Measuring and Addressing Indexical Bias in Information Retrieval

Caleb Ziems William Held Jane Dwivedi-Yu Diyi Yang

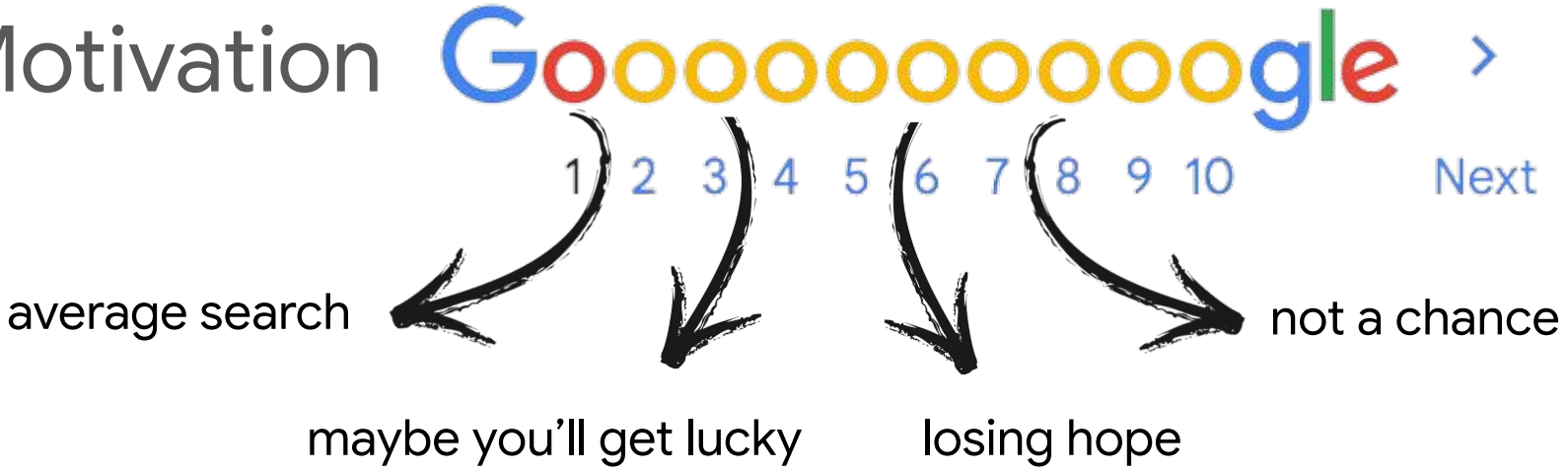
@cjziems @WilliamBarrHeld @JaneDwivedi @Diyi_Yang

Motivation

Motivation

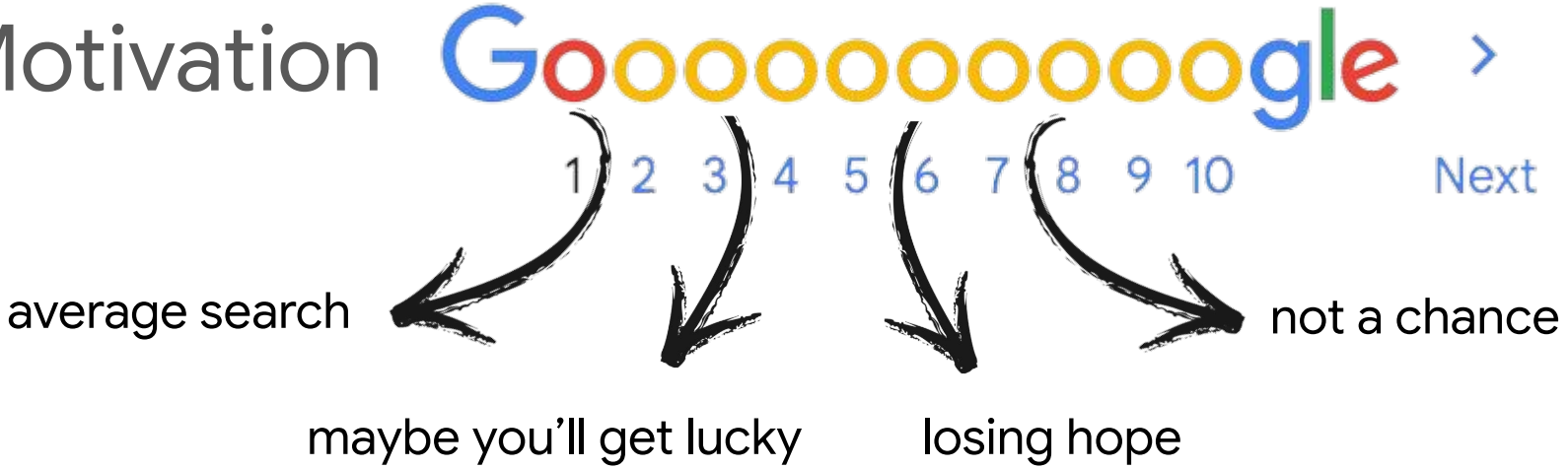
average search

Motivation


Gooooooooooogle
1 2 3 4 5 6 7 8 9 10 Next
average search
maybe you'll get lucky
losing hope
not a chance

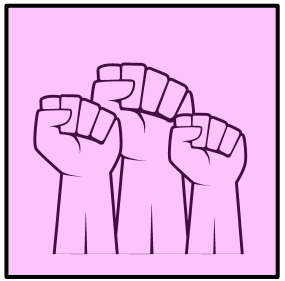
Capitalism vs. Socialism.

Motivation



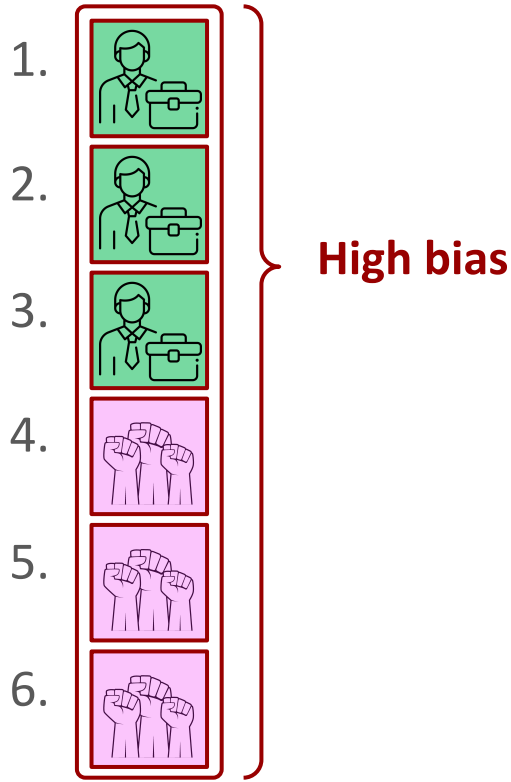
Capitalism vs. Socialism.

Motivation



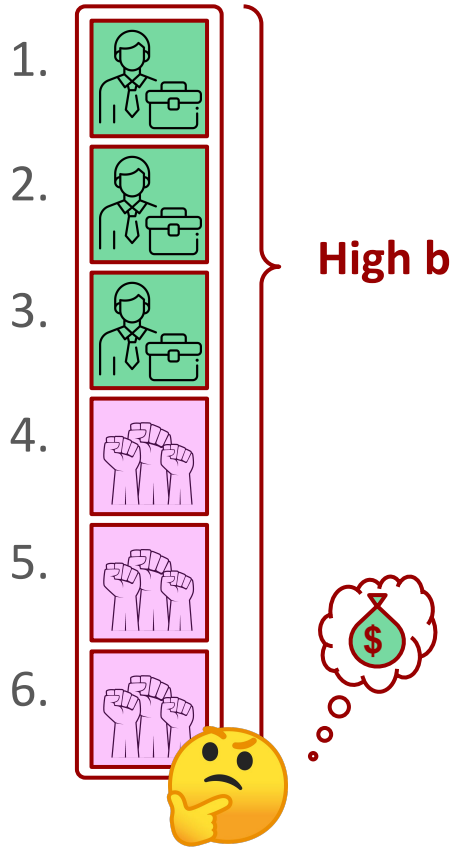
Capitalism vs. Socialism.

Motivation



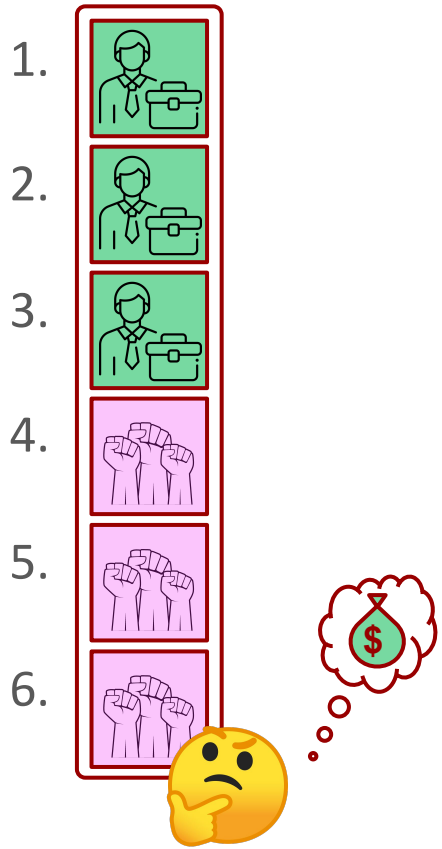
Capitalism vs. Socialism.

Motivation



Capitalism vs. Socialism.

Motivation

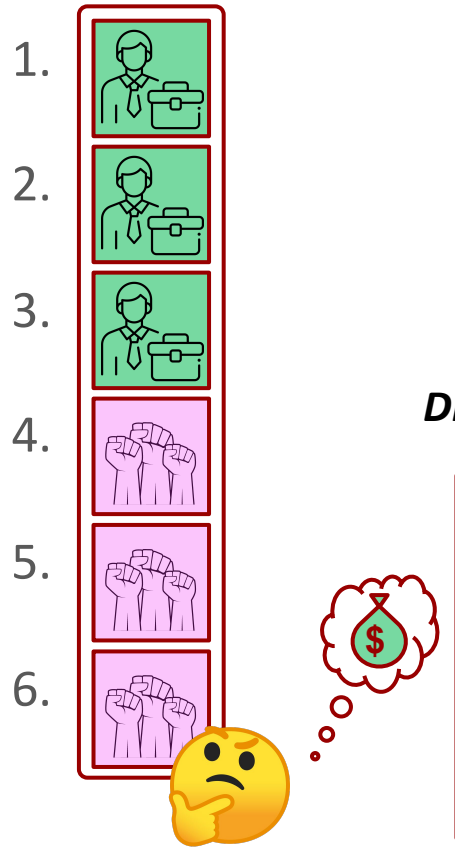


SEME

The Search Engine Manipulation Effect

Capitalism vs. Socialism.

Motivation



SEME metric:

DUO

***D**iscounted **U**niformity of **O**pinions*

$$\mathrm{DUo} = 1 - \frac{1}{Z}\sum \frac{\sigma^2_{1:i}}{\lg(i)}$$

Capitalism vs. Socialism.

$\sigma^2_{1:2}$ **zero variance**

1.
2.
3.
4.
5.
6.

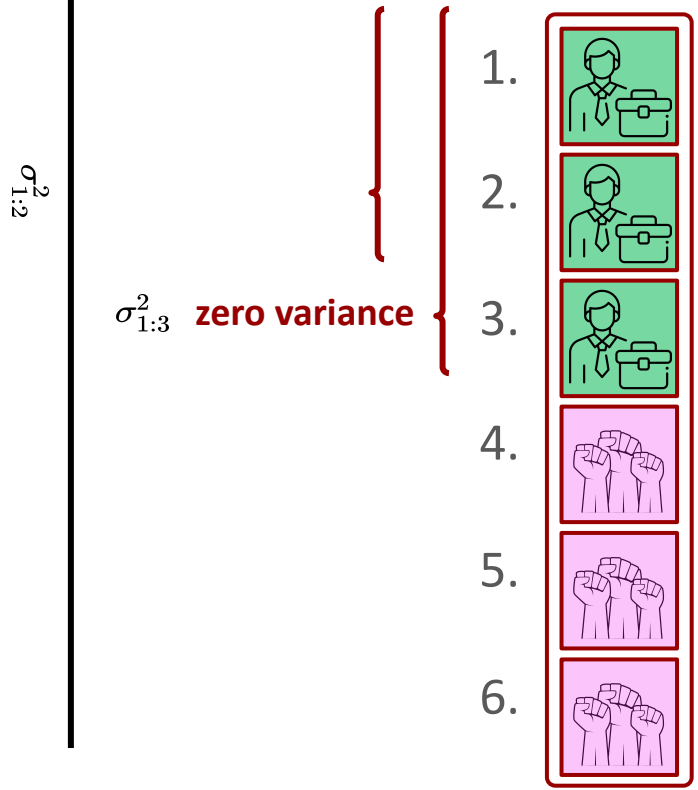
SEME metric:

DUO

***D**iscounted **U**niformity of **O**pinions*

$$\mathrm{DUo} = 1 - \frac{1}{Z}\sum \frac{\sigma^2_{1:i}}{\lg(i)}$$

Capitalism vs. Socialism.



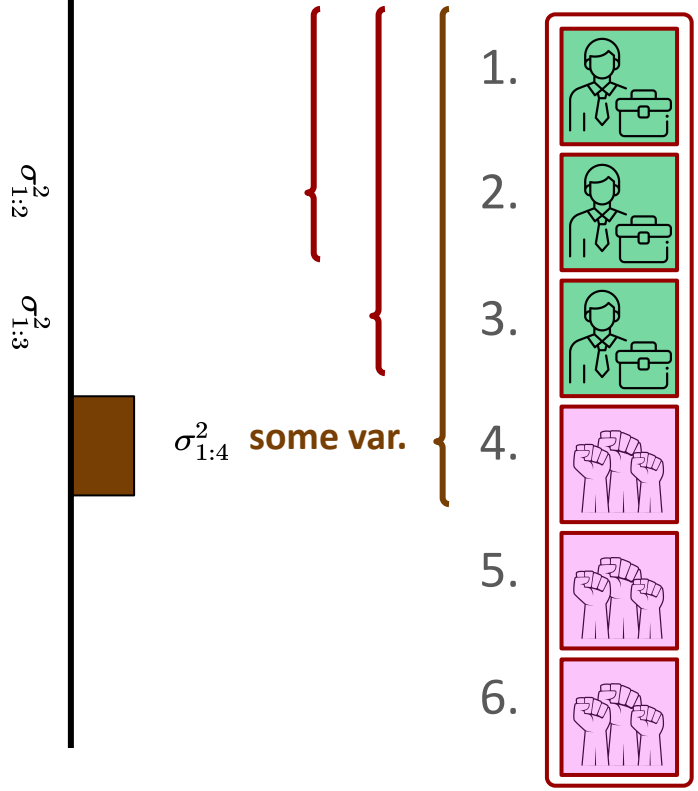
SEME metric:

DUO

***D**iscounted **U**niformity of **O**pinions*

$$\mathrm{DUo} = 1 - \frac{1}{Z}\sum \frac{\sigma^2_{1:i}}{\lg(i)}$$

Capitalism vs. Socialism.



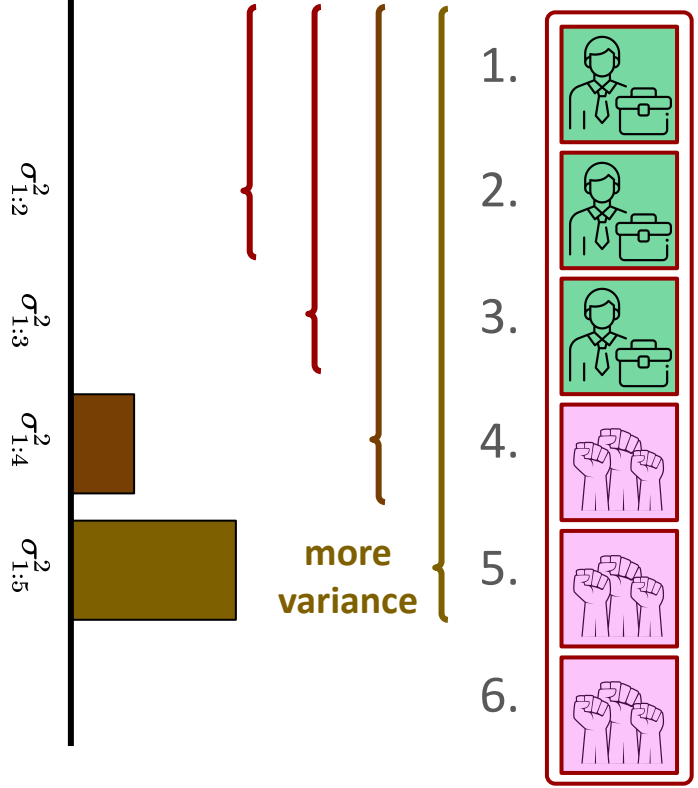
SEME
metric:

DUO

***D**iscounted **U**niformity of **O**pinions*

$$\mathrm{DUo} = 1 - \frac{1}{Z}\sum \frac{\sigma^2_{1:i}}{\lg(i)}$$

Capitalism vs. Socialism.



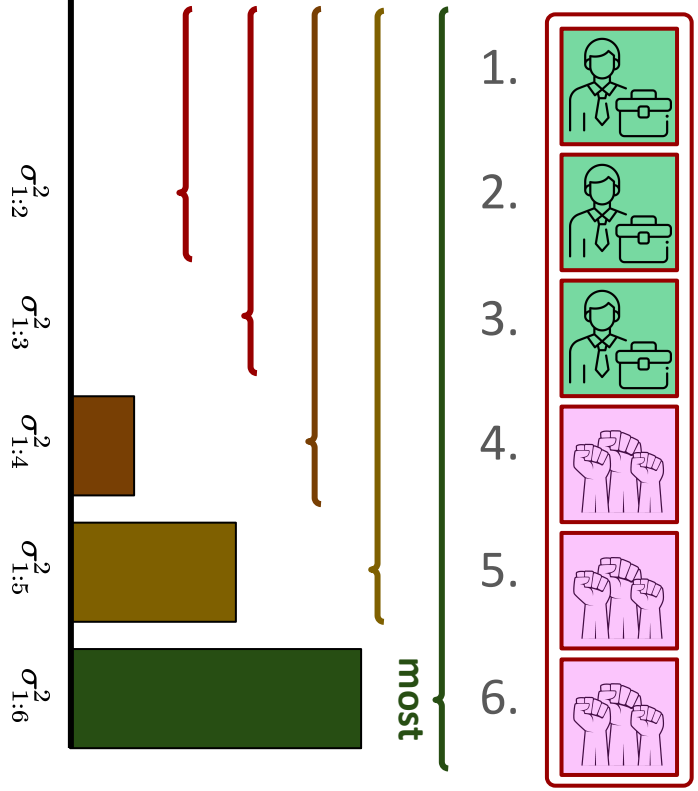
SEME metric:

DUO

***D**iscounted **U**niformity of **O**pinions*

$$\mathrm{DUo} = 1 - \frac{1}{Z}\sum \frac{\sigma^2_{1:i}}{\lg(i)}$$

Capitalism vs. Socialism.



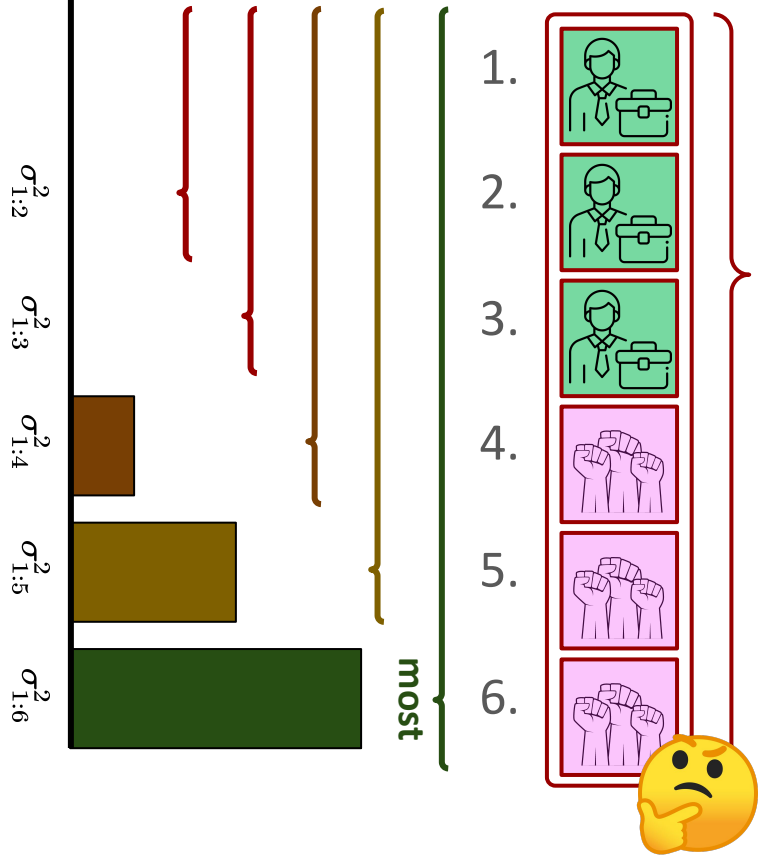
SEME metric:

DUO

***D**iscounted **U**niformity of **O**pinions*

$$\mathrm{DUo} = 1 - \frac{1}{Z}\sum \frac{\sigma^2_{1:i}}{\lg(i)}$$

Capitalism vs. Socialism.



High bias

DUO = 1

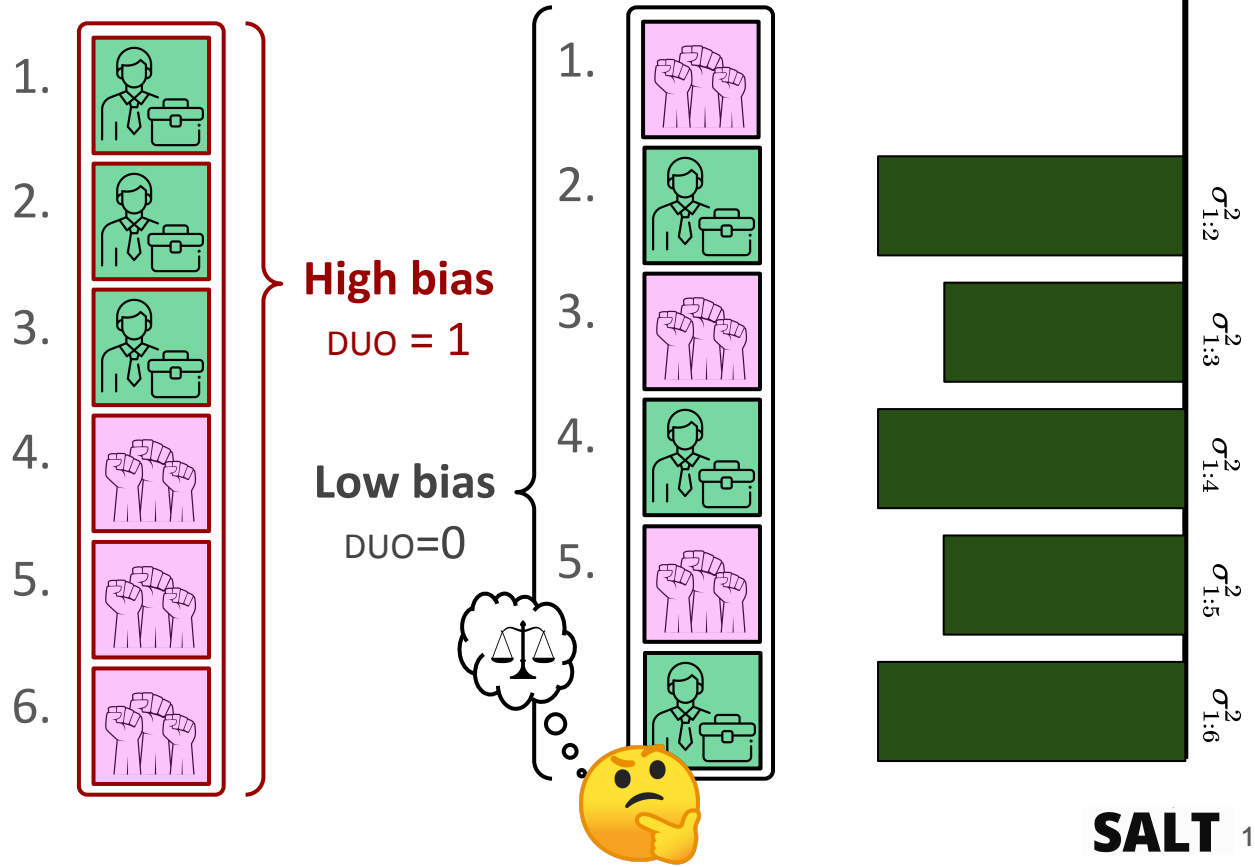
SEME metric:

DUO

***D**iscounted **U**niformity of **O**pinions*

$$\mathrm{DUo} = 1 - \frac{1}{Z}\sum \frac{\sigma^2_{1:i}}{\lg(i)}$$

Capitalism vs. Socialism.



Contributions

1) The DUO Positional Bias **Metric**

Contributions

1) The DUO Positional Bias **Metric**
2) Wiki-Balance Bias **Corpora**

Contributions

1) The DUO Positional Bias **Metric**
2) Wiki-Balance Bias **Corpora**
3) SEME **Human Behavioral Study**

Contributions

1) The DUO Positional Bias **Metric**
2) Wiki-Balance Bias **Corpora**
3) SEME **Human Behavioral Study**
4) Bias **Audit** Evaluations

Contributions

1) The DUO Positional Bias **Metric**
2) Wiki-Balance Bias **Corpora**
3) SEME **Human Behavioral Study**
4) Bias **Audit** Evaluations

PAIR

perspective-aligned information retrieval

DUO
Positional Bias Metric

DUO *Positional Bias Metric*

$$\mathrm{DCG}(r,u) = \sum_{i=1}^{|r|} \frac{u(i,r)}{\log_2 i}$$

⇦ discount

lower ranks

i document index

DUO *Positional Bias Metric*

$$\mathrm{DCG}(r, u) = \sum_{i=1}^{|r|} \frac{u(i, r)}{\log_2 i}$$

⇐ discount

lower ranks

i document index

u utility function

DUO *Positional Bias Metric*

$$\mathrm{DCG}(r, u) = \sum_{i=1}^{|r|} \frac{u(i, r)}{\log_2 i}$$

⇐ discount lower ranks

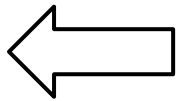
i document index

u utility function

r rank ordering

DUO *Positional Bias Metric*

$$\mathrm{DCG}(r,u) = \sum_{i=1}^{|r|} \frac{u(i,r)}{\log_2 i}$$



utility
based on
idea-balance

$$u_V(i,r) = \frac{1}{i}\sum_{j=1}^{i}(p_j - \bar{p})^2$$

DUO *Positional Bias Metric*

$$\mathrm{DCG}(r,u) = \sum_{i=1}^{|r|} \frac{u(i,r)}{\log_2 i}$$

p_j polarization score

$\bar{p}$ (avg)

$$\boxed{u_V(i,r)} = \boxed{\frac{1}{i}\sum_{j=1}^{i}(p_j - \bar{p})^2}$$

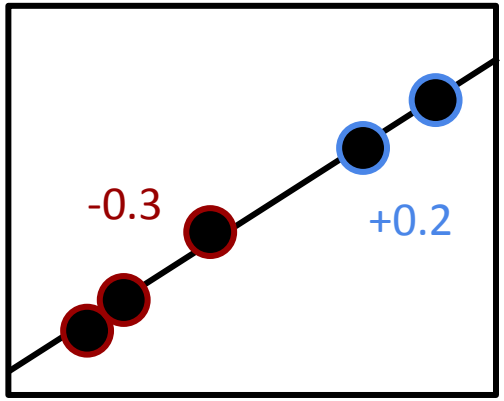
idea-balance

is the variance in document polarization

DUO *Positional Bias Metric*

$$\mathrm{DCG}(r,u) = \sum_{i=1}^{|r|} \frac{u(i,r)}{\log_2 i}$$

$$u_V(i,r) = \frac{1}{i} \sum_{j=1}^{i} (p_j - \bar{p})^2$$



polarization scores

come *automatically* from PCA on polarized doc embeddings

Wiki-Balance Corpus of Polarized Documents

Wiki-Balance SYNTHETIC

Wiki-Balance *Synthetic Corpus*

topics

- **Politics**
- **History**
- **Environment**
- **Culture**
- **Language**
- **People**

Wiki-Balance *Synthetic Corpus*

topics

- **Politics**
  - China
  - gun control
- **History**
- **Environment**
- **Culture**
- **Language**
- **People**

243 topics

Wiki-Balance *Synthetic Corpus*

topics

- **Politics**
  - China
  - gun control
- **History**
  - Cuban Missile Crisis
- **Environment**
- **Culture**
- **Language**
- **People**

122 topics

Wiki-Balance *Synthetic Corpus*

topics

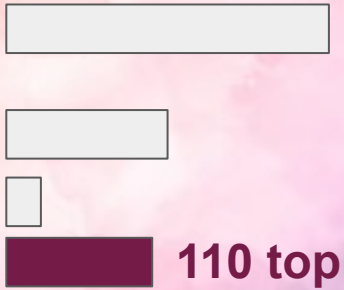
- **Politics**
  - China
  - gun control
- **History**
  - Cuban Missile Crisis
- **Environment**
  - recycling
- **Culture**
- **Language**
- **People**

26 topics

Wiki-Balance *Synthetic Corpus*

topics

- **Politics**
  - China
  - gun control
- **History**
  - Cuban Missile Crisis
- **Environment**
  - recycling
- **Culture**
  - Nike Air Max
- **Language**
- **People**

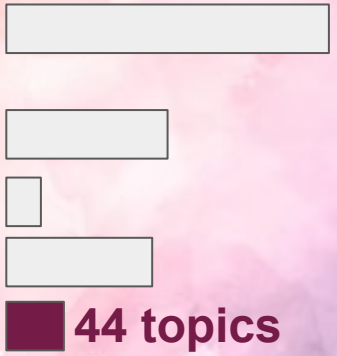


110 topics

Wiki-Balance *Synthetic Corpus*

topics

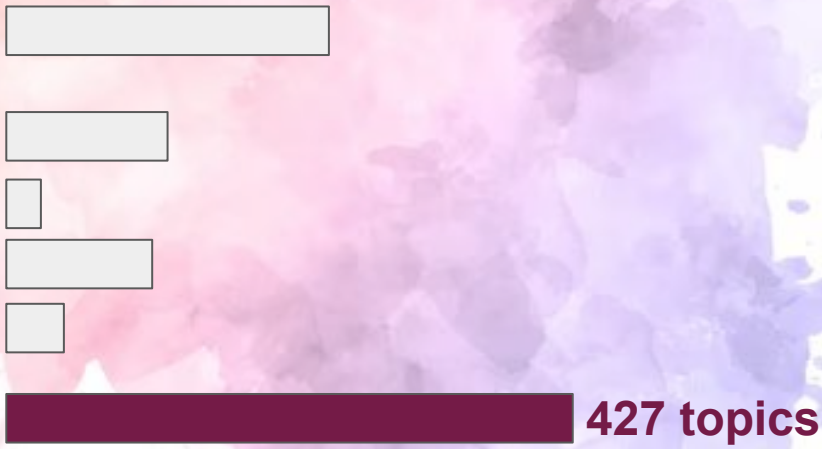
- **Politics**
  - China
  - gun control
- **History** — Cuban Missile Crisis
- **Environment** — recycling
- **Culture** — Nike Air Max
- **Language** — Sapir-Whorf
- **People**



Wiki-Balance *Synthetic Corpus*

topics

- **Politics**
  - China
  - gun control
- **History** — Cuban Missile Crisis
- **Environment** — recycling
- **Culture** — Nike Air Max
- **Language** — Sapir-Whorf
- **People**
  - Taylor Swift
  - Freud



Wiki-Balance *Synthetic Corpus*

topics

- **Politics**
  - China
  - gun control
- **History** — Cuban Missile Crisis
- **Environment** — recycling
- **Culture** — Nike Air Max
- **Language** — Sapir-Whorf
- **People**
  - Taylor Swift
  - Freud

Wiki-Balance *Synthetic Corpus*

topics

Give me 10 debate questions about ${topic}

- **Politics**
  - **China**
  - **gun control** — *background checks enough?*
- **History** — **Cuban Missile Crisis** — *US response appropriate?*
- **Environment** — **recycling** — *Effective at reducing waste?*
- **Culture** — **Nike Air Max** — *Are they a symbol of status and wealth?*
- **Language** — **Sapir-Whorf** — *Can it be tested empirically?*
- **People**
  - **Taylor Swift** — *Has her image been consistent?*
  - **Freud**

Wiki-Balance *Synthetic Corpus*

topics

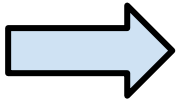
Give me 10 debate questions about ${topic}

${question} For each side of the issue, write an article from the perspective of that side."

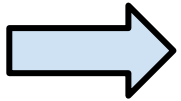
- **Politics**
  - **China**
  - **gun control** — *background checks enough?*
- **History** — **Cuban Missile Crisis** — *US response appropriate?*
- **Environment** — **recycling** — *Effective at reducing waste?*
- **Culture** — **Nike Air Max** — *Are they a symbol of status and wealth?*
- **Language** — **Sapir-Whorf** — *Can it be tested empirically?*
- **People**
  - **Taylor Swift** — *Has her image been consistent?*
  - **Freud**

Wiki-Balance *Synthetic Corpus*

topics



Give me 10 debate questions about ${topic}



${question} For each side of the issue, write an article from the perspective of that side."

Pos. 1

Recycling is a proven, **effective method for reducing waste** and conserving natural resources. By recycling materials, we are significantly reducing the amount of waste sent to landfills,

Environment — **recycling** — *Effective at reducing waste?*

Pos. 2

*it is **not a sustainable** way to reduce waste and conserve natural resources. Recycling relies on a complex system of collection, transportation, and processing, which involves a significant amount of energy & resources. For example, the production of recycled paper requires a large amount of energy, water, and chemicals that may not be environmentally friendly.*

Wiki-Balance *Evaluation Corpora*

topics

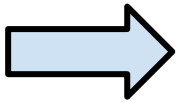
Give me 10 debate questions about ${topic}

${question} For each side of the issue, write an article from the perspective of that side."

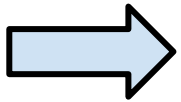
Corpus	Mean Relevance	Mean Faithfulness	Mean Coherence	Mean Fluency	Non-Toxic
Synthetic	**4.7** / 5.0	**4.9** / 5.0	**4.9** / 5.0	**4.9** / 5.0	✔

Wiki-Balance *Evaluation Corpora*

topics



Give me 10 debate questions about ${topic}

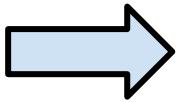


${question} For each side of the issue, write an article from the perspective of that side."

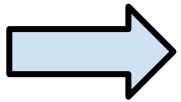
Corpus	Domains	Topics	Queries	Docs
Synthetic	15	1.4k	4k	32k

Wiki-Balance *Evaluation Corpora*

topics



Give me 10 debate questions about ${topic}



${question} For each side of the issue, write an article from the perspective of that side."

Corpus	Domains	Topics	Queries	Docs
Synthetic	15	1.4k	4k	32k
Natural	15	1.4k	450	4.6k

SEME
Human Behavioral Study

SEME *Human Behavioral Study*

Goals:

1) show that **DUO measures** salient **indexical bias**
2) show that measurable **bias leads to SEME**

SEME *Human Behavioral Study*

Goals:

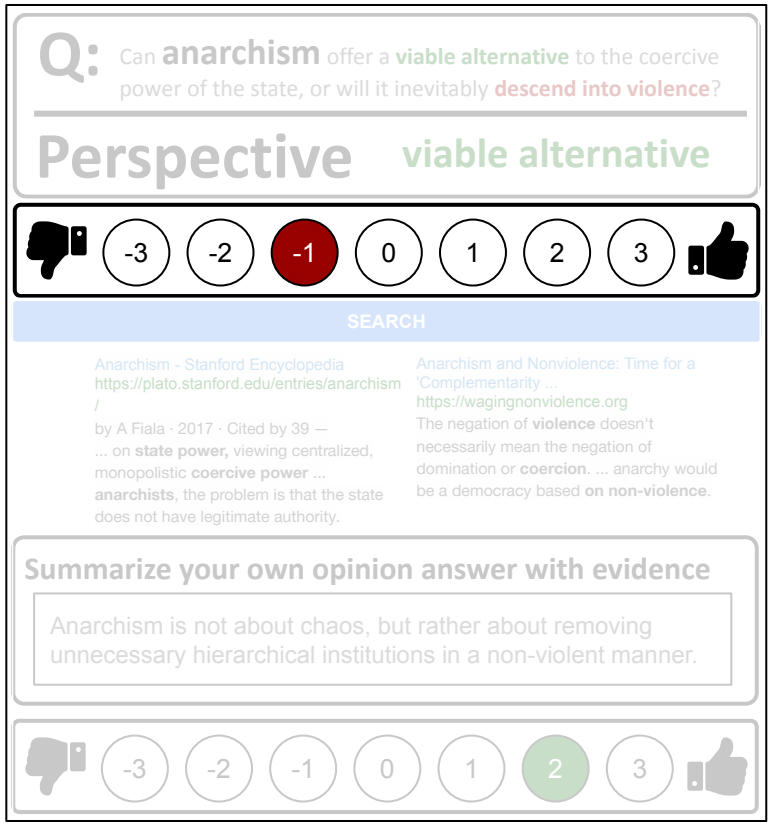
1) show that **DUo-P measures** salient **indexical bias**
2) show that measurable **bias leads to SEME**

(more biased rankings should shift opinion more)

SEME *Human Behavioral Study*

⇦ controversial topic in focus

SEME *Human Behavioral Study*



user provides prior opinion o_{prior}

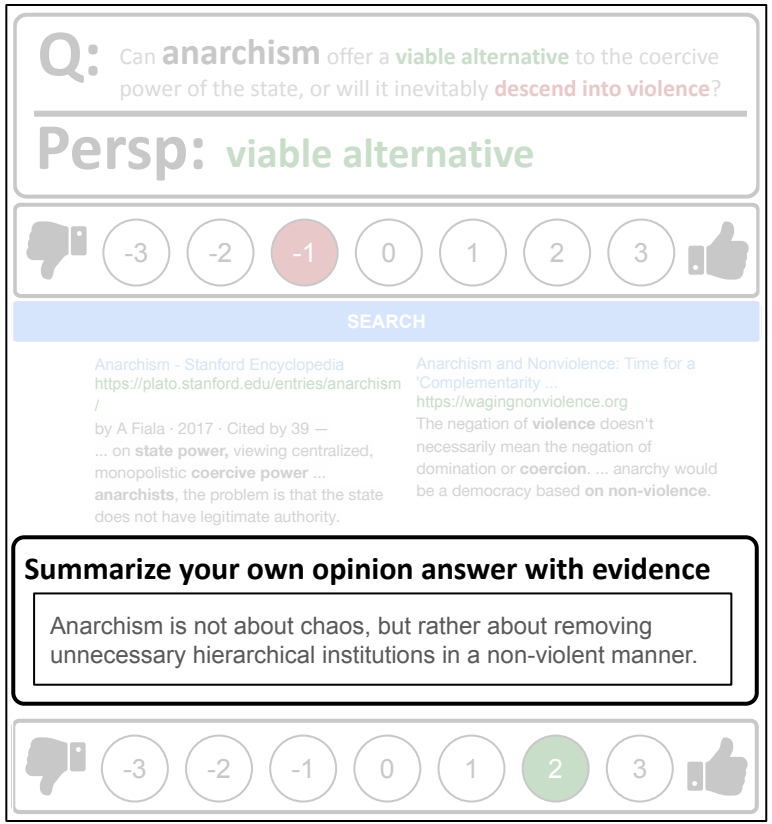
SEME *Human Behavioral Study*

o_{prior}

biased search

yields either most or least biased ranking according to metric μ^{+}_{DUo}

SEME *Human Behavioral Study*



o_{prior}

μ^{+}_{DUo}

after the user gives

close consideration

SEME *Human Behavioral Study*

o_{prior}

μ^{+}_{DUo}

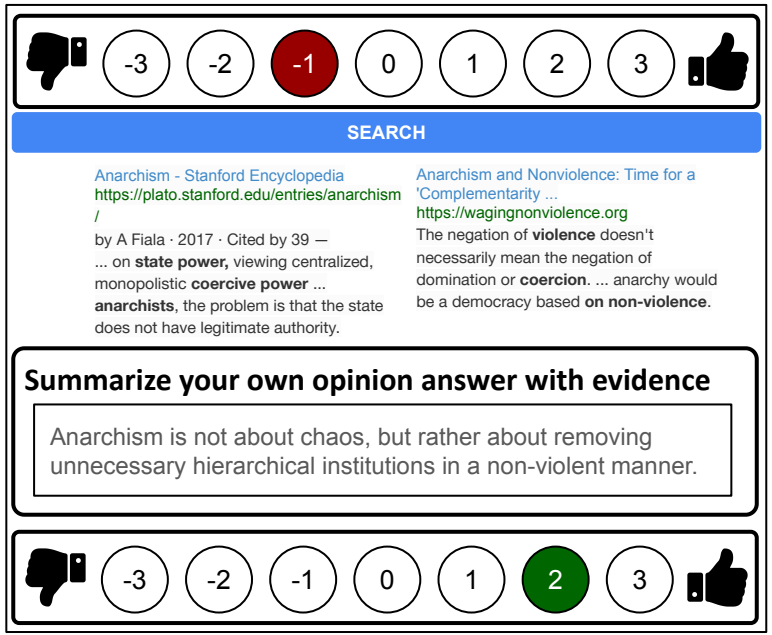
they offer a

posterior

opinion $o_{\text{posterior}}$

SEME *Human Behavioral Study*

Question: ***does more bias change opinion more?***



o_{prior}

μ^{+}_{DUo}

$o_{\text{posterior}}$

SEME *Human Behavioral Study*

Question: ***is*** β_2 ***non-zero (w/ significance)?***

$$o_{\text{posterior}} = \beta_0 + \beta_1 o_{\text{prior}} + \beta_2 \mu^{+}_{\text{DUo}} + \epsilon$$

SEME *Human Behavioral Study*

Question: ***is*** β_2 ***non-zero (w/ significance)?***

$$o_{\text{posterior}} = \beta_0 + \beta_1 o_{\text{prior}} + \beta_2 \mu^{+}_{\text{DUo}} + \epsilon$$

Answer: ***yes!***

SEME *Human Behavioral Study*

Question: ***is*** β_2 ***non-zero (w/ significance)?***

$$o_{\text{posterior}} = \beta_0 + \beta_1 o_{\text{prior}} + \beta_2 \mu^{+}_{\text{DUo}} + \epsilon$$

Answer: ***yes!*** *

** β_2=0.392; R^2=0.489; p=0.036*

SEME *Human Behavioral Study*

Question: ***is*** β_2 ***non-zero (w/ significance)?***

$$o_{\text{posterior}} = \beta_0 + \beta_1 o_{\text{prior}} + \beta_2 \mu^{+}_{\text{DUo}} + \epsilon$$

Answer: ***yes!*** **

** β_2=0.392; R^2=0.489; p=0.036*
** only when the user clicks article links*

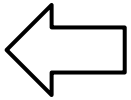
DUO *Validations*

PAIR

Bias **Audit** Evaluations

PAIR *Bias Audit Evaluations*

	Natural
Corpus	**Relevance**
BM-25	0.75



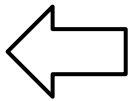
BM-25 is a **strong relevant baseline**

PAIR *Bias Audit Evaluations*

	Natural
Corpus	**Relevance**
BM-25	0.75
SPARTA	0.65
SPLADE	0.75
ANCE	0.71
SBERT	0.75
Use-QA	0.68
ColBERT	0.72

PAIR *Bias Audit Evaluations*

	Natural	
Corpus	Relevance	DUO Bias
BM-25	0.75	0.66
SPARTA	0.65	0.65
SPLADE	0.75	0.67
ANCE	0.71	0.67
SBERT	0.75	0.65
Use-QA	0.68	0.64
ColBERT	0.72	0.65



least biased ≠ most relevant

PAIR *Bias Audit Evaluations*

	Natural	
Corpus	Relevance	DUO Bias
BM-25	0.75	0.66
SPARTA	0.65	0.65
SPLADE	0.75	0.67
ANCE	0.71	0.67
SBERT	0.75	0.65
Use-QA	0.68	0.64
ColBERT	0.72	0.65

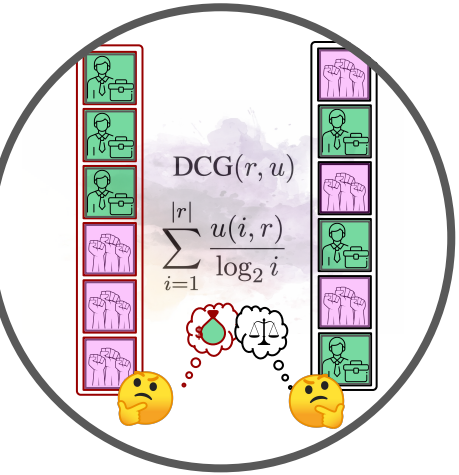
most relevant = most biased

PAIR *Bias Audit Evaluations*

Corpus	Relevance	DUO Bias
SPLADE	0.75	0.67

Domain	DUO Bias
Entertainment	0.68
Environment	0.67
Languages	0.71
Media & Culture	0.67
Psychiatry	0.62
Sex & Gender	0.59

PAIR *Summary*

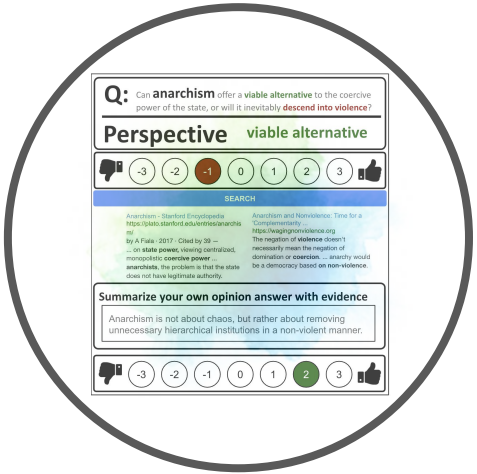


DUO

Metric

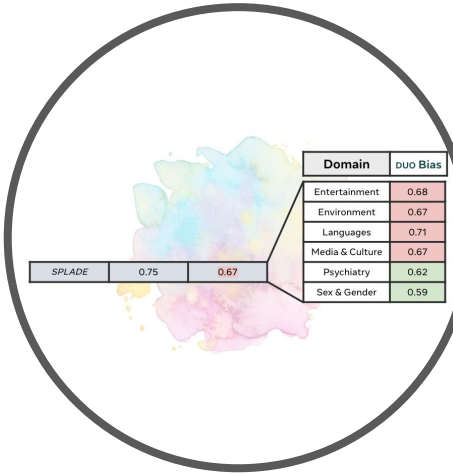
Wiki-Balance

Corpus



Experimental

Outcomes



Bias

Audit

Measuring and Addressing Indexical Bias in Information Retrieval

Caleb Ziems

@cjziems

William Held

@WilliamBarrHeld

Jane Dwivedi-Yu

@JaneDwivedi

Diyi Yang

@Diyi_Yang