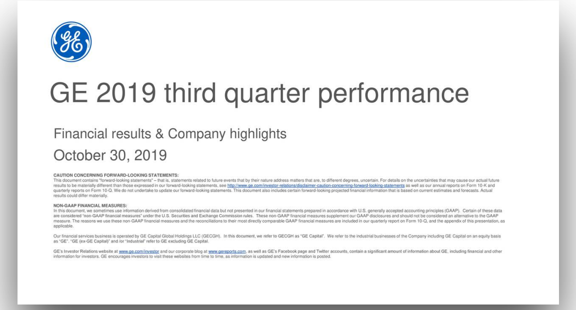
Automated Event Extraction for News-Based Counterdata

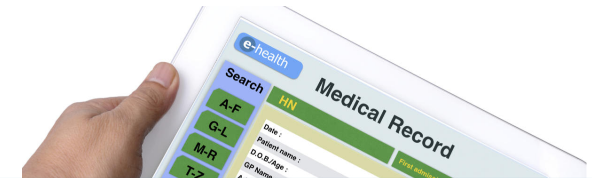
Katie Keith
Williams Statistics Colloquium

October 19, 2022

Age of abundant digitized texts

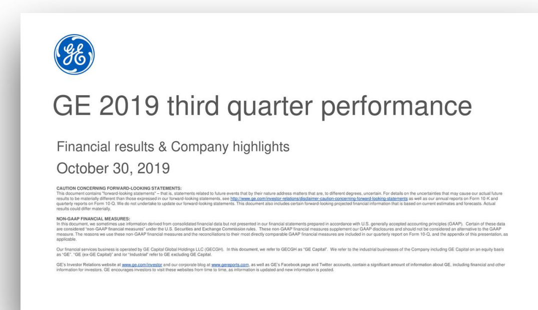


Text data for social sciences questions



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

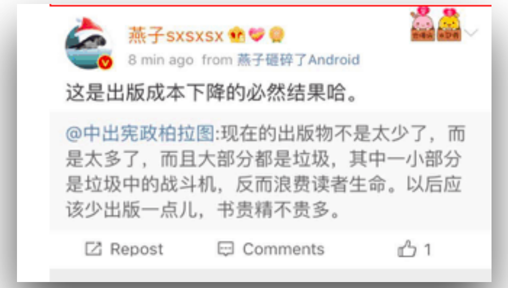
What drives newspapers' political slant?

Gentzkow and Shapiro, Econometrica, 2010

All articles for 400 news outlets

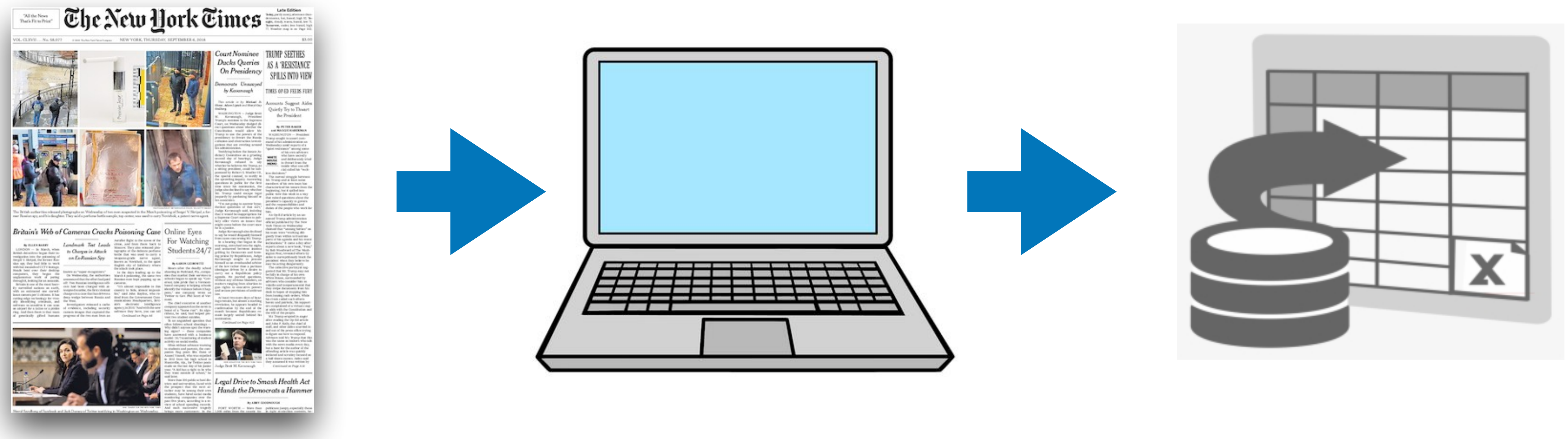
What is the nature of online censorship in China?

King et al., American Political Science Review, 2013

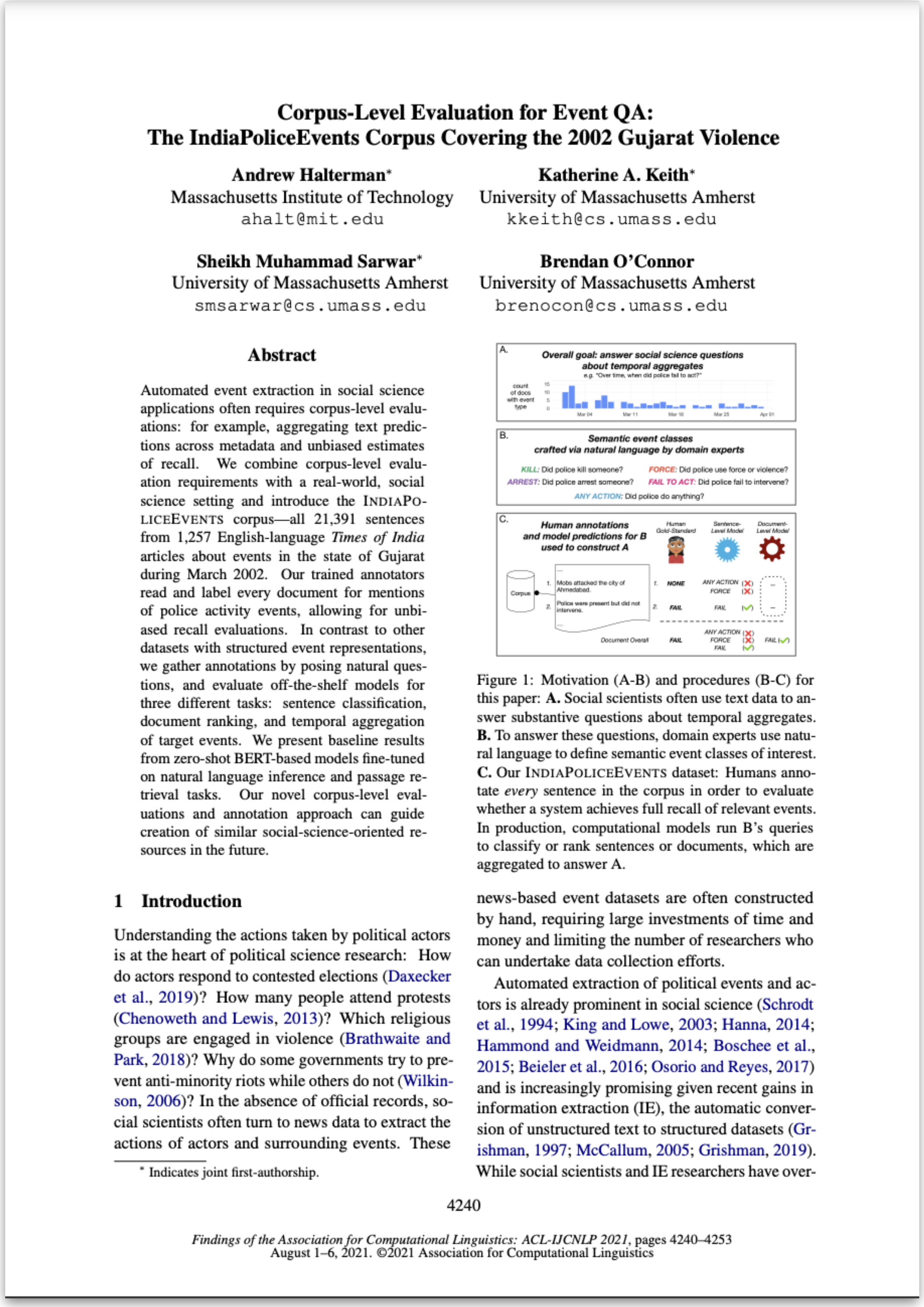


11 million posts

Natural language processing (NLP)



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.

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

Political science-motivated research questions

Andy Halterman
Political Science

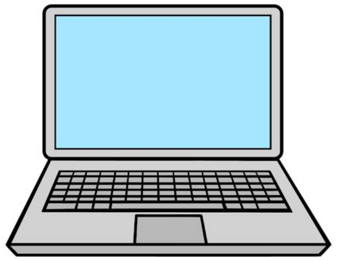
1. **Q:** Does variation in party control of state government affect whether police failed to intervene in communal violence?

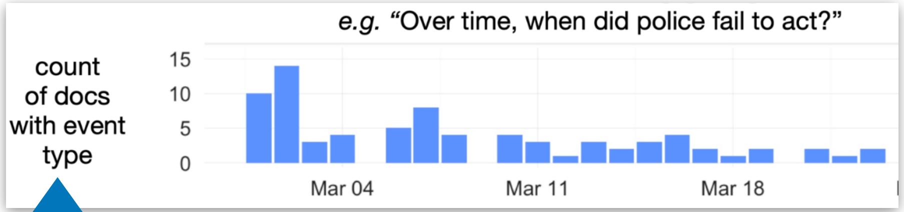
Violence in Gujarat, India 2002

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.

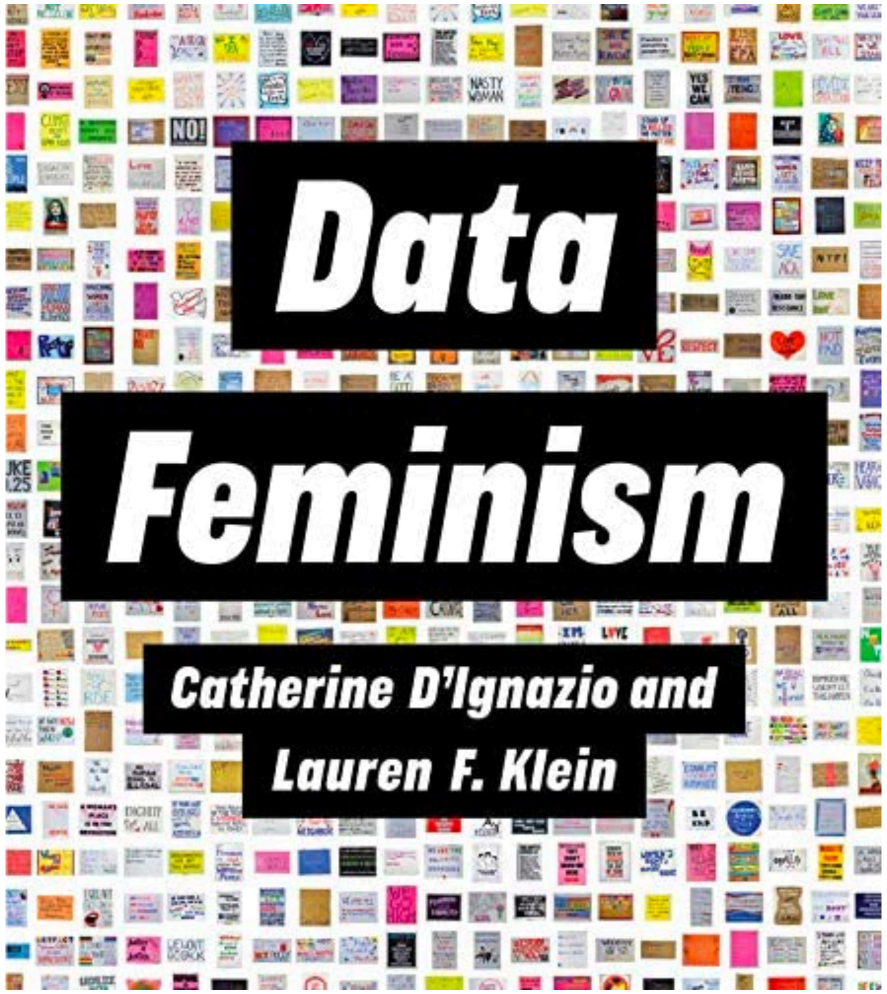
3. **Use NLP to** automate extracting events





Media bias outside the scope of this talk

Counterdata is the grassroots collection of missing datasets



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 Rapport. The State of the Art in Semantic Representation. ACL, 2017.

Automated event extraction has a large academic literature…

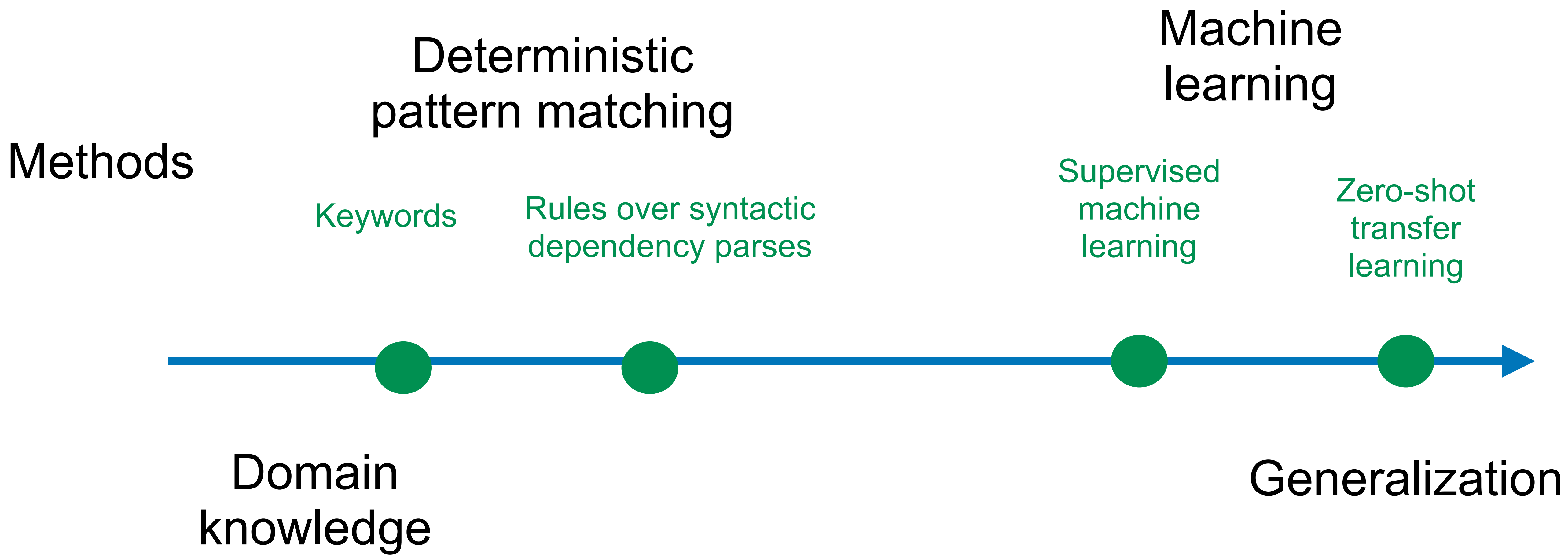
in the social sciences

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

Approaches to Automated Event Extraction



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

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

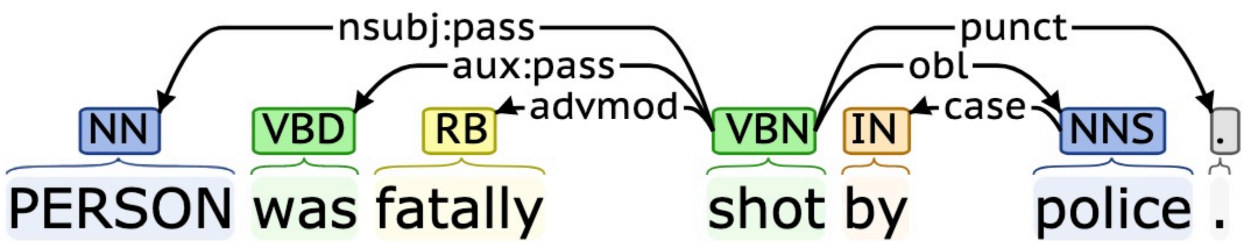
Yes

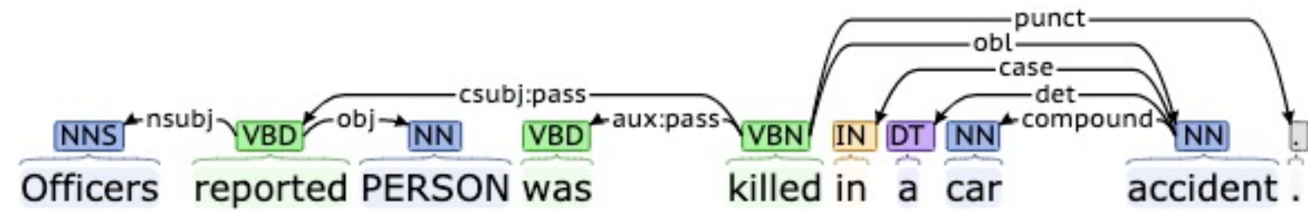
~~Yes~~

Issue: many false positives *(low precision)*

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

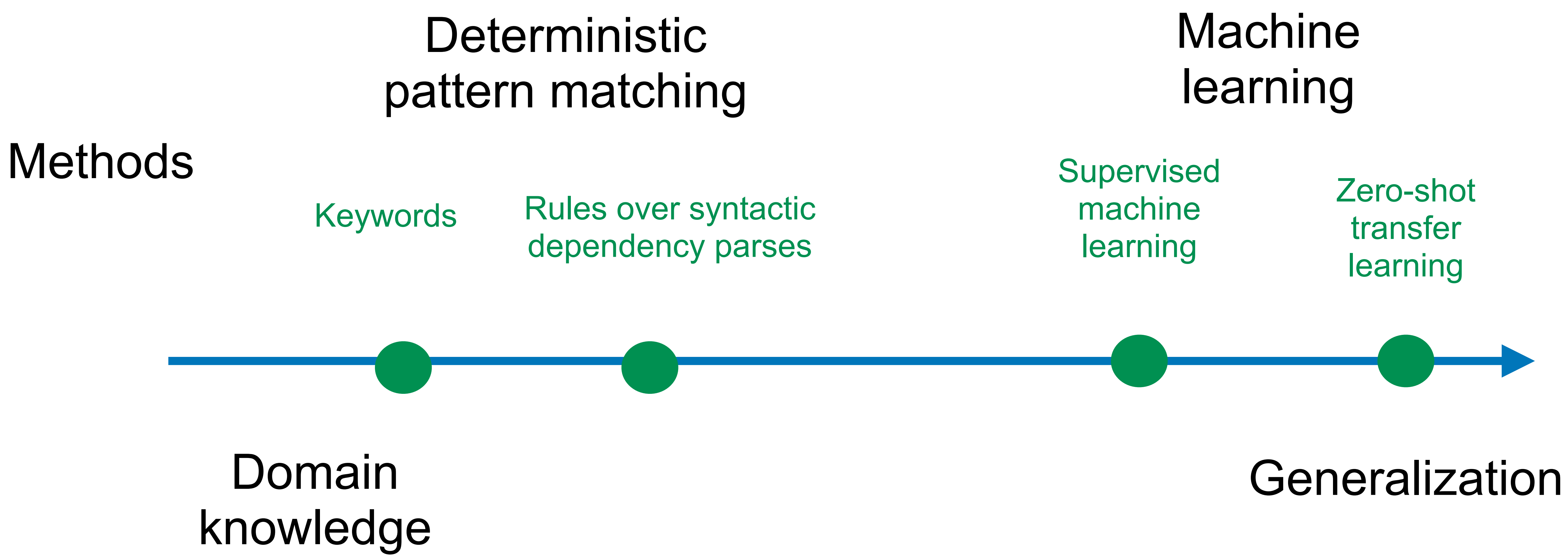
Yes

~~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

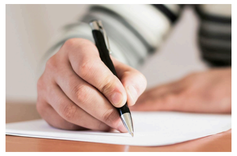


Mitchell. The Need for Biases in Learning Generalizations. 1980.

Supervised Machine Learning

1. **Gather** training data

Police killed PERSON.



Yes/No

2. **Humans label** training data

Issue: Costly

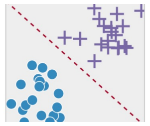
3. **Train model:** statistical pattern matching between inputs and labels

Our 2017 work:

- *logistic regression* with bag of words features
- ***convolutional neural networks*** initialized with pre-trained word embeddings

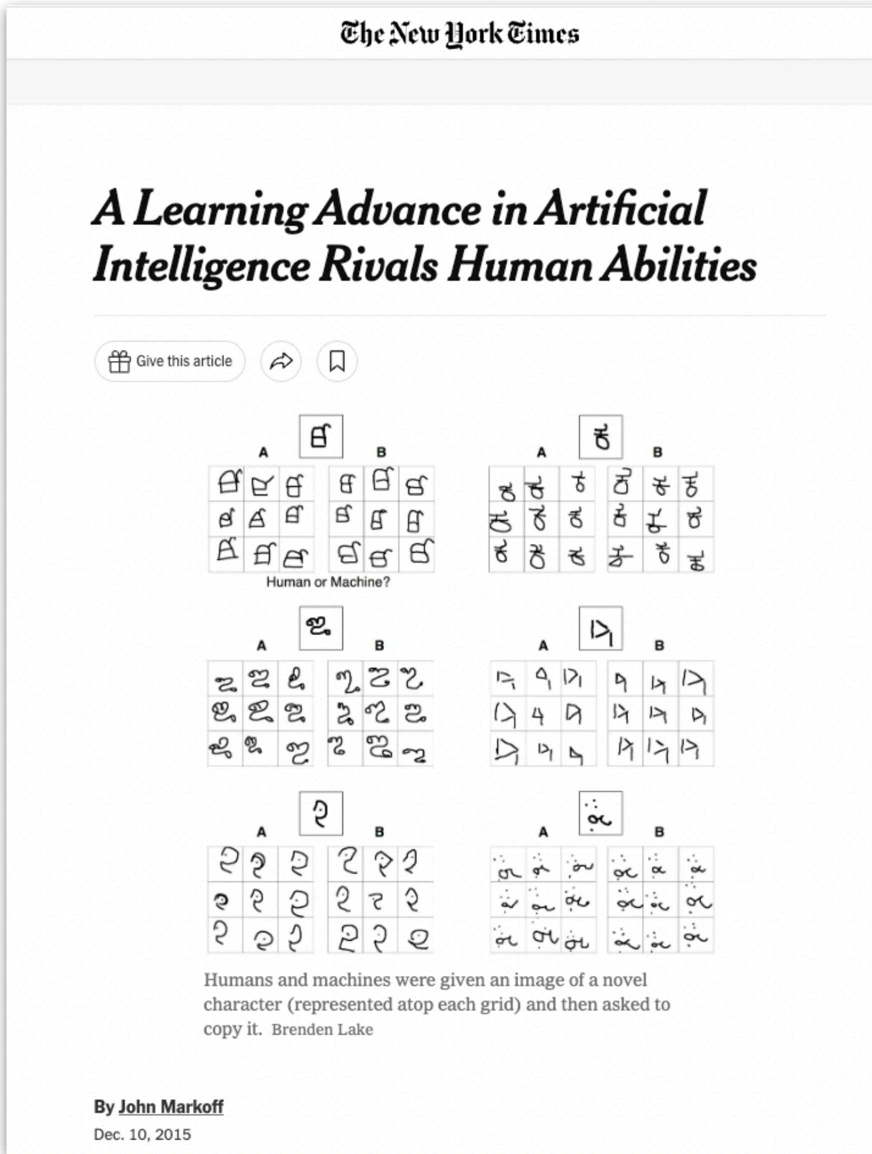
4. **Inference:** (generalization) apply trained model on unseen inputs

PERSON died in a police homicide.



Yes

AI Hype


The New York Times

A Learning Advance in Artificial Intelligence Rivals Human Abilities

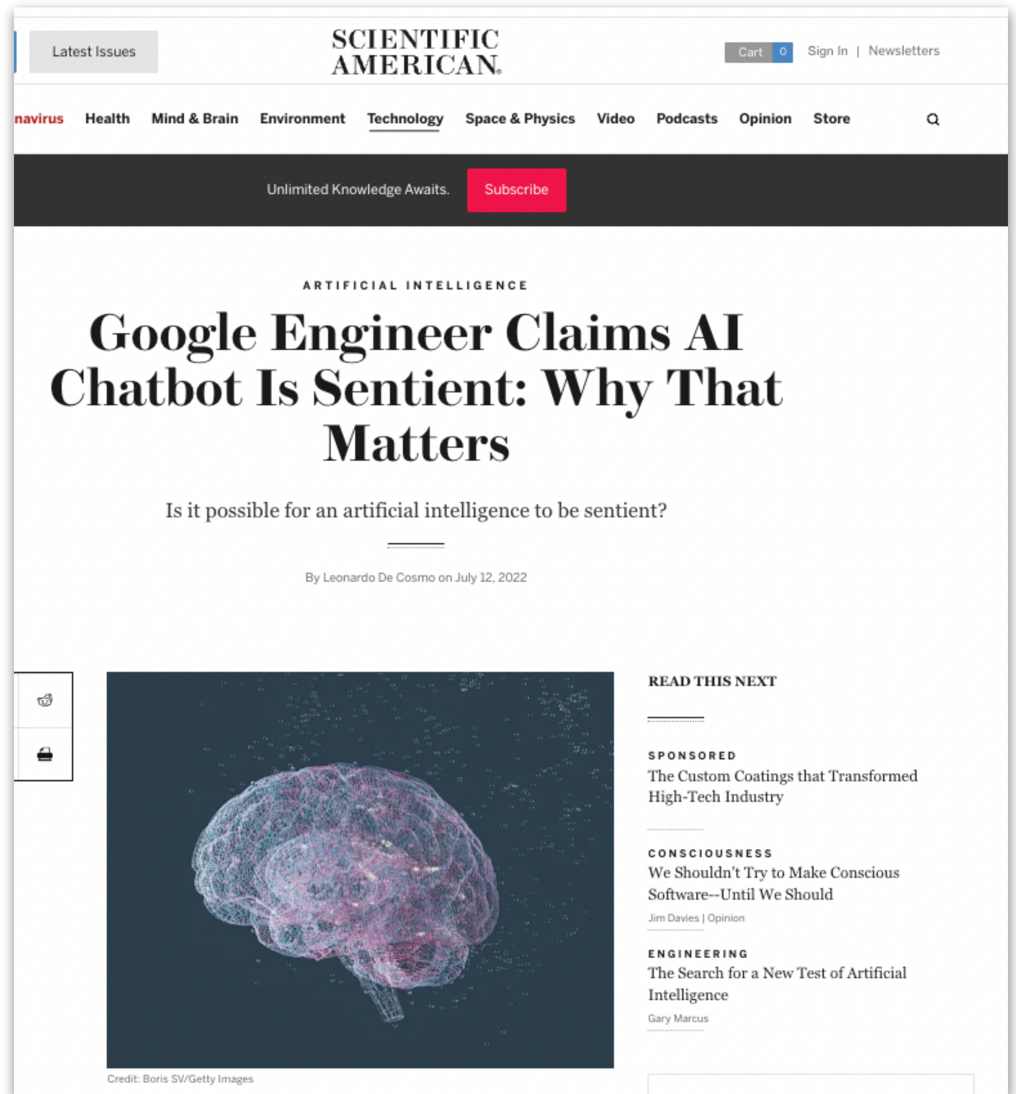
Give this article

Human or Machine?

Humans and machines were given an image of a novel character (represented atop each grid) and then asked to copy it. Brenden Lake

By John Markoff

Dec. 10, 2015


Latest Issues

SCIENTIFIC AMERICAN.

Cart 0 Sign In | Newsletters

navirus Health Mind & Brain Environment Technology Space & Physics Video Podcasts Opinion Store

Unlimited Knowledge Awaits. Subscribe

ARTIFICIAL INTELLIGENCE

Google Engineer Claims AI Chatbot Is Sentient: Why That Matters

Is it possible for an artificial intelligence to be sentient?

By Leonardo De Cosmo on July 12, 2022

Credit: Boris SV/Getty Images

READ THIS NEXT

SPONSORED

The Custom Coatings that Transformed High-Tech Industry

CONSCIOUSNESS

We Shouldn't Try to Make Conscious Software--Until We Should

Jim Davies | Opinion

ENGINEERING

The Search for a New Test of Artificial Intelligence

Gary Marcus

Pre-training with large-scale language models

Huge performance gains in recent years with **large-scale language models** trained on scrapes of the web

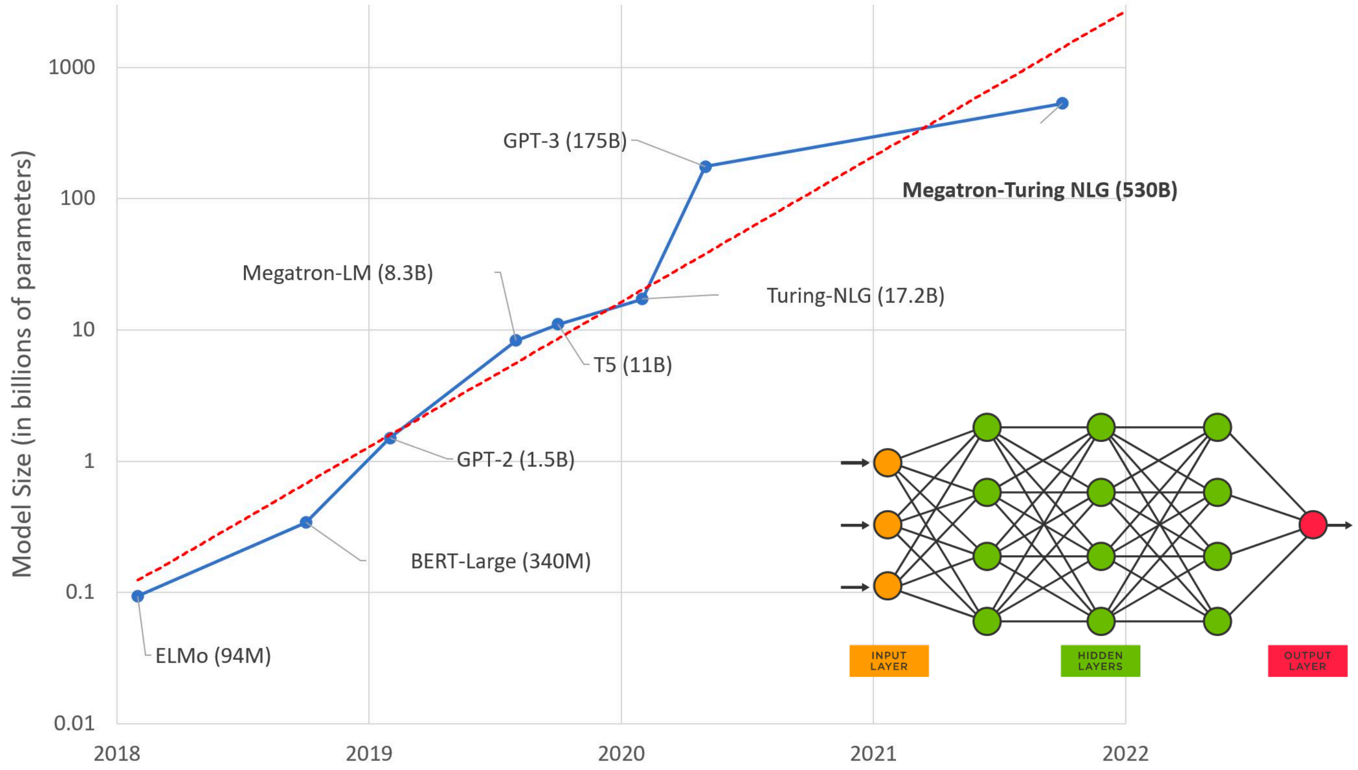


Figure credit: Hugging Face

Self-supervision: Masked language modeling (MLM) objective function

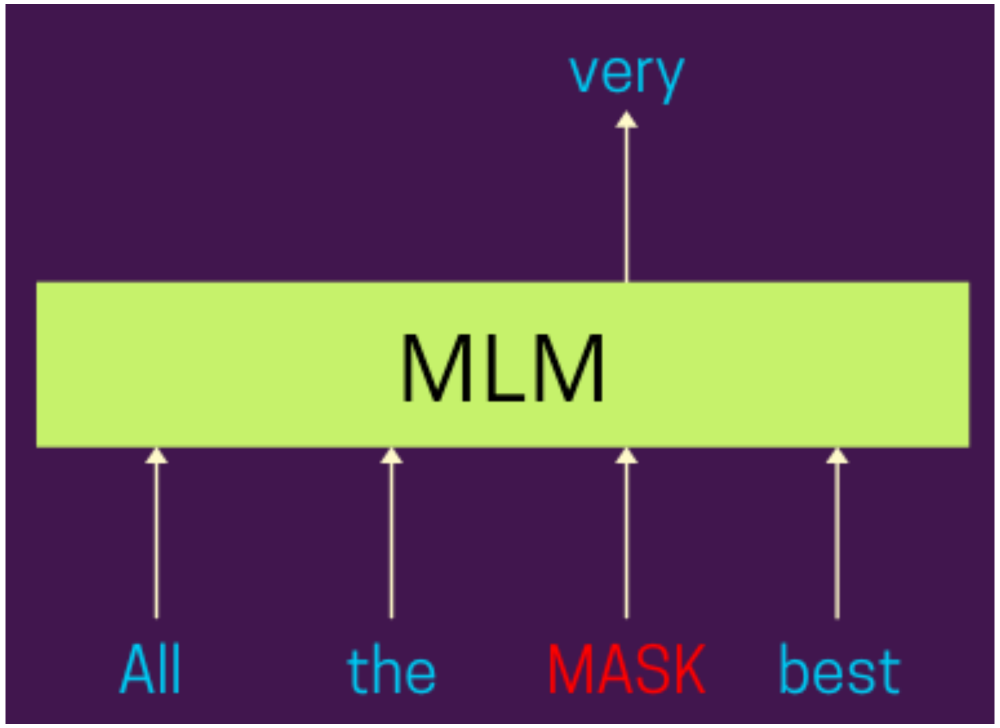


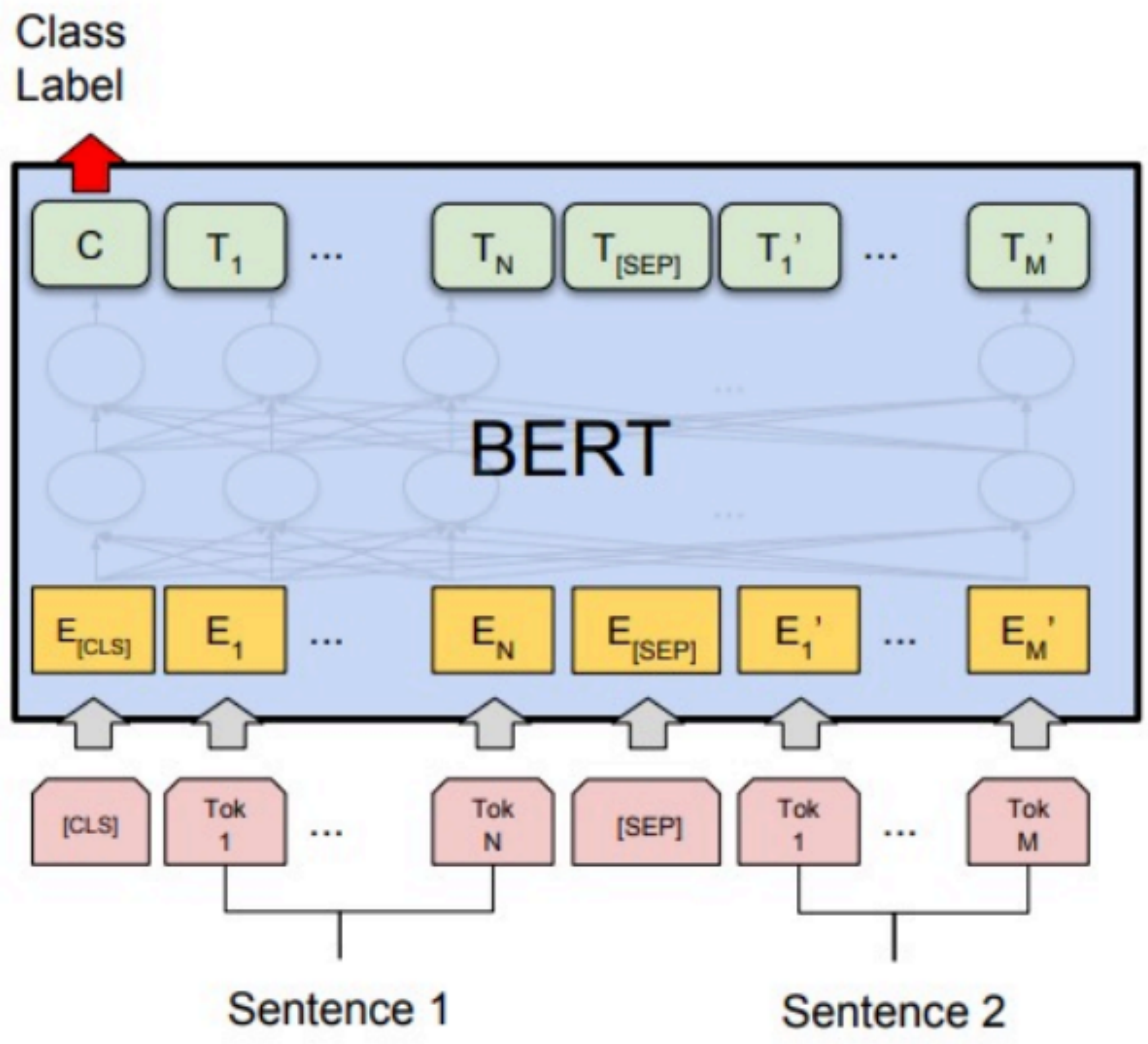
Figure credit: Prakhar Mishra, [blog]

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

2. **Fine-tuning** on a task with labeled data

Entailment
Neutral
Contradiction



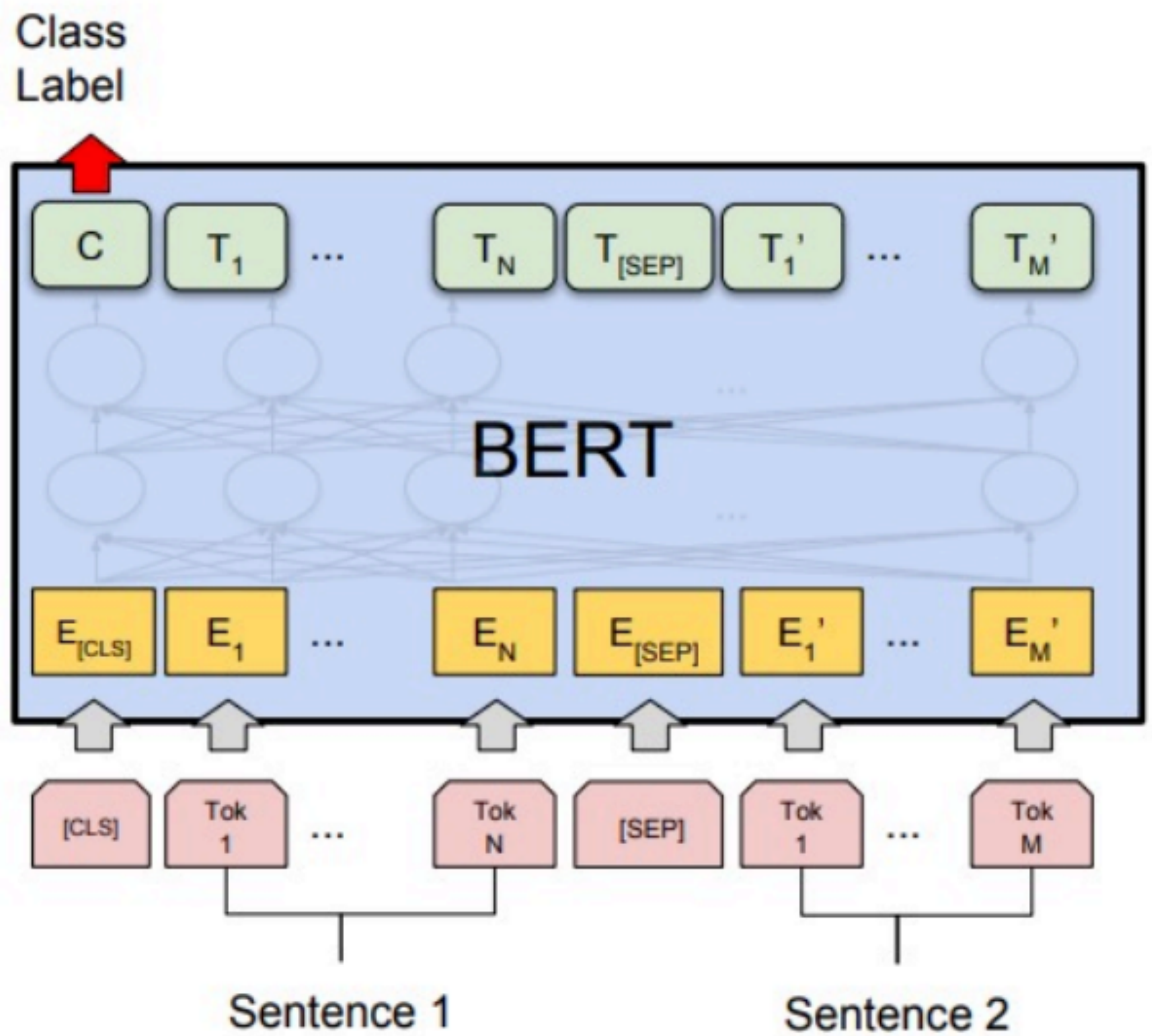
A soccer game with multiple males playing.

Some men are playing a sport.

Bowman et al. ACL, 2015

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

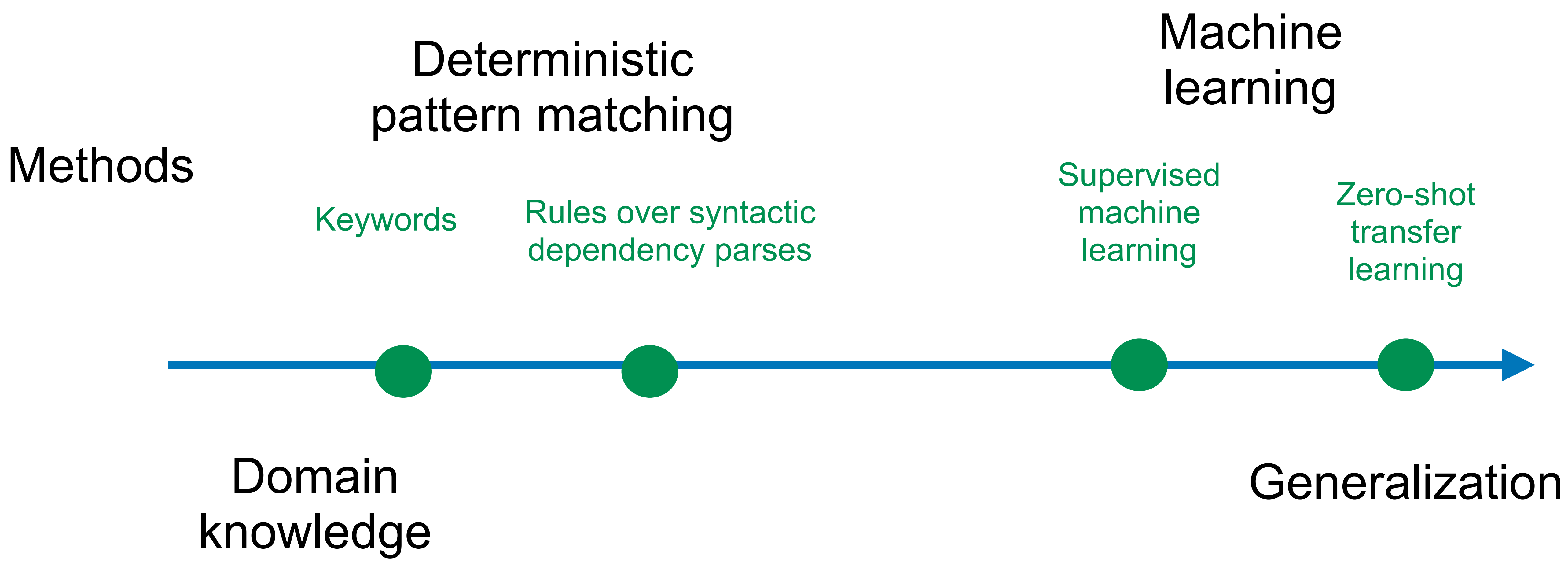
Entailment
Neutral
Contradiction



Prompt Police killed someone.

Sentence from our dataset Yesterday, 97 died in police firing.

Approaches to Automated Event Extraction



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

What was our original problem again?

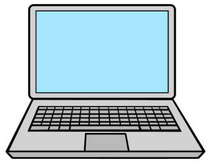
International-relations motivated research questions

Andy Halterman
Political Science

1. **Q:** Does variation in party control of state government affect whether police failed to intervene in communal violence?

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

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



Violence in Gujarat, India 2002

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

The New York Times

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

WILD SCENES OF JUBILATION IN DELHI

NEW CABINET OF INDIA

Fourteen Members

PANDIT NEHRU TO BE PREMIER

STATE VISIT TO KARACHI

LORD MOUNTBATTEN GREETS PAKISTAN

Mr. Jinnah Re-Affirms Firm Friendship With Britain

FRENZIED ENTHUSIASM IN BOMBAY

Crowds In Festive Mood

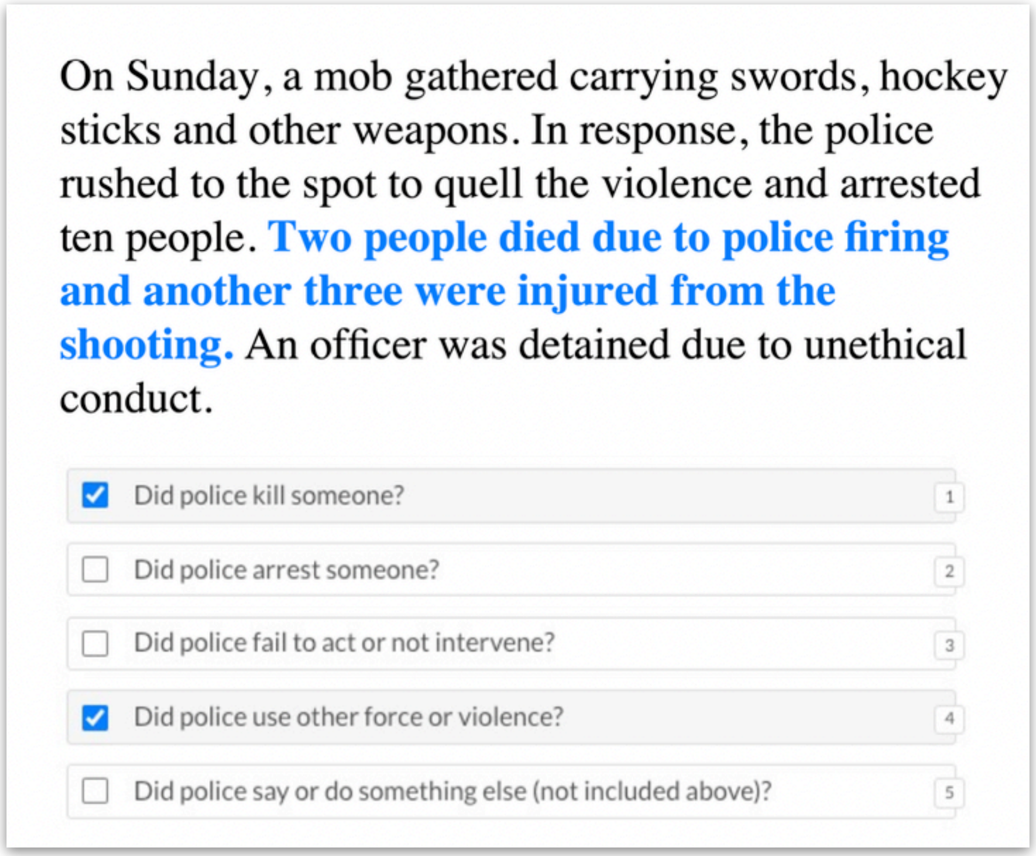
"MAY BOMBAY PROSPER"

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

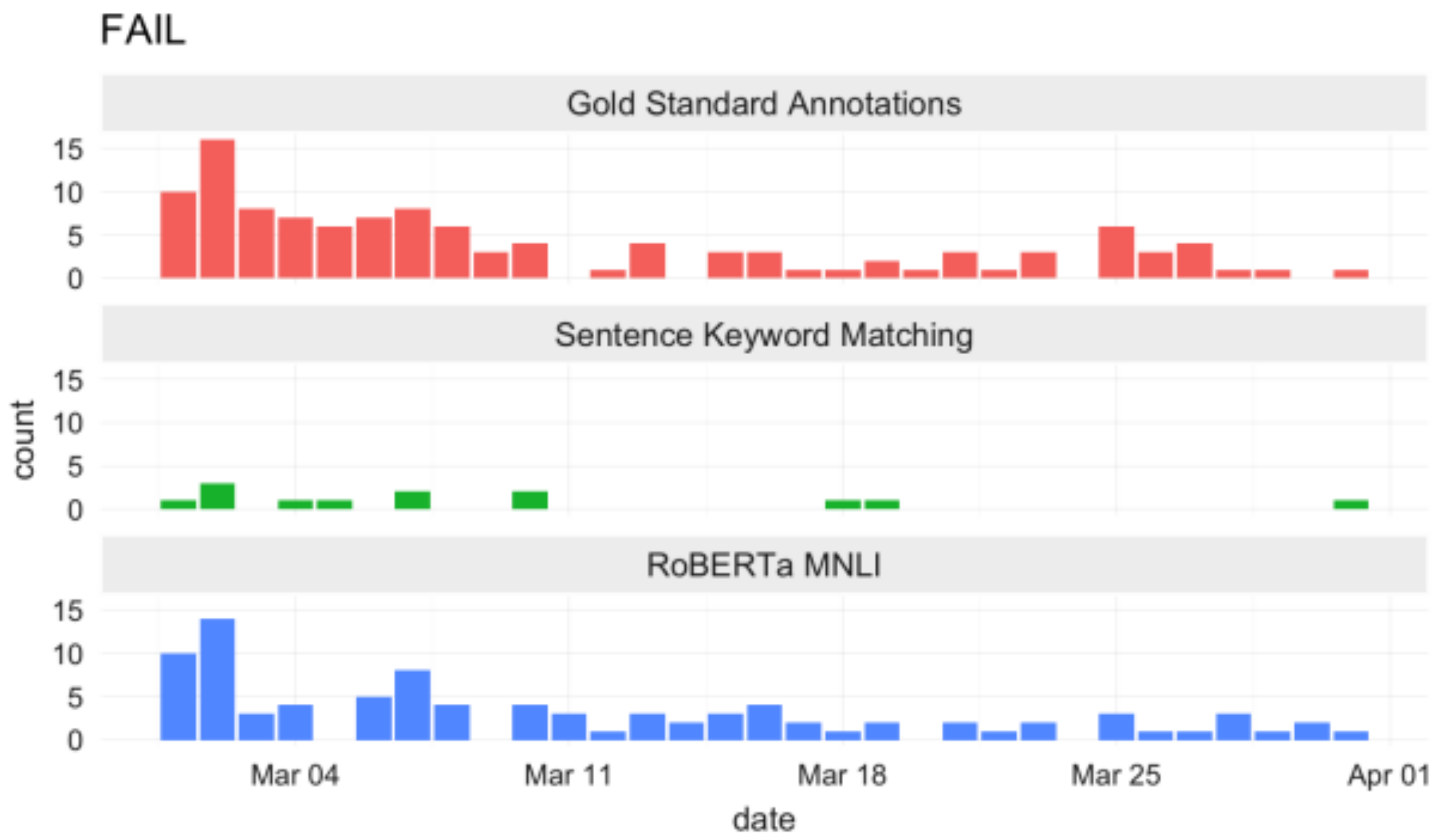
Annotation interface



Dataset publicly available

https://github.com/slanglab/IndiaPoliceEvents

Evaluation highlights



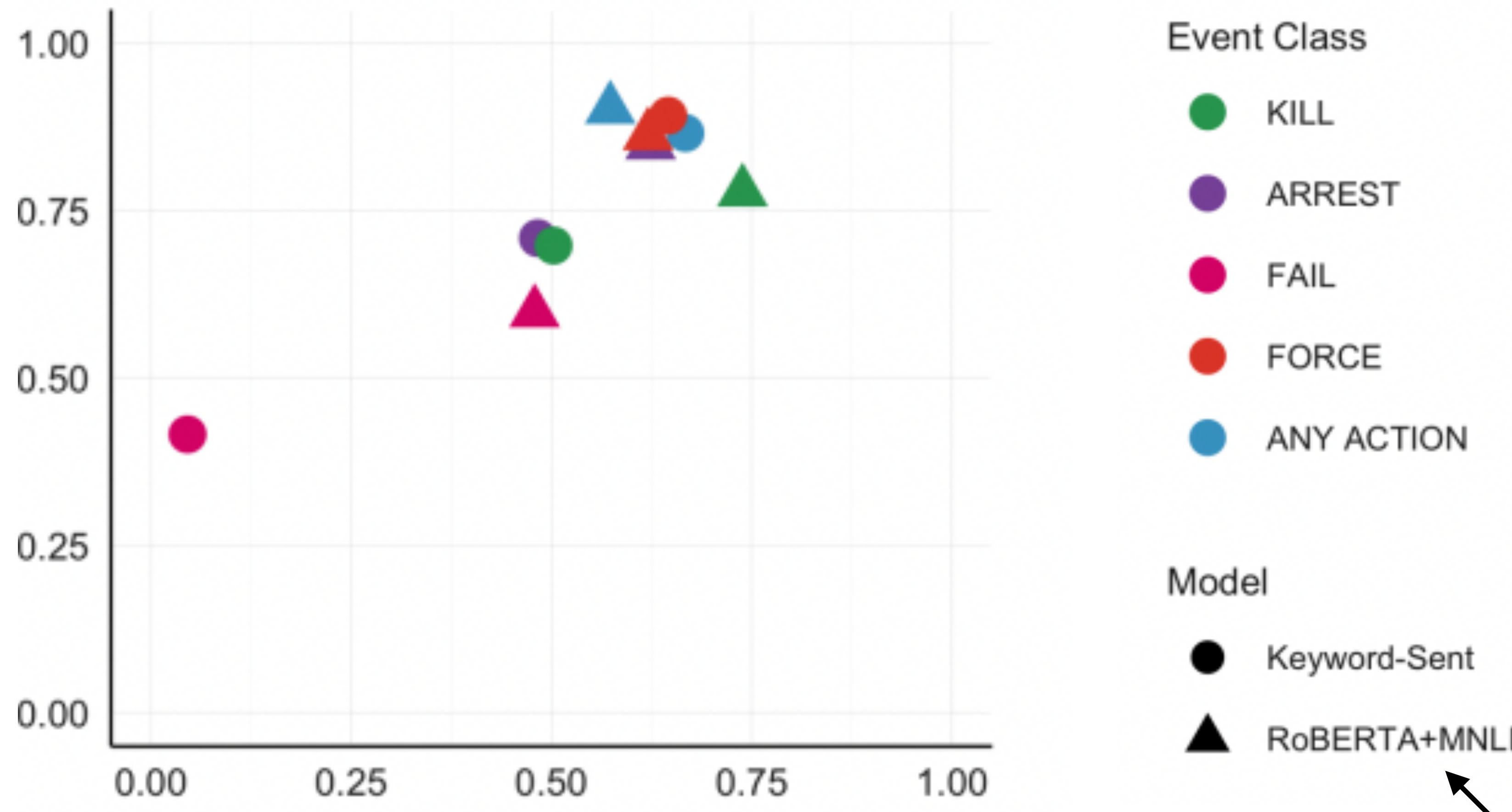
correlation = 0.42

Zero-shot language model previously mentioned

correlation = 0.6

Evaluation highlights

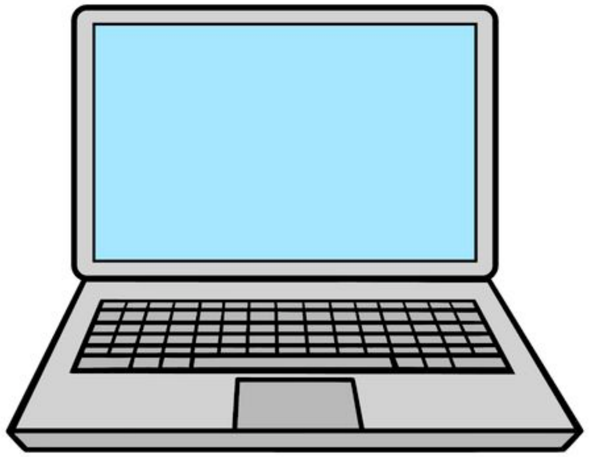
Temporal aggregates: correlation between **human gold-standard** and model



Sentence-level model F1

Zero-shot language model previously mentioned

Manual error analysis

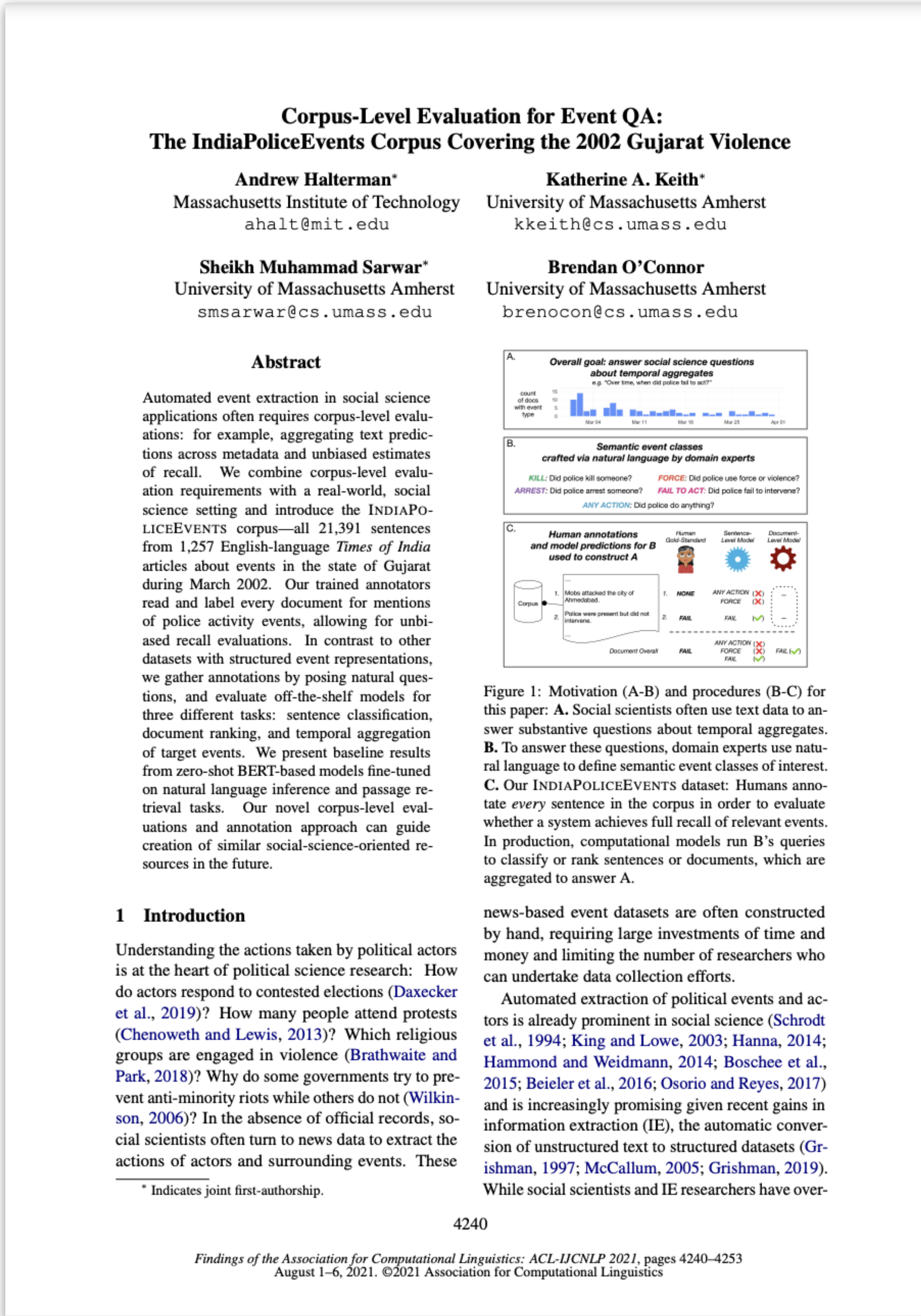


“ […] scores of people have been killed in rural Gujarat due to police failure…

Model assigns high positive probability to sentences that should be classified as **negatives**

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

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

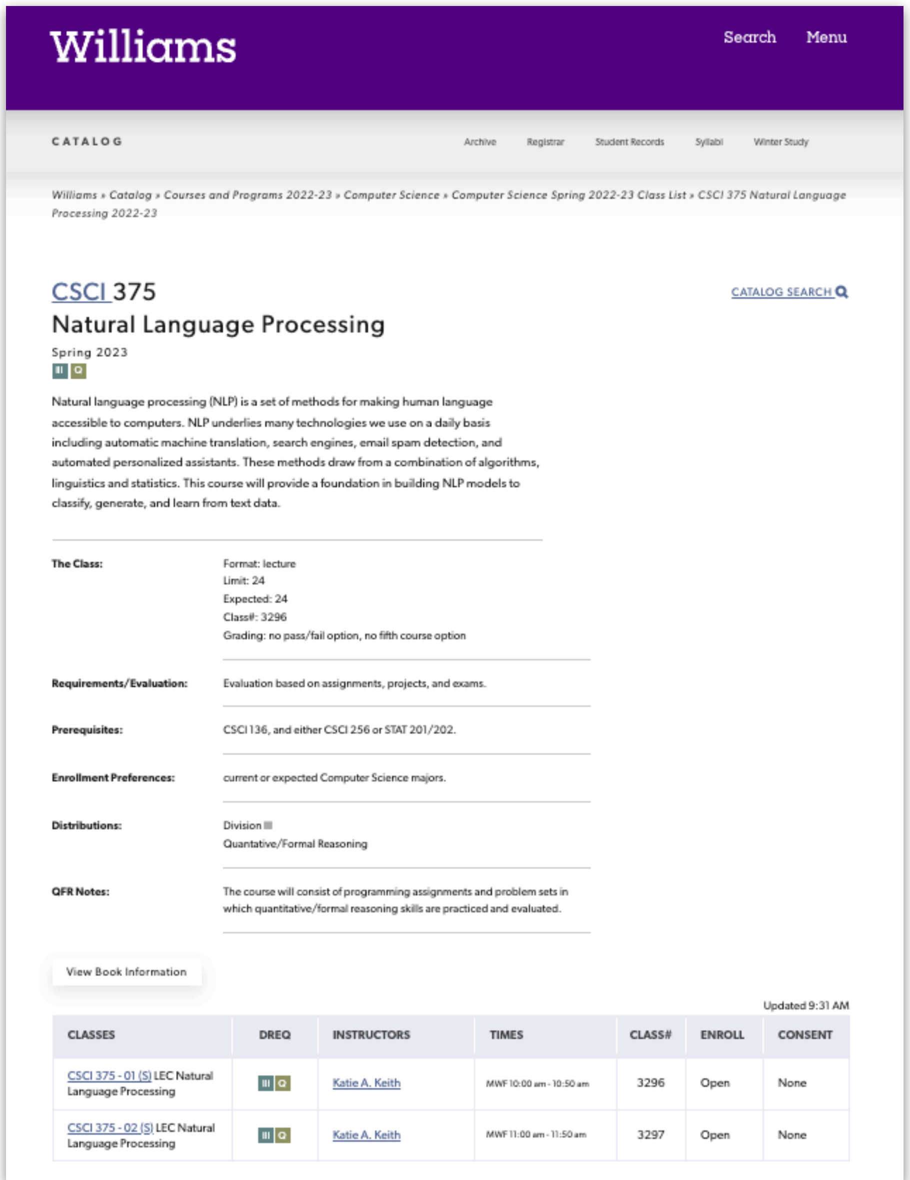
Andy Halterman
Political Science

Katie Keith
Computer Science

Sheikh Sarwar
Computer Science

Brendan O'Connor
Computer Science

Two shameless plugs



Thanks!

Collaborators




Kaggle Data Science
Research Grant

Bloomberg Data Science
PhD Fellowship