



Lecture 5:

Research on Digital Media

Seminar 'Foundations of Data Science'

Prof. Dr. Karsten Donnay, Assistant: Marcel Blum



Course Outline

- Part 1: Foundations
 - *Day 1: Mon. 14.06.2021:* Information Coding & Data
 - *Day 2: Tue. 15.06.2021:* Programming & Algorithms
 - *Day 3: Wed. 16.06.2021:* Complexity & Efficiency
- Part 2: Applications
 - *Day 4: Thu. 17.06.2021:* Data Collection & Quality
 - ***Day 5: Fri. 18.06.2021: Research on Digital Media***



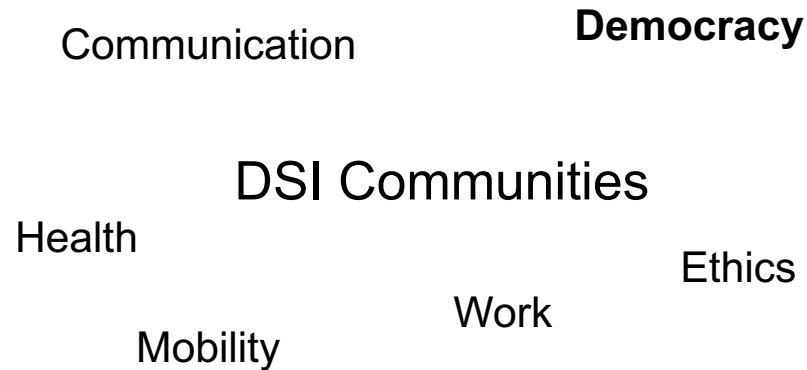
Overview of this Session

- Illustrate real data science applications in research on digital media
 - Examples from my past and ongoing work
 - Background
 - Prior Research
 - WIN Project
 - Spark your curiosity for pursuing this kind of work
 - Many of you enrolled in “Political Data Journalism”
 - Media coverage more salient influence than ever
 - Role of digital media has only grown during COVID



Digital Society Initiative (DSI)

Interdisciplinary Competence Center on Digitalization



Digital Democracy Lab

digdemlab.io



- Political implications of digital technology
- Researchers from ETH, UZH, Hertie (Berlin)
- Research infrastructure for data collection and analysis

One of 16 new **DSI professors**

- Assistant Professor in the Department of Political Science (since April 2020)
- Affiliated with the DSI and its new graduate program

Research Area “Political Behavior and Digital Media”

- Substantive research and development of new methods for analysis of online behavior and implications for politics
- Details: www.ipz.uzh.ch/en/forschung/lehrstuehle/bdm.html



**University of
Zurich** ^{UZH}

Department of Political Science

Background

Framing of News Matters...

freedom fighters

Crimean crisis, 2015

terrorists

war in Iraq

Iraq war, 2003

war on Iraq



White People "FIND"



Two residents waded through chest-deep water after **finding** bread and soda from a local grocery store after Hurricane Katrina came through the area in New Orleans, Louisiana.

(AFP/Getty Images/Chris Graythen)

Black People "LOOT"



A young man walks through chest deep flood water after **looting** a grocery store in New Orleans on Tuesday, Aug. 30, 2005. Flood waters continue to rise in New Orleans after Hurricane Katrina did extensive damage when it **Ap** Associated Press

Hurricane Katrina, 2005



Jewel Samad/Getty

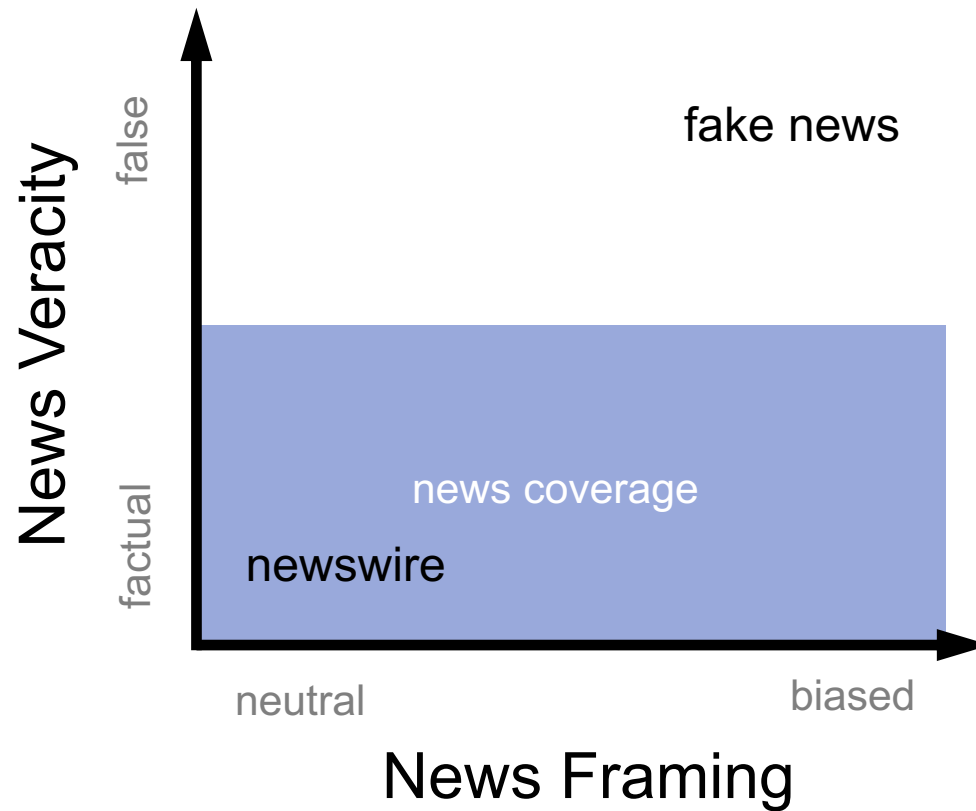
G7 summit, 2014



Jewel Samad/Getty



Framing vs. Fake News



Fake News

Collins English Dictionary

“false, often sensational,
information disseminated
under the guise of
news reporting”



Framing vs. Opinions

- There is a vibrant current debate about the role of the media
 - Media does not have to be neutral
 - Opinions on current events are valuable
 - Journalists can and should take a stance
 - Question is how it is done
 - Clearly marked opinion pieces
 - Stated opinions as part of coverage
 - Framing instead subtly transports opinion
 - Reflects own (un-)intentional biases
 - May not be clearly noticeable to readers but still influences them
 - Extreme case: “spinning” a (news) story



Relevant Academic Research I

Selection/omission
of news topics

News framing
and its impact

Political relevance
of news coverage

Social Sciences

Computer Science

Reader
perceptions

Social influence
online/offline

Information
echo chambers

Conceptual Model of News Production Process

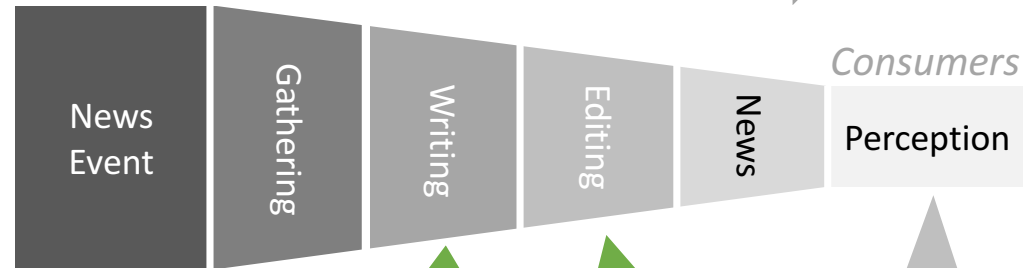
Constraints



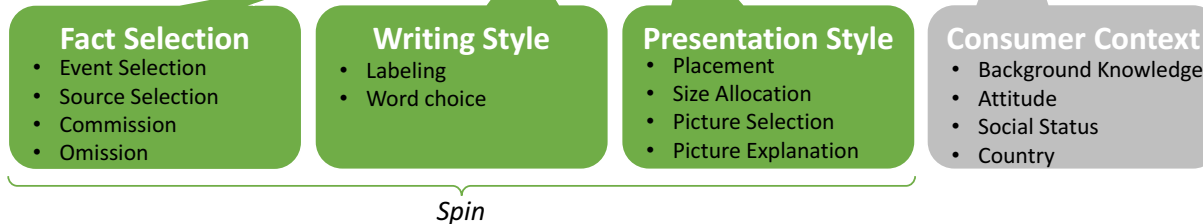
Political Interest
Reputation
...

Business Interest
Funding
...

Reality

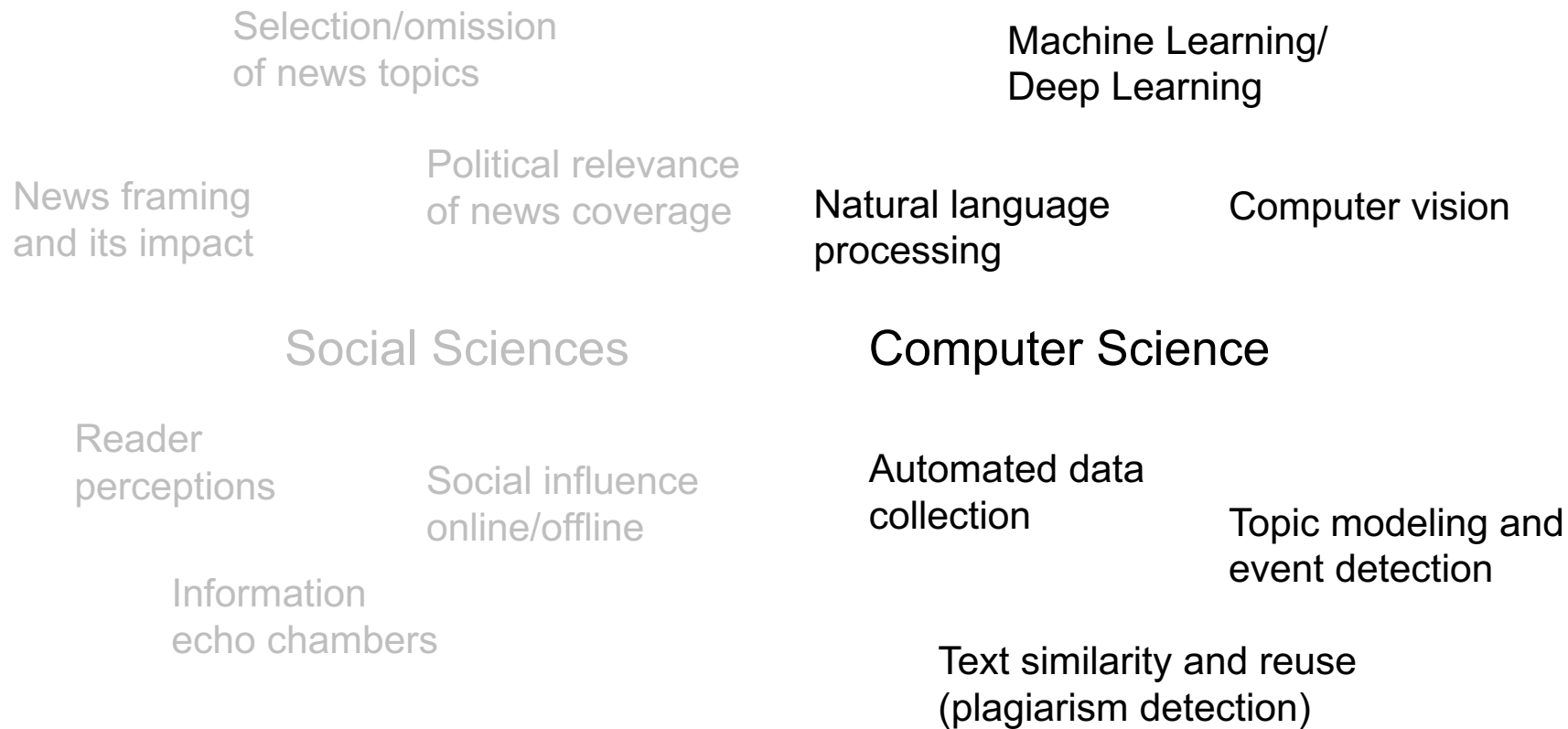


Decisions





Relevant Academic Research II





**University of
Zurich** ^{UZH}

Department of Political Science

Prior Research



Biased Information Environments

- News Media
 - Framing affects how we perceive factual information
 - Look at two key dimensions
 - Tone of coverage
 - Choice of images
- Social Media
 - Information sharing online can be partisan and biased
 - Are social media users aware of those potential biases?
- Biased information environments affect political behavior
 - Degree of political engagement
 - Voting behavior, policy implementation etc.



Media Bias

- Framing by word choice (= article tone) and labeling

freedom fighters

Crimean crisis, 2015

terrorists

immigrant

European refugee crisis, 2015

economic migrant

undocumented immigrants

US President Trump ended DACA, 2017

illegal aliens

labels

war in Iraq

Iraq war, 2003

war on Iraq

U.N. arms inspectors said they had
withdrawn two U-2 reconnaissance
planes over Iraq for safety reasons

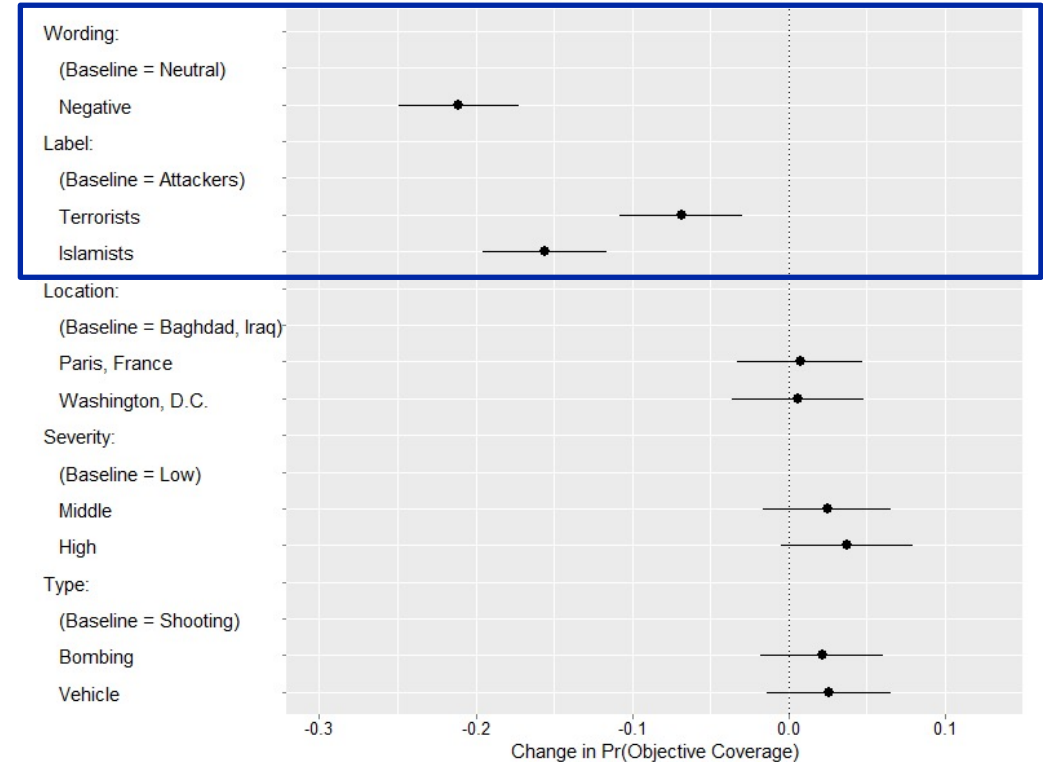
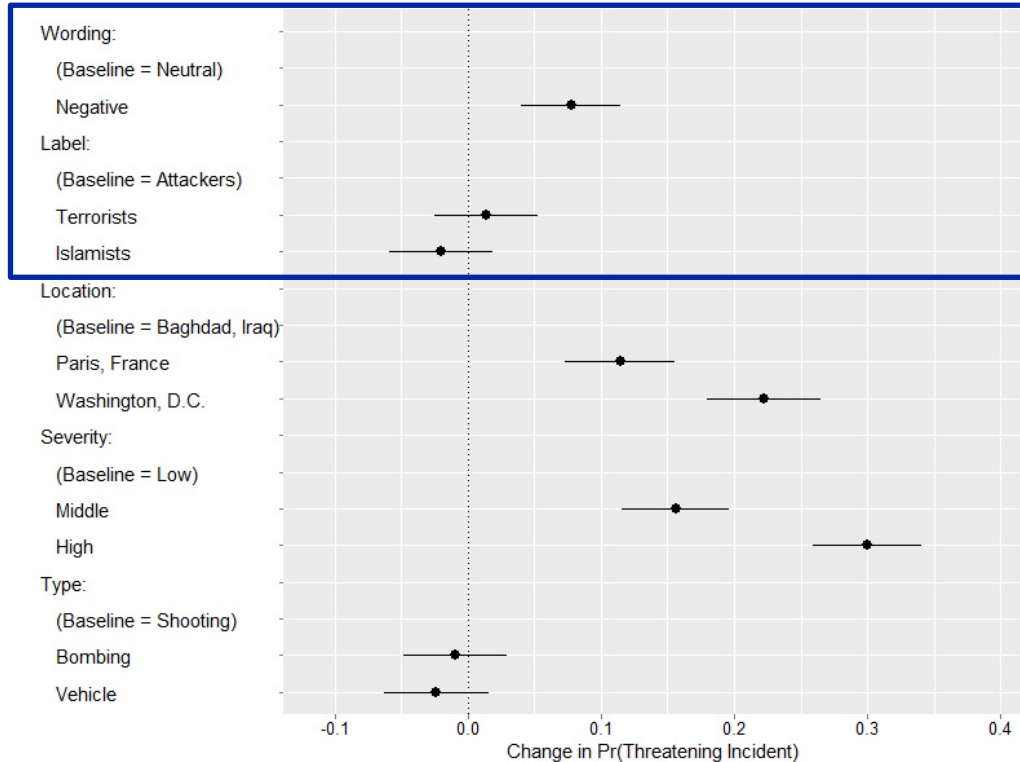
Iraq war, 2003

Iraqi fighter jets threatened two
American U-2 surveillance planes,
forcing them to abort their mission
and to return.

word choice

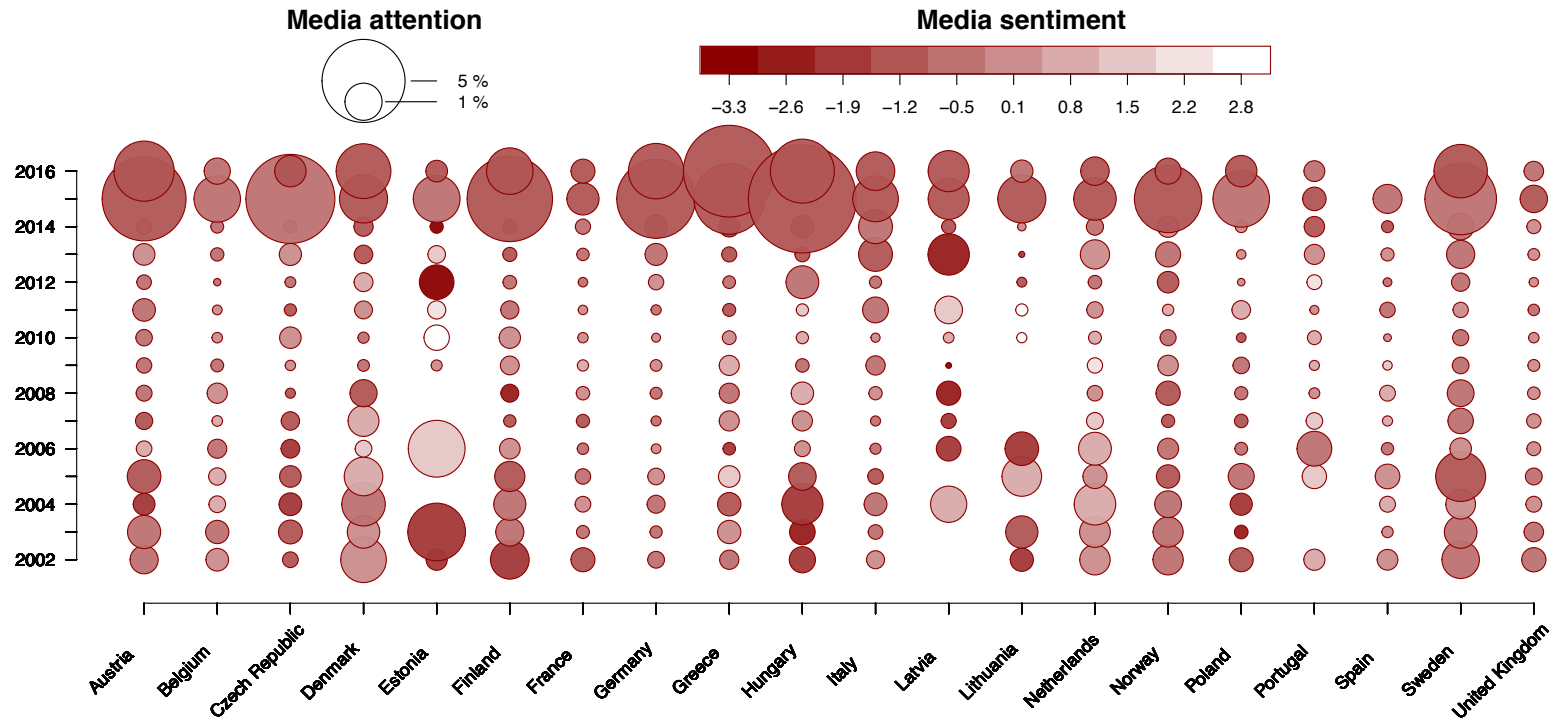
Media Bias

– Subtle word choice vs. recognizable labels



Media Bias

- Negative article tone can affect policy



Media Bias

- Choice of Images Matters: On-the-ground vs. media coverage of Charlottesville on left and right
 - August 11, 2017 Unite the Right rally in Charlottesville, Virginia
 - James Fields murders Heather Heyer in car attack on protestors that injures 28 more



Image 1



Image 2



Image 3



Image 4



Image 5



Image 6



Image 7



Image 8



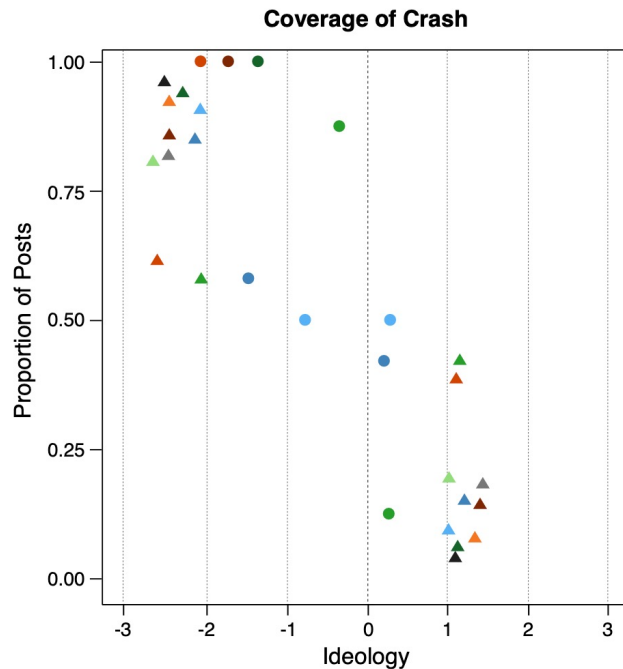
Image 9



Image 10

Media Bias

- Usage of images from car attack: media vs. social media, left vs. right



Account Type:

○ Media △ Non-Media

Images:

Image 1 Image 2 Image 3 Image 4 Image 5 Image 6 Image 7 Image 8 Image 9 Image 10

- Coverage of media vs. social media differs
 - Media crash coverage selectively omits
 - Chaos in crowd (image 6)
 - Rescue workers and injured (image 7)
 - Differences greater for news on the right
 - Show even less images
 - Pattern holds for broader coverage, largely showed official shots of president
- Coverage on social media also differs
 - Image sharing heavily dominated from left



Awareness of Bias

- Information sharing online is partisan
 - Whom you get information from matters because of the implicit selection of what information they share/ you see
 - Echo chambers, i.e., information environment is biased towards your own views
 - Knowing how your information environment looks like is crucial
- Social Media (Twitter)
 - Whom you follows, defines your information environment
 - Idea: measure your own perceptions of political leanings of your friends vs. reality
 - Linked survey and Twitter API measurement:
www.newsfeedanalysis.app

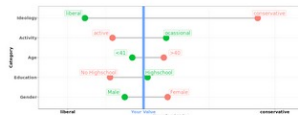
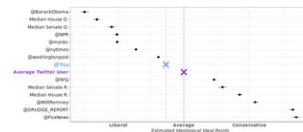
Awareness of Bias

TWITTER NEWSFEED ANALYSIS

Home About Methodology Results

Get an analysis of your Twitter newsfeed!

Are you interested in the overall **ideological slant** of the Tweets you receive? Do you want to know how your newsfeed compares to those of **major political figures**?



With the demographic information you provide we can also tell you how your newsfeed compares to users of similar **age** or the same **gender**. This **non-profit Ph.D.-research project** provides these statistics and more **for free**. Head over to **the app** and see for yourself!

Analyze my Twitter-newsfeed

© Philipp Kling 1919

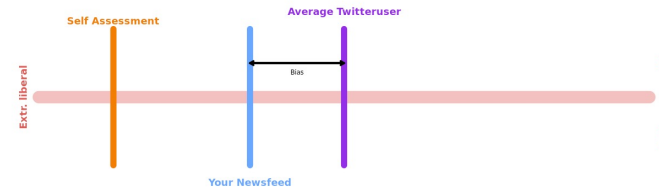
Share

www.newsfeedanalysis.app

TWITTER NEWSFEED ANALYSIS

Your Results Pol. Elites Peers Suggestions Methodology

Hello karstendonnay! The following graphic visualizes your basic results. The **blue bar** represents the **numeric result** for your Twitter newsfeed, while the **orange bar** signals your **self assessment** provided by you in the survey. To have a reference, the **violet bar** represents the **average Twitter user**.



According to the **analysis of the political U.S. elites you are following** - indicated by the **blue bar** -, the algorithm calculated a value that suggests your newsfeed is leaning towards a **liberal** direction. That means your newsfeed is biased towards the political point of view of **Democrats**.

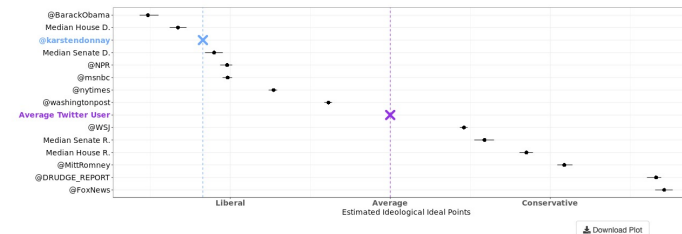
Compare to political elites!

[newsfeedanalysis.app](#)
Share

TWITTER NEWSFEED ANALYSIS

Your Results **Pol. Elites** Peers Suggestions Methodology

This graphic displays indicators and certainty estimates for various **U.S. political elites**, the **average Twitter-user** and your **calculated value** from the **Results-page**.



Compare to similar users!



**University of
Zurich** ^{UZH}

Department of Political Science

WIN Project

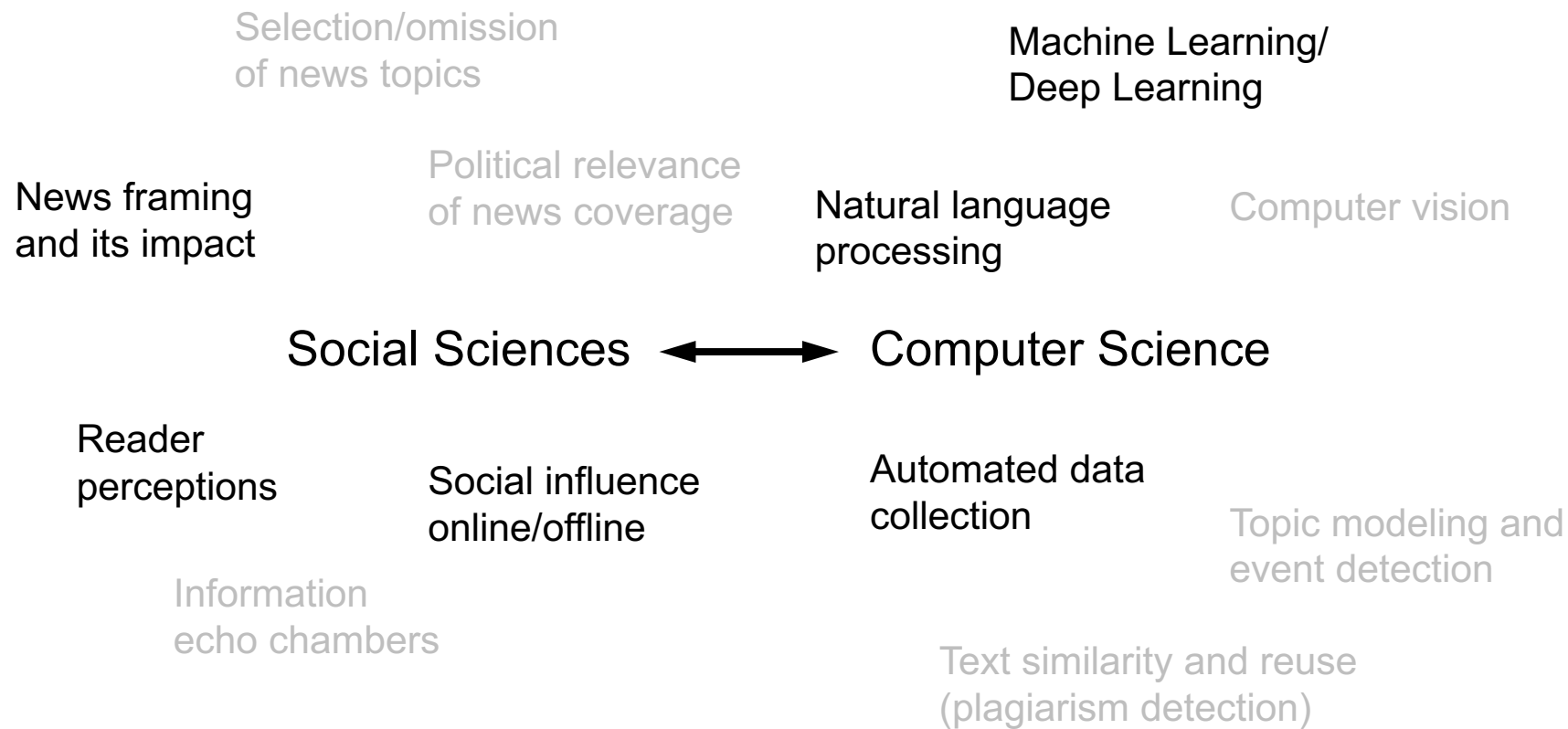


WIN Project

- “Fake News and Collective Decision Making – Rapid Automated Assessment of Media Bias”
 - Three-year project funded by Heidelberger Akademie der Wissenschaften
 - Interdisciplinary collaboration with team of computer scientists
 - Project team:
 - Gipp Team (Wuppertal) – dke.uni-wuppertal.de
Felix Hamborg, Anastasia Zhukova, Bela Gipp, Moritz Bock
 - Donnay Team (Zürich/Konstanz) – ipz.uzh.ch
Timo Spinde, Angelica Becerra, Marc Lüttecke
 - More information: <https://www.karstendonnay.net/research/projects/win/>



Contribution of this Project



Contribution of this Project – Two Pillars

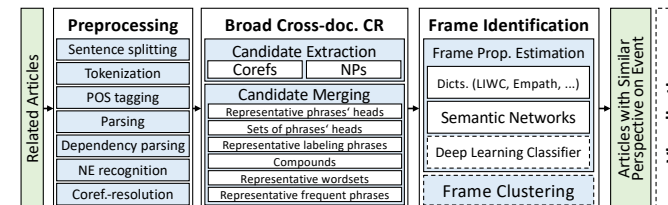
Testing the impact of
readers' perceptions

Automated classification
of degree of bias in coverage

Social Sciences ↔ Computer Science

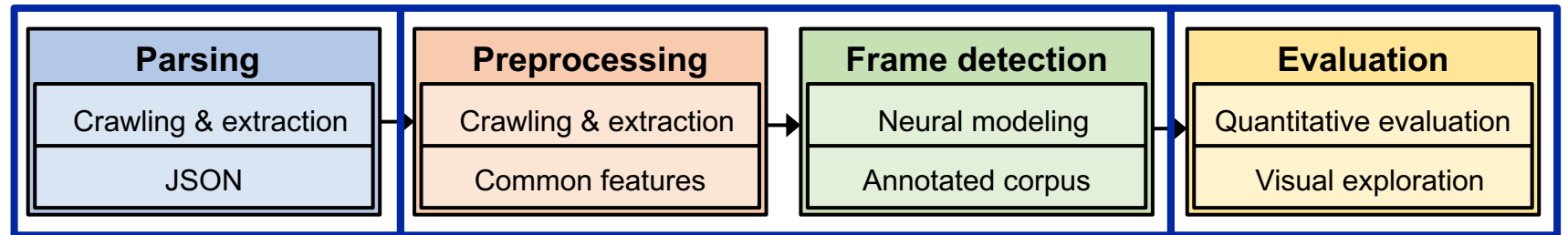
Noncitizen voter fraud allegations
According to a 2012 Pew Research Center (Pew) report, which Donald Trump has cited as proof of voter fraud, millions of voter registration records were out of date as people were either deceased or had moved.

<p>From the left</p> <p>FOLLOWING TRUMP VOTER FRAUD ALLEGATIONS, CLAIM THAT 5.7 MILLION NONCITIZENS VOTED IS WRONG</p> <p>President Donald Trump's unfounded allegations that millions voted illegally in 2016 is back in the news, with...</p> <p>Source: Tampa Bay Times</p> <p>Strength of slanted language bar: </p>	<p>From the center</p> <p>Report: Trump commission did not find widespread voter fraud</p> <p>PORTLAND, Maine (AP) — The now-disbanded voting integrity commission launched by the Trump ...</p> <p>Source: Associated Press</p> <p>Strength of slanted language bar: </p>	<p>From the right</p> <p>Hillary Says Voter Registration Cost Her the Election</p> <p>President Trump was criticized last year when he tweeted out an accusation that he would have won the popular vote were it ...</p> <p>Source: National Review</p> <p>Strength of slanted language bar: </p>
---	--	--



Technical Approach

- Overview of processing pipeline



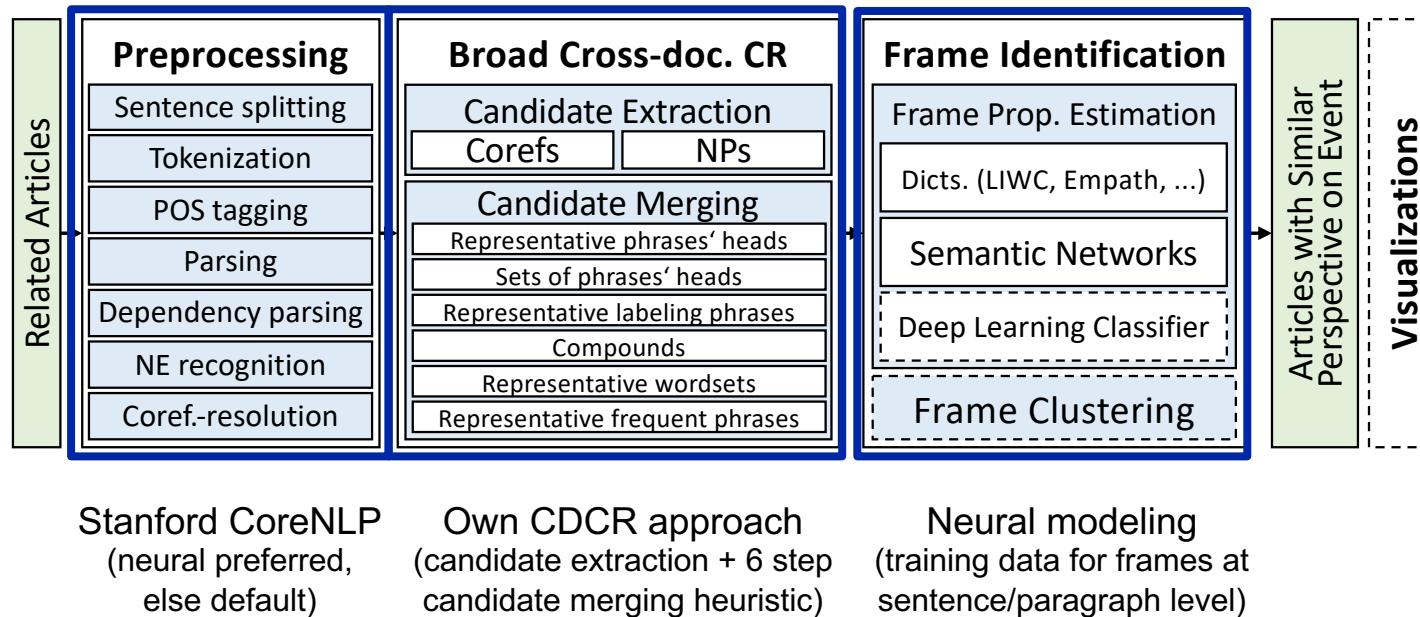
Online news media
(extraction of current
articles across domains)

Automated frame annotation pipeline

NewsWCL50 Data
(Hamborg, Zhukova, Gipp 2019 JCDL)

Technical Approach – Details

- Automated frame annotation pipeline



Hamborg, Zhukova, Donnay Gipp 2020 JCDL



Target-dependent Sentiment

– NewsTSC data

In the sentence below, what do you think is the affection of the sentence's author towards the underlined target? Consider the affection only towards the underlined target, not the event itself or other people.

Target: Mr. Turnbull

Pressed about Mr. Trump's tone, and whether the president ended the call by hanging up, Mr. Turnbull refused to comment.

The attitude of the sentence's author towards the underlined target is...

☐ **strong dislike** ☐ ☐ ☒ **neutral / can't tell** ☐ ☐ **strong affection**

Class frequencies of NewsTSC sets

	negative	neutral	positive	total
training	530	1600	171	2301
test	167	487	47	701
total	697	2087	218	3002

Hamborg, Donnay, Gipp (2021) iConf

Target-dependent Sentiment

- Fine-tuned BERT model for NewsTSC

LM	Method	AvgRec	acc	F1 _m	F1 _{pn}
base	AEN-BERT	59.7	62.9	55.0	47.3
	BERT-SPC	62.1	62.1	53.3	44.9
	LCF-BERT	67.3	61.3	54.4	46.5
news	AEN-BERT	59.8	62.9	54.5	46.2
	BERT-SPC	66.7	63.5	55.0	45.8
	LCF-BERT	69.8	66.0	58.8	51.4

State-of-the-art TSC methods:

AEN-BERT (Song et al., 2019)

BERT-SPC (Devlin et al., 2018)

LCF-BERT (Zeng et al., 2019)

(developed on Laptop, Restaurant, Twitter TSC data)

Hamborg, Donnay, Gipp (2021) iConf



Ground-truth Annotation

Structural Challenges

- High class imbalance
- No academic benchmark dataset in news domain
- Neural model requires very large datasets
- Proper sampling across domains and news sources

Coding Challenges

- Attitude of coder towards target might affect coding
- Low attention of coders on MTurk
- ICR not very high for all coding tasks
- Difficult to generalize beyond sentiment (“frame properties”)



Ground-truth Annotation

- Building a new comprehensive **gold standard** dataset for **frame detection**
 - 10,000s of instances annotated on Amazon MTurk
 - Optimized coding protocol to minimize annotation bias
 - Only consider high quality coding (4/5 or 5/5 agreement)
 - Balanced across classes and news domains
 - Articles from across the US political spectrum
 - See also new POLUSA dataset: <https://zenodo.org/record/3813664>
 - Systematic sampling of paragraphs/sentences from many articles
 - Validation of MTurk coding against human coders
 - 5 IPZ students performing same task on subset of sentences
 - Provide benchmark comparison of online annotation quality

Experimental Design I

- News aggregator with bias indication similar to [allsides.com](https://www.allsides.com)



Noncitizen voter fraud allegations

According to a 2012 Pew Research Center (Pew) report, which Donald Trump has cited as proof of voter fraud, millions of voter registration records were out of date as people were either deceased or had moved.

From the left

FOLLOWING TRUMP VOTER FRAUD ALLEGATIONS, CLAIM THAT 5.7 MILLION NONCITIZENS VOTED IS WRONG

President Donald Trump's unfounded allegations that millions voted illegally in 2016 is back in the news, with...

Source: Tampa Bay Times

Strength of slanted language bar:



From the center

Report: Trump commission did not find widespread voter fraud

PORTLAND, Maine (AP) — The now-disbanded voting integrity commission launched by the Trump ...

Source: Associated Press

Strength of slanted language bar:



From the right

Hillary Says Voter Registration Cost Her the Election

President Trump was criticized last year when he tweeted out an accusation that he would have won the popular vote were it ...

Source: National Review

Strength of slanted language bar:



Experimental Design II

– Text level annotation of framing of targeted named entity

The boxes in the following article are showing how named entities (such as persons or institutions) are framed. Red color illustrates negative framing, green color illustrates positive framing. If you hover over the grey text, the box that it relates to is highlighted.

FOLLOWING TRUMP VOTER FRAUD ALLEGATIONS,
CLAIM THAT 5.7 MILLION NONCITIZENS VOTED IS
WRONG

President Donald Trump's unfounded allegations that millions voted illegally in 2016 is back in the news, with his supporters pointing to a new analysis that claims millions of undocumented immigrants voted in 2008. Fox and Friends co-host Ainsley Earhardt talked about it on the morning show recently. "5.7 million -- that's how many illegal immigrants might have voted" in 2008, she said. Her comments referenced an article in the Washington Times, a conservative newspaper. Trump has made repeated claims about massive voter fraud and election rigging, which we've debunked again and again and again and again and again and again and again (and we debunked a claim by his spokesman Sean Spicer). The claim made on Fox and Friends is based on an extrapolation of a controversial study that relied on a very small number of responses. Researchers involved in the underlying survey of voters have cautioned against using their data to reach conclusions about noncitizen voters. We emailed a spokeswoman for Fox News and did not get a reply; however, the Washington Times article showed that the information came from Just Facts, a think tank that describes itself as conservative/libertarian and was founded by James D. Agresti, a mechanical engineer in New Jersey.

Online survey experiment

- Conjoint experimental design with variation of
 - Overview
 - Target annotation
 - Type of annotation
- Subjects see overview and then one article
- Questions about perceptions of news reporting after each article
- Clean inference from full randomization of variations



**University of
Zurich** ^{UZH}

Department of Political Science

Course Administration



Reminder – Final Project

- Idea:
 - Work on a topic you are interested in or that is related to your research/thesis
 - Identify a concrete **research question** you are trying to address
 - Apply the lessons-learned from this class in the context of this question
- Format:
 - Full data science pipeline from initial data collection/processing to data wrangling, extracting relevant information, analysis and presentation of results
 - Should also be done in Markdown and doubles as your project report; emphasis is on coding but we also expect
 - Motivation for your research question (& short overview of related research)
 - Text narrative that leads through all parts of the data science pipeline
 - Discussion of results and potential shortcomings



Questions & Feedback

- Please reach out to us if you have any questions (or use the forum)
 - Working on your final project next week
 - But also afterwards if you continue to work with data science approaches
- We are looking forward to your feedback and comments
 - The course evaluation is now live and you should have received an invitation
 - There will be time to fill this out during the exercise this afternoon but it also stays open until June 30, 2021
 - The evaluation form might be a bit different this term again given the online-only teaching format..



Up Next

- Exercise (this afternoon)
 - Course Evaluation
 - Accessing APIs
 - Q&A Final Project