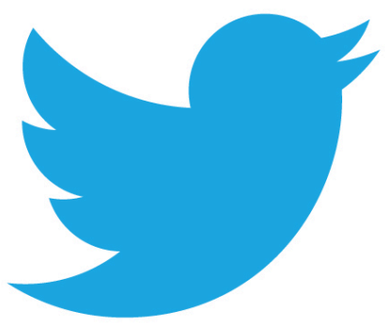
Quantifying Search Bias: Investigating Sources of Bias for Political Searches in Social Media

Muhammad Bilal Zafar

with Juhi Kulshrestha, Motahhare Eslami, Johnnatan Messias, Saptarshi Ghosh, Krishna Gummadi and Karrie Karahalios

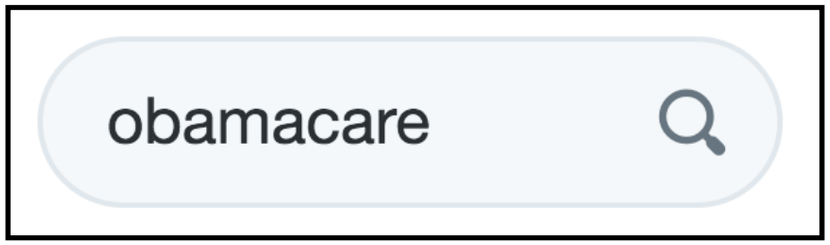
Social media as search platform

- Two-thirds of Americans on social media use it to get news (informational queries)



Social media as search platform

- Two-thirds of Americans on social media use it to get news (informational queries)



Social media as search platform

- Two-thirds of Americans on social media use it to get news (informational queries)

Ranked list (importance)

Potential bias in search results

obamacare

Steven Crowder
@scrowder

Follow

Ouch => Tucker Carlson Obliterates ObamaCare Architect: 'Nobody believes you, doctor'

Potential bias in search results

obamacare

Steven Crowder
@scrowder

Follow

Ouch => Tucker Carlson Obliterates ObamaCare Architect: 'Nobody believes you, doctor'

mia farrow
@MiaFarrow

Follow

My daughter in law is a pediatric cardiologist in a Detroit hosp. Says Obamacare is a lifeline for her patients

Search rankings can shape user opinion

- Users place greater trust in higher ranked items *[Pan et al., 2007]*
- Biased search results can influence voting patterns *[Epstein & Robertson, 2015]*

This talk:

Study sources of bias in social media search

This talk: Study sources of bias in social media search

- Identifying sources of search bias
- Quantifying bias in each source
- Studying bias in political searches in Twitter

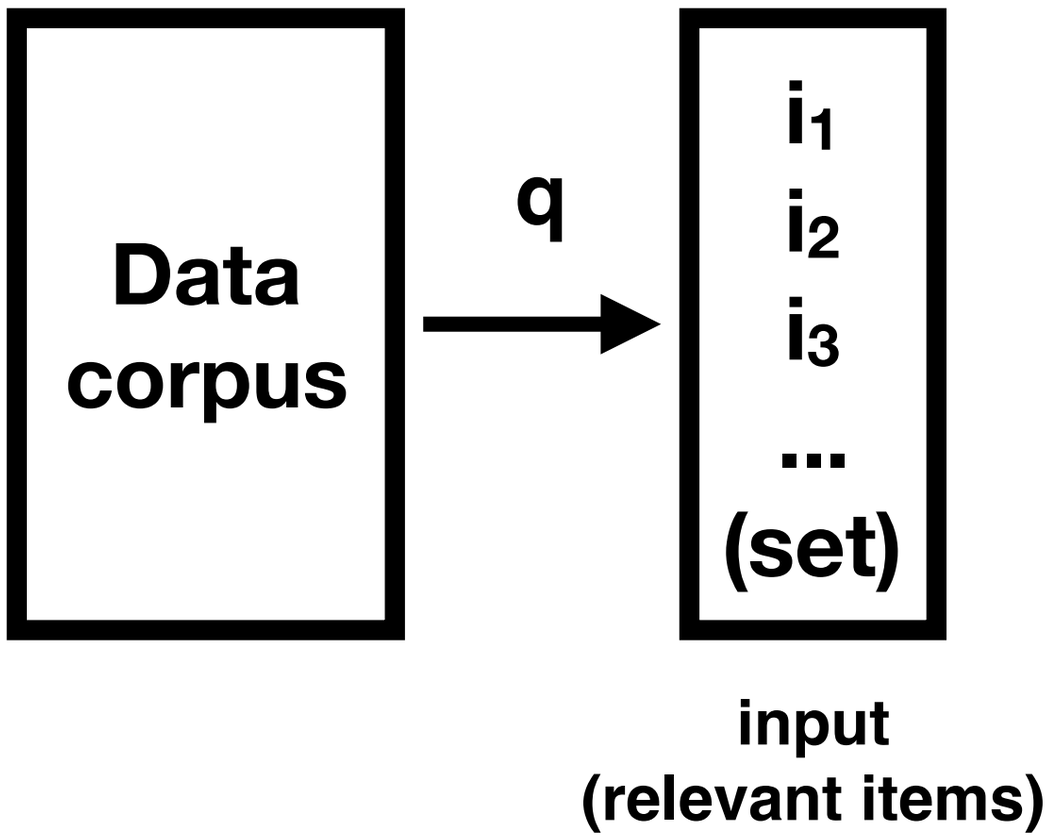
This talk: Study sources of bias in social media search

- **Identifying sources of search bias**
- Quantifying bias in each source
- Studying bias in political searches in Twitter

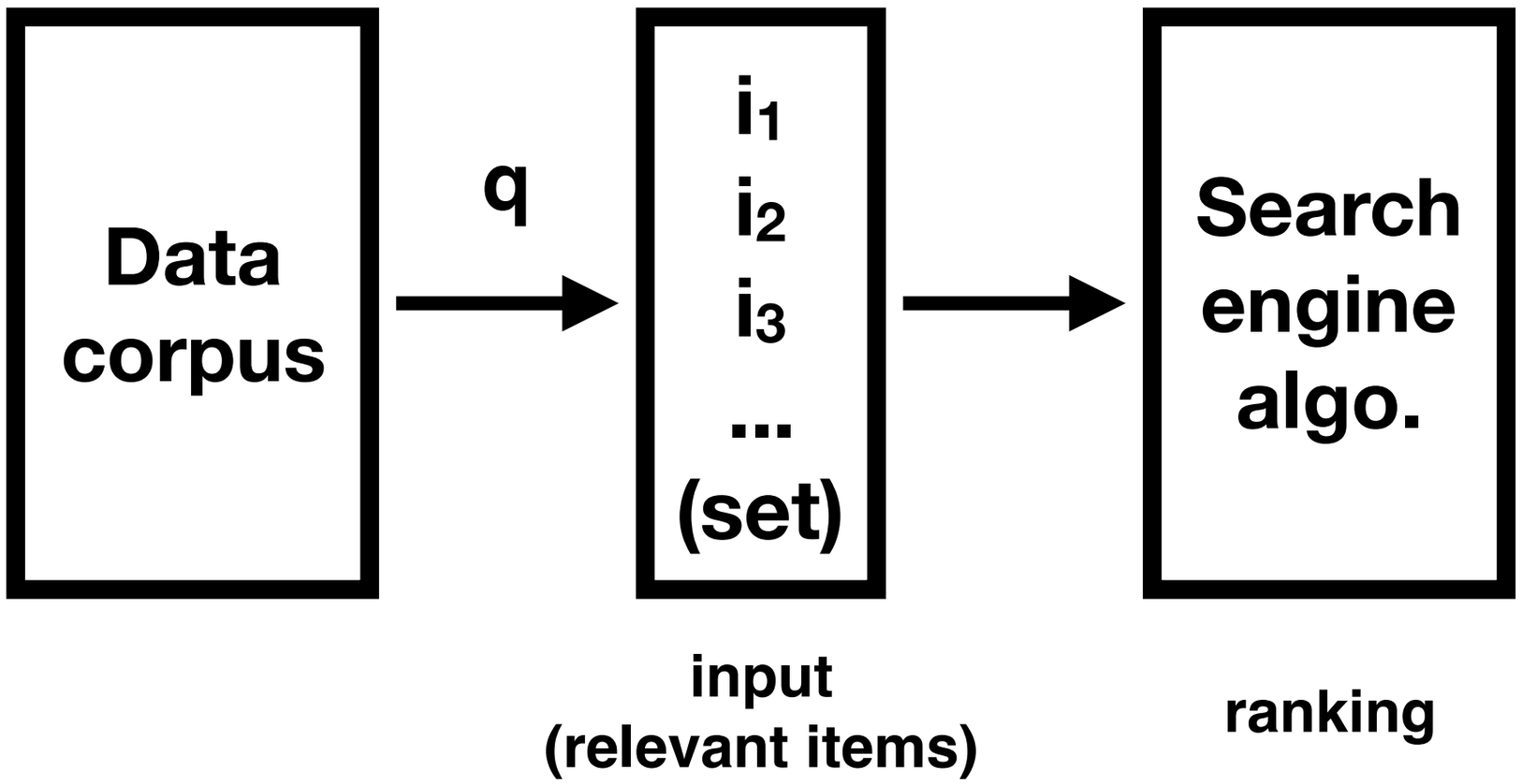
Social media search engine design

Data corpus

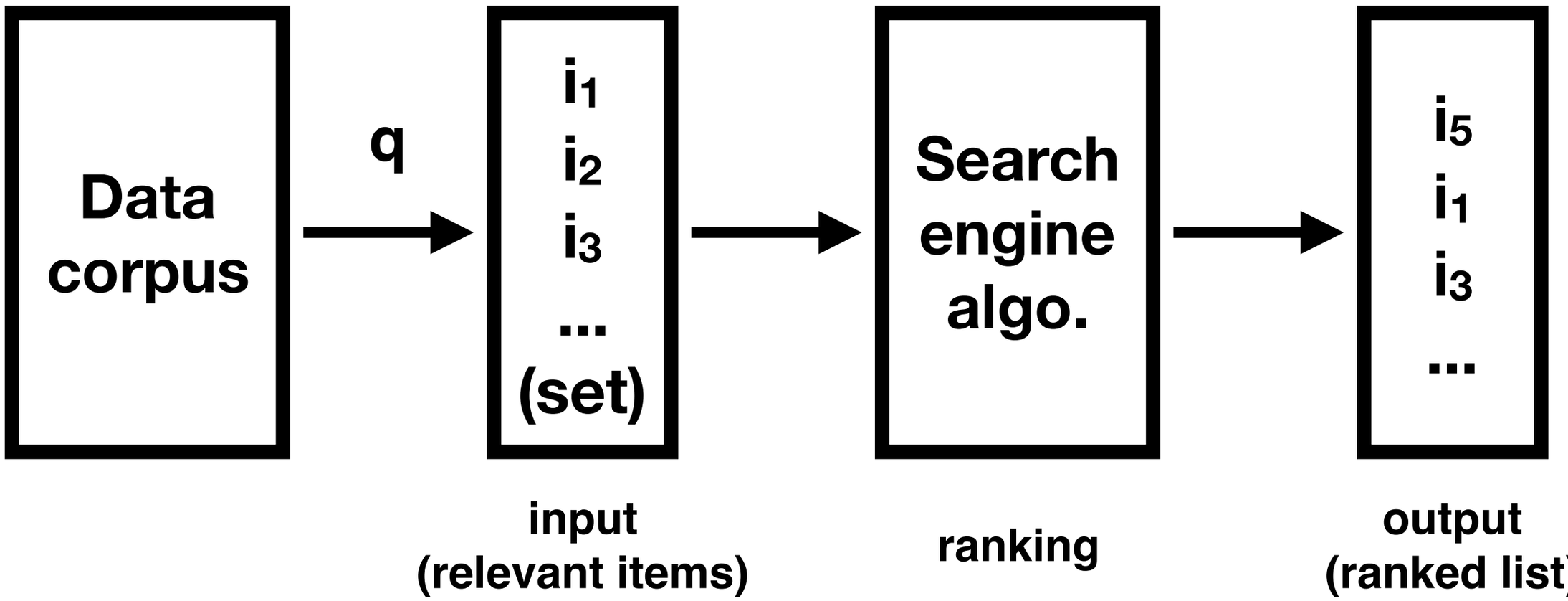
Social media search engine design



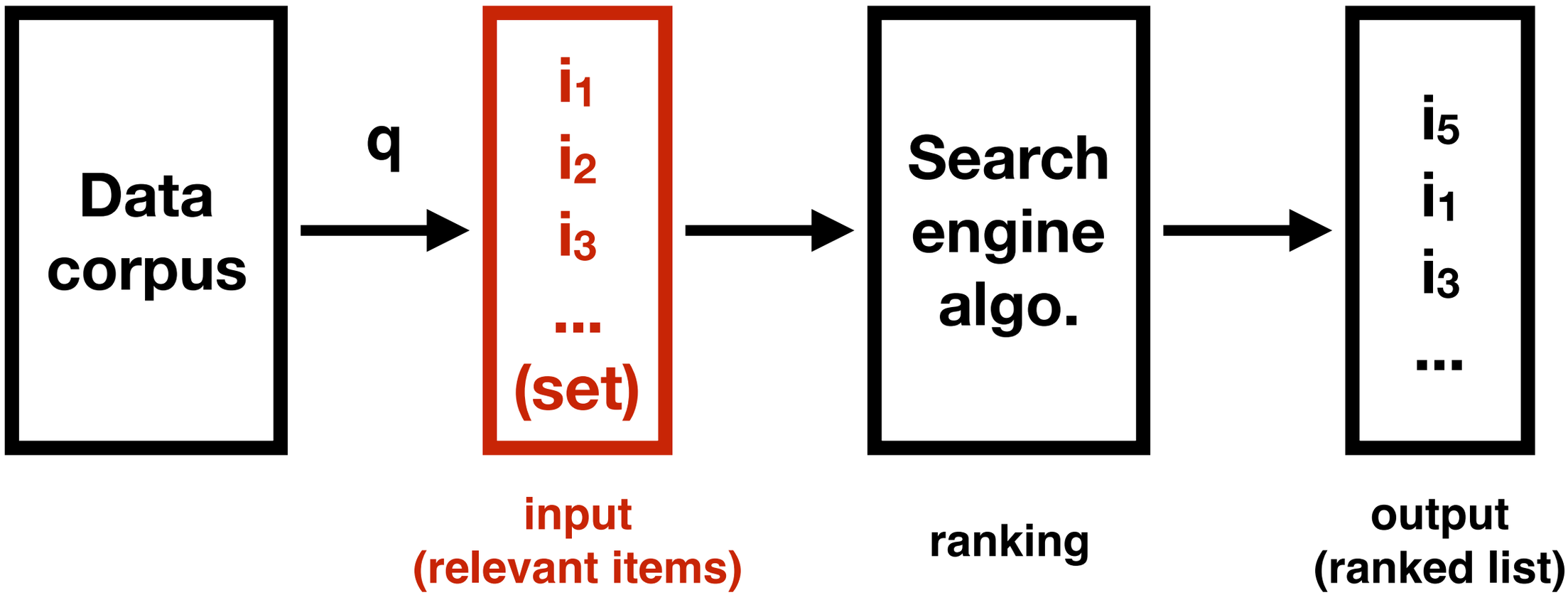
Social media search engine design



Social media search engine design

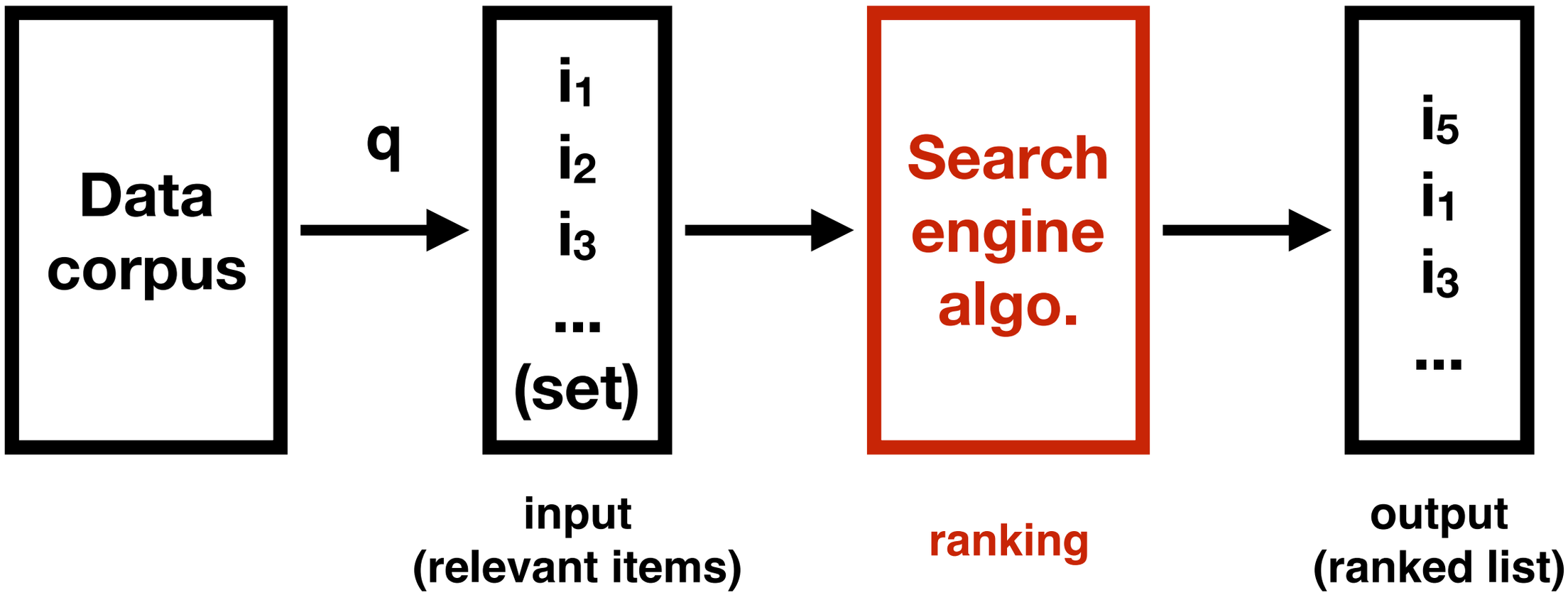


Social media search engine design



- Output bias may stem from
  - Bias in the (input) relevant item set

Social media search engine design

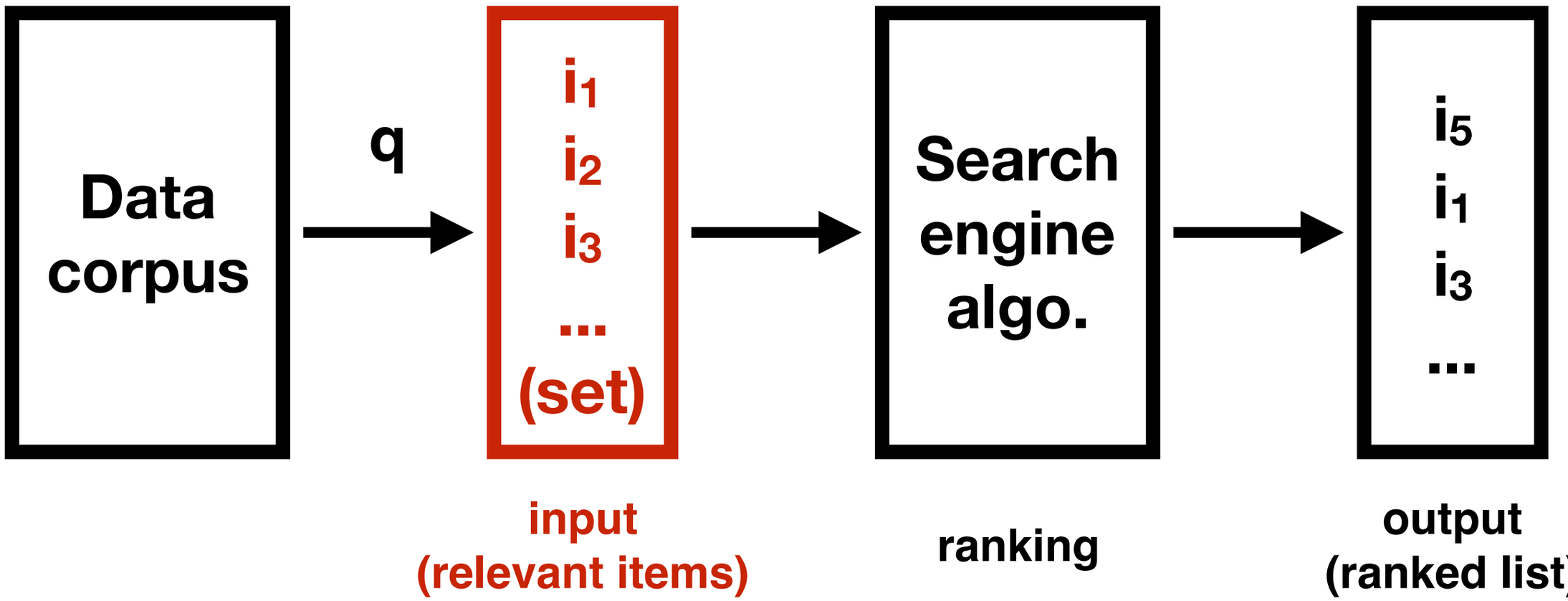


- Output bias may stem from
  - Bias in the (input) relevant item set
  - Bias introduced by the ranking algorithm

This talk: Study sources of bias in social media search

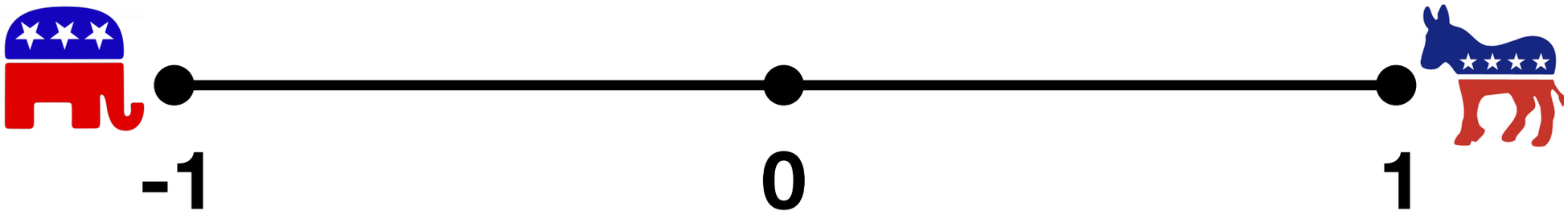
- Identifying sources of search bias
- **Quantifying bias in each source**
- Studying bias in political searches in Twitter

Quantifying bias in each source

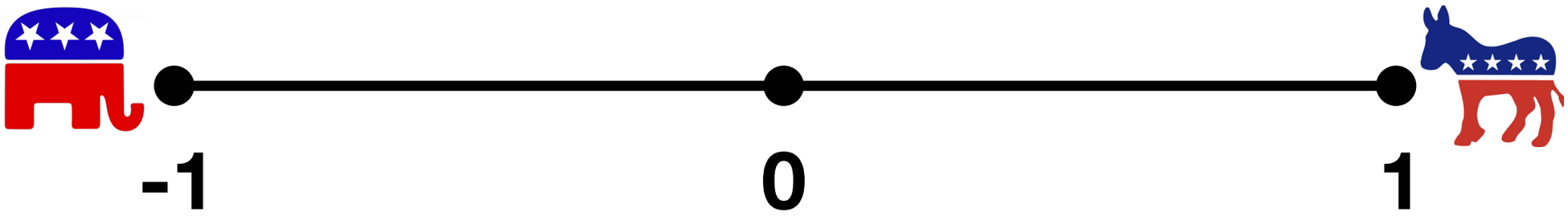


- Quantify bias of each individual item

Quantifying bias of individual items

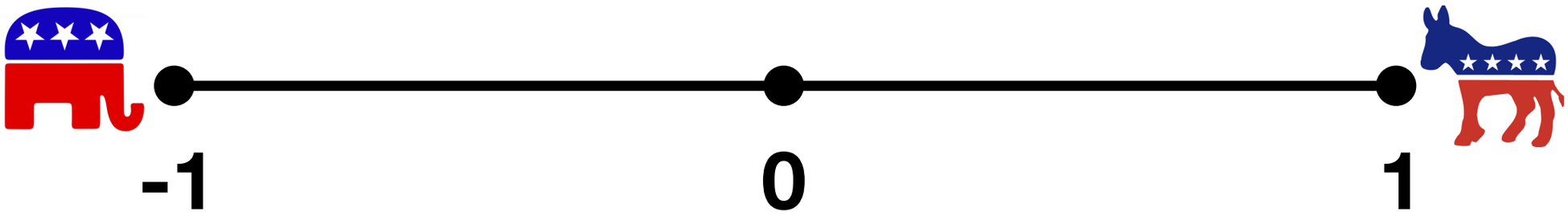


Quantifying bias of individual items



- Our method: Infer bias of each individual item from the bias of the author

Quantifying bias of individual items



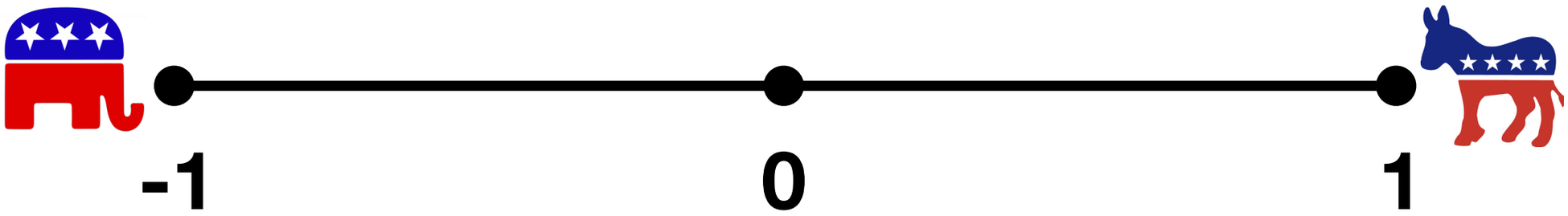
- Our method: Infer bias of each individual item from the bias of the author

Steven Crowder
@scrowder

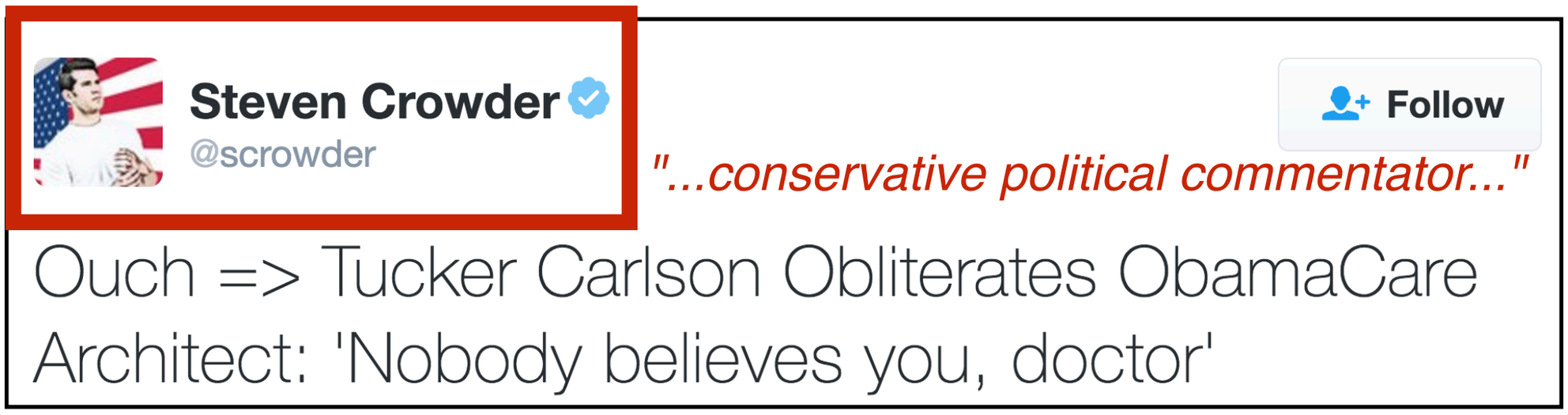
Follow

Ouch => Tucker Carlson Obliterates ObamaCare Architect: 'Nobody believes you, doctor'

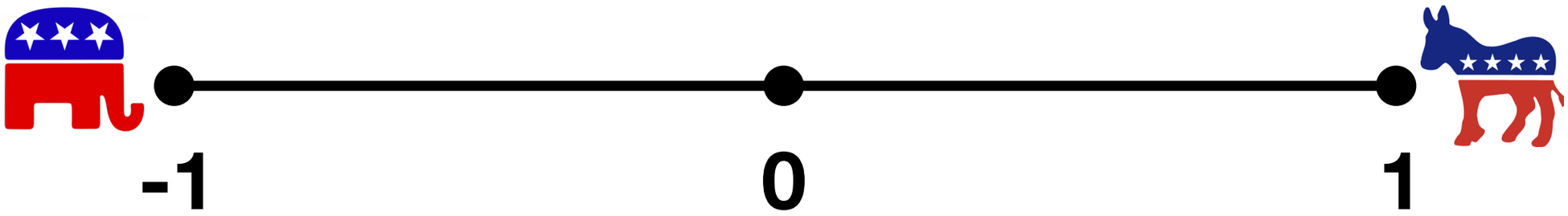
Quantifying bias of individual items



- Our method: Infer bias of each individual item from the bias of the author

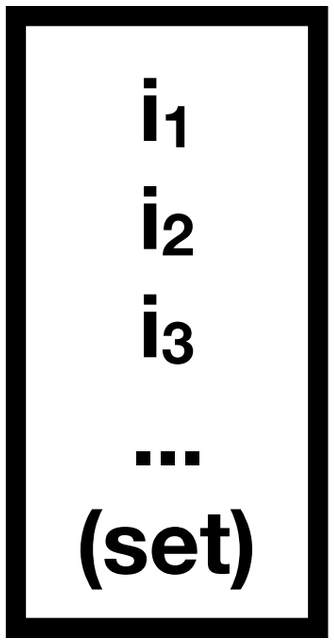


Quantifying bias of individual items



- Our method: Infer bias of each individual item from the bias of the author
- 75% accuracy
- Highly scalable, high coverage
- Other measures can be used [Zafar et al., 2016]

Quantifying bias in inputs (relevant item set)

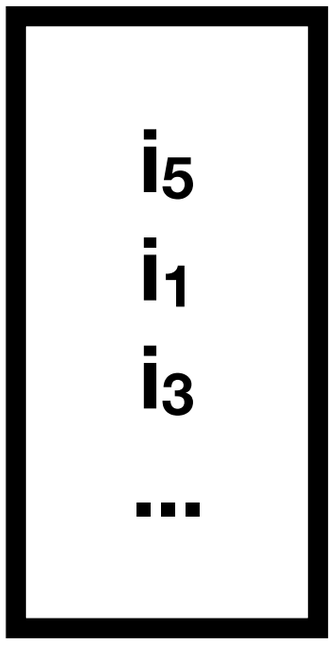


input
(relevant items)

- Compute bias for each item
- Take the average over the whole set

$$IB(q) = \frac{\sum_{i=1}^{n} s_i}{n}$$

Quantifying bias in outputs (ranked list)



output (ranked list)

- MAP-style measure
- Bias till rank *r*

$$B(q,r) = \frac{\sum_{i=1}^{r} s_i}{r}$$

- Output bias

$$OB(q,r) = \frac{\sum_{i=1}^{r} B(q,i)}{r}$$

Quantifying bias in ranking algorithm

Search engine algo.

ranking

Ranking bias = Output bias - Input bias

This talk: Study sources of bias in social media search

- Identifying sources of search bias
- Quantifying bias in each source
- **Studying bias in political searches in Twitter**

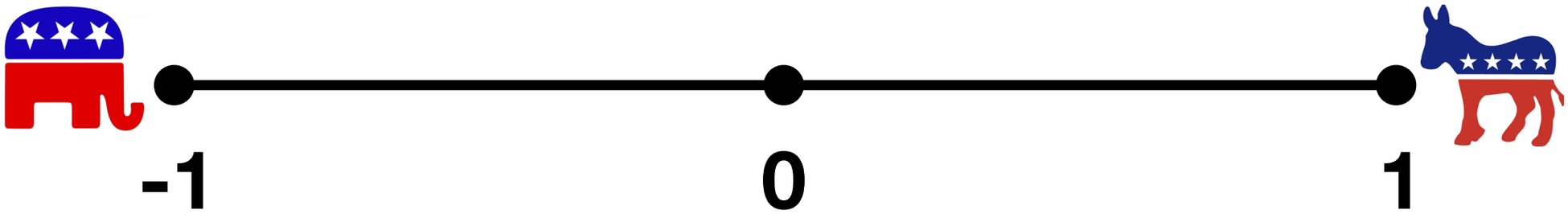
Studying bias in political searches in Twitter

- Search queries for
  - 2016 Democratic and Republican debates (dem debate, #demdebate, ...)
  - Presidential candidates (Hillary Clinton, Donald Trump, ...)

Key takeaways

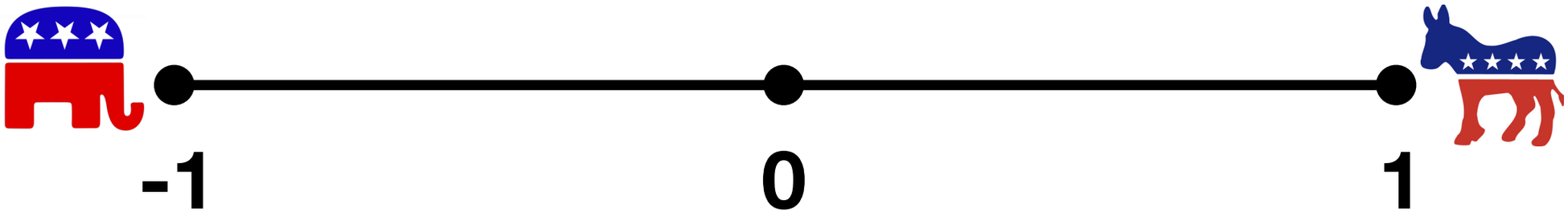
- It is not just the algorithm, input data also matters

High ranking bias



Query	Input Bias	Ranking bias
Hillary Clinton	0.03	0.18
Ted Cruz	-0.11	-0.37

High input bias



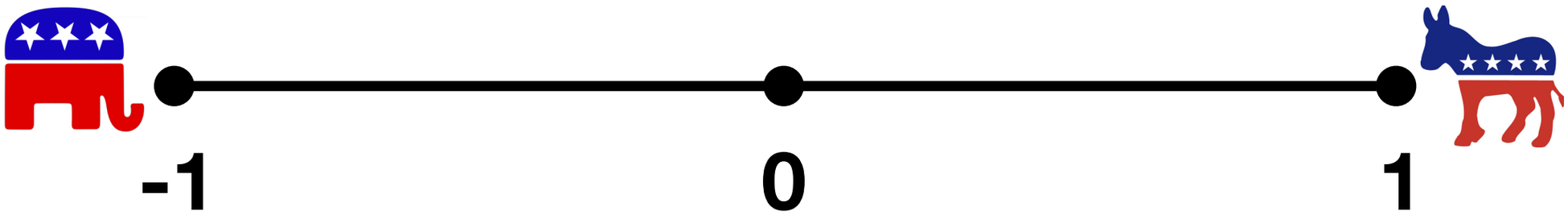
Query	Input Bias	Ranking bias
Bernie Sanders	0.55	0.16
Donald Trump	0.19	0.10

Key takeaways

- It is not just the algorithm, input data also matters

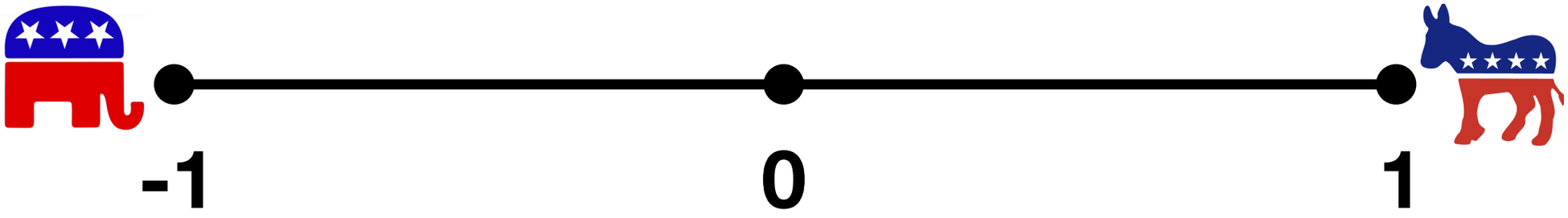
- Analyzing the input bias
  - Search topic matters

Effect of search topic



Query	Input Bias
Hillary Clinton (D)	0.03
Bernie Sanders (D)	0.55
Martin O'Malley (D)	0.64
Donald Trump (R)	0.19
Ted Cruz (R)	-0.11
Marco Rubio (R)	-0.12

Effect of search topic



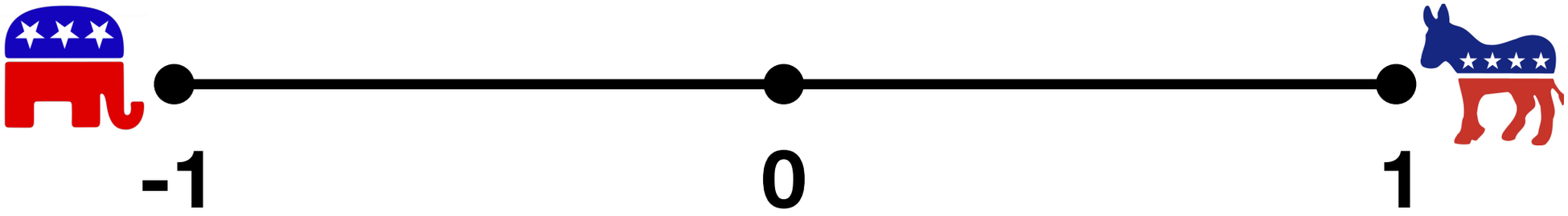
Query	Input Bias
Hillary Clinton (D)	0.03
Bernie Sanders (D)	0.55
Martin O'Malley (D)	0.64
Donald Trump (R)	0.19
Ted Cruz (R)	-0.11
Marco Rubio (R)	-0.12

Key takeaways

- It is not just the algorithm, input data also matters

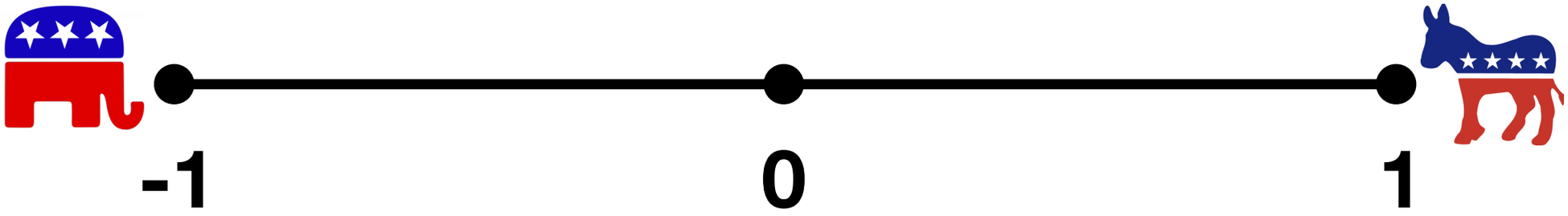
- Analyzing the input bias
  - Search topic matters
  - Query phrasing can make a big difference

Query phrasing and search bias



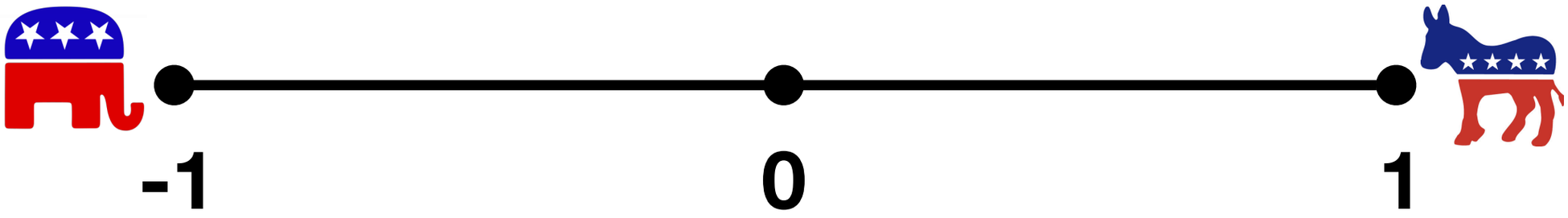
Query	Input Bias
democratic debate	0.38
#democraticdebate	0.19
#demdebate	0.56

Query phrasing and search bias



Query	Input Bias
democratic debate	0.38
#democraticdebate	0.19
#demdebate	0.56

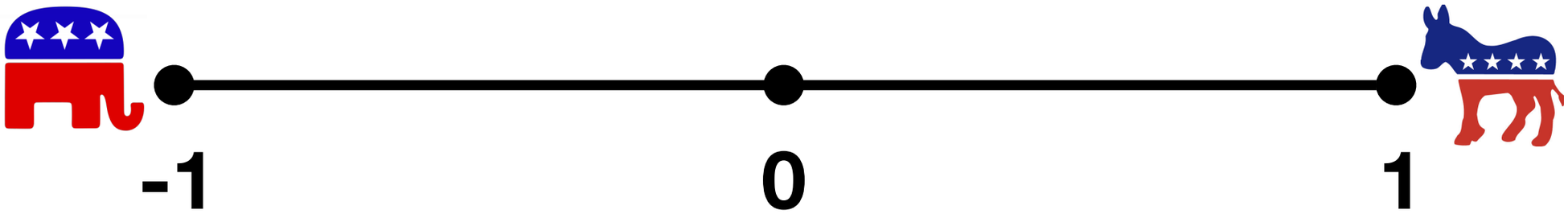
Query phrasing and search bias



Query	Input Bias
republican debate	0.27
#gopdebate	0.10
rep debate	0.40

Even for the same topic, query phrasing can greatly affect the input bias

Query phrasing and search bias



Query	Input Bias
republican debate	0.27
#gopdebate	0.10
rep debate	0.40

Even for the same topic, query phrasing can greatly affect the input bias

Future work: Addressing search bias

- Alter the ranking algorithm
  - Might lead to reduction in quality of results or lead to unforeseen biases
- Make the bias transparent
  - Keep the current ranking
  - Inform the users about what they are seeing
  - Make biases related to query phrasing transparent

Questions

Paper at:

http://tinyurl.com/se-bias