CS 640 Lecture on

AI for Multimodal Analysis of Political News

Margrit Betke
Co-Director of the AI Research Initiative at Boston University and Professor of Computer Science

BOSTON UNIVERSITY

Communication Research

What are the effects of news media on public opinion ?

Traditional Media

"Manual Content Analysis"

???

Online Media

Artificial Intelligence and Emerging Media (AIEM) Research Group

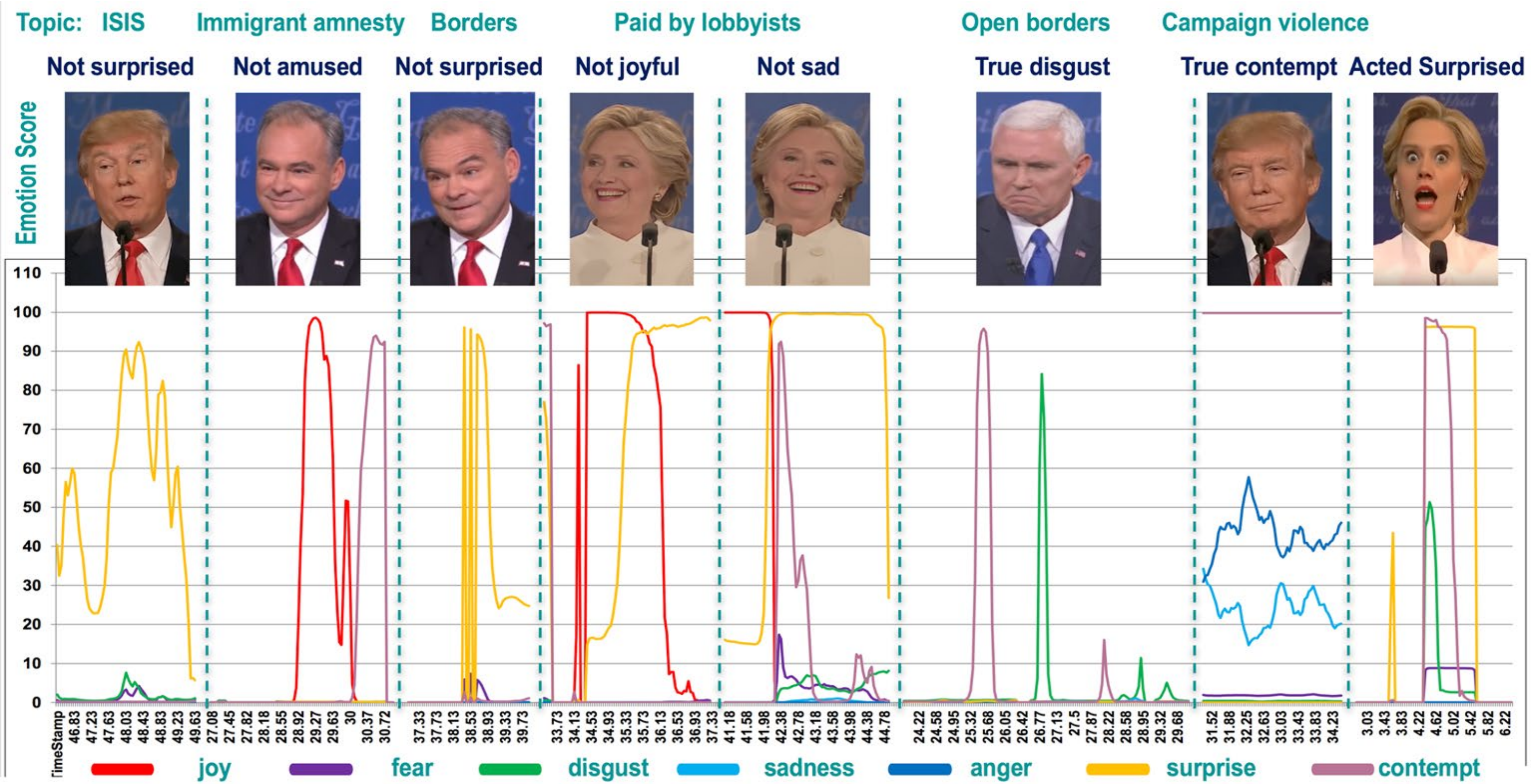
Communication Research

The ways in which the news media depict a certain political candidate, textually or visually, influence how the public perceives the candidate

Political candidates in various democracies have increasingly realized the power of their images in races and endeavored to put their best foot forward to appeal to voters

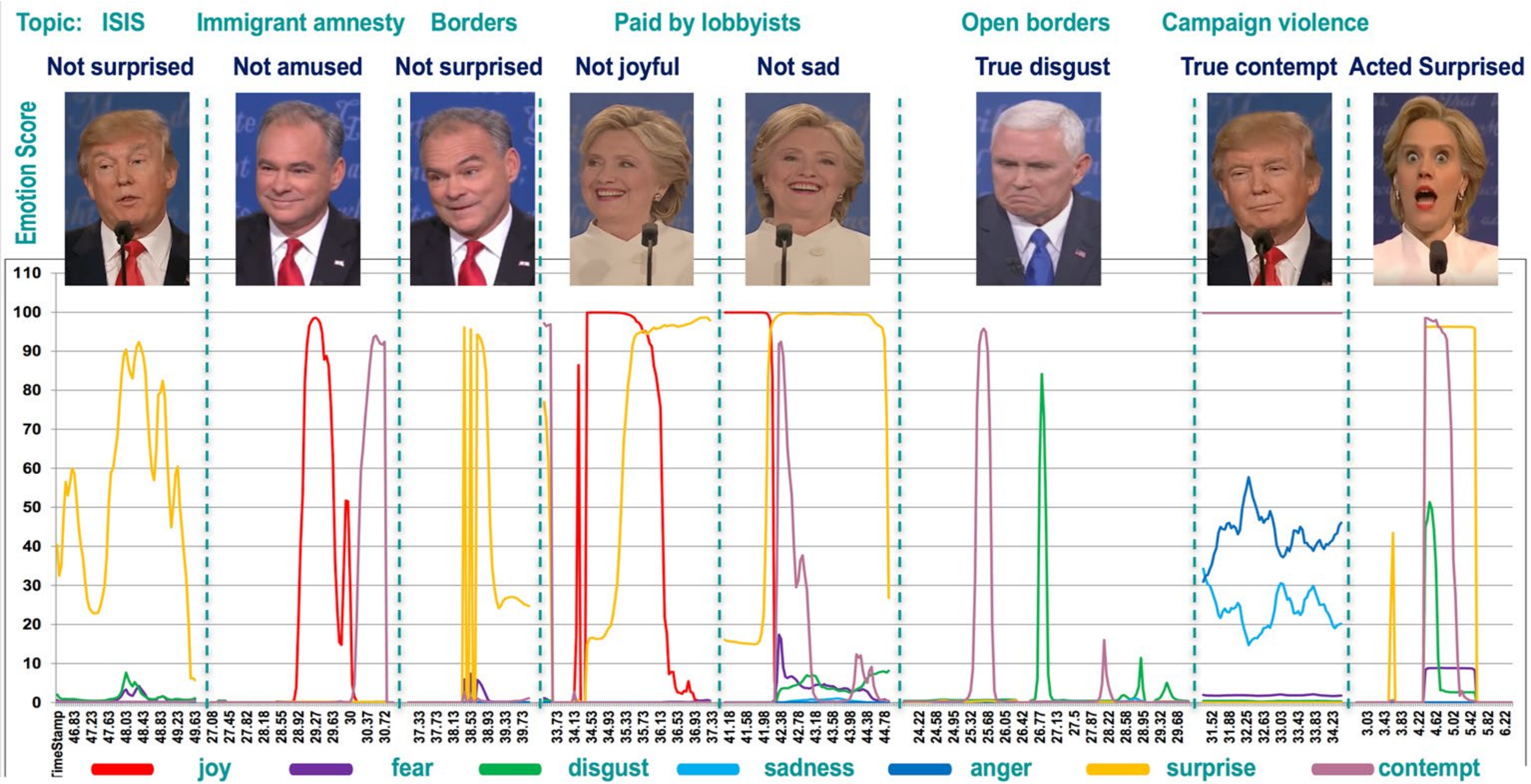
Candidates’ nonverbal language, facial expressions, and physical attractiveness have exhibited persuasion power

Debates between 2016 President/Vice President Candidates SNL



State-of-the-art Facial Expression Recognition: 7 Emotions & Neutral

Debates between 2016 President/Vice President Candidates SNL



State of the art in expression recognition is not sufficient
Datasets with spontaneous (vs. posed) expressions are lacking [Dupree et al, 2020]

2020 TV News Portrayals of Political Candidates

Dataset of TV News Portrayals of Presidential Candidates

- Six leading candidates of the 2020 Democratic Party presidential primaries in the United States—Biden, Buttigieg, Klobuchar, Sanders, Warren, and Yang
- Candidate videos uploaded to archive.org between January 1, 2020 and March 2, 2020
- A total of 28,297 news clips were downloaded
- ABC, CBS, NBC, PBS, CNN, MSNBC, and Fox News
- 26,125 video segments of length 1-5 seconds were generated
- Communication scholars provide expert labels

Video Analysis Method

- Face detection with ArcFace [Deng et al, 2019]
- Facial landmark detection with MTCNN [Zhang et al, 2016]
- Temporal alignment of facial landmarks
- Instead of 7 emotions & neural, we grouped expressions into 3 classes: [negative, neutral, positive]
- Two-stage Frame Attention Network [Meng et al., 2019]
  - Stage 1:   Compute high-dimensional representation of the facial characteristics for each input facial image
  - Stage 2: . Use a frame attention module to learn the attention weights that adaptively aggregate the feature vectors from the previous module into a single discriminative video representation
- Video representation is fed into multiple fully connected layers for final classification

BOSTON UNIVERSITY

CV Results

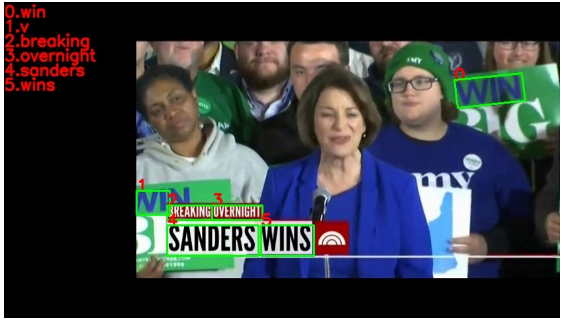
Performance of Personalized Models based on 4-fold cross validation

	Mean Accuracy	Mean F-score
Sanders	0.83	0.82
Yang	0.89	0.89
Biden	0.79	0.78
Warren	0.86	0.86
Buttigieg	0.83	0.82
Klobuchar	0.80	0.80
Overall	0.83	0.82

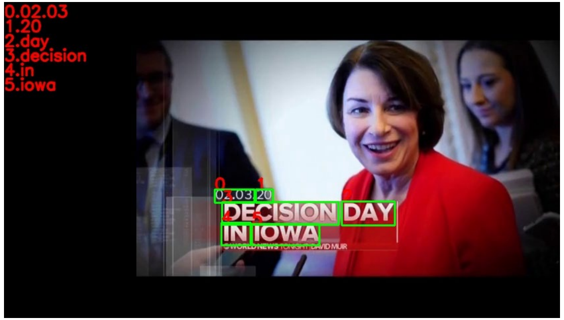
COM Results: Connecting Survey Data and CV Results

- Among all the candidates, Buttigieg received the greatest amount of airtime across all the seven television networks.
- The proportion of Buttigieg' positive facial expression in his airtime is among the highest.
- Sanders' facial expression tended to be more negative in television news compared with other candidates.
- Media effect varies by the candidate and by the media source: For Biden and Buttigieg, for example, the exposure to the two candidates' positive images in CBS and NBC makes surveyed viewers feel more positive toward the candidates.

Text Recognition for Context Analysis




```
*****************************
v breaking overnight
*****************************
sanders wins
*****************************
win
*****************************
```




```
******************************
02.03 20
******************************
decision day
in iowa
******************************
```


```
*****************************
sen amy klobuchar
democratic presidential candidate
*****************************
for new
*****************************
pbs9
*****************************
hour
*****************************
iv
*****************************
re
*****************************
```

BOSTON UNIVERSITY

A method for detecting text of arbitrary shapes in natural scenes that improves text spotting

CVPR 2020 Workshop, June 2020

Qitong Wang, Yi Zheng, and Margrit Betke

BOSTON UNIVERSITY

Framework of our model: UHT (Abbrevaited as: UNet, Heatmaps, Texfill Algorithm)

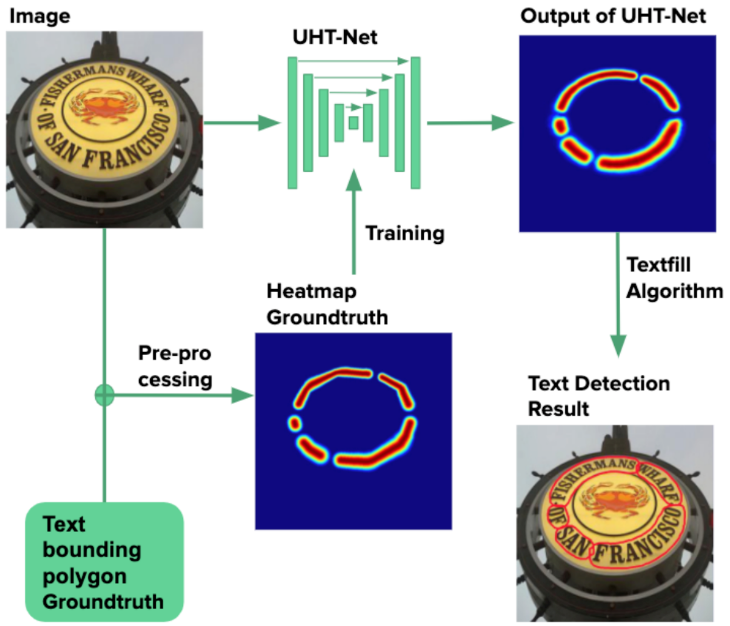


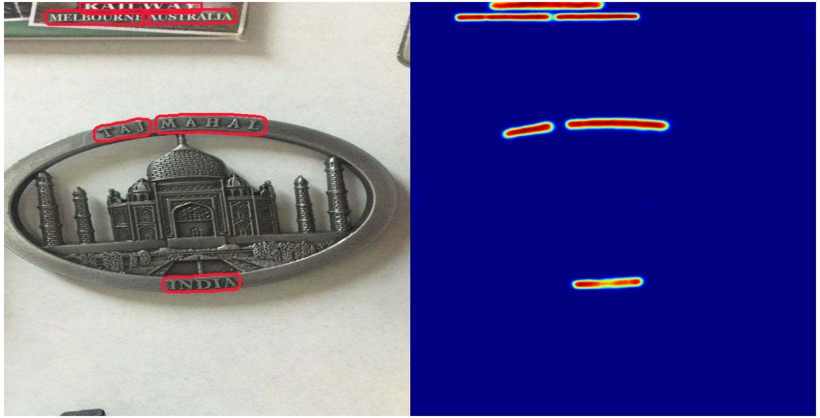
Figure 2. Pipeline of UHT. The process of text detection of UHT can be divided into three steps: 1) Pre-processing is used to generate a heatmap text region ground truth, which is used as a training label of UHT-Net. 2) A trained UHT-Net can output predicted text region heatmaps. 3) In the post-processing step, the Textfill algorithm outputs the final predicted text bounding polygons interpreting the outputs of the UHT-Net.

BOSTON UNIVERSITY

Selected visualization on text detection

Methodology	Venue	P (%)	R (%)	F (%)
Single-scale Testing				
Poly-FRCNN-3 [8]	IJDAR-2019	78.0	68.0	73.0
TextSnake [24]	ECCV-2018	82.7	74.5	78.4
CSE$^+$ [22]	CVPR-2019	81.4	79.7	80.2
TextField [43]	TIP-2019	81.2	79.9	80.6
PSENet-1s$^+$ [40]	CVPR-2019	84.02	77.96	80.87
FTSN [9]	ICPR-2018	84.7	78.0	81.3
ICG [38]	PR-2019	82.9	80.9	81.5
LOMO [46]	CVPR-2019	88.6	75.7	81.6
CRAFT$^+$ [7]	CVPR-2019	87.6	79.9	83.6
PSENet_v2 [41]	ICCV-2019	89.3	81.0	85.0
CharNet H-88 [42]	ICCV-2019	**89.9**	81.7	**85.6**
UHT V16 (Ours)	-	88.8	**82.6**	**85.6**
UHT R50 (Ours)	-	88.2	81.8	84.9
Multi-scale Testing				
LOMO MS [46]	CVPR-2019	87.6	79.3	83.3
CharNet H-88 MS [42]	ICCV-2019	**88.0**	85.0	**86.5**
UHT V16 MS (Ours)	-	85.0	**85.6**	85.3
UHT R50 MS (Ours)	-	85.4	84.2	84.8

Table 1. Experimental results on the Total-Text dataset. "P" means Precision, "R" Recall, "F" F-measure, * denotes results on updated groundtruth annotations, and "MS" multi-scale testing.



Text detection demo of UHT.
Left: polygonal text detection result.
Right: deep learned heatmap from UHT-Net

BOSTON UNIVERSITY

Framework of our model: UHTA (Abbrevaited as: UHT+ASTER)

Methodology	Venue	F-measure (%)
Textboxes [18]	AAAI-2017	36.3
Mask TextSpotter [25]	ECCV-2018	52.9
TextNet [37]	ACCV-2018	54.0
CharNet H-88 [42]	ICCV-2019	66.6
TSA [24, 33]	ECCV-2018	58.1
UHTA V16 (Ours) [33]	-	75.7
UHTA R50 (Ours) [33]	-	**77.6**

Table 6. Experimental results of TSA and UHTA on the Total-Text Dataset. Pretrained ASTER [33] model is downloaded from official pytorch reimplementation [5]. Evaluation method for UHTA and TSA is end-to-end recognition from [6]. Annotations are horizontal text-bounding rectangles because UHT and TextSnake outputs horizontal text-bounding rectangles in UHTA and TSA. No distinction between uppercase and lowercase was made when we evaluated UHTA and TSA. The listed F-measures of the prior works were reported in their original papers. UHTA V16 denotes UHTA with VGG-16backbone; UHTA R50 denotes UHTA with ResNet-50 backbone.

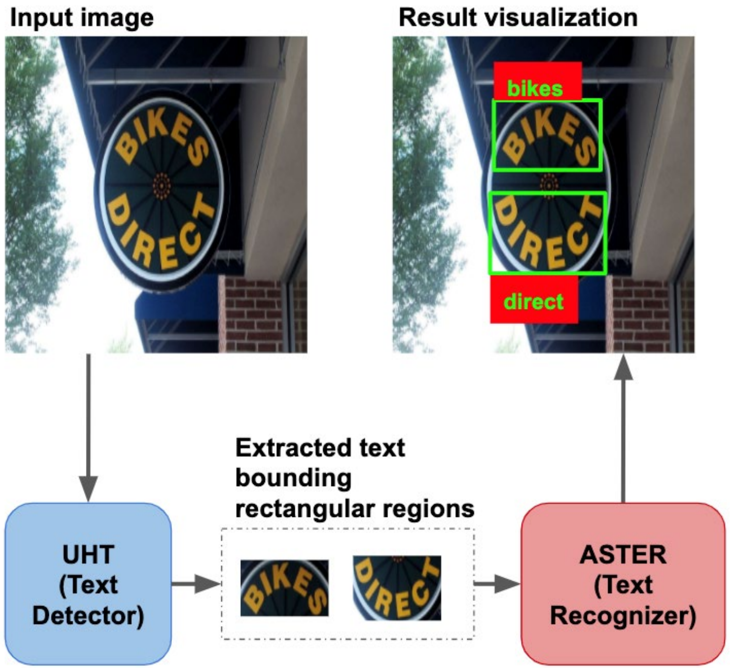


Figure 5. Pipeline of the proposed text spotting model UHTA. The model first calls the proposed UHT Detector and then converts UHT's polygonal output regions into horizontally-aligned rectangular regions of text. These text regions are then passed to the state-of-the-art text recognition model ASTER [33], which can accurately recognize the text in these regions and output text strings. So, like a person, UHTA does not only know where the text regions are, but also recognize the content of each text region.

BOSTON UNIVERSITY

LAL: Linguistically Aware Learning for Scene Text Recognition

MM '20: Proceedings of the 28th ACM International Conference on Multimedia
October 2020 Pages 4051–4059

Yi Zheng, Wenda Qin, Derry Wijaya, and Margrit Betke

BOSTON UNIVERSITY

Motivation

Problem: Incorrectly recognized characters caused by embeddings of text that has ambiguous or misleading visual appearance (Fig.1a).

Solution: Combine embeddings of visual & linguistic information (Fig.1b).

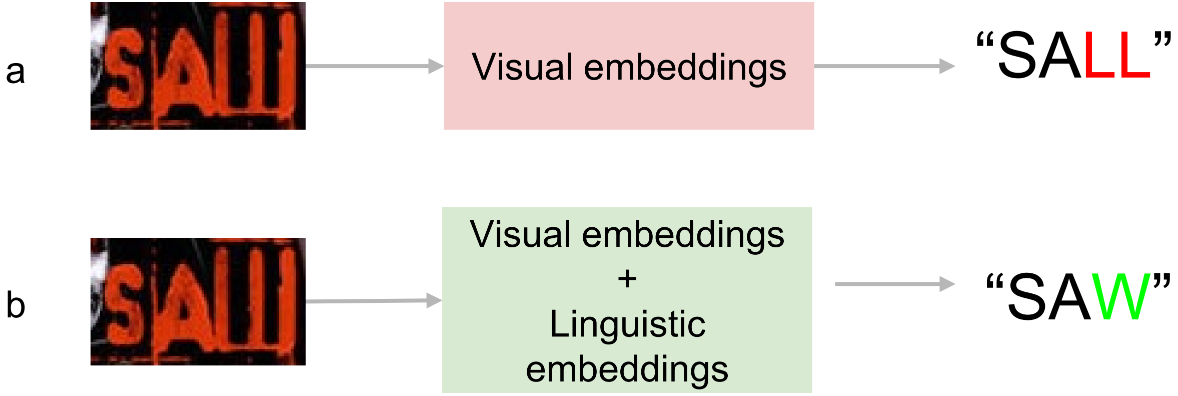


Figure 1: Problem and solution for scene text recognition

Method

Visual embeddings: Rectifier + Encoder + Attention Decoder

Linguistic embeddings: Next-Character Predictor (Transformer-based)

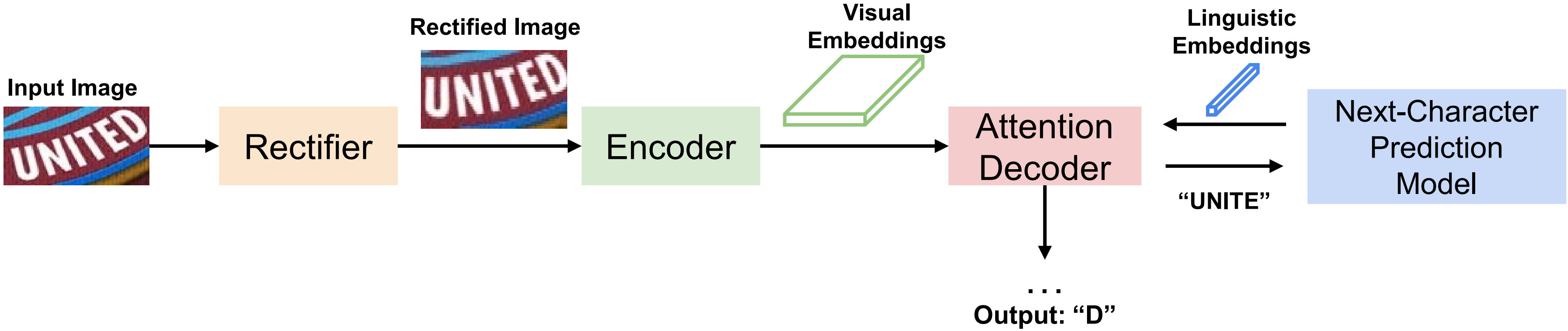


Figure 2: Overview of the proposed LAL method.

Combined Visual and Textual Analysis

rectified image

CV			NCP			CV+NCP		
h	m	l	-	-	-	h	m	l
97.2	1.15	0.38	-	-	-	97.2	1.15	0.38
o	e	a	e	a	o	o	e	a
99.17	0.25	0.18	26.29	23.12	22.88	98.17	1.32	0.39
t	w	l	u	w	l	t	w	m
74.43	19.01	2.3	20.33	16.52	11.95	75.50	10.07	2.74
i	!	e	e	<EOS>	s	e	h	m
38.15	31.18	12.73	51.89	40.23	2.50	58.83	21.28	4.64
l	<EOS>	s	l	s	v	l	!	,
88.68	4.67	1.47	99.96	0.01	0.01	99.98	0.00	0.00
<EOS>	.	,	<EOS>	,	s	<EOS>	,	.
99.48	0.19	0.1	67.30	12.94	9.94	99.10	0.00	0.00

BOSTON UNIVERSITY

Experiments

1. **Highest accuracy** compared to others on **six of seven** benchmarks.

2. An average of **3.3 percentage points boost** on irregular text.

Method	Dimension of Feature Map	Rectification	Regular				Irregular		
			IIIT5K	SVT	IC03	IC13	IC15	SVTP	CT
CRNN [33]	1D	no	78.2	80.9	-	86.7	-	-	-
RARE [34]	1D	no	81.9	81.9	-	-	-	71.8	59.2
STAR-Net [24]	1D	no	83.3	83.6	-	89.1	-	73.5	-
FAN [5]	1D	no	87.4	85.9	94.2	93.3	-	-	-
ASTER [35]	1D	yes	93.4	89.5	-	91.8	76.1	78.5	79.5
ESIR [44]	1D	yes	93.3	**90.2**	-	-	76.9	79.6	83.3
Attentive Text Recognition [41]	2D	no	-	-	-	-	-	75.8	69.3
AON [6]	2D	no	87.0	82.8	91.5	-	68.2	73.0	76.8
CA-FCN [22]	2D	no	92.0	82.1	-	91.4	-	-	79.9
SAR [20]	2D	no	91.5	84.5	-	-	69.2	76.4	83.3
LAL without NCP	1D	yes	94.4	87.5	92.5	93.8	76.4	79.5	84.7
LAL	1D	yes	**95.0**	89.8	**94.3**	**95.1**	**79.0**	**82.9**	**87.8**

Figure 3: Scene text recognition accuracy (%) on benchmark datasets, as reported in the paper.

Experiments

3. Difference between combining visual information with and without linguistic information is **statistically significant** on the larger datasets IIIT5k and IC15, which contains over 2,000 samples.

Dataset	IIIT5K	SVT	IC03	IC13	IC15	SVTP	CT
P-value	**2.59**	9.45	6.06	10.42	**2.21**	5.84	13.81

Figure 4: P-value (×10^-2) on seven benchmark datasets.

Deep Neural Network for Semantic-based Text Recognition in Images

arXiv

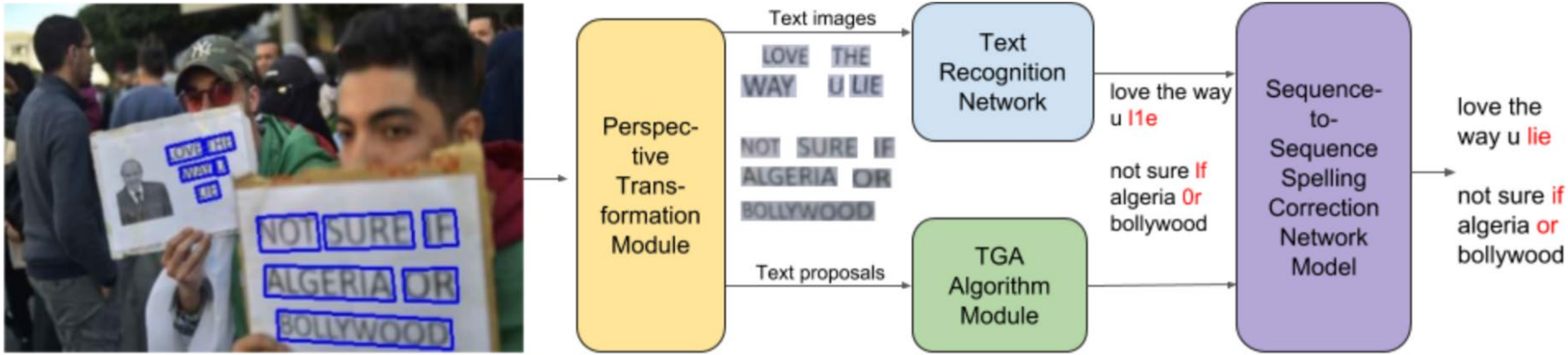
Yi Zheng, Qitong Wang,
and Margrit Betke

BOSTON UNIVERSITY

Idea: Use Context

BOSTON UNIVERSITY

Method

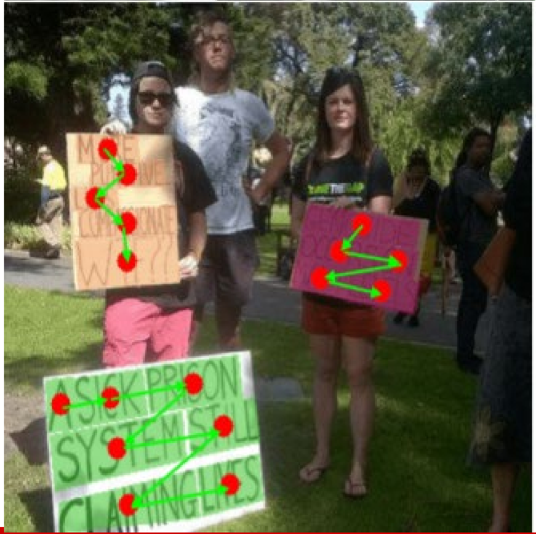


BOSTON UNIVERSITY

Results


WE ARE
ALL
MUSLIMS
NOW
RACIST


POLICE VIOLENCE IS A
HUMAN RIGHTS ISSUE


A SICK PRISON
SYSTEM STILL
CLAIMING LIVES


KAVA
-NAH

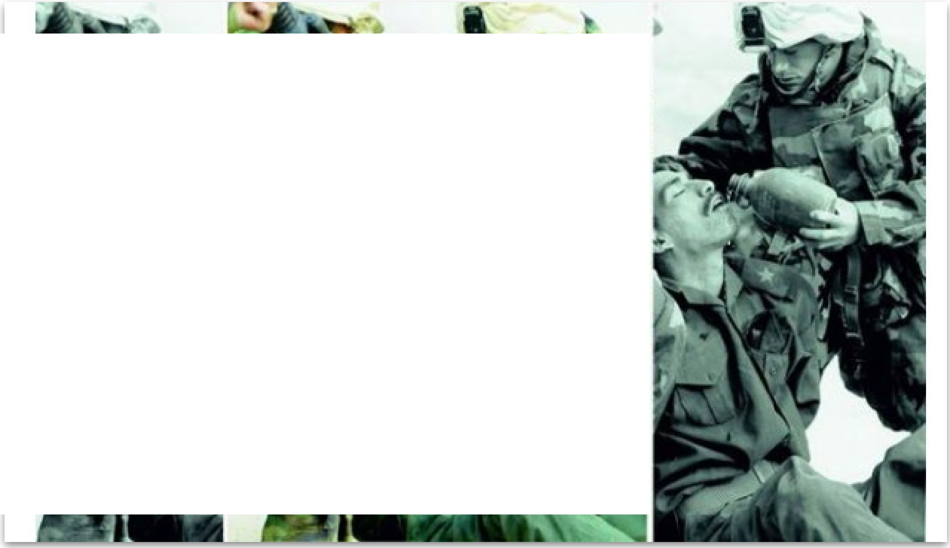

DUMP TRUMP

What is Media Framing?

Process of emphasizing certain perspectives of an issue to promote a particular interpretation, for example, a certain side of the political spectrum, thus influencing public opinion

What is Media Framing?

Process of emphasizing certain perspectives of an issue to promote a particular interpretation, for example, a certain side of the political spectrum, thus influencing public opinion



What is Media Framing?

Process of emphasizing certain perspectives of an issue to promote a particular interpretation, for example, a certain side of the political spectrum, thus influencing public opinion

What is Media Framing?

Process of emphasizing certain perspectives of an issue to promote a particular interpretation, for example, a certain side of the political spectrum, thus influencing public opinion

Detect media frames in gun violence coverage:

Textual Data

GOP lawmaker calls for age restriction on AR-15s

Gun control

Checking Facts and Falsehoods About Gun Violence and Mental Illness After Parkland Shooting

Mental health

Lindsey Graham: Both parties will suffer if Congress doesn't act on new gun bill

Politics

BOSTON UNIVERSITY

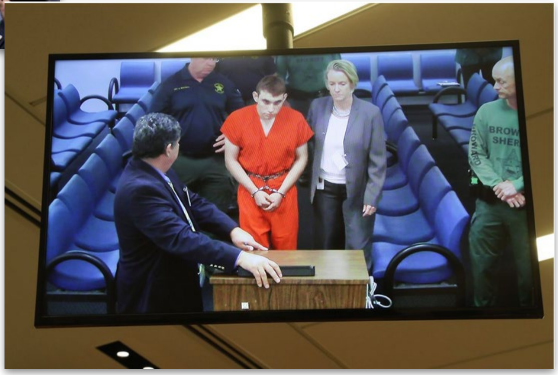
Current project:

Detect media frames in gun violence coverage

Visual Data

Gun control

Politics



Mental health

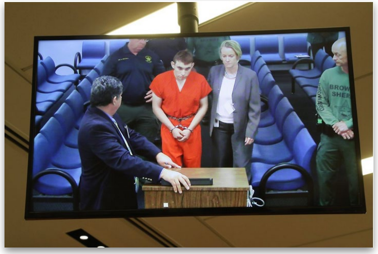
BOSTON UNIVERSITY

Detect media frames in textual and visual coverage of gun violence

Our approach: Natural Language Processing (NLP) + Computer Vision (CV)

COM Experts

Checking Facts and Falsehoods About Gun Violence and Mental Illness After Parkland Shooting



Data

News Frame	# Articles	Relevant Images Frac.
2nd Amendment/Gun Rights	*38*	*0.34*
Gun Control/Regulation	215	0.43
Politics	373	0.65
Mental Health	*65*	*0.43*
School/Public Space Safety	137	0.50
Race/Ethnicity	*114*	*0.30*
Public Opinion	237	0.62
Society/Culture	*41*	*0.10*
Economic Consequences	*80*	*0.58*
Overall	1300	0.52

BOSTON UNIVERSITY

Methods and Results

Subset	Res Net-50	BERT API	ResNet-50 + BERT API	BERT Headline
9 frames	62.5	72.7	74	85.9
4 frames	71.8	82	84.8	90.3

Communicating COVID-19

https://covid19.philemerge.com

BOSTON UNIVERSITY

Goals

Develop and apply machine learning tools to examine international news coverage of COVID-19

Create data visualizations of how journalists around the world have reported on various facets of the pandemic

BOSTON UNIVERSITY

Demo Website:

https://covid19.philemerge.com

Research Method

Methods to group COVID 19 news articles into topics:

- dictionary-based analysis
- unsupervised topic modeling
- supervised machine learning

BOSTON UNIVERSITY

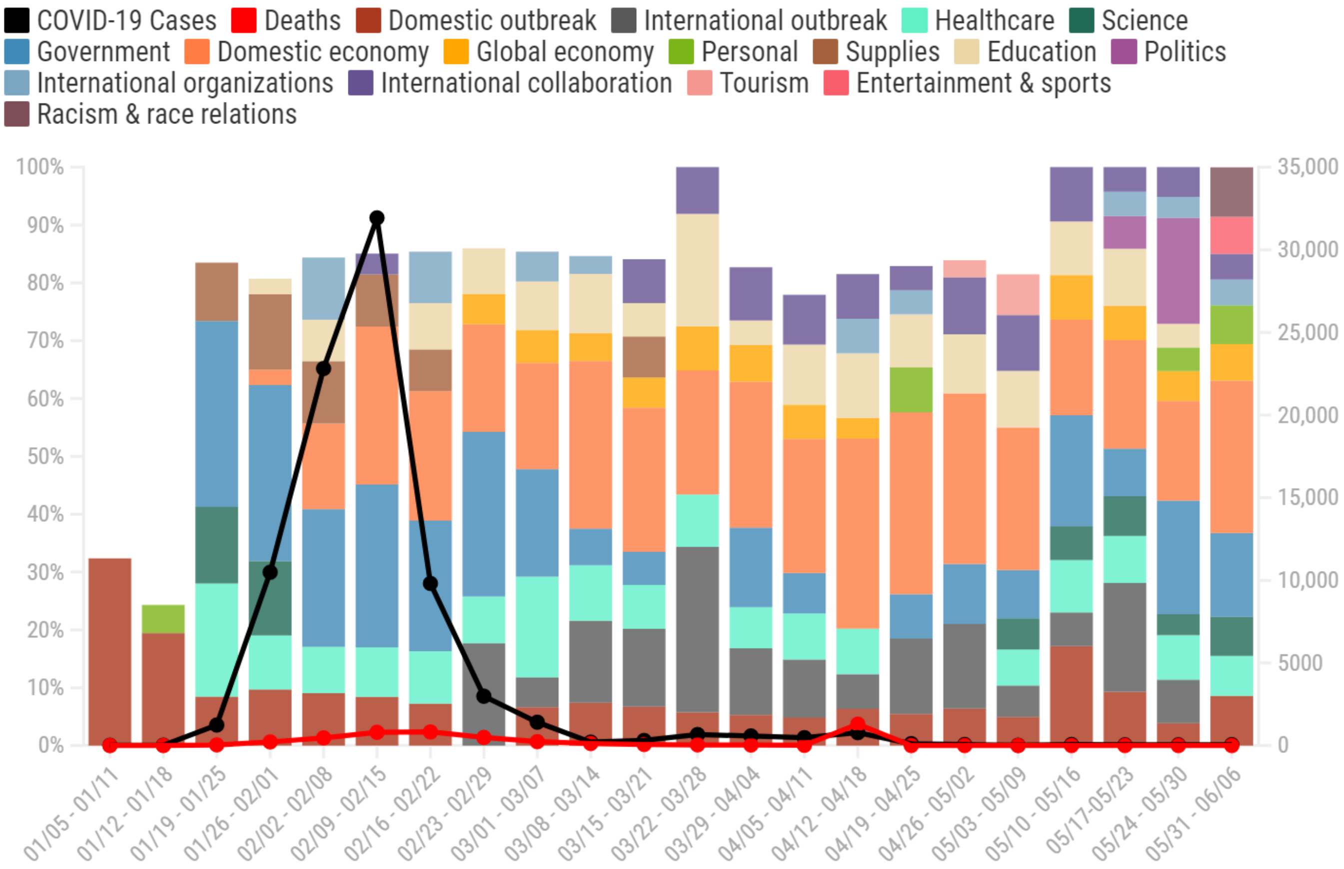
Research Method

Methods to group COVID 19 news articles into topics:

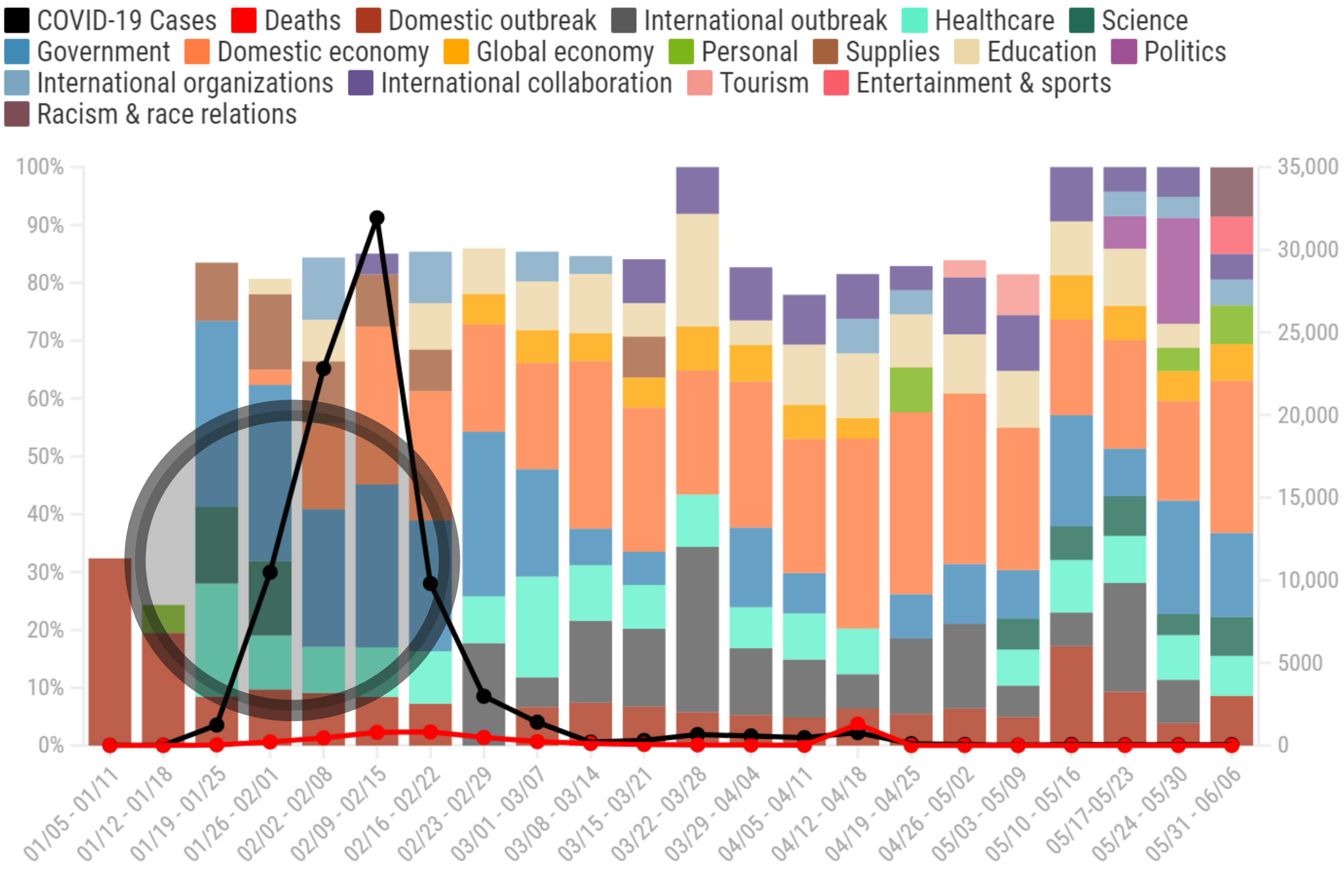
- dictionary-based analysis
- unsupervised topic modeling: Latent Dirichlet Allocation
- supervised machine learning

Ten main topics from each country’s news coverage are identified by week. Each topic is associated with 20 terms.

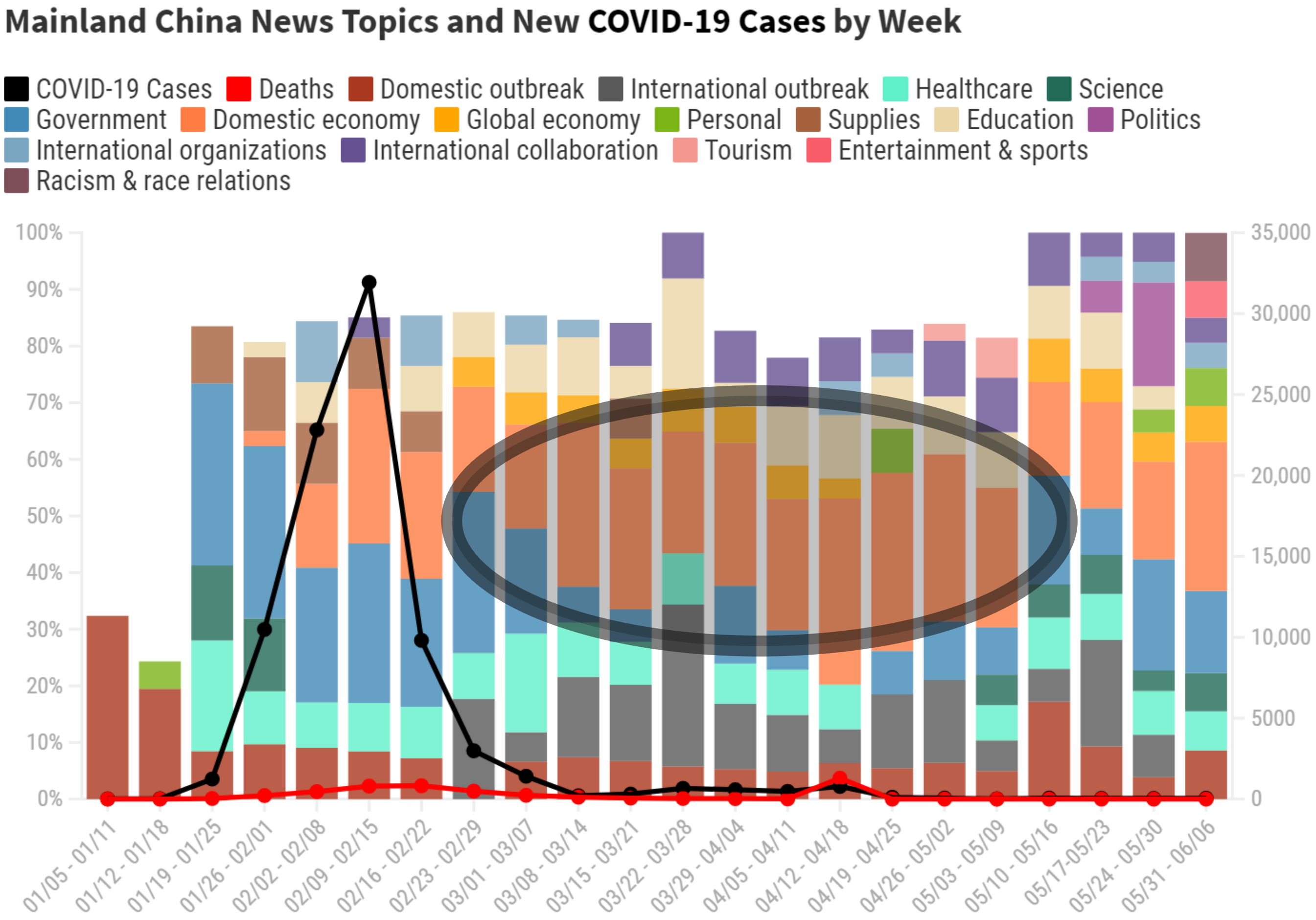
Mainland China News Topics and New COVID-19 Cases by Week



Mainland China News Topics and New COVID-19 Cases by Week



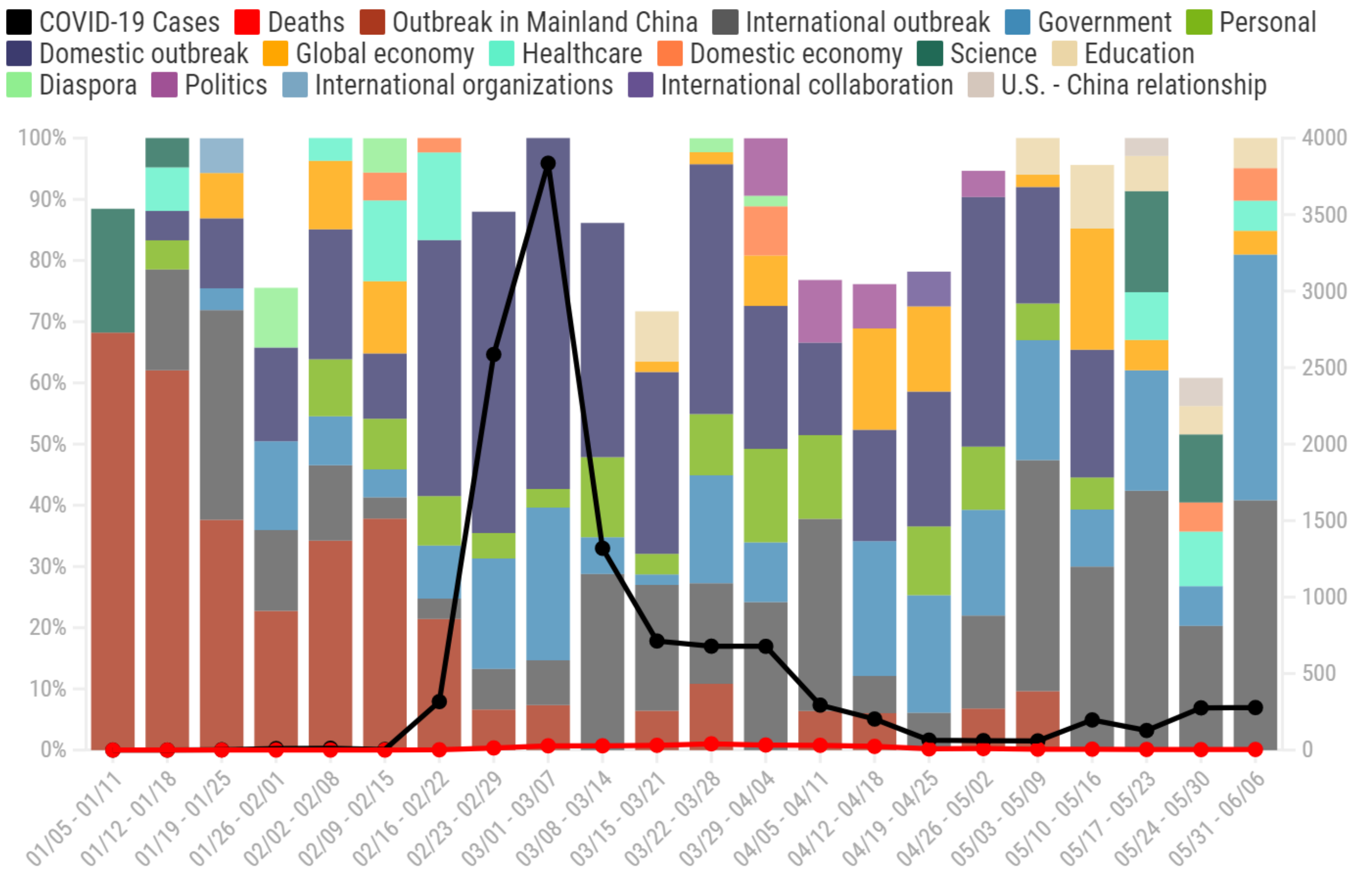
- Government Response
- Science
- Healthcare



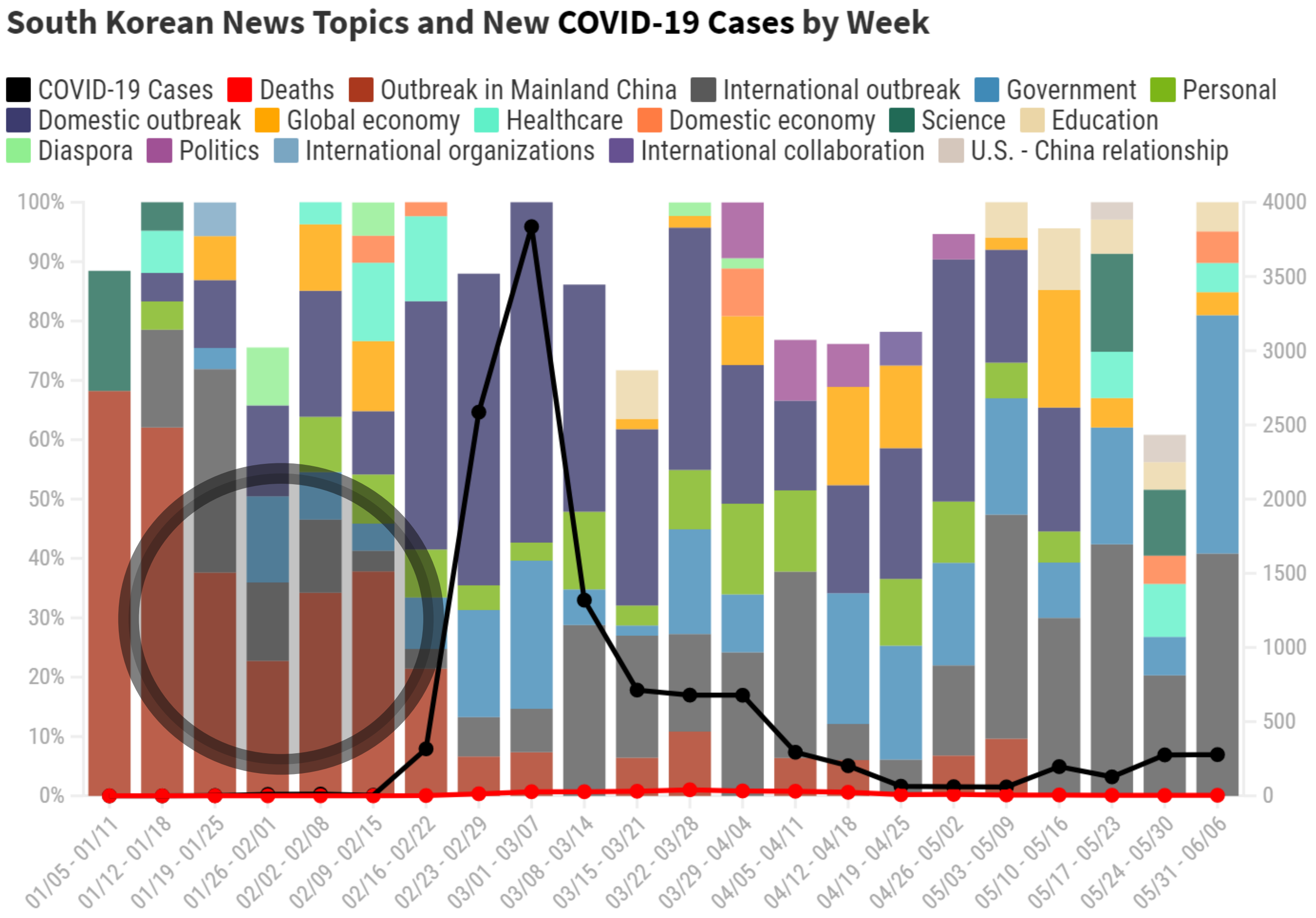
- Domestic Economy

BOSTON UNIVERSITY

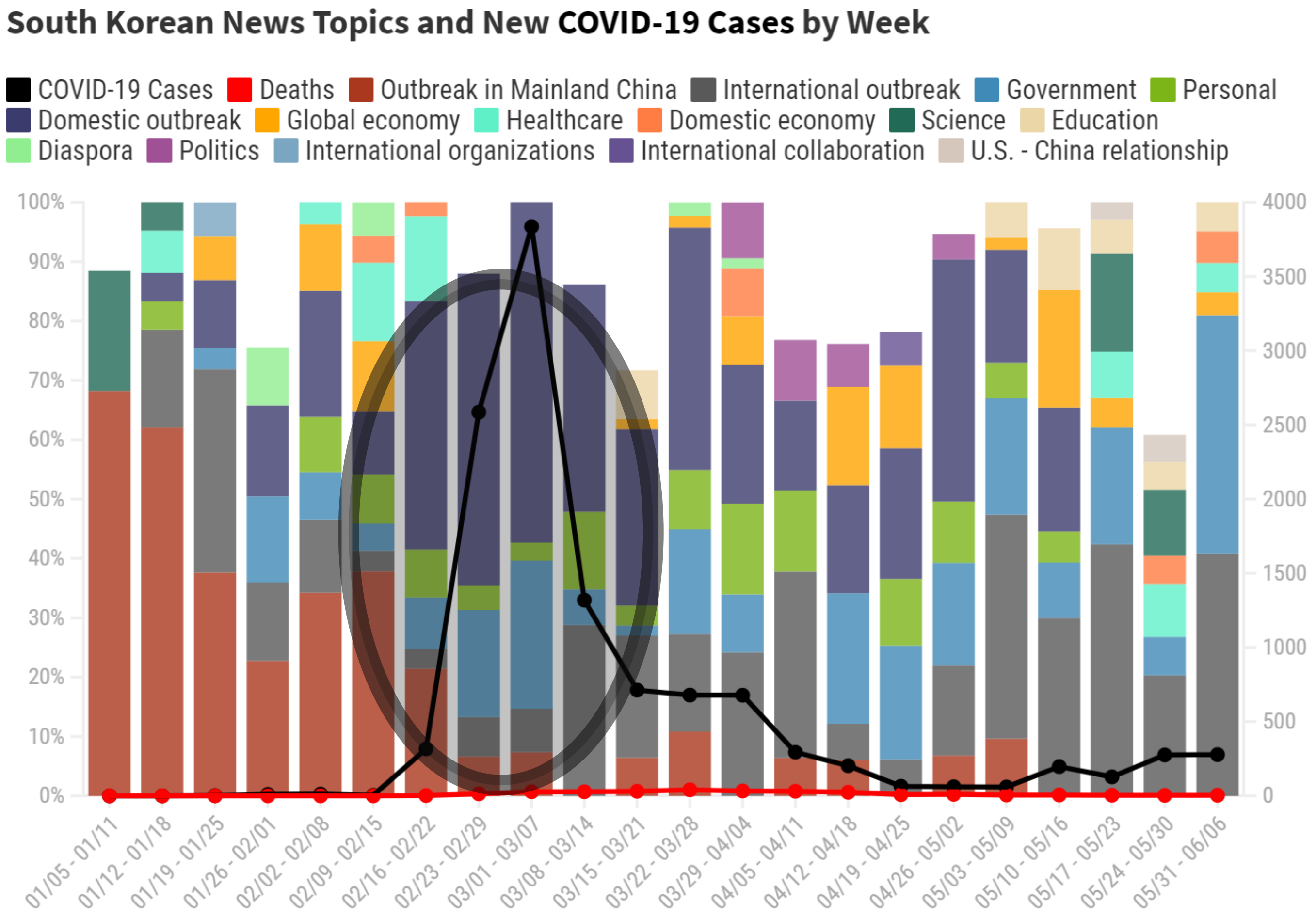
South Korean News Topics and New COVID-19 Cases by Week



BOSTON UNIVERSITY

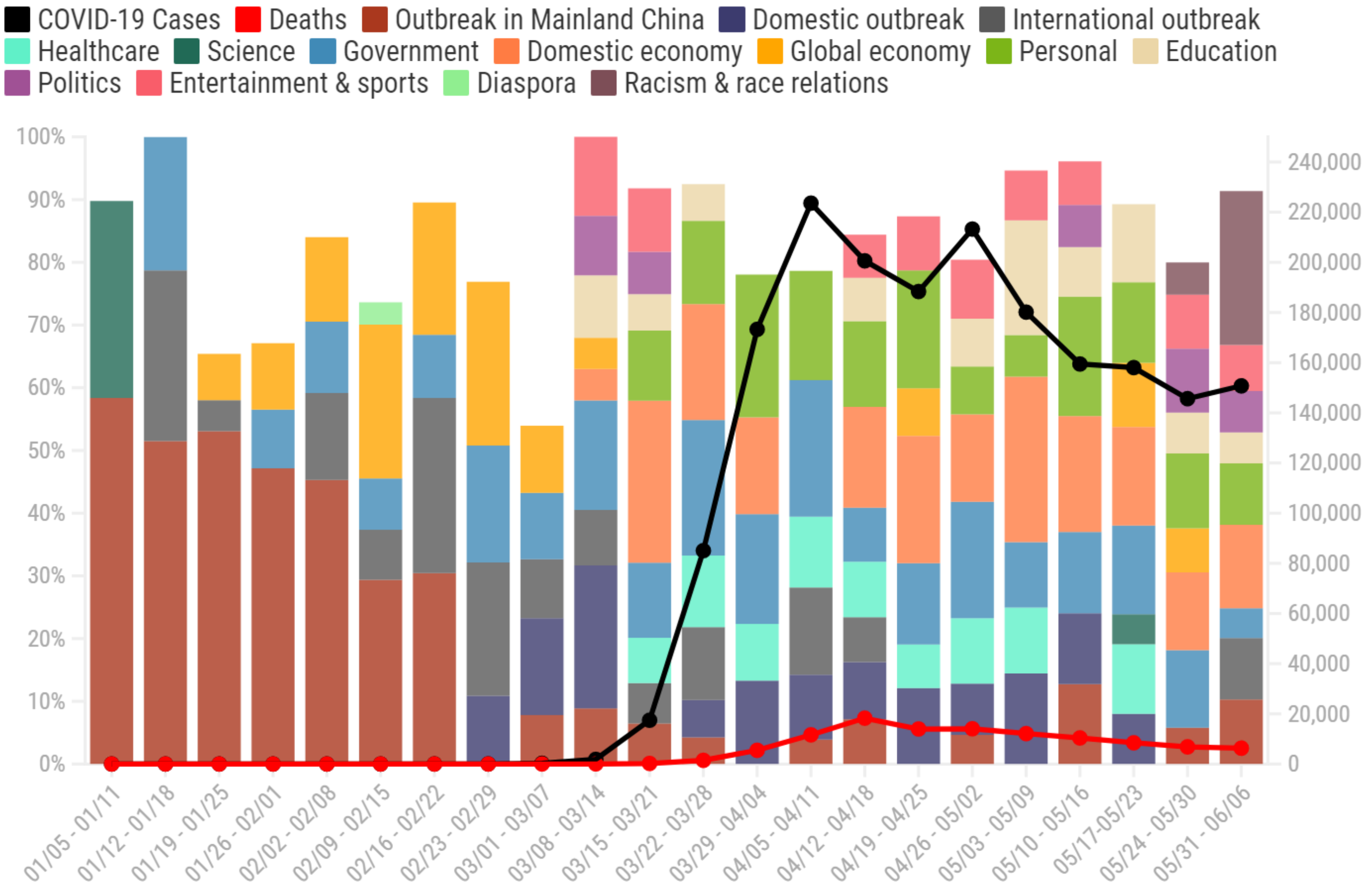


Outbreak in Mainland China

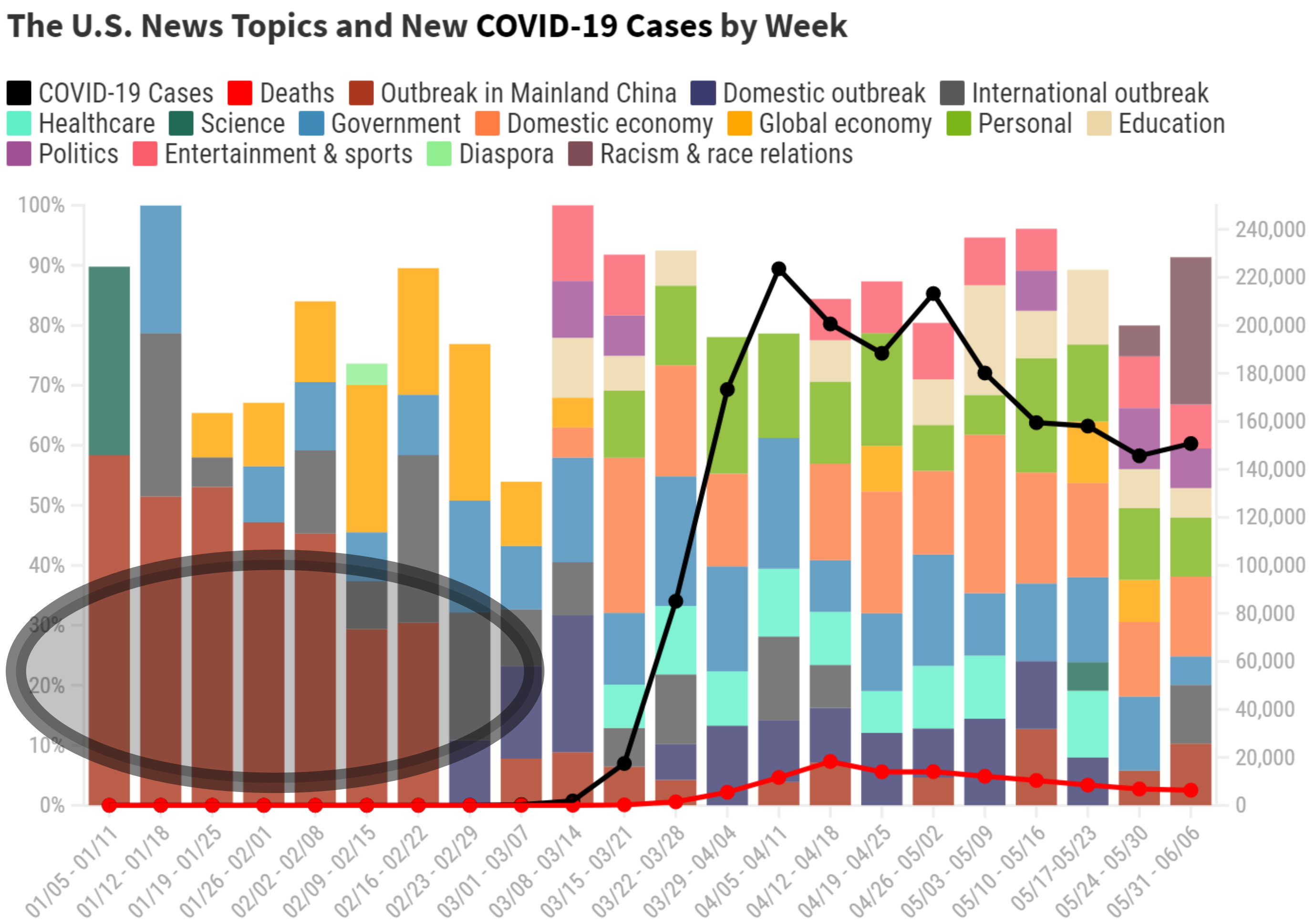


- International Collaboration
- Government Response

The U.S. News Topics and New COVID-19 Cases by Week

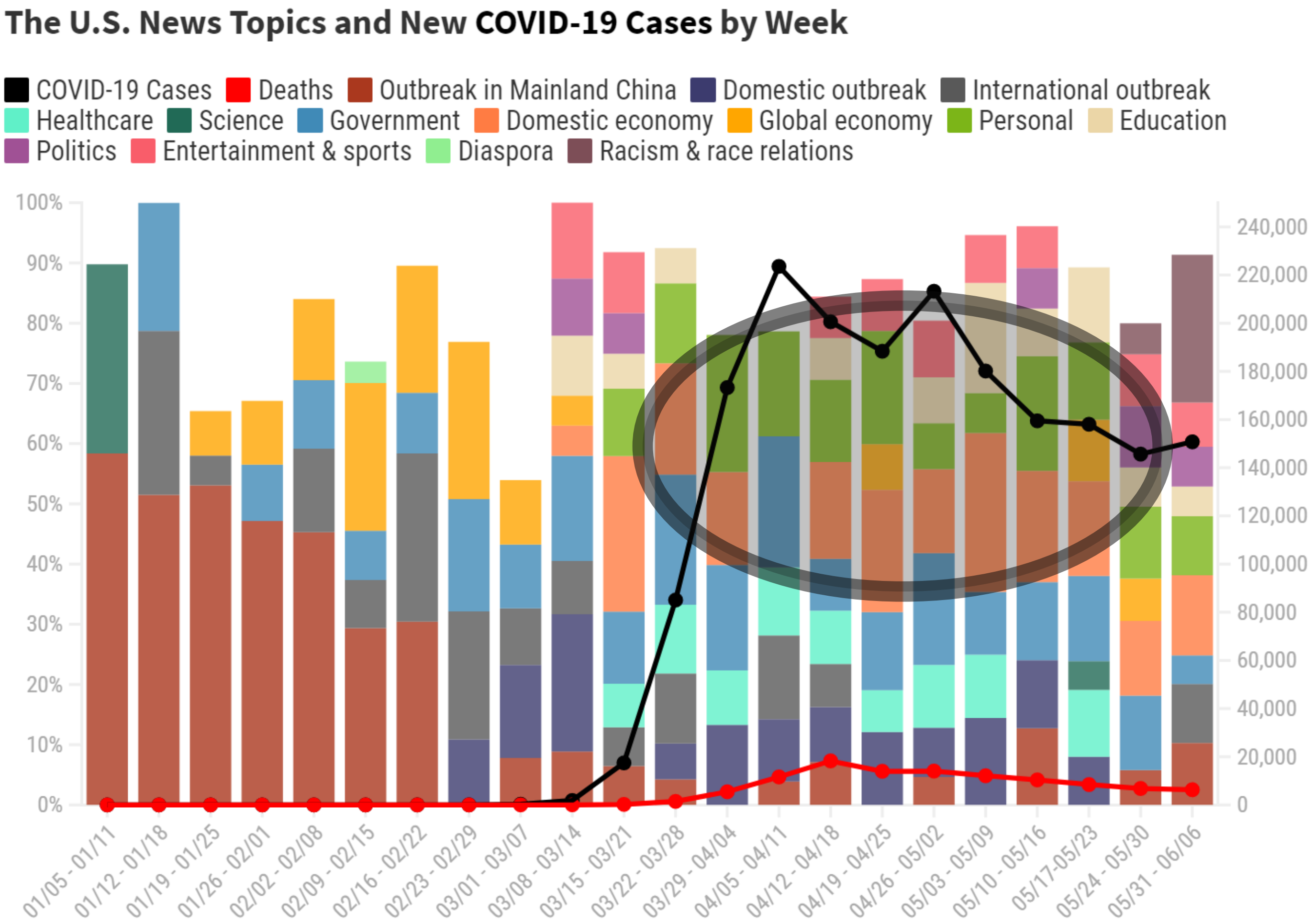


BOSTON UNIVERSITY



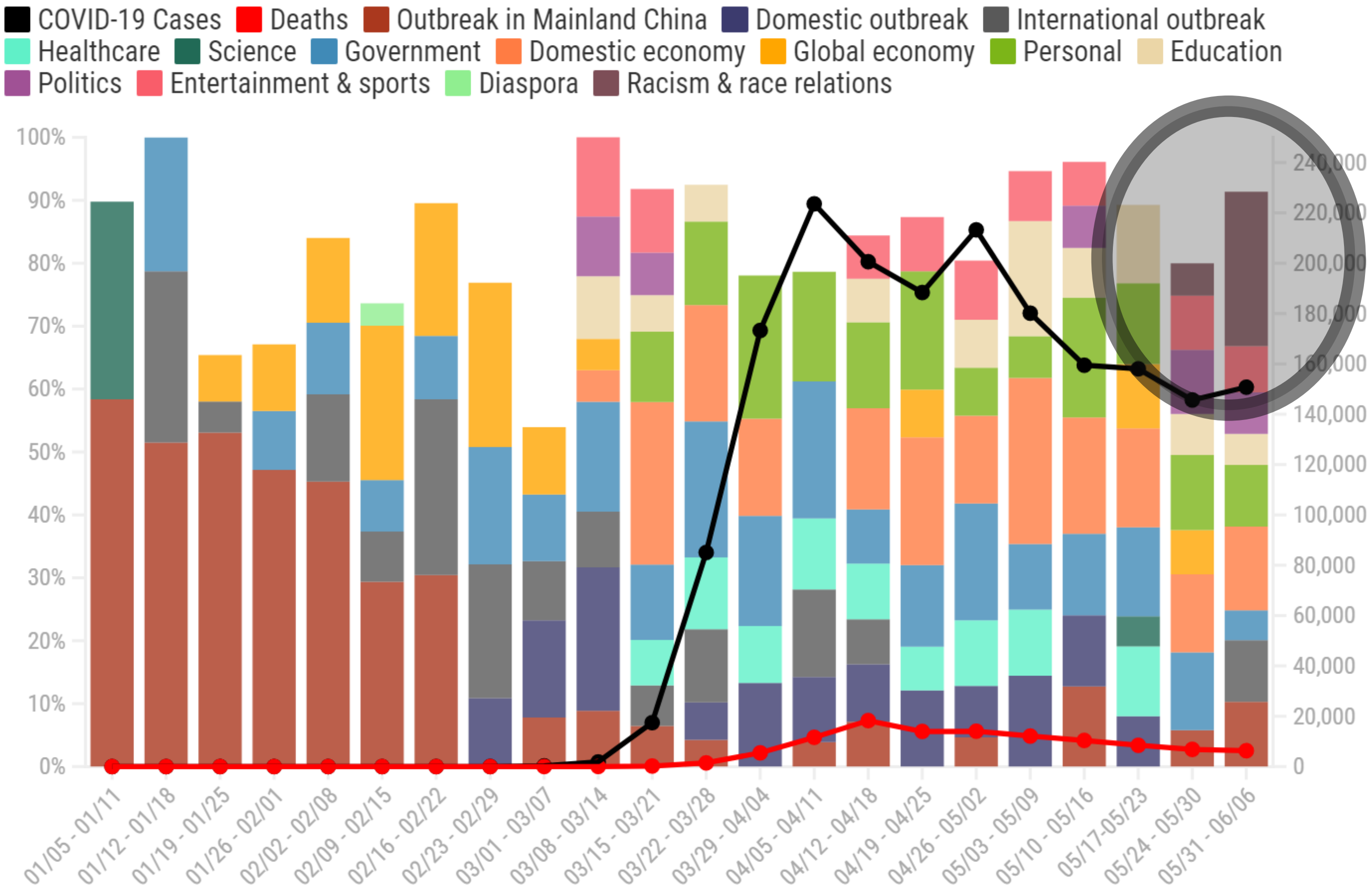
Outbreak in Mainland China

BOSTON UNIVERSITY



- Personal Preventative Actions
- Domestic Economy

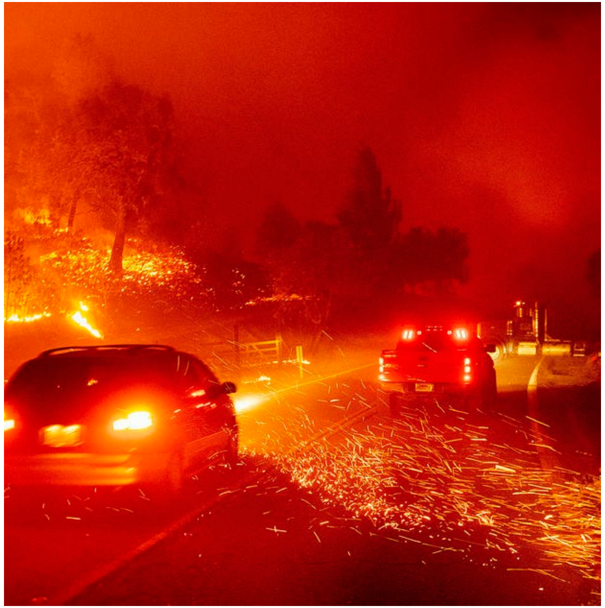
The U.S. News Topics and New COVID-19 Cases by Week



Racism & race relations

BOSTON UNIVERSITY

AI for News Analysis: What emotions do Headlines and Images Cause with News Readers?



Emotion:

Image by itself:
Fear

Headline:
Fear

This will only get worse: California fires linked to climate change

Emotion:

Image by itself:
Curiosity

Headline:
Disapproval
Admiration

A topic we cannot sugarcoat: Waukegan students join peers in global climate strike

Our ACII 2023 project:

News headline	Human - News Image	AI - News Image
Jacksonville mass killing once again proves the left's gun control 'solution' is a fleeting illusion		
Climate change protests snarl DC traffic as bizarre scenes unfold in capital		

Emotion from 🤖 image-only	**Emotion from 🤖 image + headline**
Admiration	Fear
	 Rising emissions could drain foods like rice and wheat of their nutrients, causing a slow-moving global food crisis
Emotion from 👩🏻 image-only	**Emotion from 👩🏻 image + headline**
Admiration	Fear
	 Dems say GOP focus on mental health is redirection from gun control

"Climate change is altering the nation's environment and the microbes, viruses and insects that inhabit it, potentially increasing where diseases are.

“House Dems are promising tougher gun control measures, but advocates may have lost ground in the Senate”

"To survive global warming, Mojave Desert birds will need a lot more water — and they probably won't get it"

“Florida sheriff: The Parkland massacre could have been avoided if someone else had a gun"

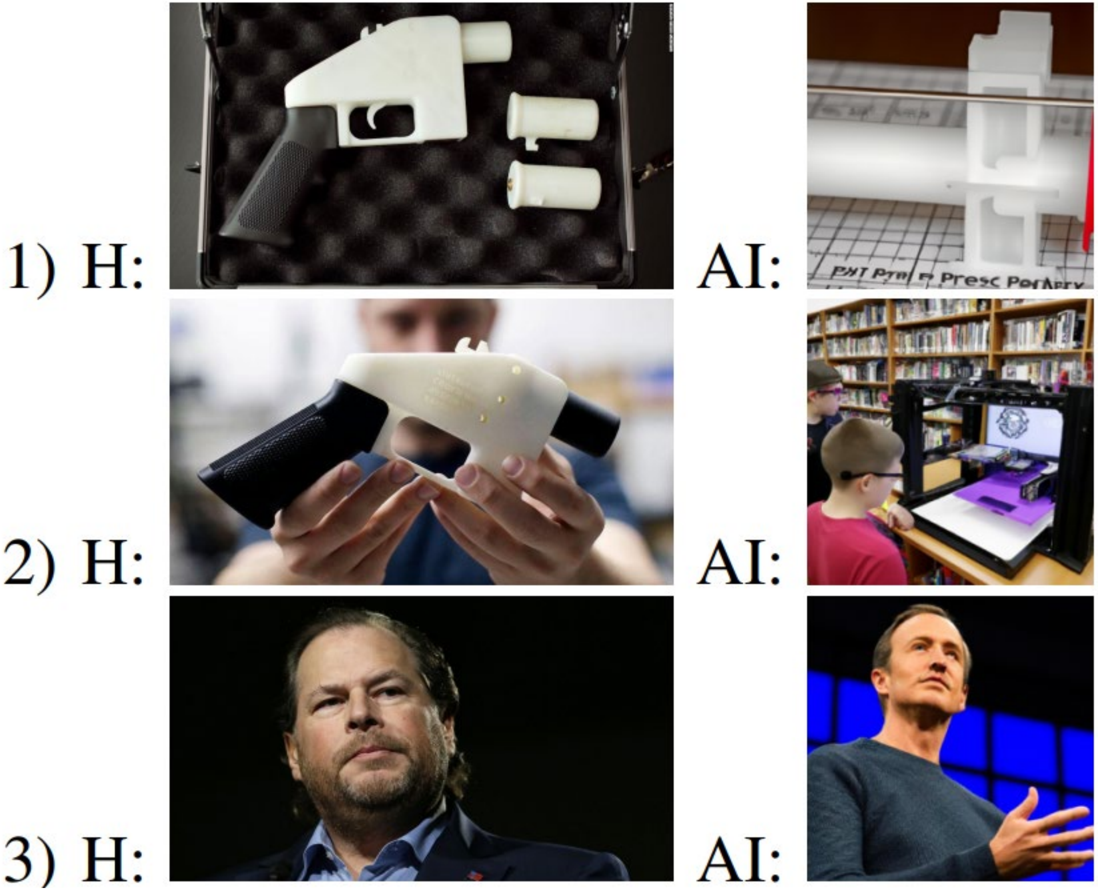


Fig. 6: Depiction of objects and figures. Human (left) and AI (right) provided images. 1) & 2) News images of 3D-printed guns are common. AI instead shows the 3D printing process (row 1) or other context (row 2, headline “Chicago, suburban libraries brace for the question: Can I print out a 3D gun?” 3) AI generates a fake, yet stereotypical and real-looking person when the headline contains phrasing such “tech mogul” and/or “CEO.” These phrases elicited a male figure, with a hand gesture similar to Steve Jobs’ gestures, and the camera angle tilted upward.

Hammer & Nail

- For computer scientists:

➤ Beat the benchmark *vs.* solve the real problem

- For communication researchers:

➤ Expand the scope of research