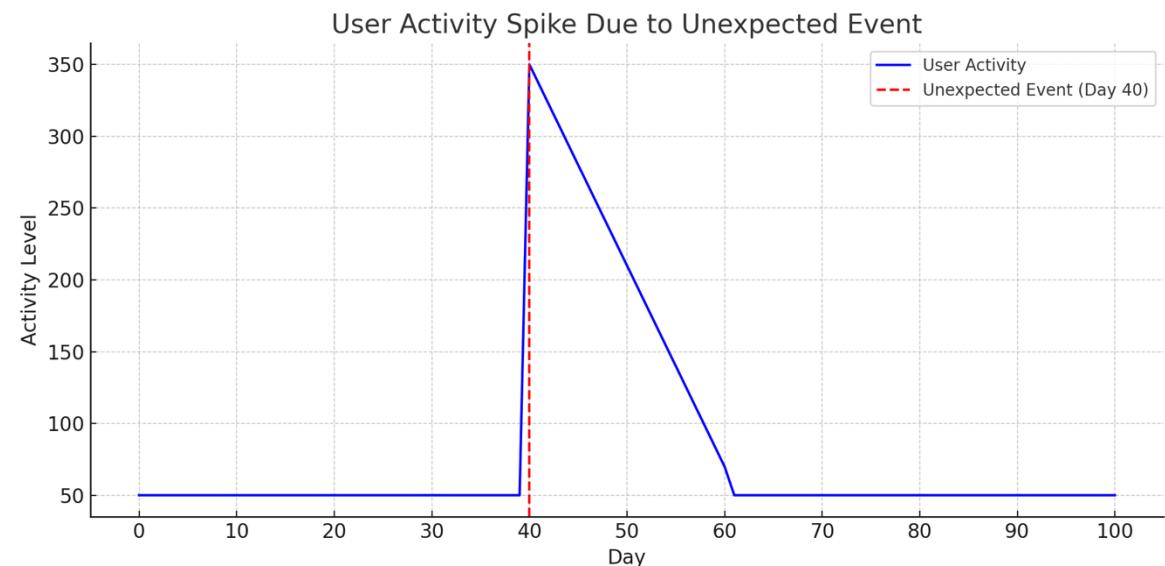
Big Data is Cheap

- Many free, public data sources
 - Reddit, Wikipedia, Common Crawl...
- Limitations.. Ethical problems..



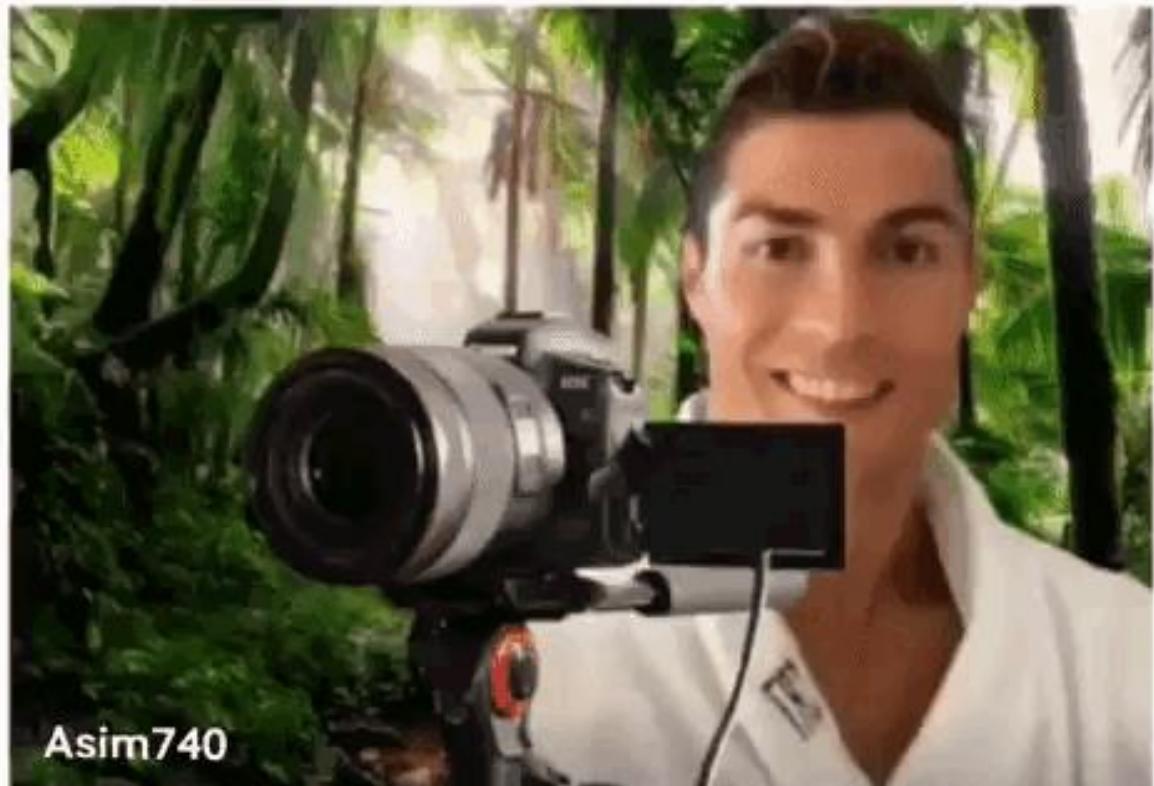
Big Data is Always-On

- Constantly being collected
- Longitudinal Studies
 - Reddit is 20 years old
- Unexpected events
 - Crisis, disasters, protests, war
 - Data before, during & after



Big Data is Non-Reactive

- Observer effect
- Natural behavior.. sort of
 - Non-Opinions e.g., hate speech
- Limitations
 - Non-researcher observer effect



Sampling from a Population of Focus

- Probability Sampling
 - True random
 - Ideal, but hard
- Non-Probability Sampling
 - Some have less/zero chance of being sampled
 - Sociodemographic stratification: age, gender, education..



Caveat: Big Data is Biased!

- Internet ≠ Real Life
- Platforms have different demographics
 - “Facebook is for the old”
- Behaviors shift over time
 - Regular folk don’t use hashtags anymore
- “Dead Internet Theory”
 - Redditors in Change My View use AI for persuasion



Methods & Analysis

- Data is not in the exact form required
 - Dirty
 - Biased
 - Unethical (e.g., not anonymous)
 - Open-ended
- Define Sample & Measures
- Data Preprocessing
 - Clean, normalize, anonymize...
- Data Labeling
 - Make it “measureable”



Data Labeling by Annotation

- Annotate data according to your measure
 - Does this Reddit comment contain evidence ? Yes / No
 - Manually read the comments and answer the question
- (If challenging) report annotation agreement
 - Have another person to annotate & compare the annotations
- Reuse the “annotation” dataset later



Caveat: Big Data is Big!

- Cannot annotate everything!



Small data



BIG DATA

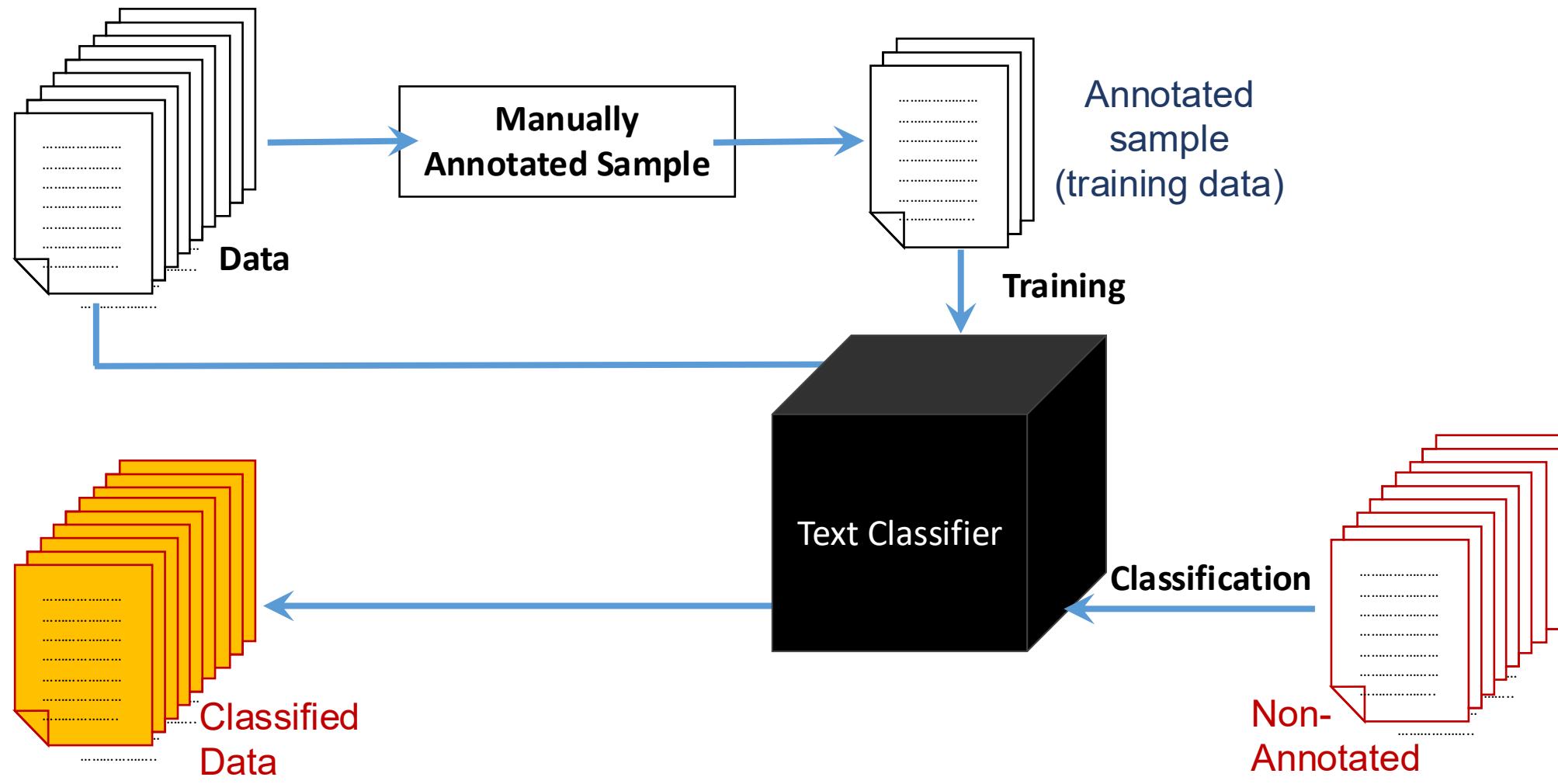


Scaling Up: Crowdsourcing

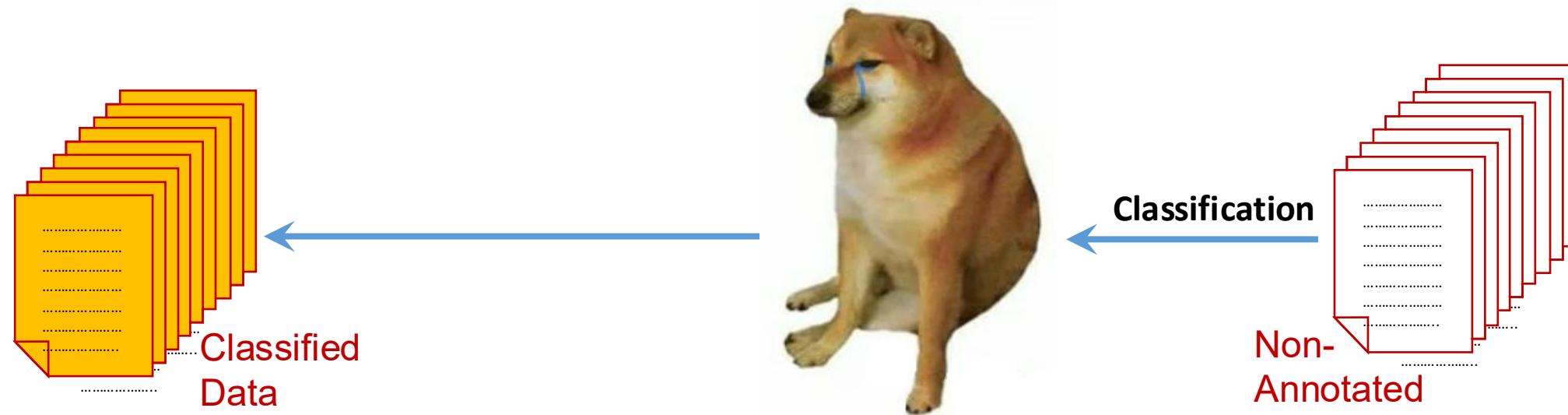
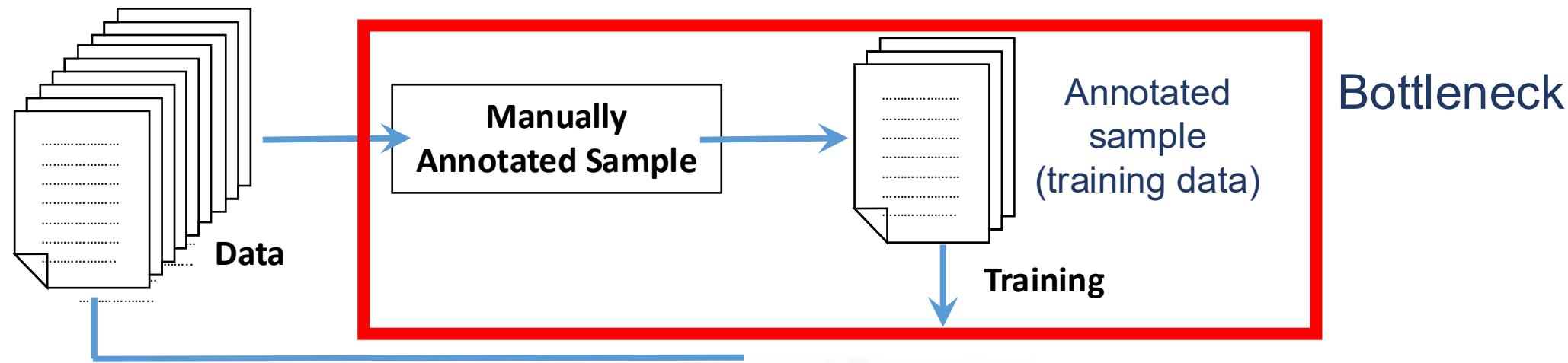
- Outsource annotations
 - Pay others to annotate
- “Crowdsource” annotations
 - Hire “crowdworkers”
 - Amazon MTurk, Prolific, Upwork
- Challenges
 - Low-quality annotations
 - **New: Annotators using AI**



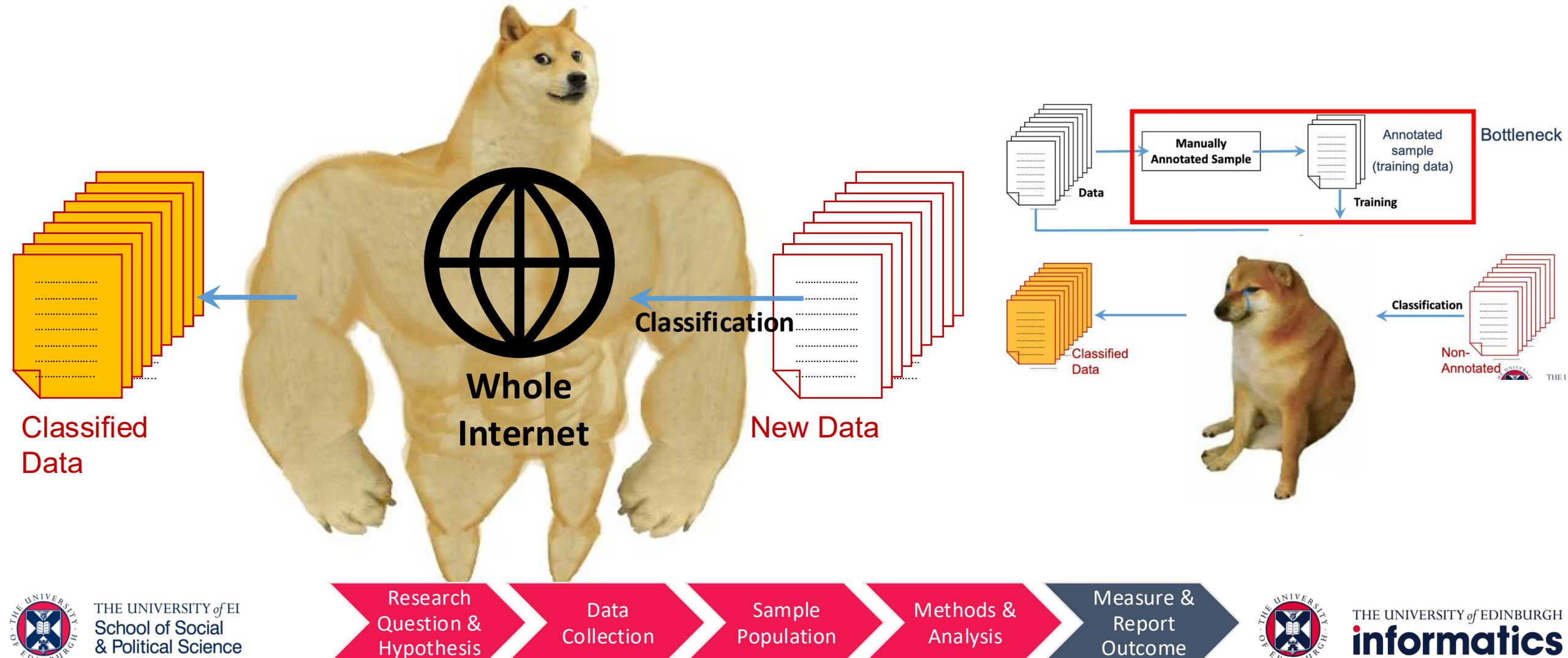
Scaling Up: Classification



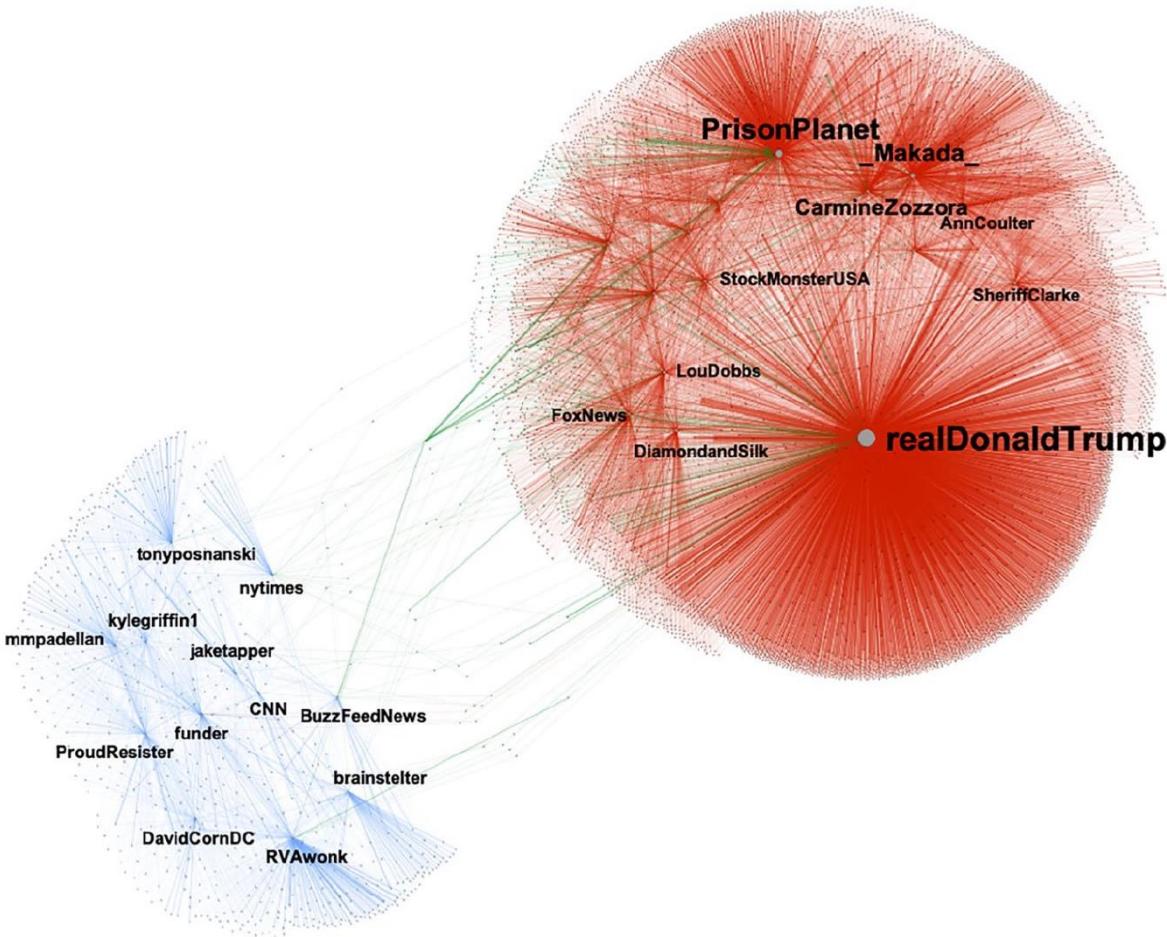
Is This A Good Classifier?



Classification Using LLMs



Caveat: Big Data is “Multimodal”



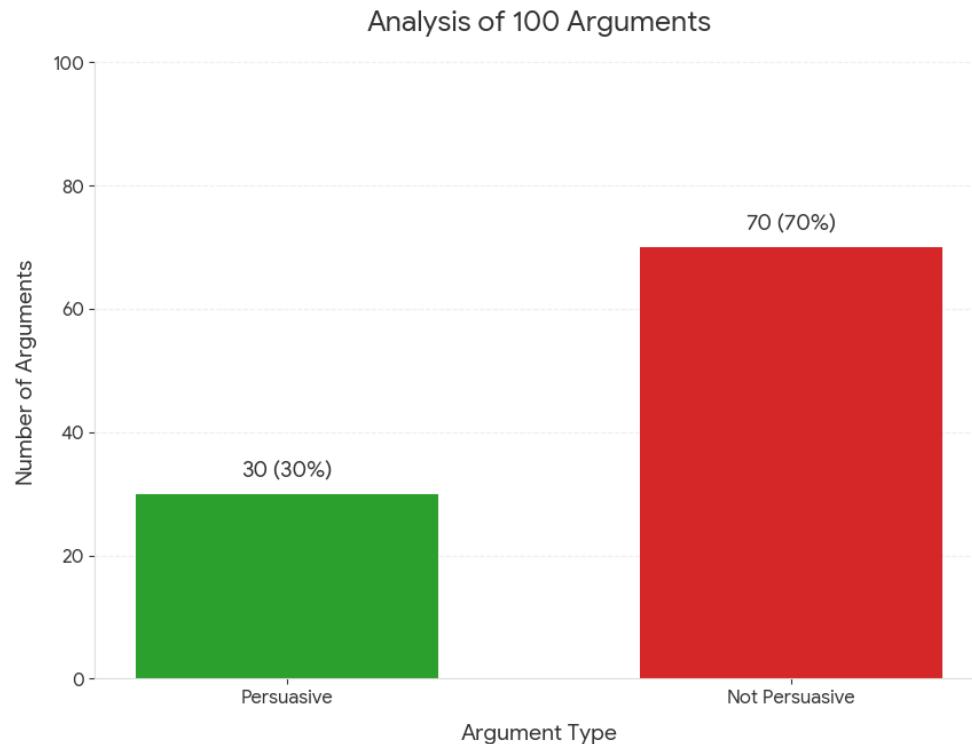
Measurement

- We analyzed. Now what?
- Measure Outcome & Report Findings
 - “Build a Data-Story”



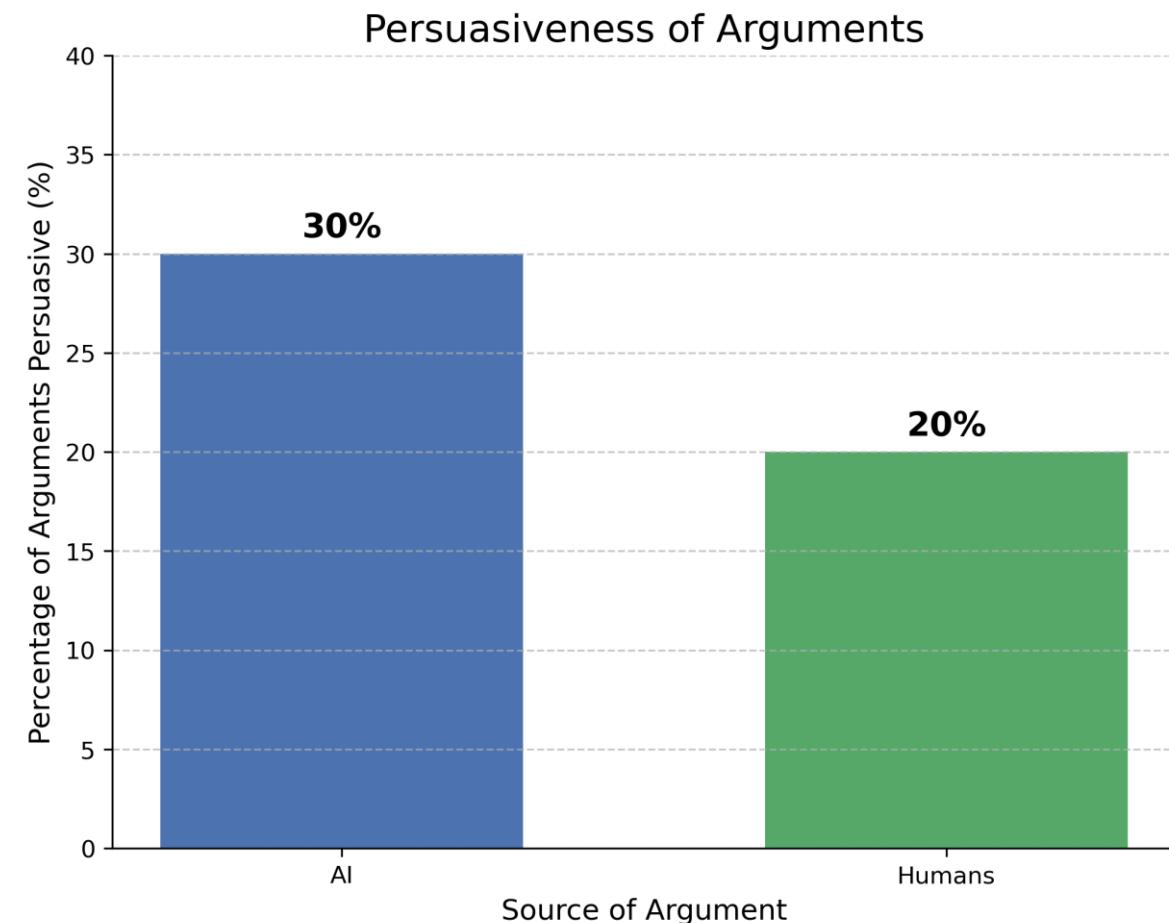
Counting

- ... analyzed 100 arguments created by AI
- ... 30 out of 100 arguments were persuasive
- ... equals to 30%



Comparative Analysis

- 30% of arguments by **AI** were persuasive
- 20% of arguments by **Humans** were persuasive
- Relative Difference: $\frac{30 - 20}{20} = 50\%$
 - AI is 50% more likely to persuade than a human



Forecasting

- Collect Data from Past
- Predict the Future

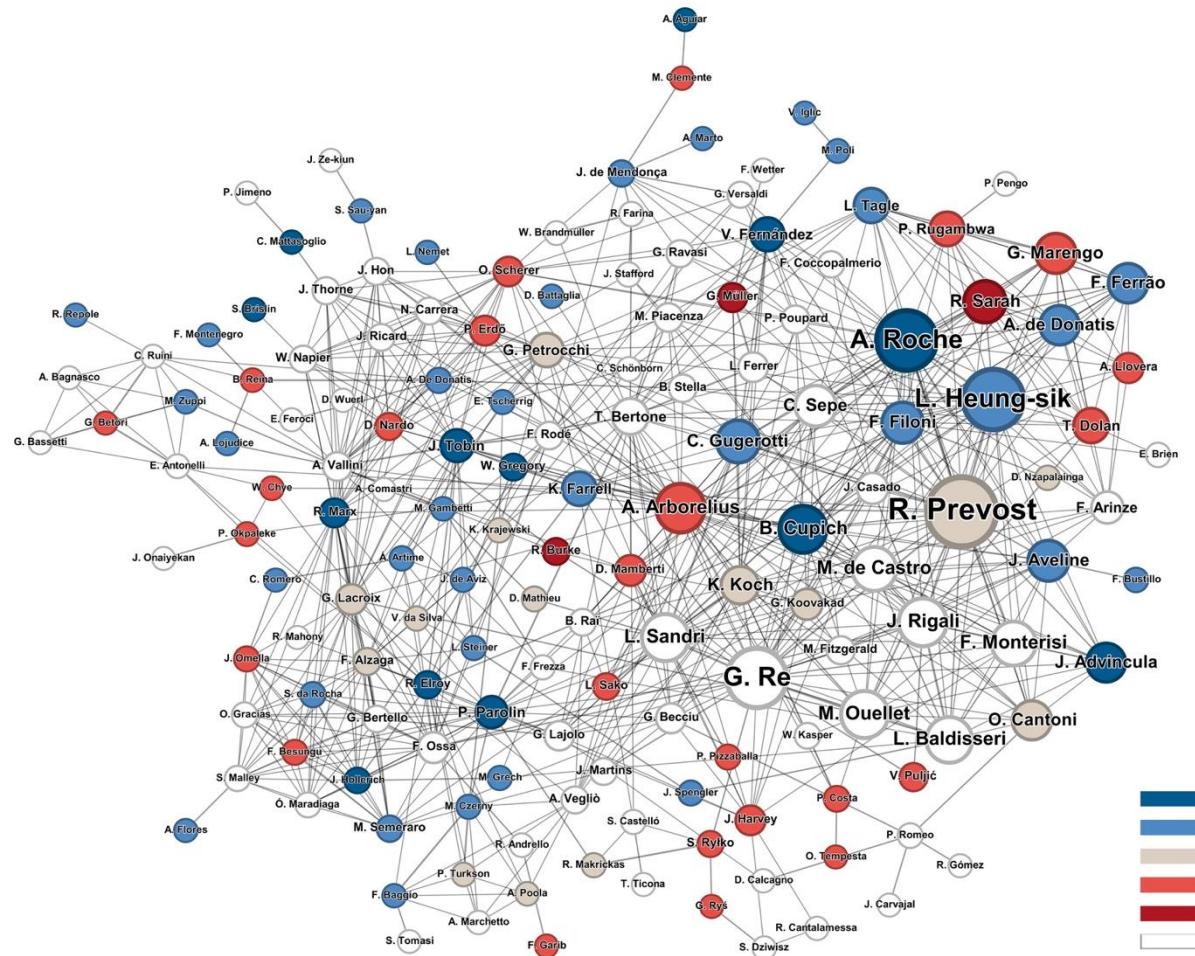


Can You Predict...

- Which arguments are more persuasive?
- The Next US President
- Spread of Corona
- Price of Bitcoin
- Trump Getting Greenland?



Can You Predict... the Pope?



AI predicted the next pope. Did it get it right?

Researchers use algorithms to study political factions within the Catholic Church

8 MAY 2025 • 1:45 PM ET • BY SARA REARDON



We Predicted. Now What?

- Make millions from algorithmic trading
- Reveal findings through your prediction method
 - Your classifier classifies AI arguments as more persuasive => AI is more persuasive
 - What information does your classifier relies on?
 - “Predictive Modelling”



University of Edinburgh Study (2023)

Predictive Modeling:

- Collect the Reddit replies
- Identify links (proxy for evidence) by looking for “http”...
- Do Magic (Statistical analysis)
 - Run logistic regression 😊
 - Compute persuasion scores (coefficients 😊)
- Links make persuasion ~1.6 times as likely
 - If 10% chance without links => 16% with links
 - Adding a link gives 60% persuasion boost

		95% CI		
	Parameter	Estimate	Lower	Upper
	Intercept	1.07	0.99	1.15
	Links	0.48	0.36	0.62
Post-comment similarity		0.22	0.14	0.31
	Length of thread	0.18	0.07	0.29
	Moral language	0.16	0.09	0.24
	Comment word count	0.13	0.03	0.24
	Use of unique words	0.03	-0.06	0.14
	Causal language	-0.05	-0.12	0.02
Second person pronouns		-0.05	-0.11	0.02
	Number of replies	-0.07	-0.13	0.00
	Questions	-0.07	-0.15	0.00
	Exclamation	-0.09	-0.19	-0.01
	Swearwords	-0.17	-0.23	-0.11

(From the paper)