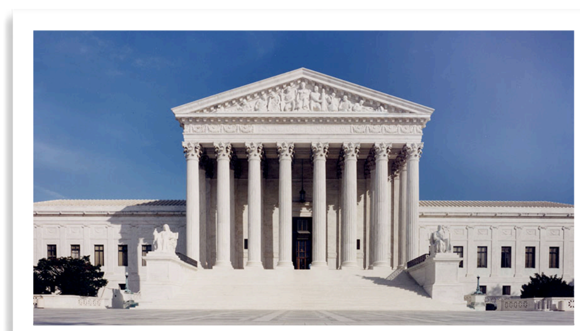
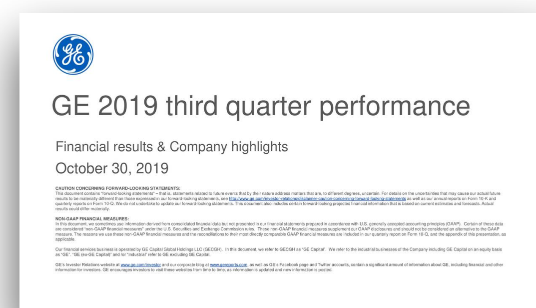
Natural Language Processing in Action:

Automated Event Extraction for News-Based Counterdata

Katie Keith
CS 104
December 2, 2022

Age of abundant digitized texts

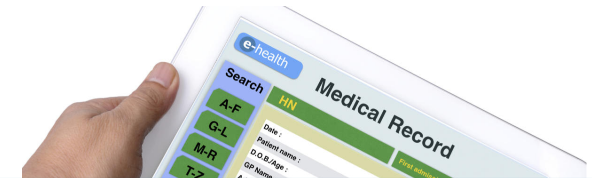



GE 2019 third quarter performance

Financial results & Company highlights

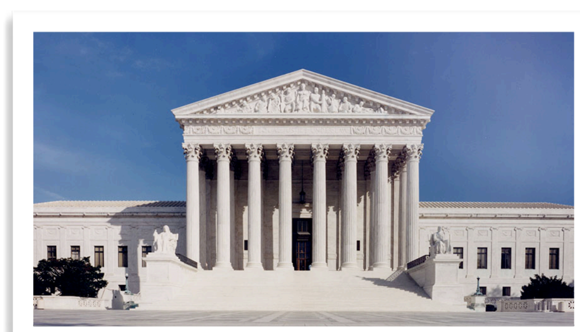
October 30, 2019

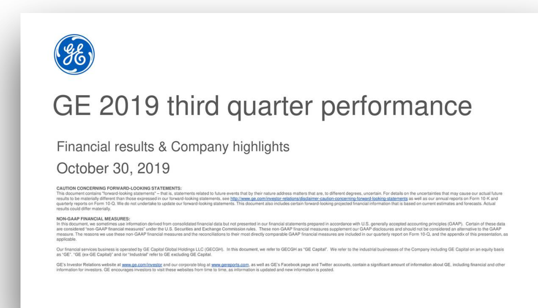
Text data for social sciences questions



What drives newspapers' political slant?

Gentzkow and Shapiro, Econometrica, 2010





What is the nature of online censorship in China?

King et al., American Political Science Review, 2013

Manual analysis is costly at scale

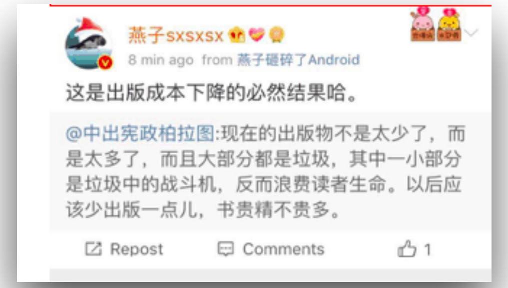
What drives newspapers' political slant?

Gentzkow and Shapiro, Econometrica, 2010

400 news outlets
x 1 year of articles

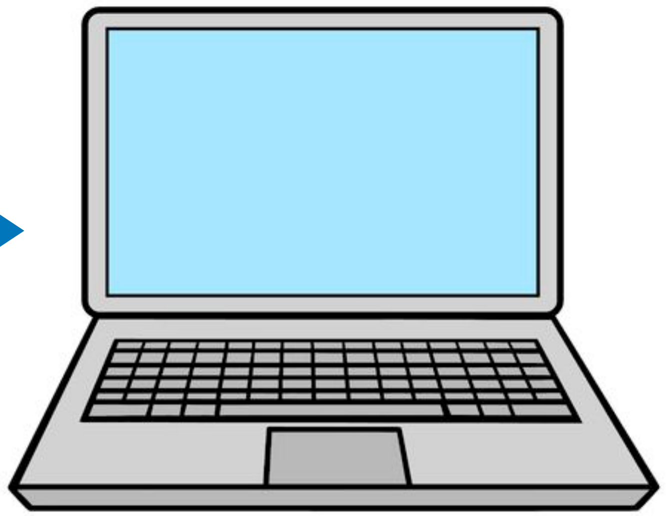
What is the nature of online censorship in China?

King et al., American Political Science Review, 2013



11 million
posts

Natural language processing (NLP)



CS 104 lingo: Table()

Focus of today's talk

Corpus-Level Evaluation for Event QA:
The IndiaPoliceEvents Corpus Covering the 2002 Gujarat Violence

Andrew Halterman* — Massachusetts Institute of Technology — ahalt@mit.edu
Katherine A. Keith* — University of Massachusetts Amherst — kkeith@cs.umass.edu
Sheikh Muhammad Sarwar* — University of Massachusetts Amherst — smsarwar@cs.umass.edu
Brendan O'Connor — University of Massachusetts Amherst — brenocon@cs.umass.edu

Abstract

Automated event extraction in social science applications often requires corpus-level evaluations: for example, aggregating text predictions across metadata and unbiased estimates of recall. We combine corpus-level evaluation requirements with a real-world, social science setting and introduce the INDIAPOLICEEVENTS corpus—all 21,391 sentences from 1,257 English-language *Times of India* articles about events in the state of Gujarat during March 2002. Our trained annotators read and label every document for mentions of police activity events, allowing for unbiased recall evaluations. In contrast to other datasets with structured event representations, we gather annotations by posing natural questions, and evaluate off-the-shelf models for three different tasks: sentence classification, document ranking, and temporal aggregation of target events. We present baseline results from zero-shot BERT-based models fine-tuned on natural language inference and passage retrieval tasks. Our novel corpus-level evaluations and annotation approach can guide creation of similar social-science-oriented resources in the future.

Figure 1: Motivation (A-B) and procedures (B-C) for this paper: **A.** Social scientists often use text data to answer substantive questions about temporal aggregates. **B.** To answer these questions, domain experts use natural language to define semantic event classes of interest. **C.** Our INDIAPOLICEEVENTS dataset: Humans annotate *every* sentence in the corpus in order to evaluate whether a system achieves full recall of relevant events. In production, computational models run B's queries to classify or rank sentences or documents, which are aggregated to answer A.

1 Introduction

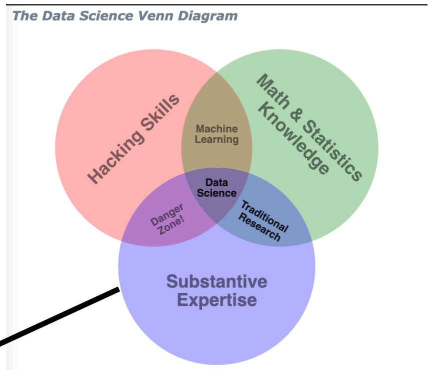
Understanding the actions taken by political actors is at the heart of political science research: How do actors respond to contested elections (Daxecker et al., 2019)? How many people attend protests (Chenoweth and Lewis, 2013)? Which religious groups are engaged in violence (Brathwaite and Park, 2018)? Why do some governments try to prevent anti-minority riots while others do not (Wilkinson, 2006)? In the absence of official records, social scientists often turn to news data to extract the actions of actors and surrounding events. These news-based event datasets are often constructed by hand, requiring large investments of time and money and limiting the number of researchers who can undertake data collection efforts.

Automated extraction of political events and actors is already prominent in social science (Schrodt et al., 1994; King and Lowe, 2003; Hanna, 2014; Hammond and Weidmann, 2014; Boschee et al., 2015; Beieler et al., 2016; Osorio and Reyes, 2017) and is increasingly promising given recent gains in information extraction (IE), the automatic conversion of unstructured text to structured datasets (Grishman, 1997; McCallum, 2005; Grishman, 2019). While social scientists and IE researchers have over-

* Indicates joint first-authorship.

4240

Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pages 4240–4253
August 1–6, 2021. ©2021 Association for Computational Linguistics



Andy Halterman
Political Science

Katie Keith
Computer Science

Sheikh Sarwar
Computer Science

Brendan O'Connor
Computer Science

- Will **mention violence and death** but nothing graphic.
- Feel free to discretely **leave the room** at any time, for any reason.
- Much more context and **nuance** surrounding the social issues than I’ll cover in today’s lecture. Feel free to come chat!

Andy Halterman
Political Science

1. **Q:** Does variation in party control affect whether state actors (e.g. police) fail to intervene during communal violence?

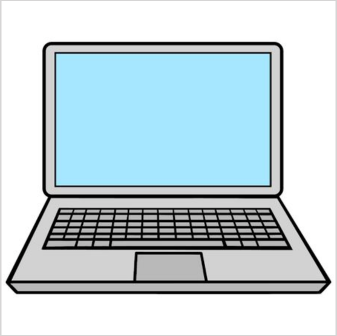
Case Study: Violence in Gujarat, India 2002

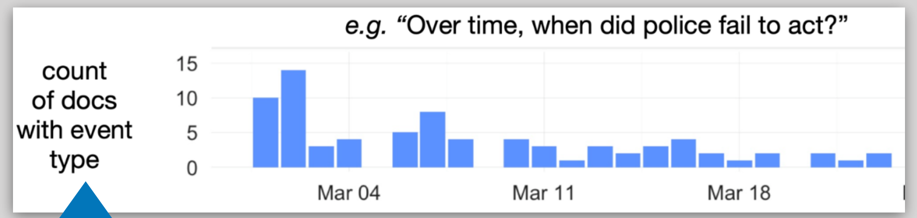
Train fire kills Hindu Pilgrims, Feb. 27, 2002
Photo Credit: New York Times

2. **Challenges**
   - No official records.
   - Only news articles
   - Reading documents manually is costly.

Many events of interest: failure to act, killing, other violence

3. **Use NLP to** automate extracting events

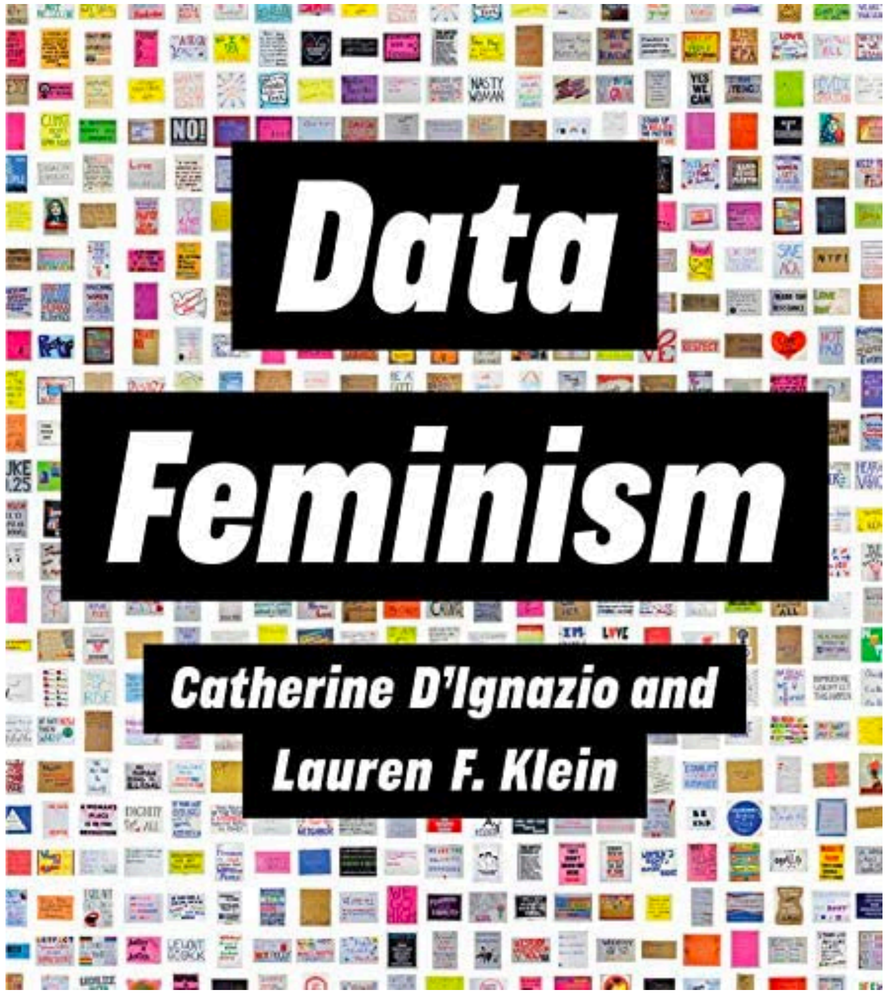




Media bias outside the scope of this talk

Counterdata is the grassroots collection of missing datasets

Structured datasets necessary for further data-driven analysis and policy proposals



7 Principles of Data Feminism

Examine power
Challenge power
Rethink binaries and hierarchies
Elevate emotion and embodiment
Embrace pluralism
Consider context
Make labor visible

Events

Who did what to whom?

Police killed [PERSON].

Even simple event types present challenges

Police killed PERSON.

Police officers spotted the butt of a handgun in **Alton Sterling**'s front pocket and saw him reach for the weapon before **opening fire**, according to a Baton Rouge Police Department search warrant filed Monday that offers the first police account of the events leading up to **his fatal shooting**.

Keith et al. *Identifying civilians killed by police with distantly supervised entity-event extraction. EMNLP, 2017.*

Even simple event types present challenges

Police killed PERSON.

long-range dependencies

Police officers spotted the butt of a handgun in **Alton Sterling**’s front pocket and saw him reach for the weapon before **opening fire**, according to a Baton Rouge Police Department search warrant filed Monday that offers the first police account of the events leading up to **his fatal shooting**.

Even simple event types present challenges

Police killed PERSON.

long-range dependencies

Police officers spotted the butt of a handgun in **Alton Sterling**'s front pocket and saw him reach for the weapon before **opening fire**, according to a Baton Rouge Police Department search warrant filed Monday that offers the first police account of the events leading up to **his** **fatal shooting**.

coreference

Even simple event types present challenges

Police killed PERSON.

long-range dependencies

Police officers spotted the butt of a handgun in **Alton Sterling**'s front pocket and saw him reach for the weapon before **opening fire**, according to a Baton Rouge Police Department search warrant filed Monday that offers the first police account of the events leading up to **his fatal shooting**.

coreference

event coreference

Events

Who did what to whom?

Hovy et al. Events are Not Simple: Identity, Non-Identity, and Quasi-Identity. Workshop on EVENTS, 2013.

Abend and Rappport. The State of the Art in Semantic Representation. ACL, 2017.

Automated event extraction has a large academic literature…

in the social sciences

Schrodt et al., 1994; King and Lowe, 2003; Hanna, 2014; Hammond and Weidmann, 2014; Boschee et al., 2015; Beieler et al., 2016; Osorio and Reyes, 2017

in computer science

Grishman, 1997; McCallum, 2005; Aguilar et al., 2014; Hovy et al., 2013; Levy et al., 2017; Abend and Rappoport, 2017; Grishman, 2019; Liu et al., 2020; Du and Cardie, 2020

Events

Who did what to whom?

Police killed [PERSON].

Deterministic Keyword Matching

Input: sentences

*PERSON was **fatally shot** by **police**.*

***Officers** reported PERSON was **killed** in a car accident.*

Method:

Keyword matching

Agents
officers, police, cops, troopers, deputy, …

&

Event triggers
kill, killing, shoot, shooting, murder, homicide …

Output: Classification

Yes

~~Yes~~

Issue: many false positives *(low precision)*

Approaches to Automated Event Extraction

Deterministic
pattern matching

Methods

Keywords

Rules over syntactic
dependency parses

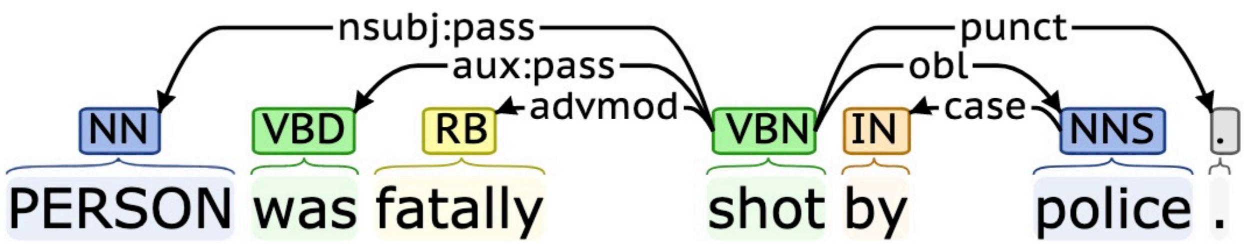
Hard code
domain knowledge

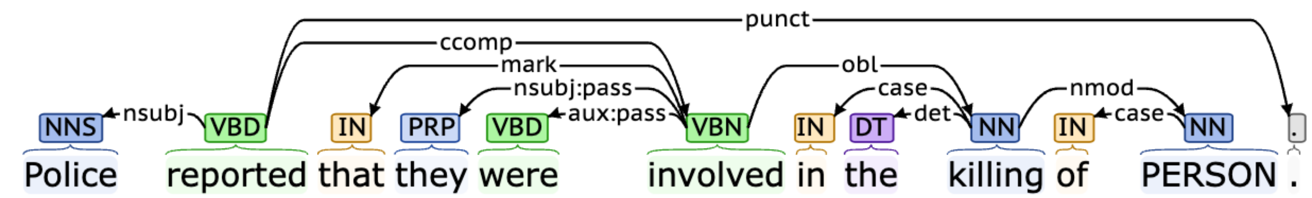
Generalization

Mitchell. The Need for Biases in Learning Generalizations. 1980.

Deterministic Syntax Matching

Input: automatically infer dependency parse trees over sentences





Method: Rules over dependency paths

PERSON <-nsubj:pass <- kill, killing, shoot, shooting, murder, homicide … ->obl -> officers, police, cops, troopers, deputy, …

Output: Classification

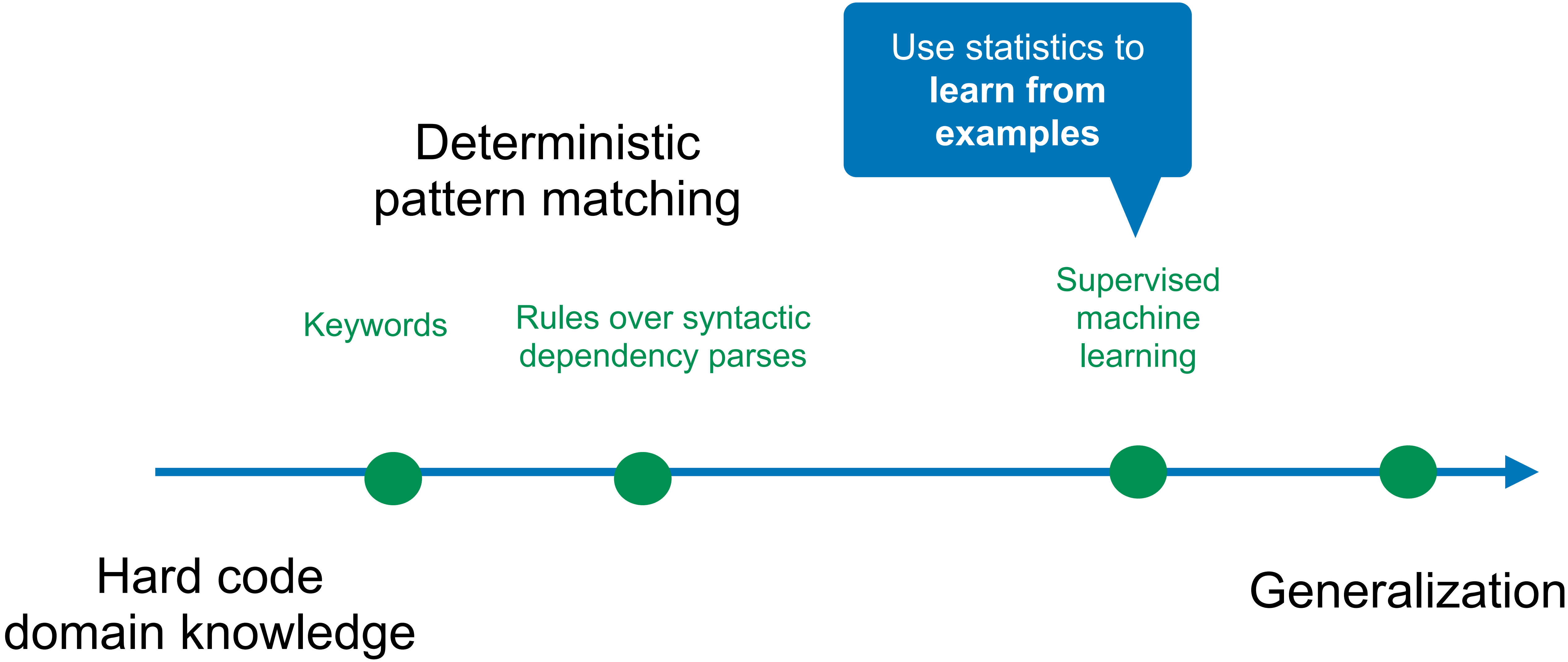
Yes

~~No~~

Issue: Difficult for a domain expert to list all possible rules *(low recall)*

Chen and Manning, EMNLP, 2014; Nivre et al. LREC, 2016; ***Keith*** *et. al, NAACL, 2018*

Approaches to Automated Event Extraction



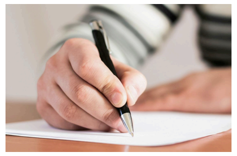
Mitchell. The Need for Biases in Learning Generalizations. 1980.

Supervised Machine Learning

1. **Gather** training data

Police killed PERSON.

x 10,000+

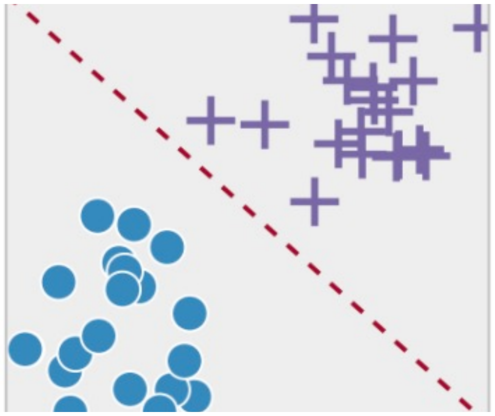


Yes/No

2. **Humans label** training data

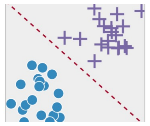
Issue: Costly

3. **Train model:** statistical pattern matching between inputs and labels



4. **Inference:** (generalization) apply trained model on unseen inputs

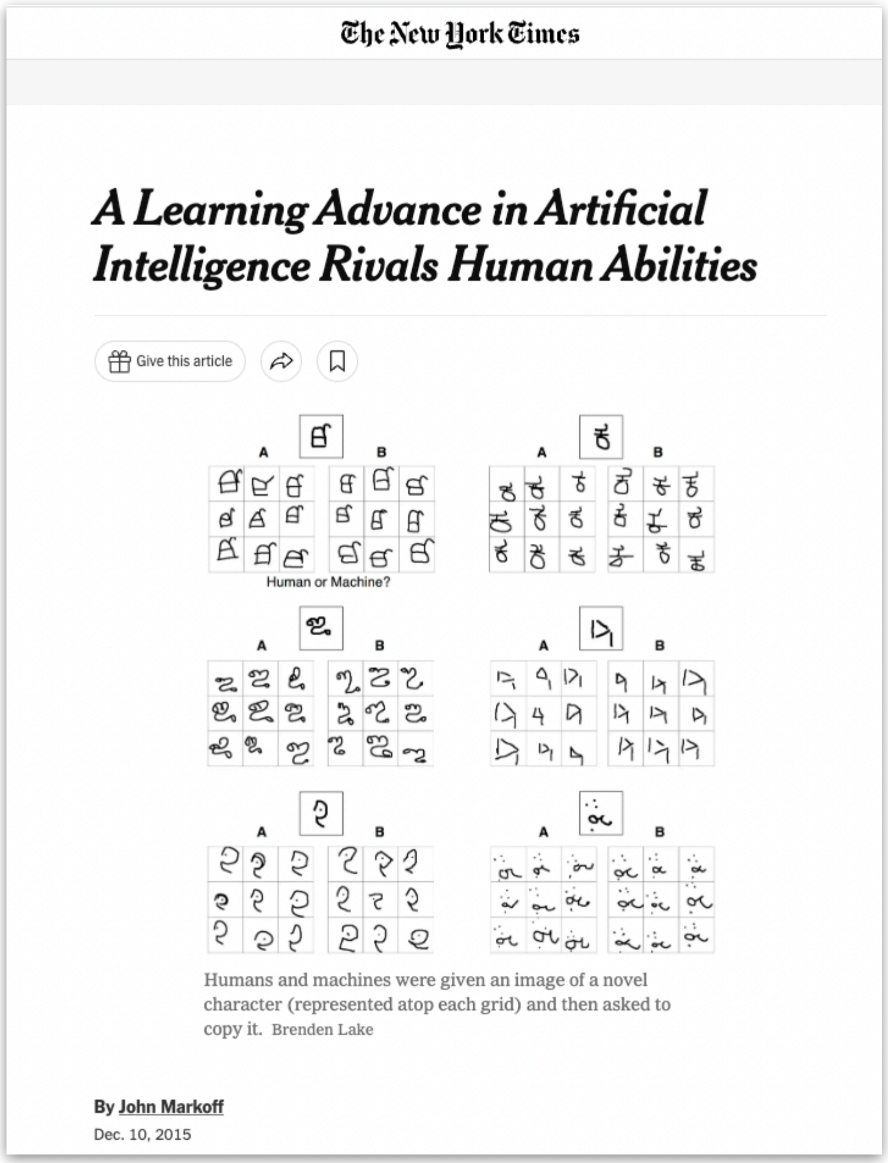
PERSON died in a police homicide.

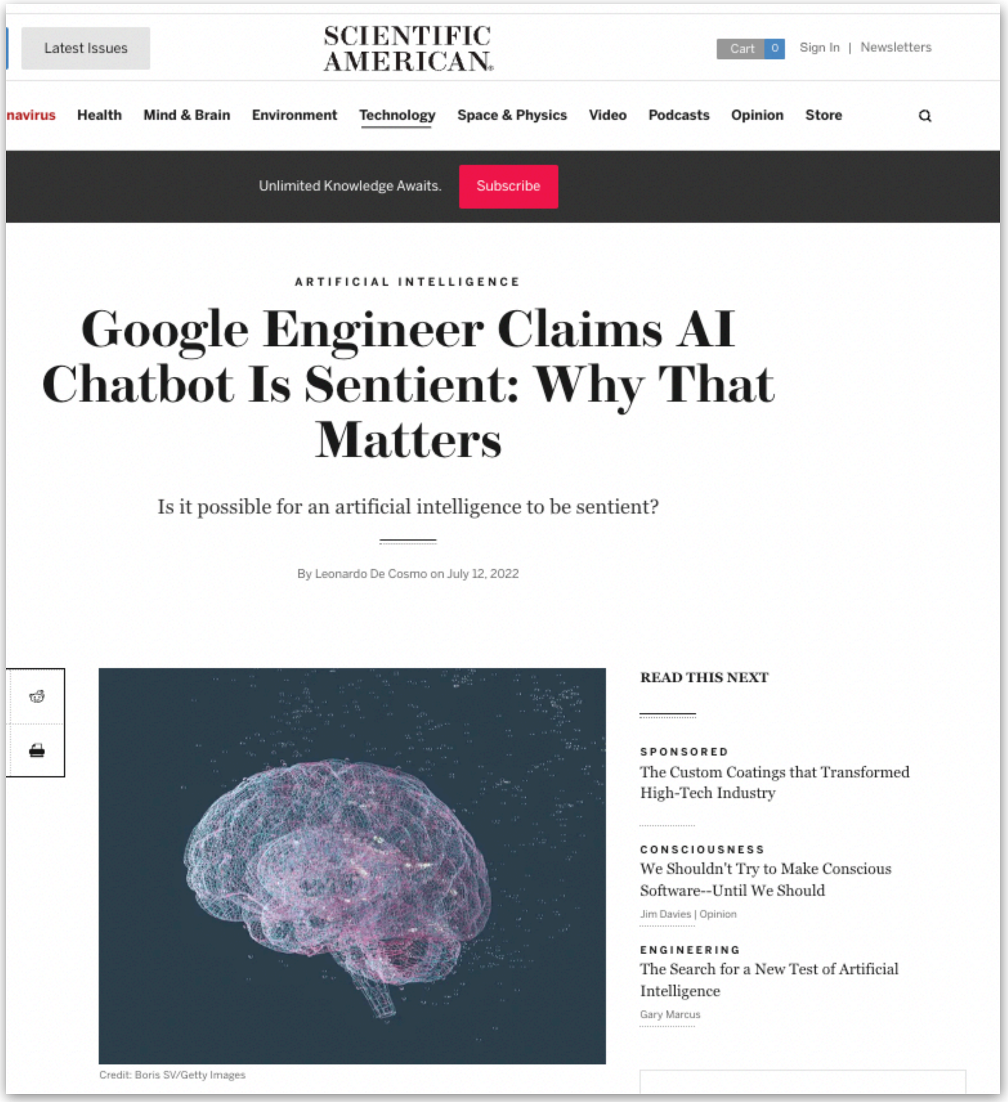


Yes

Questions?

Machine Learning (AI) Hype

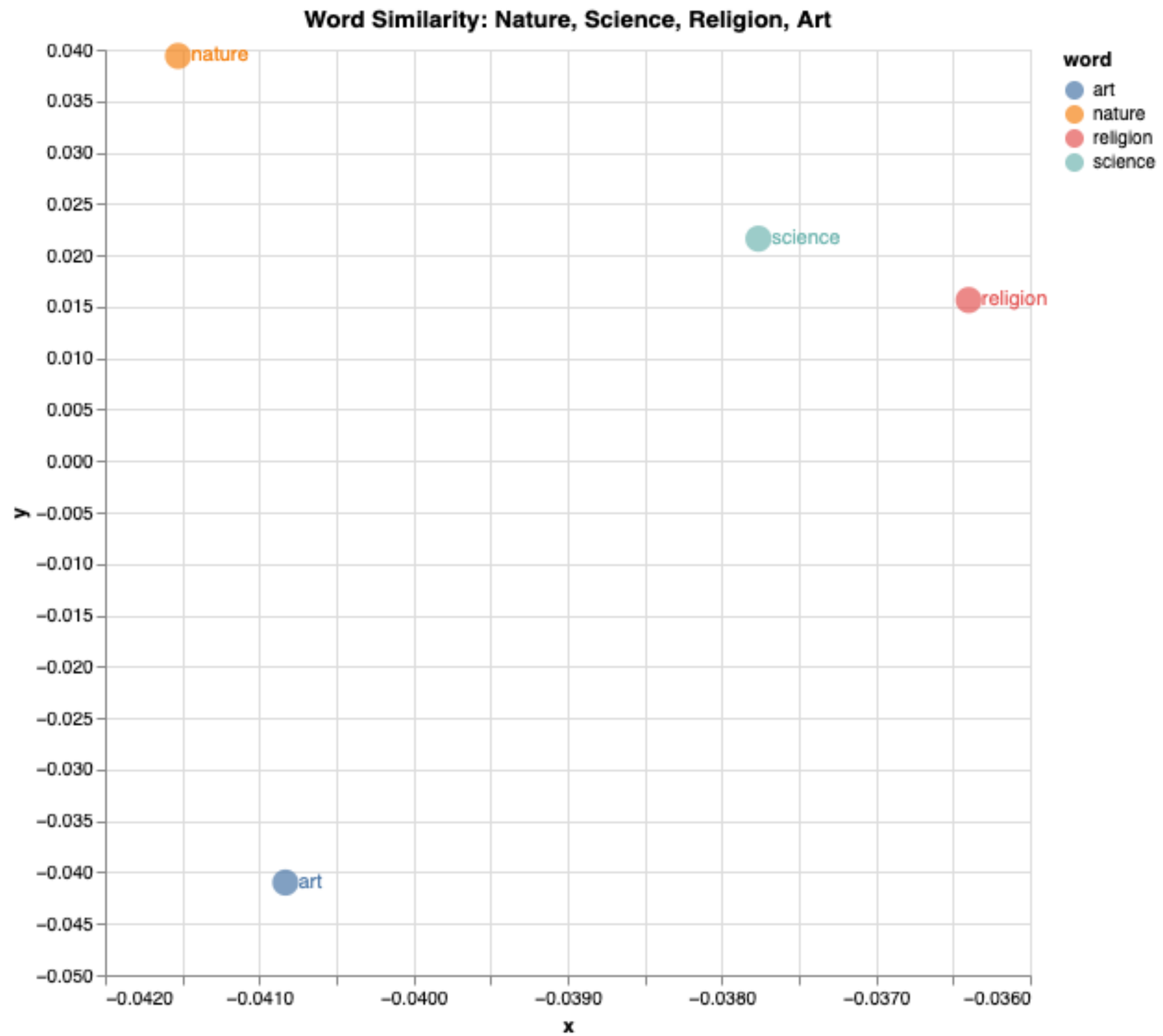




Aside: What’s behind these hyped models?

Goal: Turn words into numbers

Old way:
One per
word type

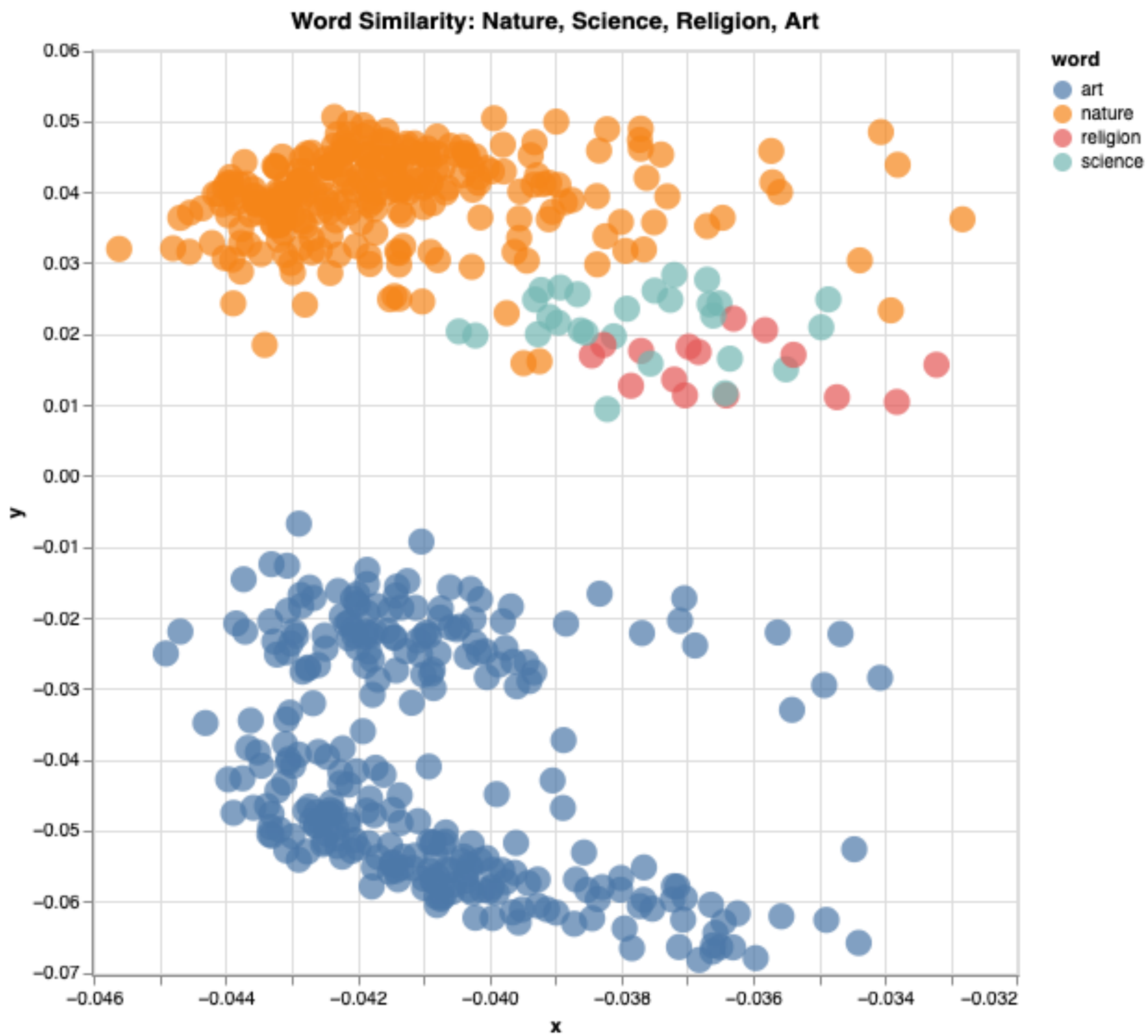


Slide credit: Maria Antoniak

Goal: Turn words into numbers

New way: One per instance of a word in context

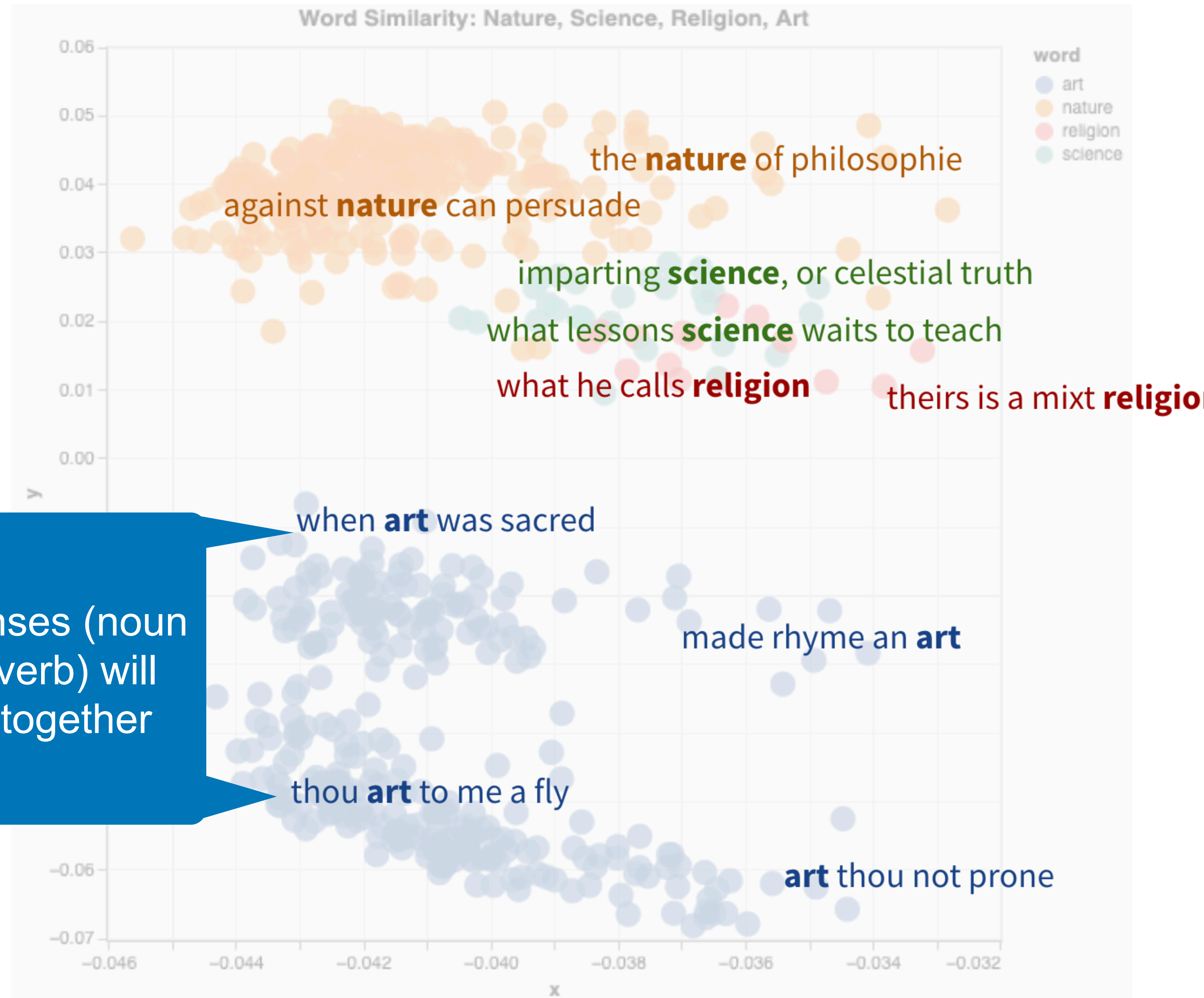
Linguistics: context matters a lot for meaning



Slide credit: Maria Antoniak

Goal: Turn words into numbers

New way: One per instance of a word in context



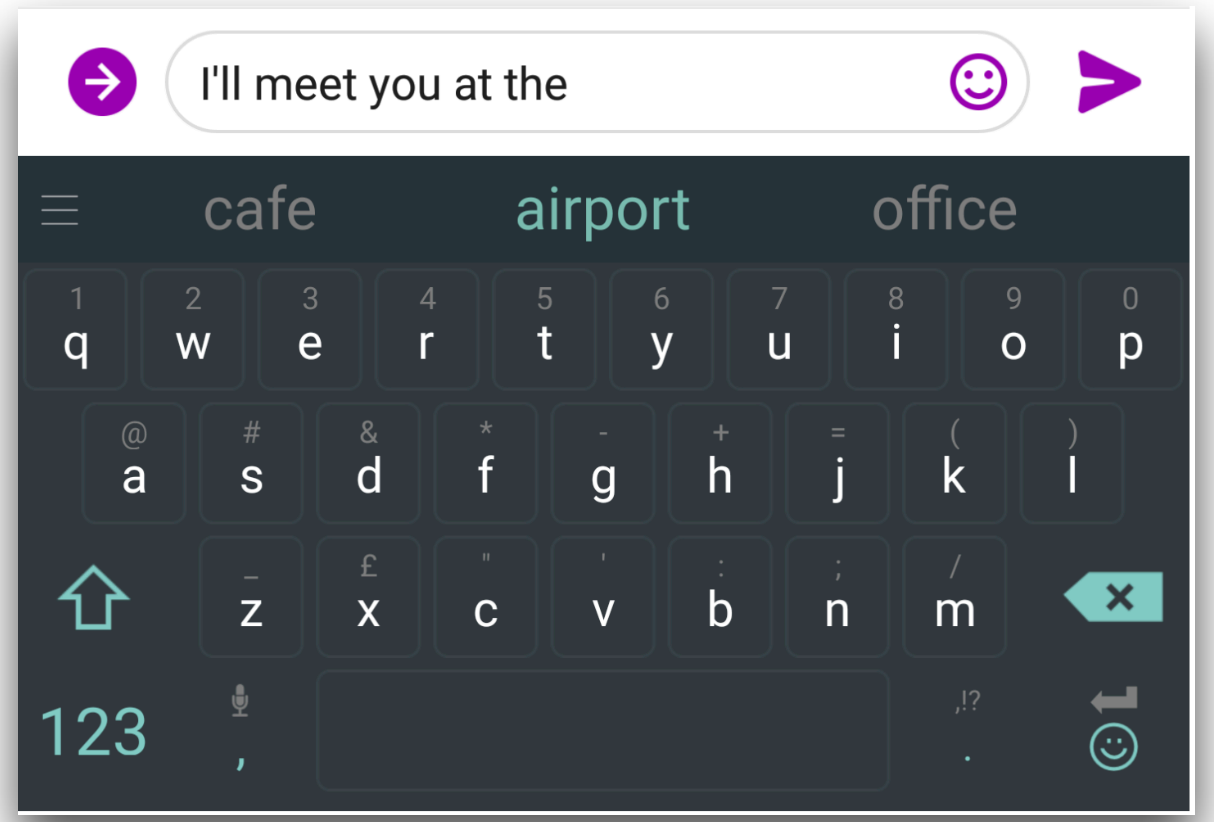
Word senses (noun versus verb) will cluster together

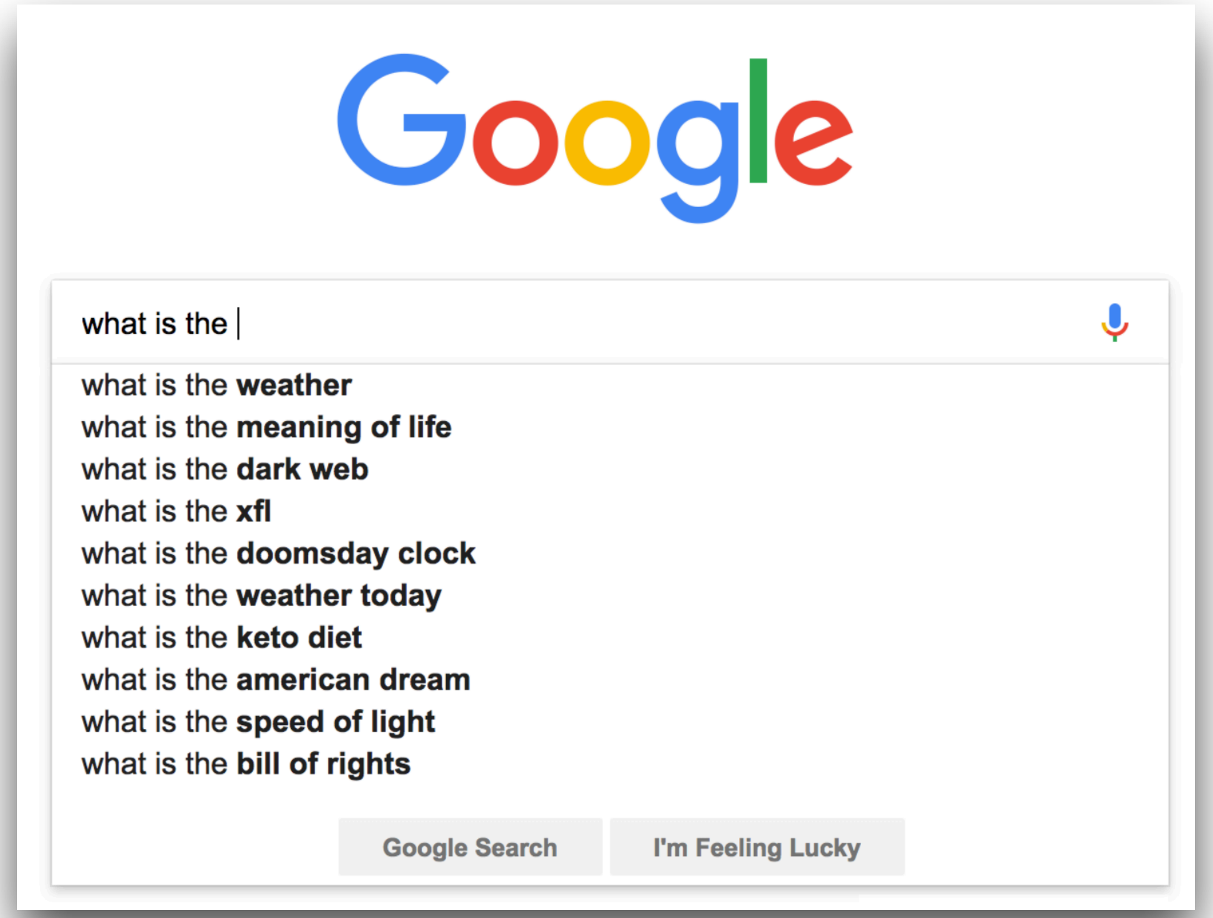
Slide credit: Maria Antoniak

Goal: Turn words into numbers.
How? "Language modeling"

Term has specific meaning in NLP.

You’ve probably seen language modeling before!





Slide credit: Mohit Iyyer

Goal: Turn words into numbers.
How? "Language modeling"

Predict probabilities over each word in the vocabulary

Model

Input: Context words

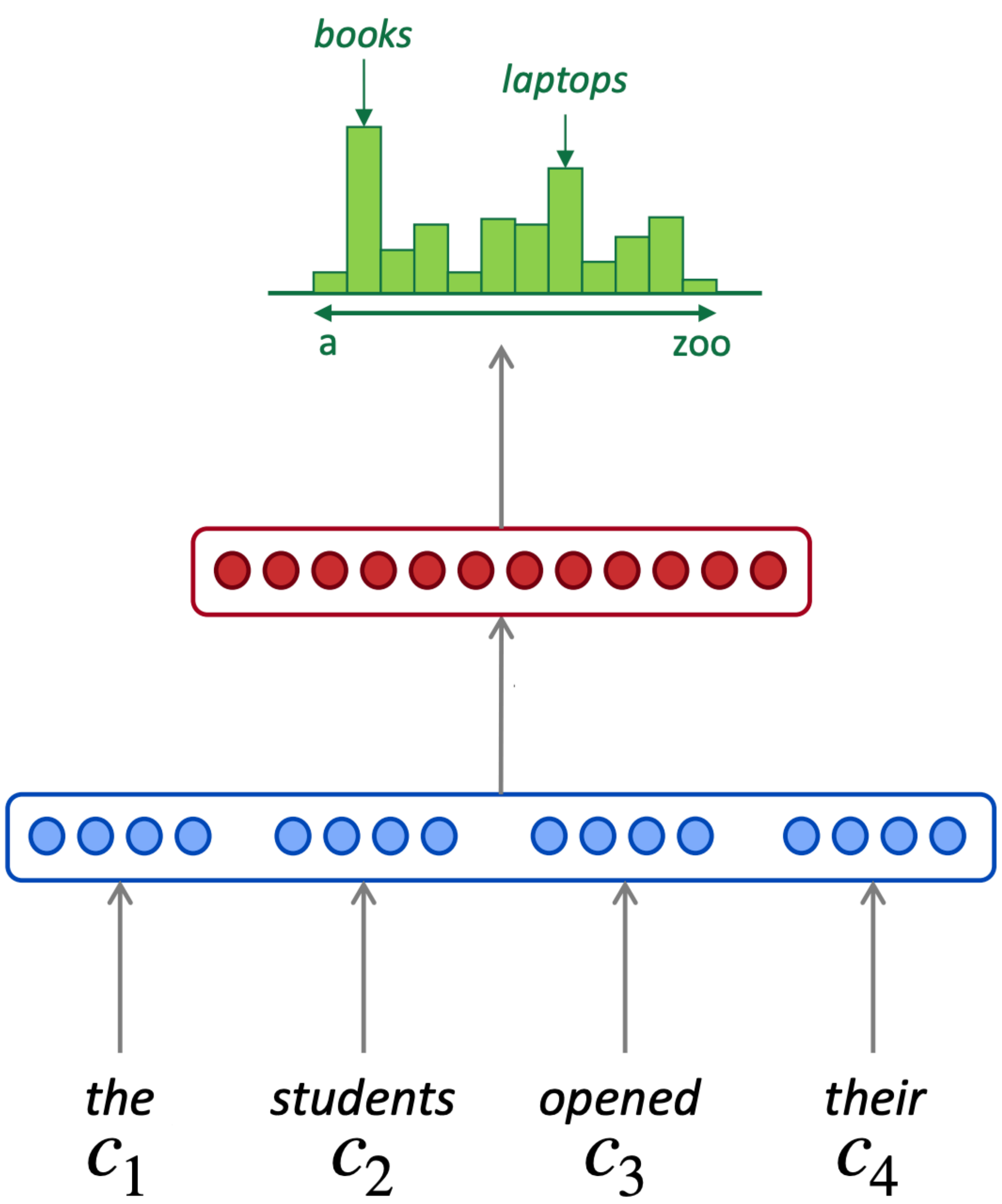
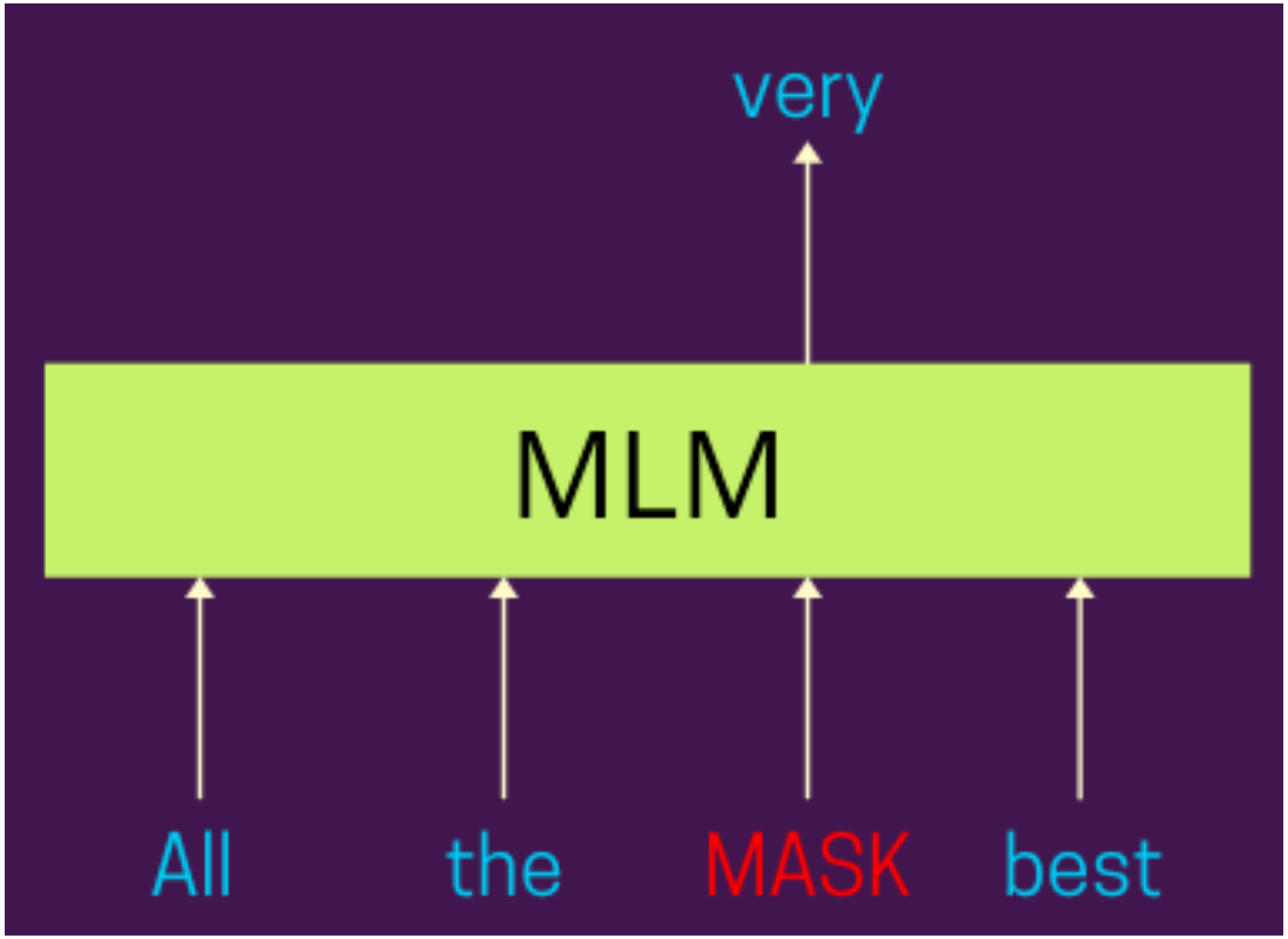


Figure credit: Mohit Iyyer

Loss Function: Masked Language Modeling (MLM)

- Randomly mask out words
- Model predicts masked words given context
- Check if the model is correct and update



Advantage: Don't need humans to create training data! Just gather all text data lying around on the internet…

Figure credit: Prakhar Mishra, [blog]

In reality… the models are really **complicated…**

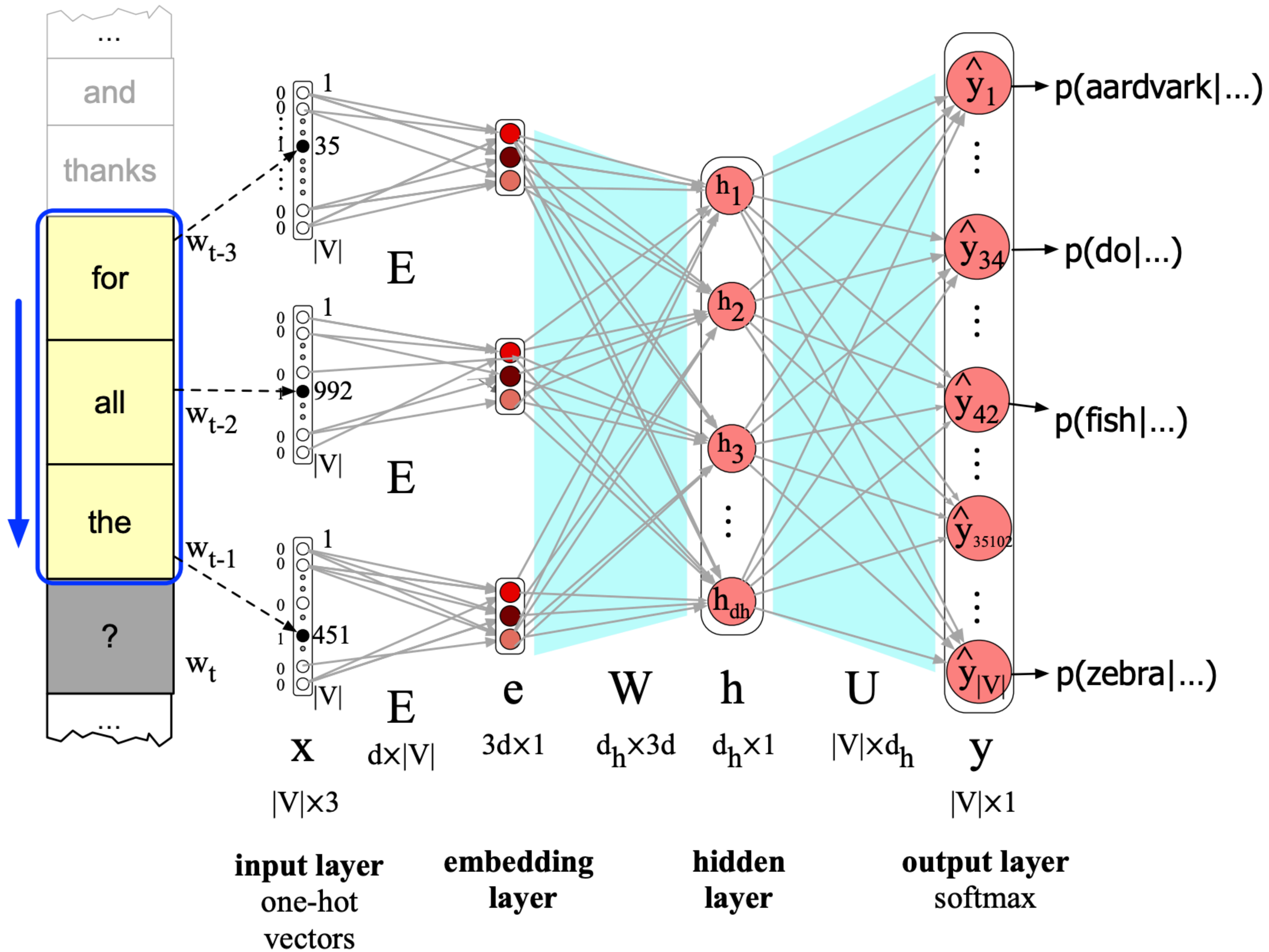


Figure credit: Jurafsky and Martin

In reality… the models are really complicated **and big...**

Linear regression: two parameters (slope and intercept)

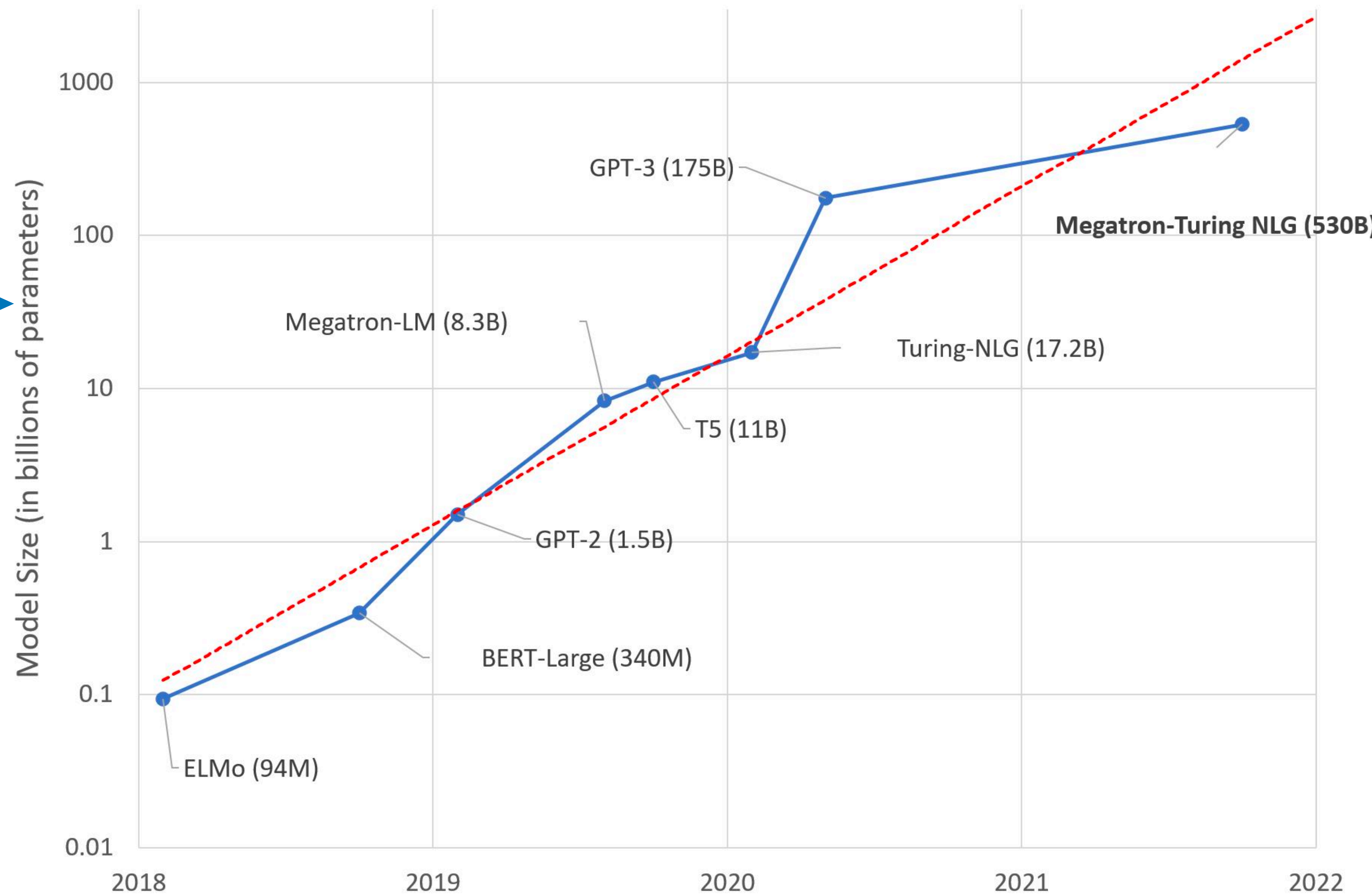


Figure credit: Hugging Face

After pre-training, models applied to new tasks

NLP Task: Community has collectively agreed upon a task definition, inputs and outputs and has collected labeled data

Large language model (LLM)

Task specific “heads”

NLP Task: Natural Language Inference

Humans label examples

Sentence1:
A soccer game with multiple males playing.

Entailment
Neutral
Contradiction

Sentence 2: Some men are playing a sport.

Bowman et al. ACL, 2015

NLP Task: Natural Language Inference

Humans label examples

Sentence1:
A soccer game with multiple males playing.

Entailment
Neutral
Contradiction

Sentence 2: The chicken crossed the road.

Bowman et al. ACL, 2015

NLP Task: Natural Language Inference

Humans label examples

Sentence1:
A soccer game with multiple males playing.

Entailment
Neutral
Contradiction

Sentence 2:
The men did not play soccer.

Train model on tens of thousands of examples

Bowman et al. ACL, 2015

Zero-Shot Transfer Learning

1. **Pre-train** large-scale language model
2. **Fine-tune** on a task with labeled data
3. Apply trained model **zero-shot** to our dataset

Apply trained model **zero-shot** to our dataset

Prompt

Police killed someone.

Sentence from our dataset

Yesterday, 97 died in police firing.

Trained model predicts

Entailment

Neutral

Contradiction

Apply trained model **zero-shot** to our dataset

Prompt

Police killed someone.

Sentence from our dataset

Police reported deaths in the area.

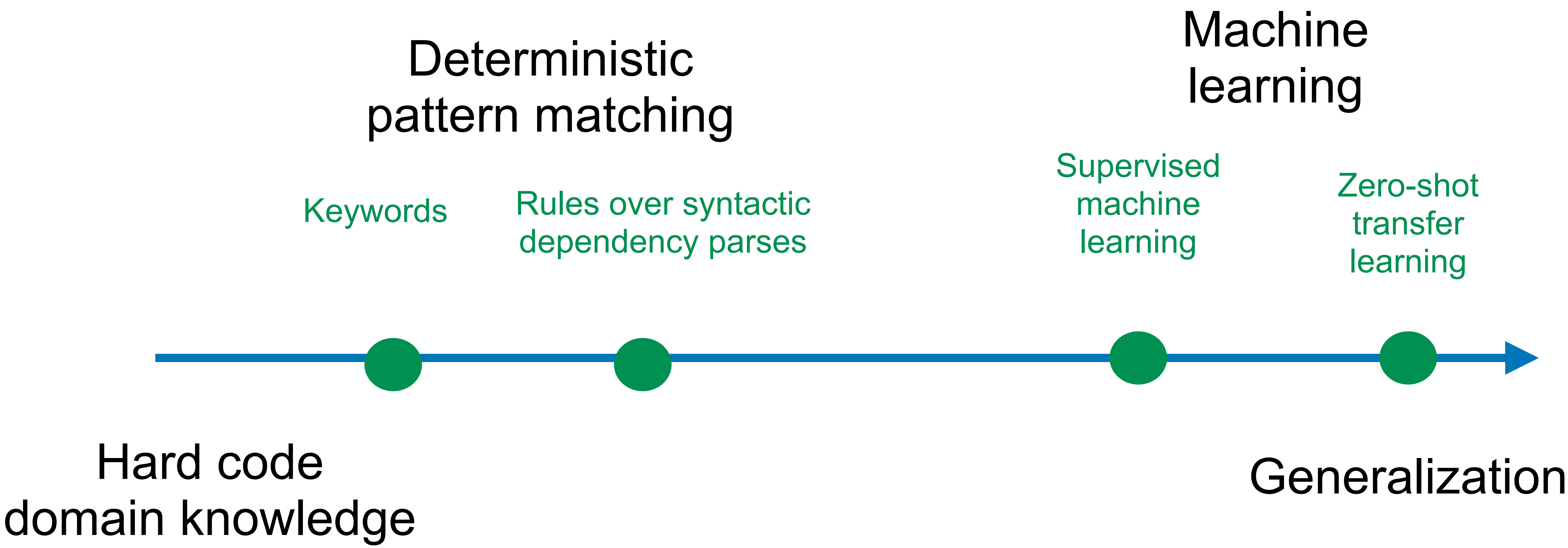
Trained model predicts

Entailment

Neutral

Contradiction

Approaches to Automated Event Extraction



Mitchell. The Need for Biases in Learning Generalizations. 1980.

Questions?

What was our original problem again?

Andy Halterman
Political Science

1. **Q:** Does variation in party control affect whether state actors (e.g. police) fail to intervene during communal violence?

Case Study: Violence in Gujarat, India 2002

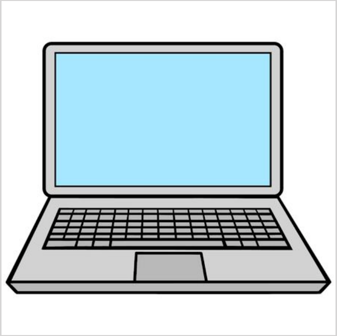
Train fire kills Hindu Pilgrims, Feb. 27, 2002
Photo Credit: New York Times

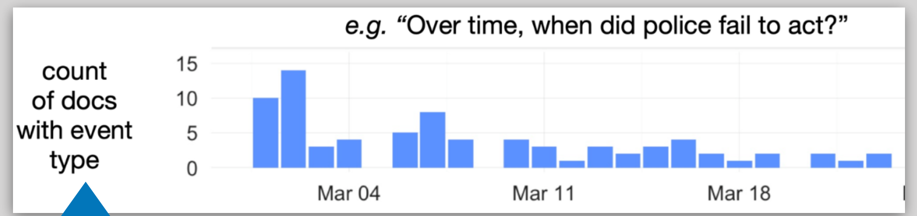
2. **Challenges**
- No official records.
- Only news articles
- Reading documents manually is costly.

The Times of India
BIRTH OF INDIA'S FREEDOM
NATION WAKES TO NEW LIFE
Mr. Nehru Calls For Big Effort From People
"INCESSANT STRIVING TASK OF FUTURE"
FRENZIED ENTHUSIASM IN BOMBAY
Crowds In Festive Mood
LORD MOUNTBATTEN GREETS PAKISTAN
Mr. Jinnah Re-Affirms Firm Friendship With Britain

Many events of interest: failure to act, killing, other violence

3. **Use NLP to** automate extracting events





Media bias outside the scope of this talk

Novel dataset created for empirical evaluation


The Times of India

BIRTH OF INDIA'S FREEDOM

NATION WAKES TO NEW LIFE

Mr. Nehru Calls For Big Effort From People

"INCESSANT STRIVING TASK OF FUTURE"

Assembly Members Take Solemn Pledge

NEW CABINET OF INDIA

WILD SCENES OF JUBILATION IN DELHI

STATE VISIT TO KARACHI

LORD MOUNTBATTEN GREETS PAKISTAN

Mr. Jinnah Re-Affirms Firm Friendship With Britain

FRENZIED ENTHUSIASM IN BOMBAY

Crowds In Festive Mood

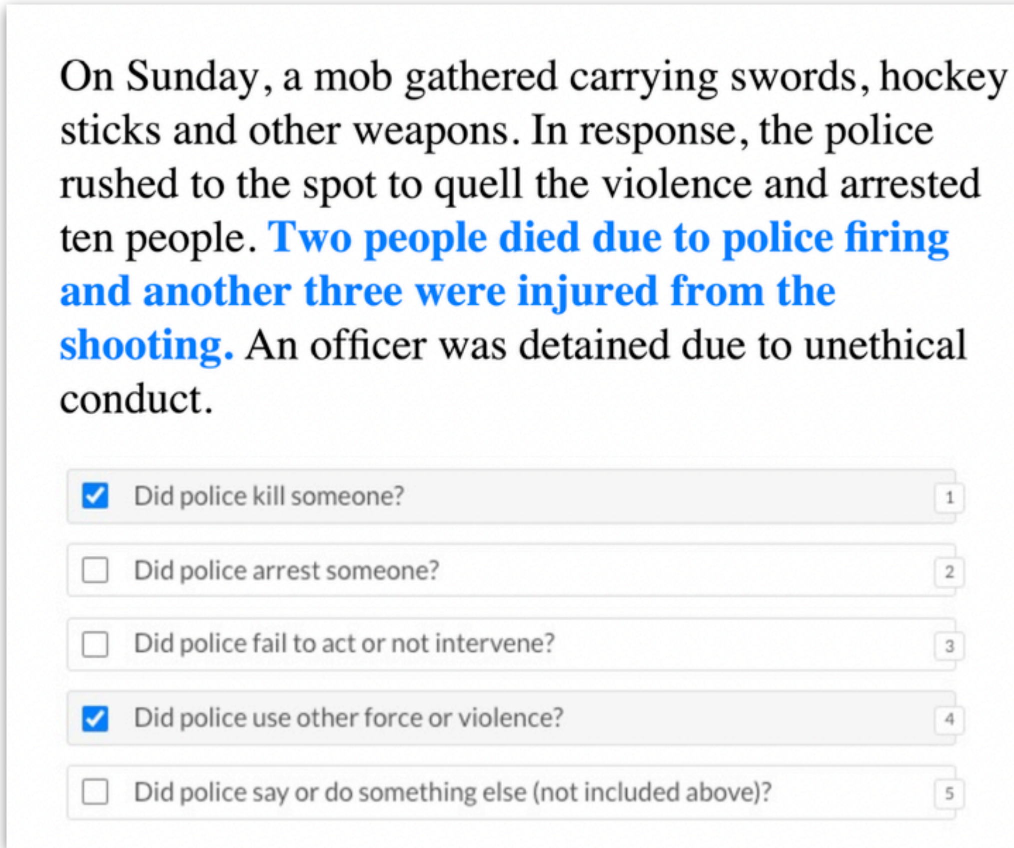
"MAY BOMBAY PROSPER"

Governor's Message

GOOD WISHES TO FREE INDIA

- *Times of India*
- Filter to March 2002 and "Ayodha" OR "Gujarat"
- **Results in 1,257 articles, 21,391 sentences**
- Every sentence annotated with 2 annotators + adjudication round

Annotation interface



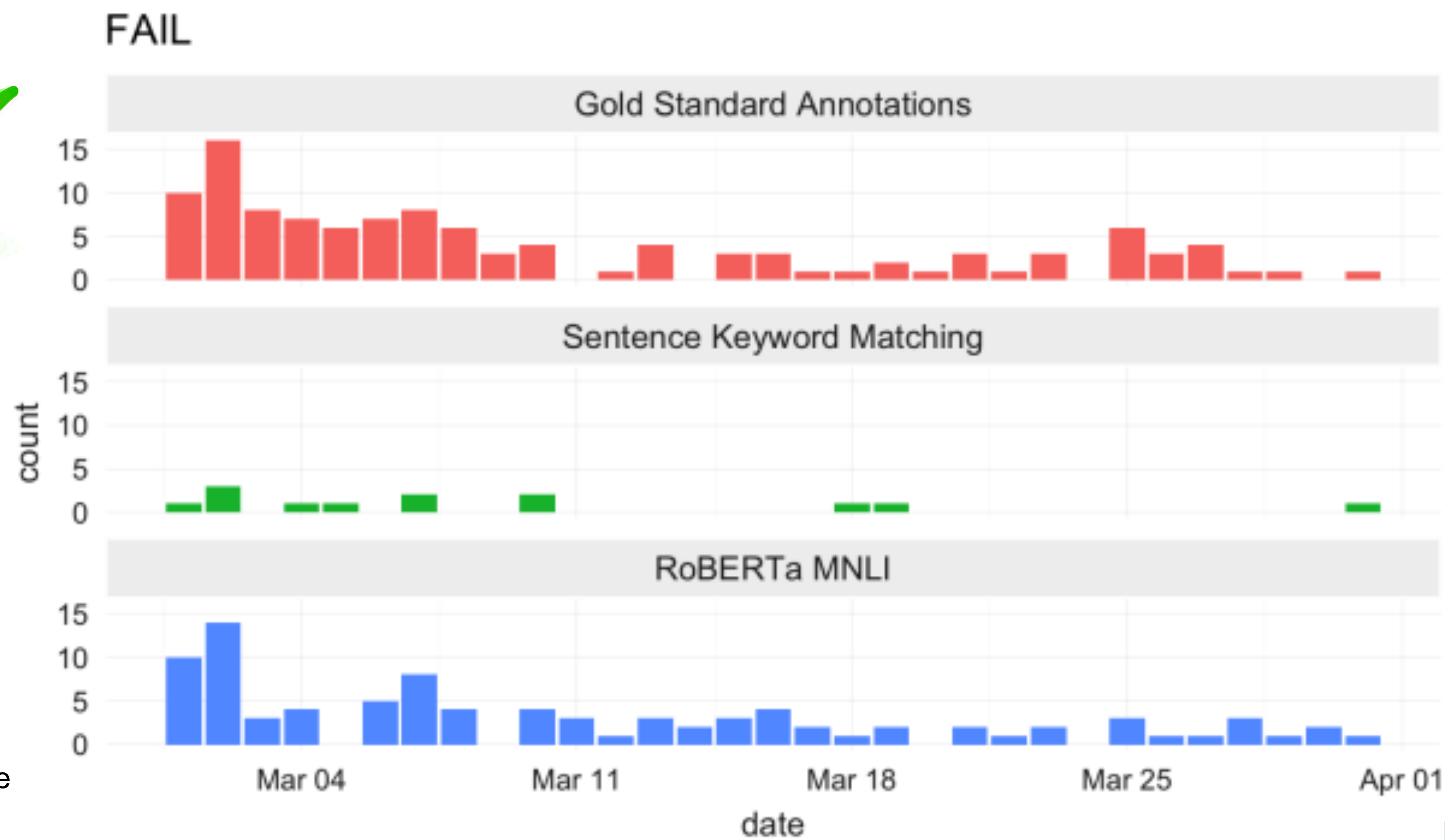
Dataset publicly available

https://github.com/slanglab/IndiaPoliceEvents

Evaluation highlights

Humans

Zero-shot language model



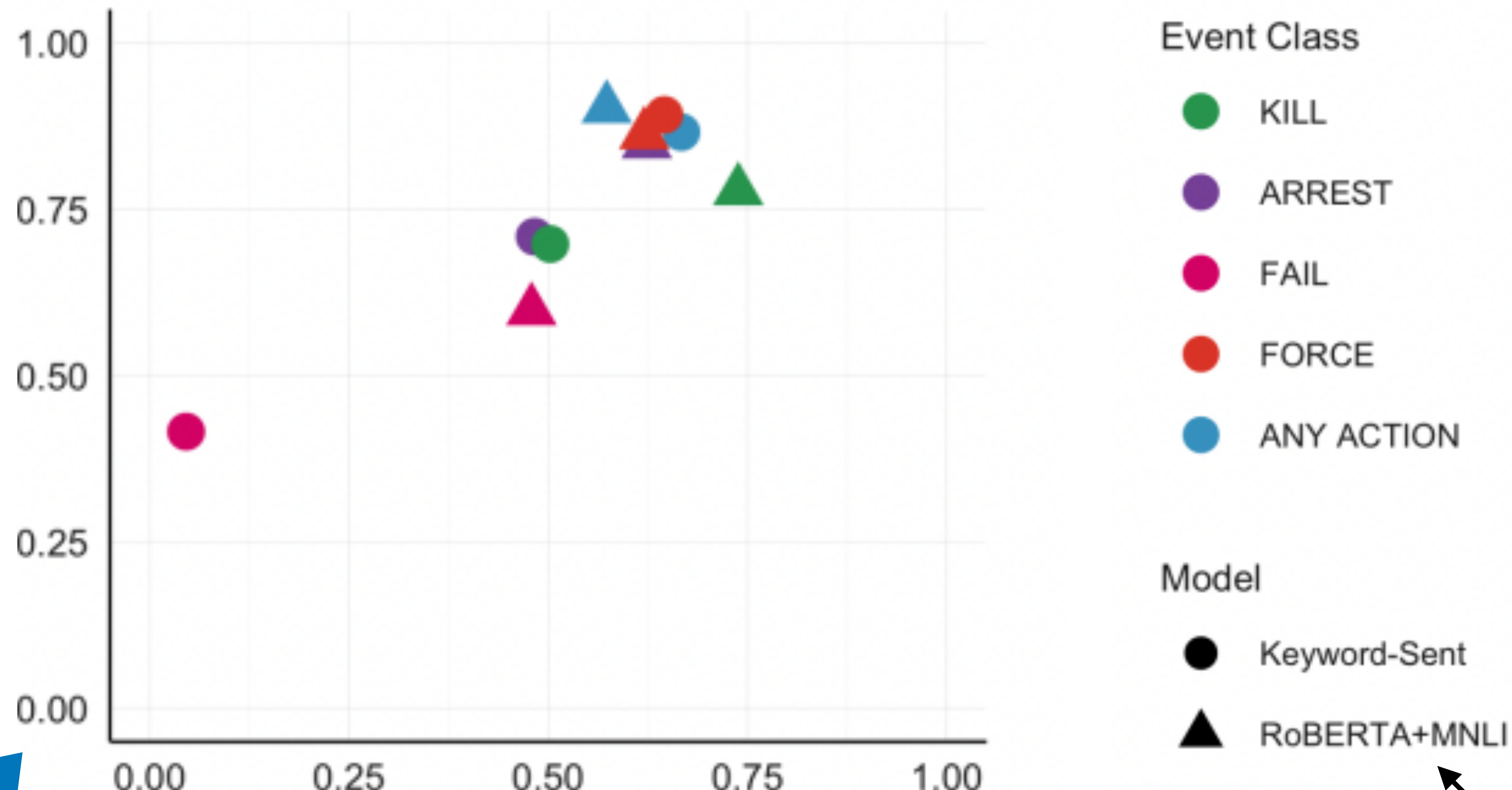
correlation = 0.42

correlation = 0.6

Takeaway: Hype doesn't quite hold up! More work to do!

Evaluation highlights

Temporal aggregates: correlation between **human gold-standard** and model

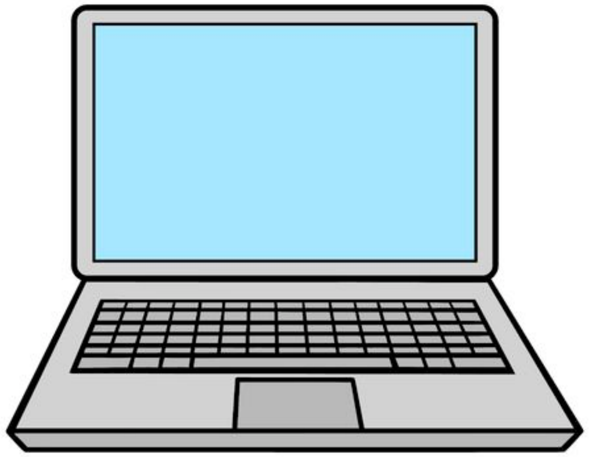


Sentence-level model F1

Zero-shot language model

Encouraging results: Focusing on sentence-level models will probably help with the social-science goals.

Manual error analysis



“ […] scores of people have been killed in rural Gujarat **due to police failure** to intervene…”

- **Negative** instances of police killing events
- **Model** assigns with high probability to the **positive class**

“Police **said** that two persons had been killed […]”

Paths forward?

- More examples with this specific linguistic phenomena
- Hybrid systems
- Human-in-the-loop

Please read our paper for more details!

Corpus-Level Evaluation for Event QA:
The IndiaPoliceEvents Corpus Covering the 2002 Gujarat Violence

Andrew Halterman*
Massachusetts Institute of Technology
ahalt@mit.edu

Katherine A. Keith*
University of Massachusetts Amherst
kkeith@cs.umass.edu

Sheikh Muhammad Sarwar*
University of Massachusetts Amherst
smsarwar@cs.umass.edu

Brendan O'Connor
University of Massachusetts Amherst
brenocon@cs.umass.edu

Abstract

Automated event extraction in social science applications often requires corpus-level evaluations: for example, aggregating text predictions across metadata and unbiased estimates of recall. We combine corpus-level evaluation requirements with a real-world, social science setting and introduce the INDIAPOLICEEVENTS corpus—all 21,391 sentences from 1,257 English-language *Times of India* articles about events in the state of Gujarat during March 2002. Our trained annotators read and label every document for mentions of police activity events, allowing for unbiased recall evaluations. In contrast to other datasets with structured event representations, we gather annotations by posing natural questions, and evaluate off-the-shelf models for three different tasks: sentence classification, document ranking, and temporal aggregation of target events. We present baseline results from zero-shot BERT-based models fine-tuned on natural language inference and passage retrieval tasks. Our novel corpus-level evaluations and annotation approach can guide creation of similar social-science-oriented resources in the future.

Figure 1: Motivation (A-B) and procedures (B-C) for this paper: **A.** Social scientists often use text data to answer substantive questions about temporal aggregates. **B.** To answer these questions, domain experts use natural language to define semantic event classes of interest. **C.** Our INDIAPOLICEEVENTS dataset: Humans annotate *every* sentence in the corpus in order to evaluate whether a system achieves full recall of relevant events. In production, computational models run B's queries to classify or rank sentences or documents, which are aggregated to answer A.

1 Introduction

Understanding the actions taken by political actors is at the heart of political science research: How do actors respond to contested elections (Daxecker et al., 2019)? How many people attend protests (Chenoweth and Lewis, 2013)? Which religious groups are engaged in violence (Brathwaite and Park, 2018)? Why do some governments try to prevent anti-minority riots while others do not (Wilkinson, 2006)? In the absence of official records, social scientists often turn to news data to extract the actions of actors and surrounding events. These news-based event datasets are often constructed by hand, requiring large investments of time and money and limiting the number of researchers who can undertake data collection efforts.

Automated extraction of political events and actors is already prominent in social science (Schrodt et al., 1994; King and Lowe, 2003; Hanna, 2014; Hammond and Weidmann, 2014; Boschee et al., 2015; Beieler et al., 2016; Osorio and Reyes, 2017) and is increasingly promising given recent gains in information extraction (IE), the automatic conversion of unstructured text to structured datasets (Grishman, 1997; McCallum, 2005; Grishman, 2019). While social scientists and IE researchers have over-

* Indicates joint first-authorship.

Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pages 4240–4253
August 1–6, 2021. ©2021 Association for Computational Linguistics

Andy Halterman
Political Science

Katie Keith
Computer Science

Sheikh Sarwar
Computer Science

Brendan O'Connor
Computer Science

Thanks!

Collaborators

Kaggle Data Science Research Grant

kaggle

Bloomberg Data Science PhD Fellowship

Bloomberg