

From Language to Information: Introductory Lecture to NLP

DSC514 / MAI623: Natural Language Processing

Demetris Paschalides, Ph.D.

dpasch01@ucy.ac.cy

About the Course

What is this class and why should you take it?

- NLP is an (increasingly) important area.
 - NLP is now good enough for real-world applications.
 - There is a huge growth in NLP companies and NLP jobs (in many industries).

What is this class and why should you take it?

- **NLP is an (increasingly) important area.**
 - NLP is now good enough for real-world applications.
 - There is a huge growth in NLP companies and NLP jobs (in many industries).
- **The rise of Large Language Models (LLMs) completely changed everything in:**
 - Natural Language Processing (NLP)
 - Artificial Intelligence (AI)
 - Information Retrieval (IR)
 - Recommendation Systems

What is this class and why should you take it?

- **NLP is an (increasingly) important area.**
 - NLP is now good enough for real-world applications.
 - There is a huge growth in NLP companies and NLP jobs (in many industries).
- **The rise of Large Language Models (LLMs) completely changed everything in:**
 - Natural Language Processing (NLP)
 - Artificial Intelligence (AI)
 - Information Retrieval (IR)
 - Recommendation Systems
- This class starts from scratch and builds up on how text is represented, and eventually, how LLMs work and how they are applied.

What will you learn in this class?

■ What is NLP?

- The core tasks (as well as data sets and evaluation metrics) that people work on in NLP.

What will you learn in this class?

■ What is NLP?

- The core tasks (as well as data sets and evaluation metrics) that people work on in NLP.

■ How does NLP work?

- The fundamental models, algorithms and representations that have been developed for these tasks.

What will you learn in this class?

■ What is NLP?

- The core tasks (as well as data sets and evaluation metrics) that people work on in NLP.

■ How does NLP work?

- The fundamental models, algorithms and representations that have been developed for these tasks.

■ Why is NLP hard?

- The relevant linguistic concepts and phenomena that have to be handled to do well at these tasks.

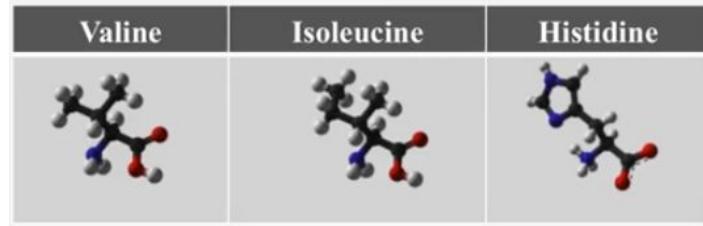
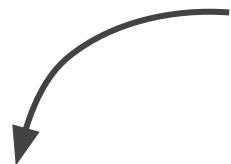
Introduction to NLP

From Language to Information

- Automatically extracting meaning and structure from:
 - Human language text and speech (news, social media, etc.)
 - Social networks
 - Genome sequences

From Language to Information

- Automatically extracting meaning and structure from:
 - Human language text and speech (news, social media, etc.)
 - Social networks
 - Genome sequences



```
AACCTGCGGAAGGATCATTACCGAGTGCGGGTCTTGGGCCAACCTCCCATCCGTGTCTA  
TTGTACCCTGTTGCTTCGGCGGGCCCGCCGCTTGTGGCCGCCGGGGGGCGCCTCTGCC  
CCGGGCCCGTGCCCGCCGGAGACCCAACACGAACACTGTCTGAAAGCGTGCAGTCTGAGTT  
GATTGAATGCAATCAGTTAAAACCTTCAACAATGGATCTCTGGTTCCGGC
```

From Language to Information

- Automatically extracting meaning and structure from:
 - Human language text and speech (news, social media, etc.)
 - Social networks
 - Genome sequences
- Interacting with humans via language
 - Dialog systems/Chatbots
 - Question Answering
 - Recommendation Systems

NLP in Industry and Commercial Applications



Microsoft ebay



You Tube



IBM



∞
Meta



Post-ChatGPT: NLP Being Integrated Everywhere!

- Enterprises, from banking to law to care, are embedding NLP into core workflows.

JPMorgan Launches AI Assistant for 60,000 Employees

By ANDREW KESSEL Published August 09, 2024 10:36 AM EDT



BUSINESS INSIDER

TECH

Inside the AI boom that's changing how Big Law attorneys work

By Melia Russell

BNY, America's Oldest Bank, Signs Multiyear Deal With OpenAI

The bank wants to combine its internal AI platform with cutting-edge tech from OpenAI

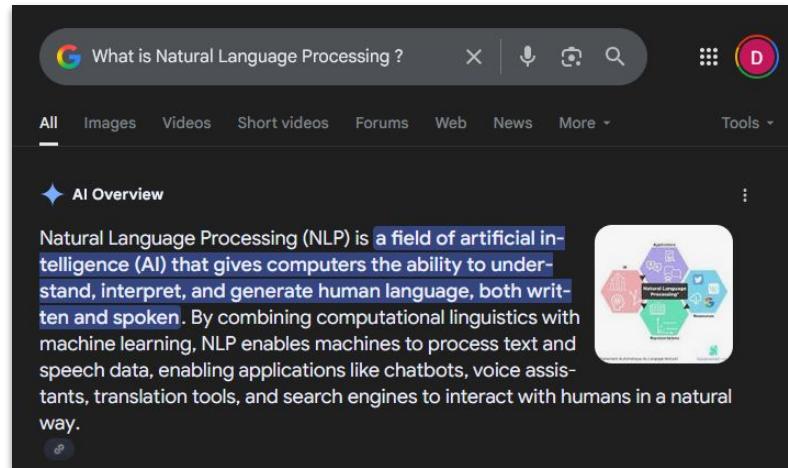
By Isabelle Bousquette [Follow](#)

Feb. 26, 2025 10:00 am ET

Extracting Information from Language

■ Information Retrieval (IR)

- **Everyday life:** Google, YouTube, Amazon, ChatGPT
→ all powered by IR.
- In 2025, Google handles **8.5 billion+** searches per day.
→ IR is the most widely used NLP application in the world.

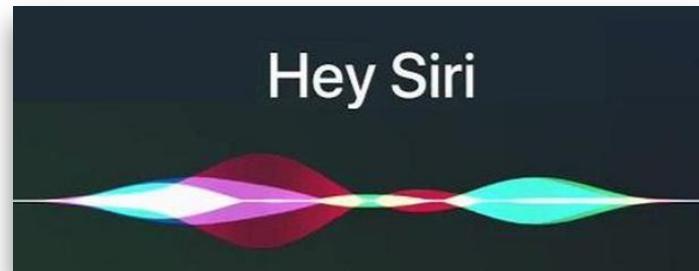


Conversational Agents

- Speech Recognition
- Language Analysis
- Dialogue Processing
- Information Retrieval
- Text-to-speech



Hi, how can I help?



What are you working on?

+ Ask anything



ChatGPT Revolution

The students of the DSC510 Introduction to Data Science are here.
Welcome them and introduce yourself.

ChatGPT Revolution

The students of the DSC510 Introduction to Data Science are here.
Welcome them and introduce yourself.

Hello everyone, and a very warm welcome to **DSC510: Introduction to Data Science!**

It's great to see you all here at the beginning of what I hope will be an exciting and eye-opening journey into the world of data.

I'm **ChatGPT**, an AI developed by OpenAI, and I'll be available to assist you throughout the course. Think of me as a helpful study companion — someone you can turn to when you're stuck on a coding problem, need help understanding a concept like regression or classification, or just want to explore ideas and practice your skills.

Data science is a powerful and evolving field that shapes decisions in everything from health and finance to sports and social media. Whether you're here to build technical skills, gain insights from real-world data, or just understand how data impacts the world around us — you're in the right place.

I'm excited to support you along the way. Let's dive in and make this semester a meaningful and successful one. Welcome aboard!

ChatGPT LLMs Revolution

Commercial



ChatGPT LLMs Revolution

Commercial



Open Source



ChatGPT LLMs Revolution

Commercial



Claude



Grok



Gemini



deepseek

Open Source



MISTRAL
AI_



Qwen3

Gemma 3



LLaMA



Small LLMs

Gemma 3n

SmolLM v2

Phi-3

Small Language
Models are the
Future of
Agentic AI



Text Classification: Disaster Response

- Classifying SMS messages during 2010 Haiti earthquake.
- **Haitian Creole:** "Mwen thomassin 32 nan pyron mwen ta renmen jwen yon ti dlo gras a dieu bo lakay mwen anfom se sel dlo nou bezwen"
- **English:** "I am in Thomassin number 32, in the area named Pyron. I would like to have some water. Thank God we are fine, but we desperately need water."



Meier, P., & Munro, R. (2010). The unprecedented role of SMS in disaster response: Learning from Haiti. *SAIS Rev. Int'l Aff.*, 30, 91.

Caragea, C., McNeese, N. J., Jaiswal, A. R., Traylor, G., Kim, H. W., Mitra, P., & Yen, J. (2011, May). Classifying text messages for the Haiti earthquake. In ISCRAM.

Text Classification: Disaster Response

- Classifying SMS messages during 2010 Haiti earthquake.
- **Haitian Creole:** "Mwen thomassin 32 nan pyron mwen ta renmen jwen yon ti dlo gras a dieu bo lakay mwen anfom se sel dlo nou bezwen"
- **English:** "I am in Thomassin number 32, in the area named Pyron. I would like to have some water. Thank God we are fine, but we desperately need water."



Meier, P., & Munro, R. (2010). The unprecedented role of SMS in disaster response: Learning from Haiti. *SAIS Rev. Int'l Aff.*, 30, 91.

Caragea, C., McNeese, N. J., Jaiswal, A. R., Traylor, G., Kim, H. W., Mitra, P., & Yen, J. (2011, May). Classifying text messages for the Haiti earthquake. In ISCRAM.

Recommendation Engines

Recommendation Engines

The Good:

- Products: Amazon, ebay
- Content: Netflix, Spotify



Recommendation Engines

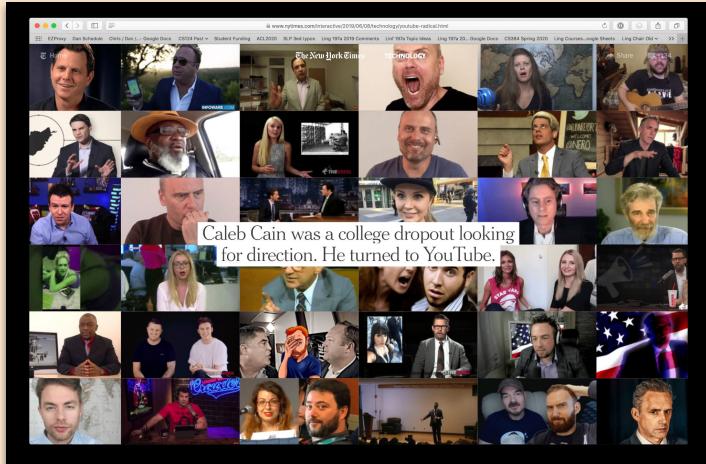
The Good:

- Products: Amazon, ebay
- Content: Netflix, Spotify



The Bad:

- Filter Bubbles and Echo Chambers
- Online Radicalization



Papadamou, K., Zannettou, S., Blackburn, J., De Cristofaro, E., Stringhini, G., & Sirivianos, M. (2021). "How over is it?" Understanding the Incel Community on YouTube. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), 1-25.

Phonetics and Phonology

- Pronunciation Modeling

- Sounds: T h i a s i e n

Words

- Language Modeling
- Tokenization
- Spelling Correction
- **Words:** This is a simple sentence

Morphology

- Morphology Analysis
- Stemming
- Lemmatization
- Words: This is a simple sentence
- Morphology: be
present

Part-of-Speech

- Part-of-Speech (PoS) Tagging

- PoS: DT VBZ DT JJ NN

- Words: This is a simple sentence

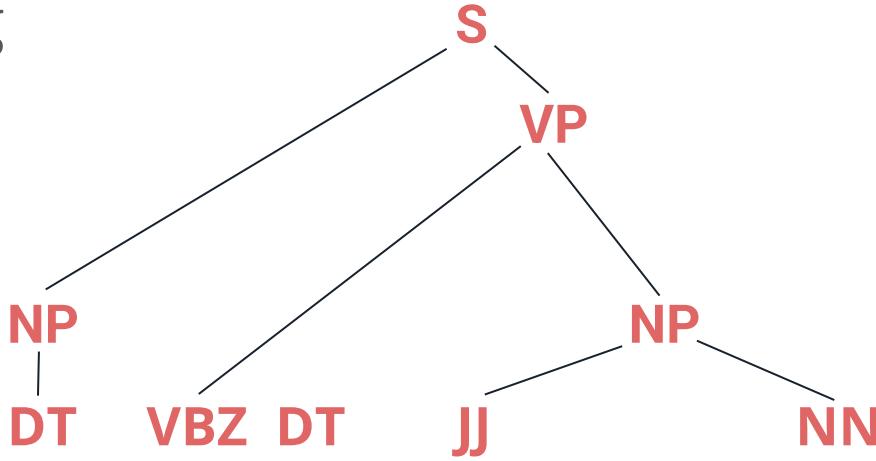
- Morphology: be

- present

Syntax

- Syntactic Parsing

- Syntax:



- PoS:

This is a simple sentence

- Words:

be

- Morphology:

present

Semantics

■ Named Entity Recognition

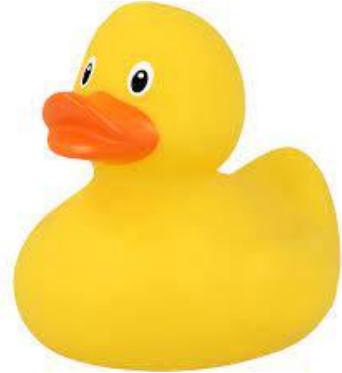
When Sebastian Thrun PERSON started at Google ORG in 2007 DATE, few people outside of the company took him seriously. "I can tell you very senior CEOs of major American NORG car companies would shake my hand and turn away because I wasn't worth talking to," said Thrun PERSON, now the co-founder and CEO of online higher education startup Udacity, in an interview with Recode ORG earlier this week DATE.

A little less than a decade later DATE, dozens of self-driving startups have cropped up while automakers around the world clamor, wallet in hand, to secure their place in the fast-moving world of fully automated transportation.

Language Interpretation is Hard

Ambiguity

- **Sentence:** "I made her duck"

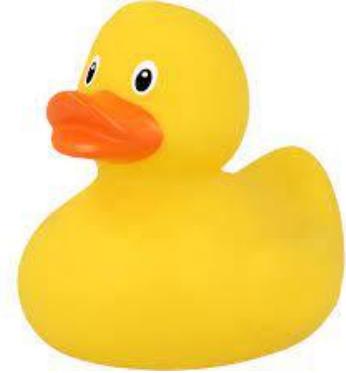


Ambiguity

■ **Sentence:** "I made her duck"

■ At least 6 different meanings:

- I cooked waterfowl for her (to eat)
- I cooked waterfowl of her
- I created the plastic waterfowl she owns
- I caused her to quickly lower her head or body



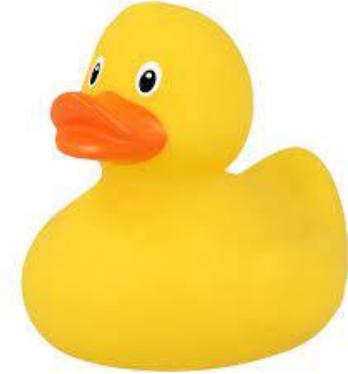
Ambiguity

■ **Sentence:** “I made her **duck**”

■ At least 6 different meanings:

- I cooked waterfowl for her (to eat)
- I cooked waterfowl of her
- I created the plastic waterfowl she owns
- I caused her to quickly lower her head or body

→ “**Duck**” can be a **Noun** or a **Verb**.

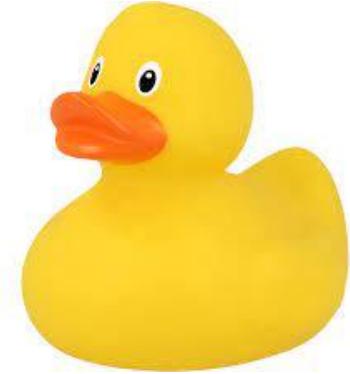


Ambiguity

■ **Sentence:** “I made **her** duck”

■ At least 6 different meanings:

- I cooked waterfowl **for her** (to eat)
 - I cooked waterfowl **of her**
 - I created the plastic waterfowl she owns
 - I caused her to quickly lower her head or body
- “**her**” can be a possessive pronoun “of her” or a dative pronoun “for her”.

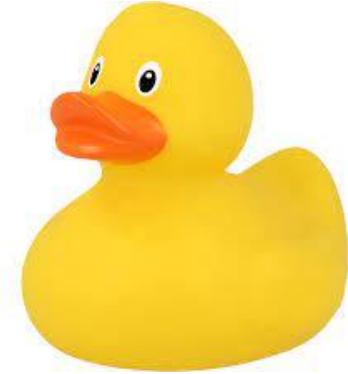


Ambiguity

■ Sentence: “I **made** her duck”

■ At least 6 different meanings:

- I **cooked** waterfowl for her (to eat)
 - I **cooked** waterfowl of her
 - I **created** the plastic waterfowl she owns
 - I **caused** her to quickly lower her head or body
- “**make**” can mean “**cooked**”, “**created**”, or “**caused**”



Challenges: Slang, Emojis and Hashtags

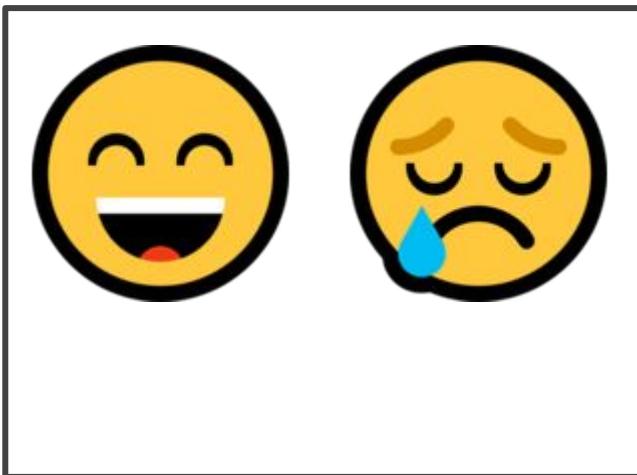
- “OMG” =
- “w8” =
- “brb” =

Challenges: Slang, Emojis and Hashtags

- “OMG” = Oh my god
- “w8” = wait
- “brb” = be right back

Challenges: Slang, Emojis and Hashtags

- “OMG” = Oh my god
- “w8” = wait
- “brb” = be right back



Challenges: Slang, Emojis and Hashtags

- “OMG” = Oh my god
- “w8” = wait
- “brb” = be right back



Challenges: Slang, Emojis and Hashtags

- “OMG” = Oh my god
- “w8” = wait
- “brb” = be right back
- #Naturallanguageprocessing = Natural Language Processing



Challenges: Slang, Emojis and Hashtags

- “OMG” = Oh my god
- “w8” = wait
- “brb” = be right back



- #Naturallanguageprocessing = Natural Language Processing

A screenshot of a tweet from JD Scott (@MrJDScott). The profile picture shows a man in a suit. The tweet reads: "My favorite thing to do at 4am is go to the airport. How about you? #Sarcasm #AutoPilot". To the right of the tweet are standard Twitter interaction icons: a gear for settings, a person icon for following, and a blue "Follow" button.

Challenges in PoS Tagging

ikr smh he asked fir yo last name

so he can add u on fb lololol

Challenges in PoS Tagging

I know, right shake my head for your
ikr smh he asked fir yo last name

you Facebook laugh out loud
so he can add u on fb lololol

Challenges in PoS Tagging

I know, right	shake my head			for	your		
ikr	smh	he	asked	fir	yo	last	name
!	G	O	V	P	D	A	N
interjection	acronym	pronoun	verb	prep.	det.	adj.	noun
		you		Facebook	laugh out loud		
so	he	can	add	u	on	lololol	
P	O	V	V	O	P	^	!
preposition				proper noun			

Challenging Morphology and Syntax



Challenging Morphology and Syntax



- A *ship*-
shipping
ship,
shipping
shipping-*ships*

How we deal with these rich meanings in LLMs?

- Neural Word Embeddings
 - A word's meaning in each sentence / context
→ A point or region in N-dimensional space.

How we deal with these rich meanings in LLMs?

- Neural Word Embeddings
 - A word's meaning in each sentence / context
 - A point or region in N-dimensional space.

“This is a simple sentence” =
[0.6, -0.2, 0.8, 0.9, -0.1, ..., -0.7]

How we deal with these rich meanings in LLMs?

■ Neural Word Embeddings

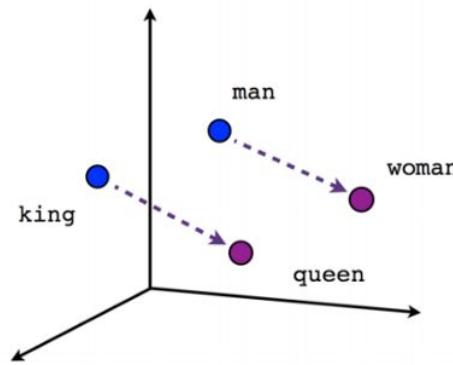
- A word's meaning in each sentence / context
→ A point or region in N-dimensional space.

"This is a simple sentence" =
[0.6, -0.2, 0.8, 0.9, -0.1, ..., -0.7]

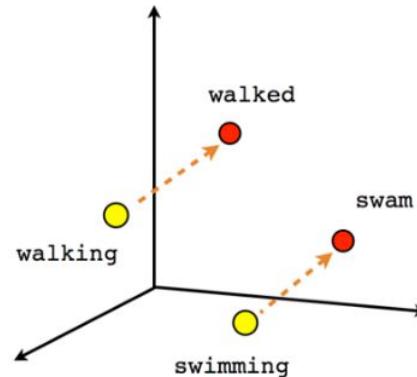


Word Embeddings

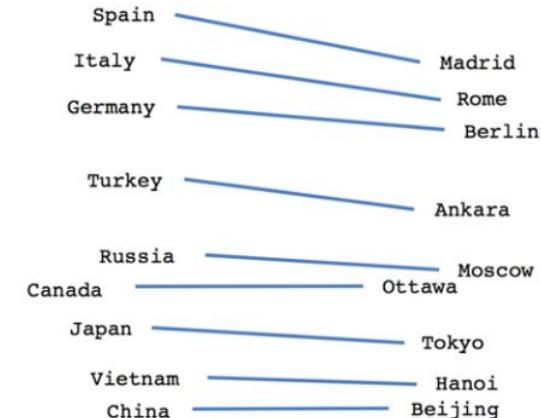
- A word's semantic meaning as a 300-dimensional vector.



Male-Female



Verb tense



Country-Capital

Embeddings and LLMs Reflect Human Biases

- LLM representation are rich but also reflect human biases!
- LLMs display stereotypes about pretty much every group (Asians, Blacks, Muslims, women, etc).
 - LLM internal representations also show these biases.
 - These representations lead LLMs to also take biased actions.
 - These biases have been clear from the very earliest papers studying LLM representations.

Bolukbasi, T., Kai-Wei C., James Y. Zou, Venkatesh S., and Adam T. K. 2016. "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings." In NeurIPS 2016, pp. 4349-4357.

Embeddings and LLMs Reflect Human Biases

- LLM representation are rich but also reflect human biases!
- LLMs display stereotypes about pretty much every group (Asians, Blacks, Muslims, women, etc).
 - LLM internal representations also show these biases.
 - These representations lead LLMs to also take biased actions.
 - These biases have been clear from the very earliest papers studying LLM representations.

We tried out DeepSeek. It worked well, until we asked it about Tiananmen Square and Taiwan
Donna Lu

Source: Guardian <https://shorturl.at/Jlzs> January 2025

Is AI chatbot Grok censoring criticism of Elon Musk and Donald Trump?



Source: Euronews
<https://www.euronews.com/my-europe/2025/03/03/is-ai-chatbot-grok-censoring-criticism-of-elon-musk-and-donald-trump> March 2025

Bolukbasi, T., Kai-Wei C., James Y. Zou, Venkatesh S., and Adam T. K. 2016. "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings." In NeurIPS 2016, pp. 4349-4357.

Course Outline

- Corpora, Tokens, and Text Pre-processing
- N-Gram Language Models
- Fundamental NLP Tasks
- Feature Engineering and Text Classification
- Vector Semantics and Word Embeddings
- Information Retrieval and Web Search
- Neural Networks in NLP
- Transformers & Contextual Embeddings
- Introduction to Large Language Models

Hands-on with Python and Jupyter Notebooks



The screenshot shows a Jupyter Notebook interface with the title "jupyter presentation". The notebook contains the following content:

Presentation from Jupyter

You can create presentation from Jupyter Notebook.
Mixing code and Markdown is just fantastic:

- computation presentations are great,
- I like to code, I don't like to manually copy-paste images,
- sharing presentation as website or PDF.

```
In [6]:  
1 from IPython.display import Markdown  
2  
3 greetings = "Hello"  
4 name = "World"  
5 Markdown(f"""  
6  
7 ## {greetings} {name}!  
8  
9 You can mix Markdown with variables thanks to `IPython.display.Markdown()`.  
10  
11 """)
```

```
Out[6]:  
Hello World!  
You can mix Markdown with variables thanks to IPython.display.Markdown().
```

Simple plot

```
In [8]:  
1 from matplotlib import pyplot as plt  
2 import random  
3  
4 points_count = 10  
5 color = "blue"  
6 x = [random.gauss(0, 1) for _ in range(points_count)]  
7 y = [random.gauss(0, 1) for _ in range(points_count)]  
8  
9 plt.figure(figsize=(10, 7))  
10 _ = plt.plot(x, y, '.', markersize=14, color=color)
```

Extracting Polarization Knowledge from News Media Articles: Techniques and Applications

DSC510 Introduction to Data Science and Analytics

November 2025

Dr. Demetris Paschalides

dpasch01@ucy.ac.cy

News Media Language Shapes Perception

[Left] From AlterNet

Employees-many of whom are undocumented immigrants from Mexico, Ecuador and elsewhere-toil seven days a week for less than minimum wage, with no overtime pay.

[Right] From Breitbart

Mass immigration has come at the expense of America's working and middle class, which suffered from poor job growth, stagnant wages, and increased public costs.

AlterNet

B BREITBART

L LEFT

LEAN LEFT

C CENTER

LEAN RIGHT

R RIGHT

Source: www.allsides.com

News Media Language Shapes Perception

[Left] From AlterNet

Employees-many of whom are undocumented immigrants from Mexico, Ecuador and elsewhere-toil seven days a week for less than minimum wage, with no overtime pay.

[Right] From Breitbart

Mass immigration has come at the expense of America's working and middle class, which suffered from poor job growth, stagnant wages, and increased public costs.

□ Terminology Differences

- Undocumented immigrants **vs.** Mass immigration
- Toil for hours **vs.** Drives down wages

AlterNet

B BREITBART

L LEFT

LEAN LEFT

C CENTER

LEAN RIGHT

R RIGHT

Source: www.allsides.com

News Media Language Shapes Perception

[Left] From AlterNet

Employees-many of whom are undocumented immigrants from Mexico, Ecuador and elsewhere-toil seven days a week for less than minimum wage, with no overtime pay.

[Right] From Breitbart

Mass immigration has come at the expense of America's working and middle class, which suffered from poor job growth, stagnant wages, and increased public costs.

❑ Terminology Differences

- Undocumented immigrants **vs.** Mass immigration
- Toil for hours **vs.** Drives down wages

❑ Framing Techniques

- Left-leaning excerpt emphasizes **labor exploitation** and **humanizes immigrants**.
- Right-leaning excerpt focuses on **economic impact** and portrays immigration as a **threat**.

AlterNet

B BREITBART

L LEFT

LEAN LEFT

C CENTER

LEAN RIGHT

R RIGHT

Source: www.allsides.com

News Media Language Shapes Perception

[Left] From AlterNet

Employees-many of whom are undocumented immigrants from Mexico, Ecuador and elsewhere-toil seven days a week for less than minimum wage, with no overtime pay.

[Right] From Breitbart

Mass immigration has come at the expense of America's working and middle class, which suffered from poor job growth, stagnant wages, and increased public costs.

AlterNet

B BREITBART

L LEFT

LEAN LEFT

C CENTER

LEAN RIGHT

R RIGHT

Source: www.allsides.com

❑ Terminology Differences

- Undocumented immigrants **vs.** Mass immigration
- Toil for hours **vs.** Drives down wages

❑ Framing Techniques

- Left-leaning excerpt emphasizes **labor exploitation** and **humanizes immigrants**.
 - Right-leaning excerpt focuses on **economic impact** and portrays immigration as a **threat**.
- ❑ Language can frame narratives in ways that align with specific political ideologies.

News Media Language Shapes Perception

[Left] From AlterNet

Employees-many of whom are undocumented immigrants from Mexico, Ecuador and elsewhere-toil seven days a week for less than minimum wage, with no overtime pay.

[Right] From Breitbart

Mass immigration has come at the expense of America's working and middle class, which suffered from poor job growth, stagnant wages, and increased public costs.

AlterNet

B BREITBART

L LEFT

LEAN LEFT

C CENTER

LEAN RIGHT

R RIGHT

Source: www.allsides.com

□ Terminology Differences

- Undocumented immigrants **vs.** Mass immigration
- Toil for hours **vs.** Drives down wages

□ Framing Techniques

- Left-leaning excerpt emphasizes **labor exploitation** and **humanizes immigrants**.
- Right-leaning excerpt focuses on **economic impact** and portrays immigration as a **threat**.

□ Language can frame narratives in ways that align with specific political ideologies.

→ Understanding these nuances is crucial for developing critical media literacy skills.

Polarized Language Impact

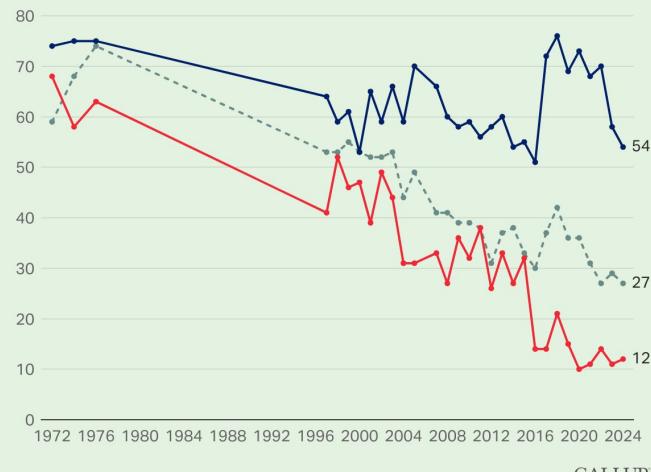
- **Shapes Public Opinion:** Phrases like "undocumented immigrants" versus "illegal aliens" or "climate crisis" versus "climate hoax" frame issues differently, influencing public perception.
- **Amplifies Division:** Individuals tend to consume media that aligns with their beliefs, leading to fragmented information environments → **Echo Chambers**.
- **Drives Engagement and Outrage:** Sensationalist and emotionally charged language increases user engagement, incentivizing media outlets to adopt more extreme framings.
- **Affects Language Models and AI:** Language models trained on polarized content may inadvertently learn and propagate biases, affecting their outputs and decision-making processes.

Partisans' Trust in Mass Media, 1972-2024

In general, how much trust and confidence do you have in the mass media -- such as newspapers, TV and radio -- when it comes to reporting the news fully, accurately and fairly -- a great deal, a fair amount, not very much or none at all?

% Great deal/Fair amount

— Republicans - - Independents — Democrats



Trust in Media Over Time 1972 - 2024, Source:
<https://news.gallup.com/poll/651977/americans-trust-media-remains-trend-low.aspx>

AI and Language Models Are Shaped by their Training Data

- LLMs (such as OpenAI ChatGPT, Anthropic Claude and Google Gemini) learn from vast web corpora, including news, forums, and social media.

AI and Language Models Are Shaped by their Training Data

- LLMs (such as OpenAI ChatGPT, Anthropic Claude and Google Gemini) learn from vast web corpora, including news, forums, and social media.
- These sources often reflect ideological and emotional bias—even when subtle.
- Biased data → Biased models:
 - Research shows LLMs can reflect political leanings, stereotypes, and framing preferences.
 - Prompt phrasing or topic framing can shift model responses.

ChatGPT may be shifting 'rightward' in political bias, study finds

The screenshot shows a news article from 'euro news.' about ChatGPT's political bias. It includes the ChatGPT logo and several examples of its responses to prompts related to quantum computing, creative ideas for a birthday, and HTTP requests in JavaScript.

Source: <https://shorturl.at/EJS6x> February 2025

Google chief admits 'biased' AI tool's photo diversity offended users

Sundar Pichai addresses backlash after Gemini software created images of historical figures in variety of ethnicities and genders

● [Human or fake? How AI is distorting beauty standards - video](#)



Source: <https://shorturl.at/0uaDE> February 2024

Language Models Can Also Amplify Polarization

- LLMs can be prompted (intentionally or not) to echo ideological viewpoints, contributing to the spread of: Misinformation, Conspiracy Theories, and Misleading Narratives.
- Creation of a feedback loop:
 - Polarized and biased content → Training data.
 - AI models generating biased outputs.
 - Biased outputs → Shared / Indexed → More polarized data.
- Use Cases:
 - News summaries with framing bias (Hao, K. 2024)
 - Partisan chatbots (Paschalides et al. 2025)
 - Deepfake political narratives (Chen & Shu 2024)

Propagandists are using AI too—and companies need to be open about it

OpenAI has reported on influence operations that use its AI tools. Such reporting, alongside data sharing, should become the industry norm.

Source: "[Propagandists are using AI!](#)", MIT Technology Review, 2024

We tried out DeepSeek. It worked well, until we asked it about Tiananmen Square and Taiwan
Donna Lu

Source: Guardian <https://shorturl.at/Jlzs> January 2025

Is AI chatbot Grok censoring criticism of Elon Musk and Donald Trump?

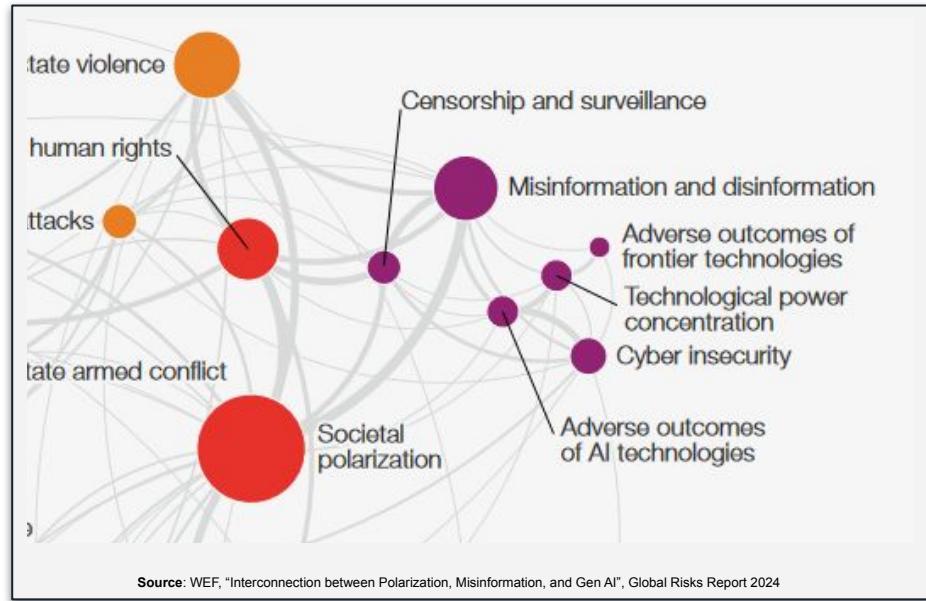
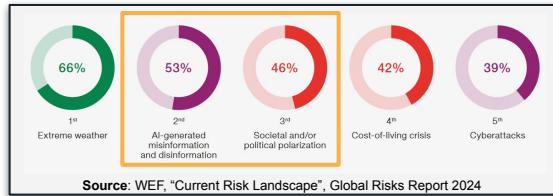


Source: Euronews

<https://www.euronews.com/my-europe/2025/03/03/is-ai-chatbot-grok-censoring-criticism-of-elon-musk-and-donald-trump> March 2025

Polarization and Misinformation: A Global Risk

- The convergence of polarization and AI-driven information dysfunction are among the most **significant global threats anticipated over the next decade**, highlighting an urgent need for intervention.



Our Research Agenda

- ❑ We develop computational approaches to answer three key questions:

Our Research Agenda

- ❑ We develop computational approaches to answer three key questions:
 - **What does polarization look like in language data?**
 - Model polarization at the level of entities, groups, and topics.
 - Extract polarization knowledge without relying on labeled political data or party annotations.

Our Research Agenda

- ❑ We develop computational approaches to answer three key questions:
 - **What does polarization look like in language data?**
 - Model polarization at the level of entities, groups, and topics.
 - Extract polarization knowledge without relying on labeled political data or party annotations.
 - **Can polarization insights help detect misinformation?**
 - Use extracted polarization structures to identify content that may be misleading or ideologically manipulative.
 - Assess the contribution of polarization cues on the performance of existing misinformation detection models.

Our Research Agenda

- ❑ We develop computational approaches to answer three key questions:
 - **What does polarization look like in language data?**
 - Model polarization at the level of entities, groups, and topics.
 - Extract polarization knowledge without relying on labeled political data or party annotations.
 - **Can polarization insights help detect misinformation?**
 - Use extracted polarization structures to identify content that may be misleading or ideologically manipulative.
 - Assess the contribution of polarization cues on the performance of existing misinformation detection models.
 - **How ideologically sensitive are Large Language Models?**
 - Assess whether models adopt beliefs or mimic styles when exposed to polarized prompts.

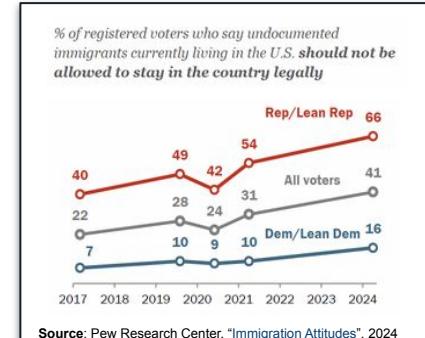
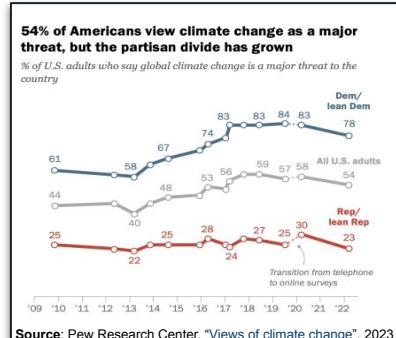
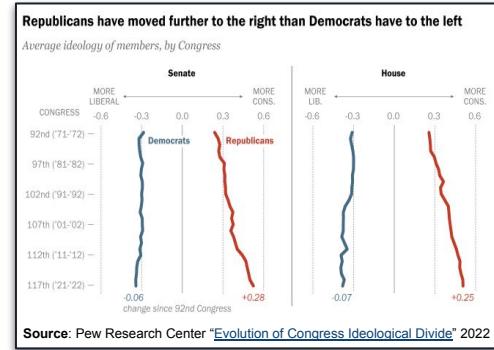
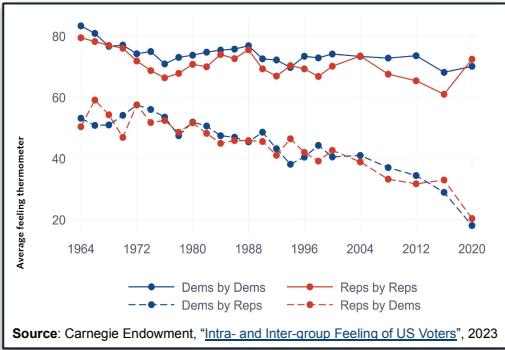
Polarization and Information Dysfunction

Polarization Pandemic

- ❑ Polarization is becoming a major concern at a global scale, with implications for social cohesion, public trust, and global stability.

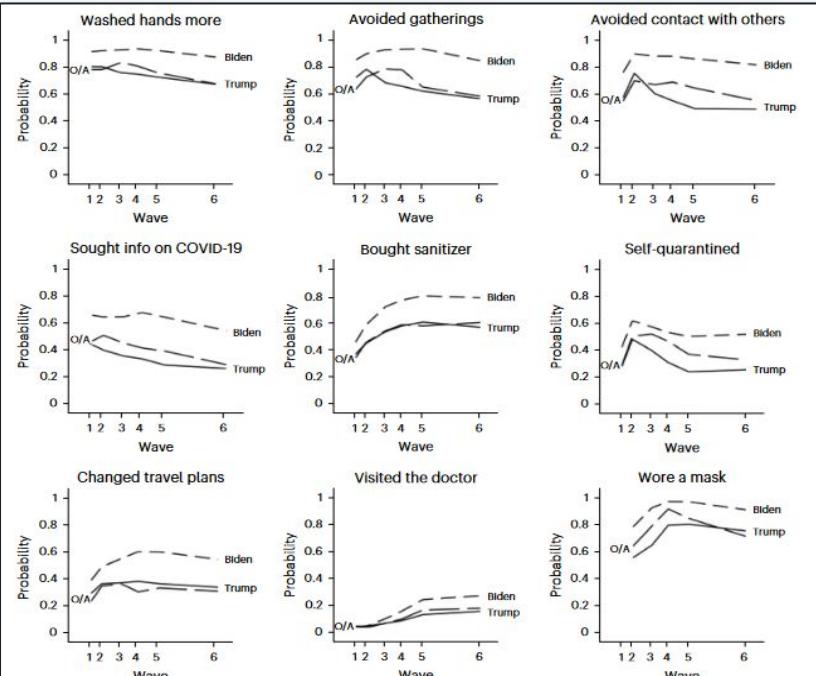
Polarization Pandemic

- ❑ Polarization is becoming a major concern at a global scale, with implications for social cohesion, public trust, and global stability.
- ❑ Evident influence in:
 - Electoral Processes
 - Social Justice Movements
 - Climate Change



Polarization Pandemic

- ❑ Polarization is becoming a major concern at a global scale, with implications for social cohesion, public trust, and global stability.
- ❑ Evident influence in:
 - Electoral Processes
 - Social Justice Movements
 - Climate Change
 - Public Health



Source: Jay J. Van Bavel, "Political polarization and Health", Nature Medicine, 2024

The Multi-level Nature of Polarization

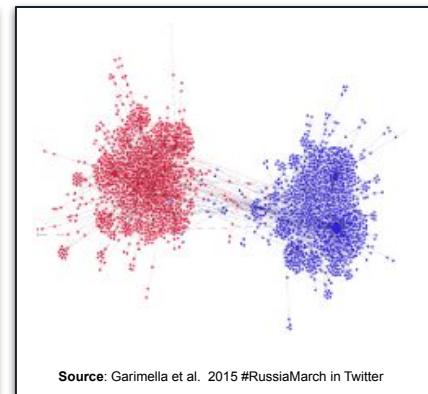
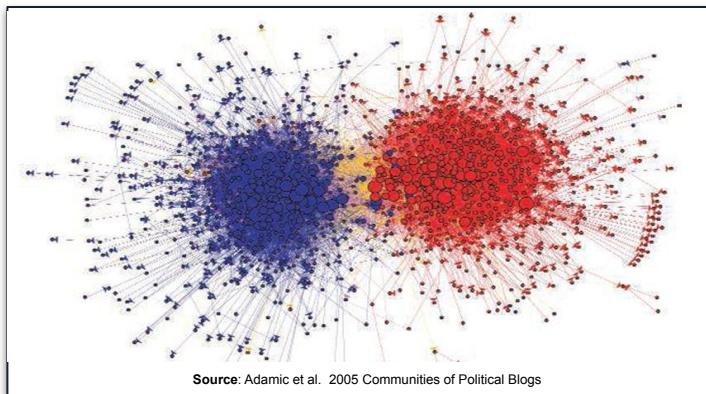
- ❑ Polarization is defined as the process where a (social or political) group is **segregated** into (two or more) **opposing sub-groups** with **conflicting beliefs**.

The Multi-level Nature of Polarization

- ❑ Polarization is defined as the process where a (social or political) group is **segregated** into (two or more) **opposing sub-groups** with **conflicting beliefs**.
- ❑ Polarization operates in multiple levels:
 - **Entity-level:** Individuals adopt **distinct ideologies and attitudes**. They cluster into **fellowships** characterized by **internal cohesion and shared viewpoints**.
 - **Group-level:** Fellowships interact and clash, with their **ideological and attitudinal differences** leading to the formation of **dipoles**.
 - **Topic-level:** Polarization manifests in **contentious topics**, on which dipole fellowships adopt **opposing positions**, leading to **heightened disagreement**.
- Polarization is a multi-dimensional phenomenon that shapes—and is shaped by—interactions within and across entities, groups, and topics.

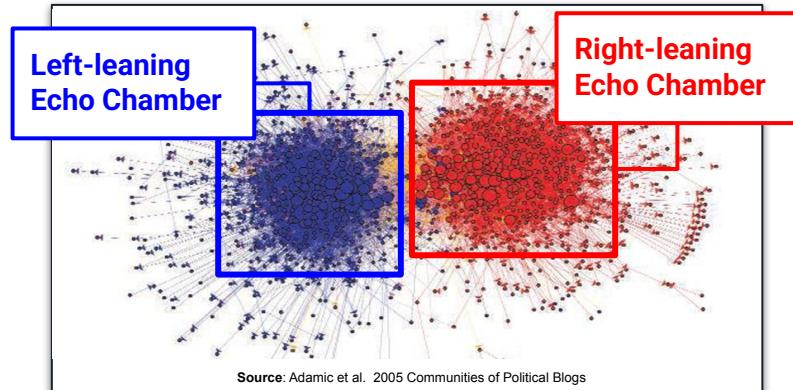
Polarization and the Digital Revolution

- ❑ Digital revolution has amplified the polarization impact, reshaping societal dynamics.
- ❑ Polarization thrives in online platforms such as Online Social Networks (OSNs), News Media, and Discussion Forums.
- ❑ These platforms foster environments for the spread of **Misinformation**, **Disinformation**, and **Hate-speech** → Referred to as **Information Dysfunction**.



Polarization and Information Dysfunction

- ❑ Polarization and Information Dysfunction are mutually reinforcing.
 - **Polarized communities seek** out and **amplify** content that **aligns with their beliefs**.
 - **False narratives** entrench people further into **opposing ideological positions**.



Polarization and Information Dysfunction

- ❑ Polarization and Information Dysfunction are mutually reinforcing.
 - **Polarized communities seek** out and **amplify** content that **aligns with their beliefs**.
 - **False narratives** entrench people further into **opposing ideological positions**.
- ❑ The rise of **Generative (GenAI)** and **Large Language Models (LLMs)** further compound these challenges.
 - Enable the **rapid** and **convincing** creation of **misleading articles**.

How Disinformation From a Russian AI Spam Farm Ended up on Top of Google Search Results

Source: WIRED, "[AI Generated Russian Disinformation](#)", 2024

Propagandists are using AI too—and companies need to be open about it

OpenAI has reported on influence operations that use its AI tools. Such reporting, alongside data sharing, should become the industry norm.

Source: "[Propagandists are using AI too—and Companies need to be Open about it](#)", MIT Technology Review, 2024

How AI-generated disinformation might impact this year's elections and how journalists should report on it

From satire to robocalls, generative AI is entering politics in a crucial year. Four experts reflect on its possible consequences and on how to cover it.

Source: Reuters, "[AI Generated Disinformation on Elections](#)", 2024

Polarization and Misinformation: A Global Risk

- ❑ The convergence of polarization and AI-driven information dysfunction are among the most **significant global threats anticipated over the next decade**, highlighting an urgent need for intervention.
- ❑ **Current Mitigation Measures:**
 - Large fines and pressure → Aggressive censorship.
 - ML models to classify content → Biased training data and cannot handle the large volume and velocity in which online content is produced.
- ❑ Transition from **Monitoring** and **Debunking** to **Measuring** and **Understanding**.



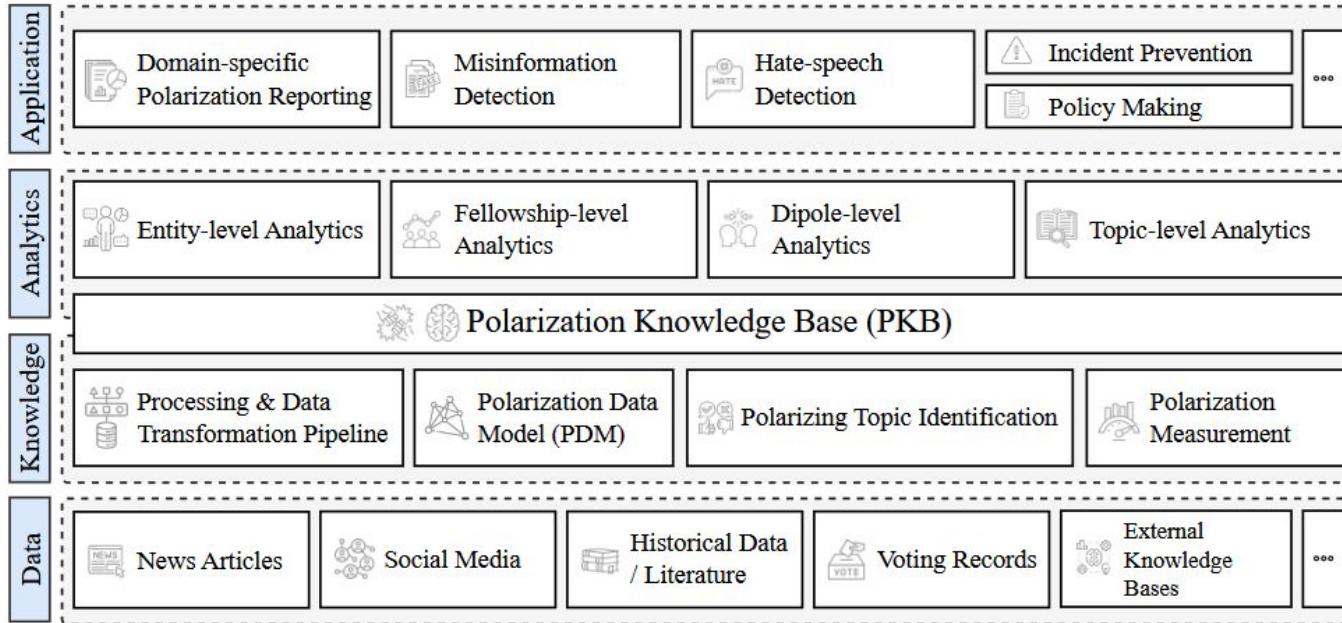
Polarization and Misinformation: A Global Risk

- ❑ The convergence of polarization and AI-driven information dysfunction are among the most **significant global threats anticipated over the next decade**, highlighting an urgent need for intervention.
- ❑ **Current Mitigation Measures:**
 - Large fines and pressure → Aggressive censorship.
 - ML models to classify content → Biased training data and cannot handle the large volume and velocity in which online content is produced.
- ❑ Transition from **Monitoring** and **Debunking** to **Measuring** and **Understanding**.



Need of effective computational and algorithmic approaches to model, extract, and analyze polarization, its dynamics, and its relationship with information dysfunction.

Abstract Comprehensive Polarization Framework



Paschalides, D., Pallis, G., Dikaiakos, D. M. "PARALLAX: Leveraging Polarization Knowledge for Misinformation Detection". ASONAM 2024

Paschalides, D., Pallis, G., Dikaiakos, D. M. "A Framework for the Unsupervised Modeling and Extraction of Polarization Knowledge from News Media" TSC 2024

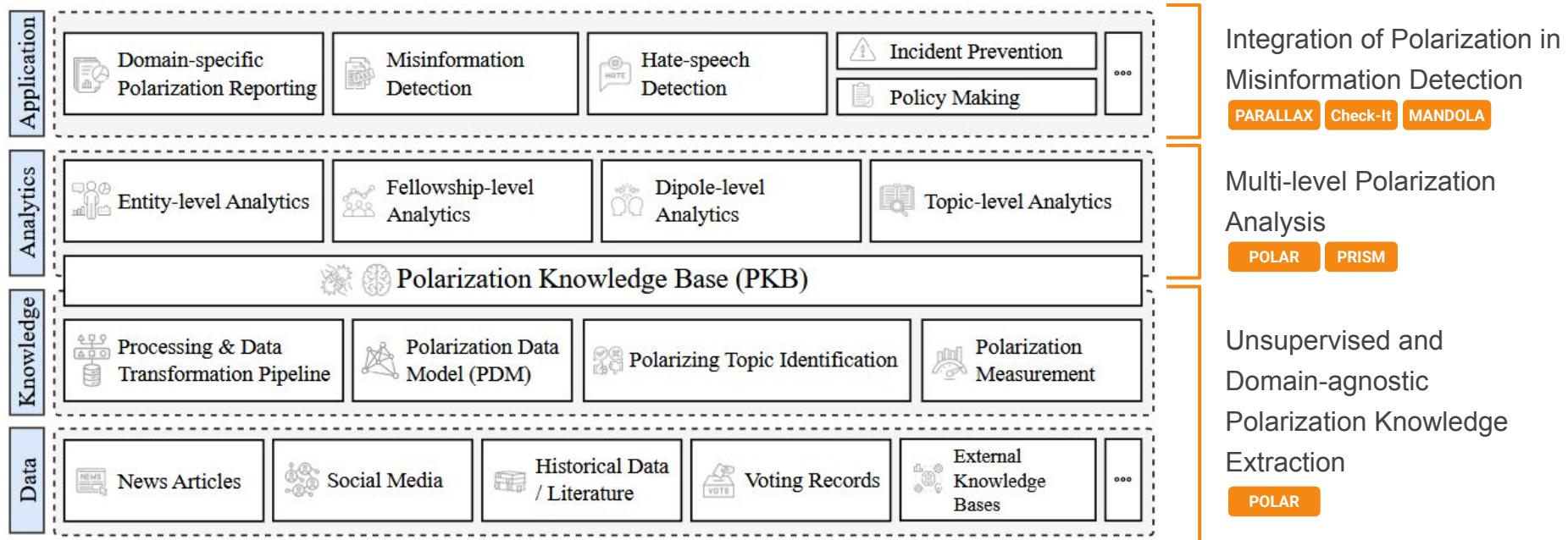
Paschalides, D., Pallis, G., Dikaiakos, D. M. "POLAR: A Holistic Framework for the Modeling of Polarization and Identification of Polarizing Topics in News Media" ASONAM 2021

Paschalides, D., Pallis, G., Dikaiakos, D. M. et al. "Check-It: A Plugin for Detecting Fake News on the Web" OSNEM 2021

Paschalides, D., Pallis, G., Dikaiakos, D. M. et al. "Mandola: A big-data processing and visualization platform for monitoring and detecting online hate speech" TOIT 2020.

Paschalides, D., Pallis, G., Dikaiakos, D. M. et al. "Check-It: A plugin for Detecting and Reducing the Spread of Fake News and Misinformation on the Web". WI 2019.

Abstract Comprehensive Polarization Framework



Paschalides, D., Pallis, G., Dikaiakos, D. M. "PARALLAX: Leveraging Polarization Knowledge for Misinformation Detection". ASONAM 2024

Paschalides, D., Pallis, G., Dikaiakos, D. M. "A Framework for the Unsupervised Modeling and Extraction of Polarization Knowledge from News Media" TSC 2024

Paschalides, D., Pallis, G., Dikaiakos, D. M. "POLAR: A Holistic Framework for the Modeling of Polarization and Identification of Polarizing Topics in News Media" ASONAM 2021

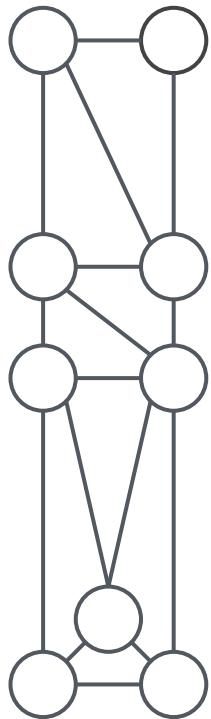
Paschalides, D., Pallis, G., Dikaiakos, D. M. et al. "Check-It: A Plugin for Detecting Fake News on the Web" OSNEM 2021

Paschalides, D., Pallis, G., Dikaiakos, D. M. et al. "Mandola: A big-data processing and visualization platform for monitoring and detecting online hate speech" TOIT 2020.

Paschalides, D., Pallis, G., Dikaiakos, D. M. et al. "Check-It: A plugin for Detecting and Reducing the Spread of Fake News and Misinformation on the Web". WI 2019.

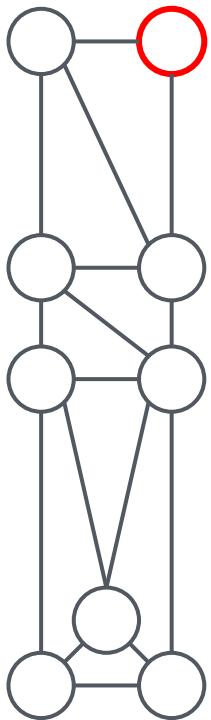
Polarization Data Model Definition

Polarization Data Model



The process whereby a **(social or political) group** is segregated into two or more opposing sub-groups with conflicting beliefs.

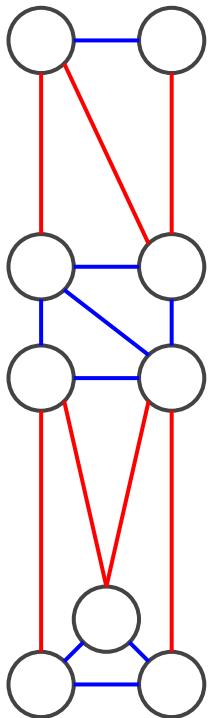
Polarization Data Model



The process whereby a **(social or political) group** is segregated into two or more opposing sub-groups with conflicting beliefs.

- ❑ **Entity:** Real world presence, with abstract or physical existence, that individually or collectively possess / represents views on various topics.
Can be a Person, Organization, Nationality, Religion, etc.

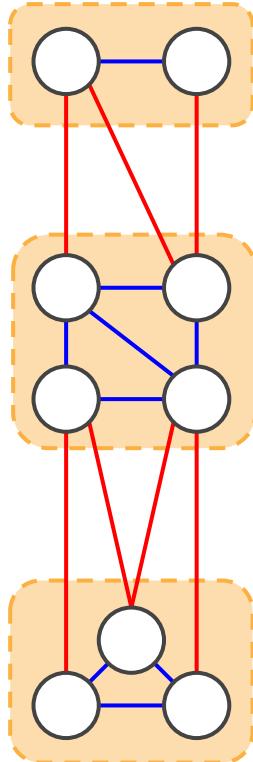
Polarization Data Model



The process whereby a **(social or political) group** is segregated into two or more opposing sub-groups with conflicting beliefs.

- ❑ **Entity:** Real world presence, with abstract or physical existence, that individually or collectively possess / represents views on various topics.
Can be a Person, Organization, Nationality, Religion, etc.
- ❑ **Entity Relationship:** **Positive** (supportive) or **Negative** (oppositional) nature.

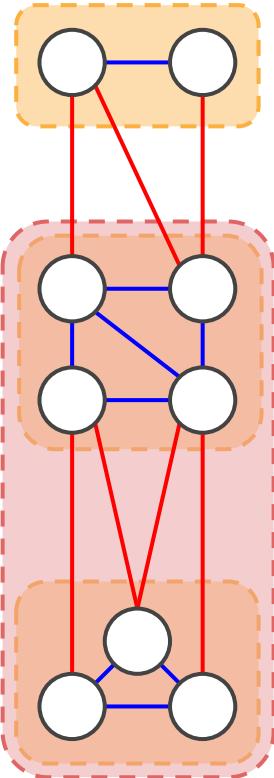
Polarization Data Model



The process whereby a **(social or political) group** is segregated into two or more opposing **sub-groups** with conflicting beliefs.

- ❑ **Entity:** Real world presence, with abstract or physical existence, that individually or collectively possess / represents views on various topics.
Can be a Person, Organization, Nationality, Religion, etc.
- ❑ **Entity Relationship:** **Positive** (supportive) or **Negative** (oppositional) nature.
- ❑ **Entity Fellowship:** Community described by the (mostly) **positive** entity relationships.

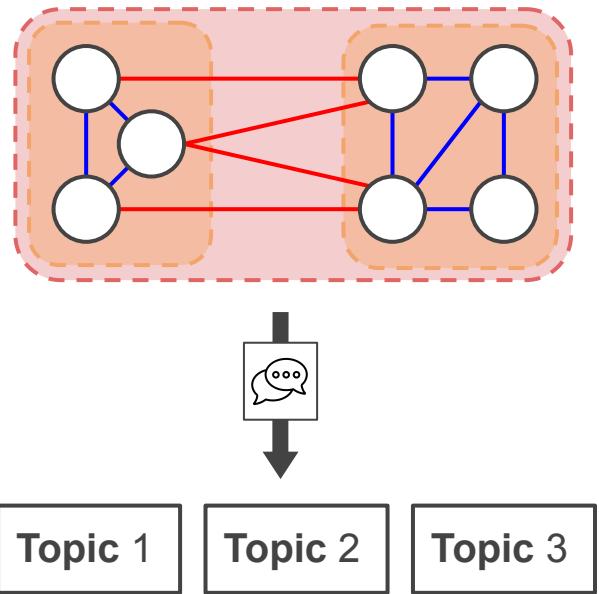
Polarization Data Model



The process whereby a **(social or political) group** is segregated into two or more **opposing sub-groups** with **conflicting beliefs**.

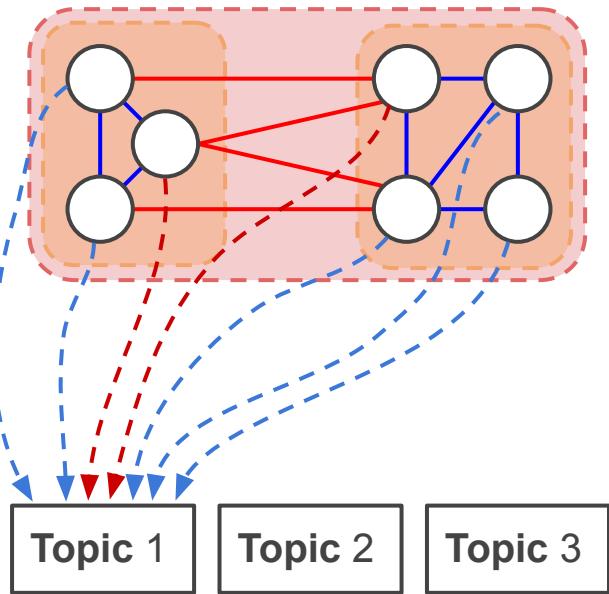
- ❑ **Entity:** Real world presence, with abstract or physical existence, that individually or collectively possess / represents views on various topics.
Can be a Person, Organization, Nationality, Religion, etc.
- ❑ **Entity Relationship:** **Positive** (supportive) or **Negative** (oppositional) nature.
- ❑ **Entity Fellowship:** Community described by the (mostly) **positive** entity relationships.
- ❑ **Fellowship Dipole:** Fellowship pairs with (mostly) **Negative** relationships between them.

Polarization Data Model

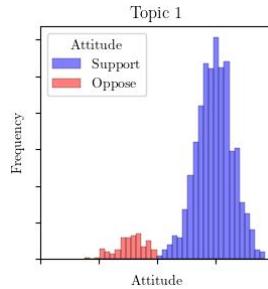


- ❑ **Dipole Discussion Topics:** Topics of discussion between dipole's conflicting fellowships.

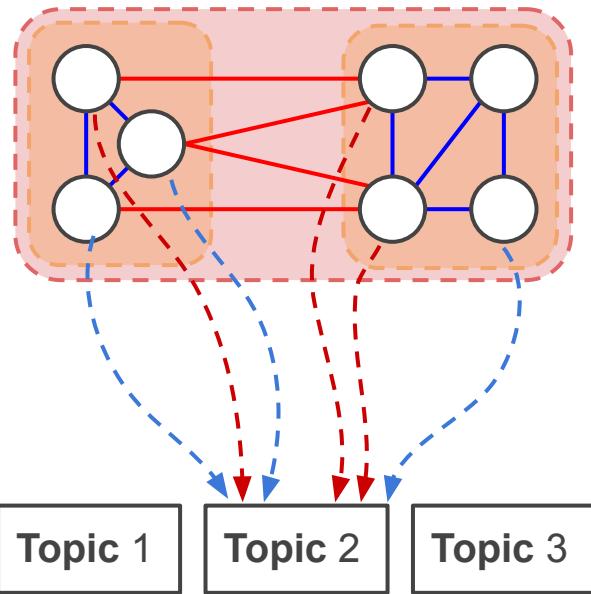
Polarization Data Model



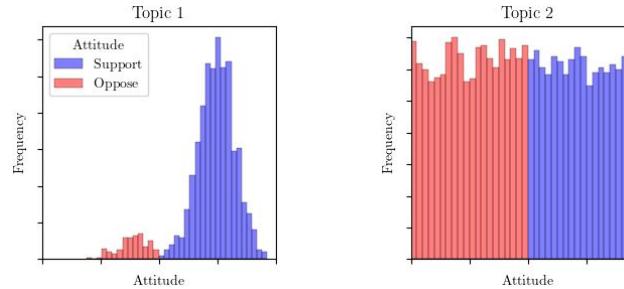
- ❑ **Dipole Discussion Topics:** Topics of discussion between dipole's conflicting fellowships.
- ❑ **Sentiment Attitudes:** **Opposition (O)** or **Support (S)** from entities toward topics.



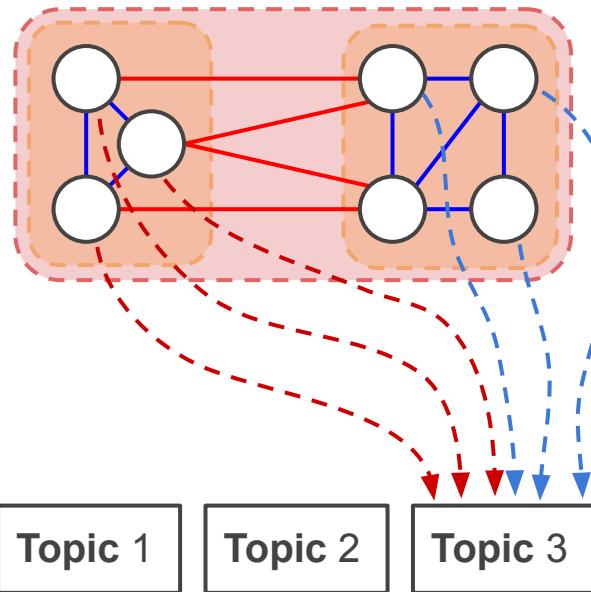
Polarization Data Model



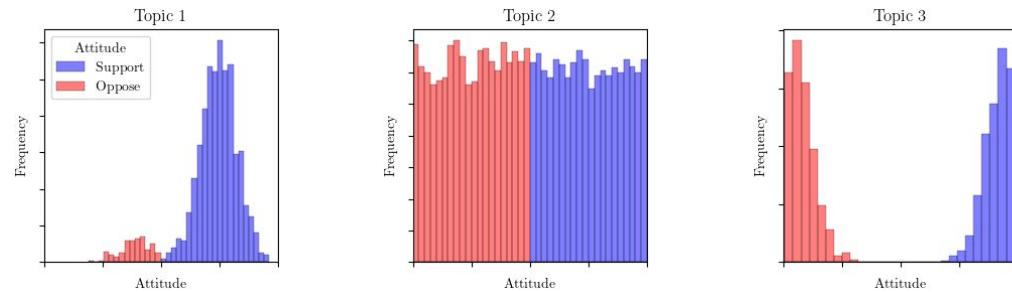
- ❑ **Dipole Discussion Topics:** Topics of discussion between dipole's conflicting fellowships.
- ❑ **Sentiment Attitudes:** **Opposition (O)** or **Support (S)** from entities toward topics.



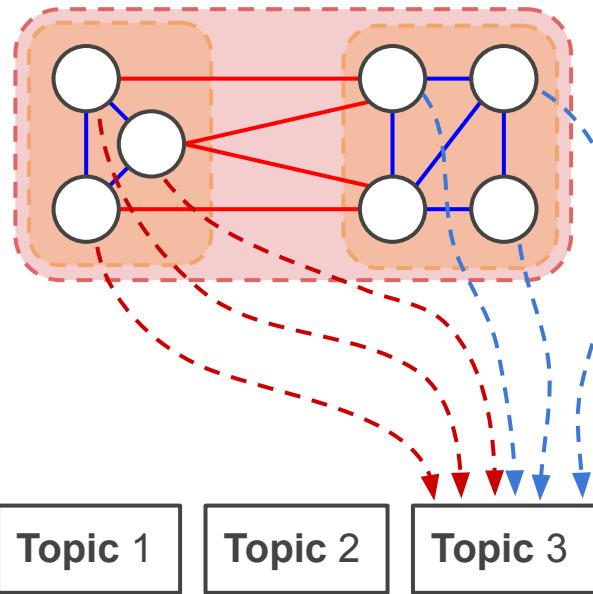
Polarization Data Model



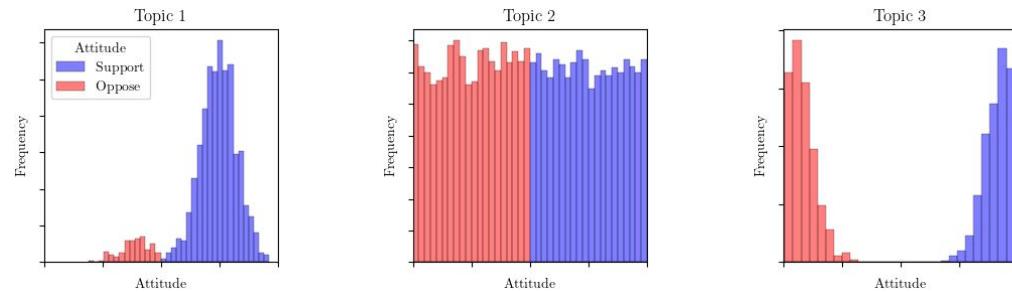
- ❑ **Dipole Discussion Topics:** Topics of discussion between dipole's conflicting fellowships.
- ❑ **Sentiment Attitudes:** **Opposition (O)** or **Support (S)** from entities toward topics.



Polarization Data Model



- **Dipole Discussion Topics:** Topics of discussion between dipole's conflicting fellowships.
- **Sentiment Attitudes:** **Opposition (O)** or **Support (S)** from entities toward topics.

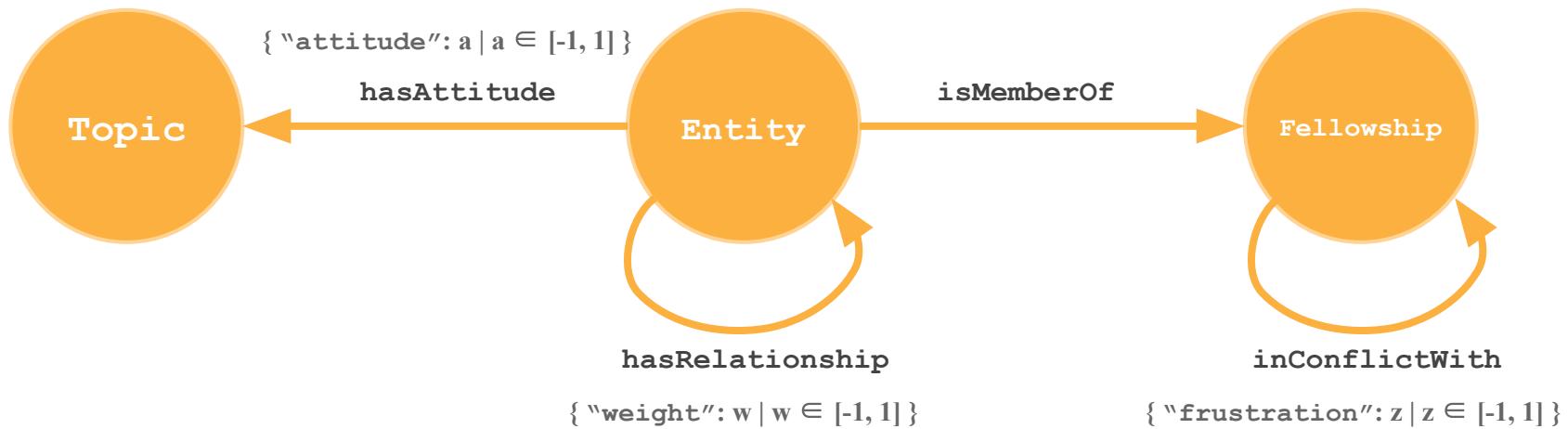


Topic 1: Consensus
Topic 2: Balance
Topic 3: Extreme Separation

→ No Polarization
→ No Polarization
→ Polarization

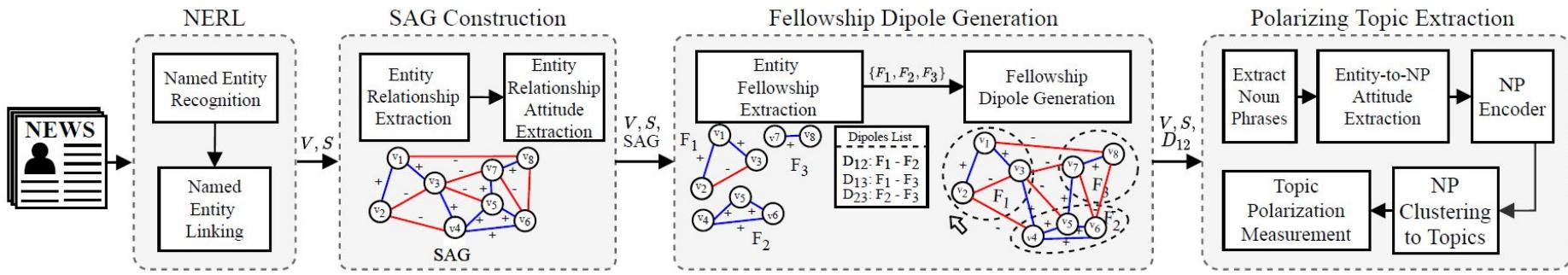
Polarization Data Model Schema

- ❑ PDM is a typed, directed, and weighted graph that encodes how entities, groups, and topics interact within a polarized space.



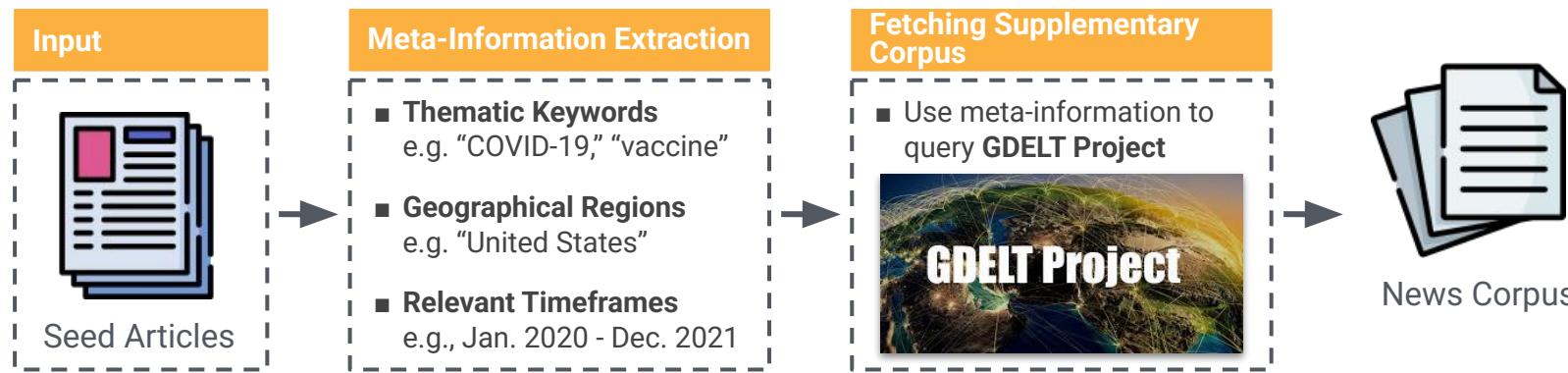
Unsupervised Polarization Knowledge Extraction

POLAR Framework



News Articles as Input

- ❑ OSN posts are short, noisy, and informal.
 - ❑ Focus on news articles:
 - Written in formal language.
 - Provide additional context.
 - Discuss specific subjects.
- Ideal for knowledge extraction.



Process News Articles

Example News Article

“Minneapolis police officer, Derek Chauvin, has been charged with the murder of Floyd. ... The Black Lives Matter movement demanded justice for the death of George Floyd from Minneapolis and president Trump, with Benjamin Crump and Erica McDonald promised to deliver. ... Trump assigns National Guard to the disposal of states’ governors as riots respond similarly to Minneapolis.”



Segment Article into Sentences

Example News Article

"Minneapolis police officer, Derek Chauvin, has been charged with the murder of Floyd. ... The Black Lives Matter movement demanded justice for the death of George Floyd from Minneapolis and president Trump, with Benjamin Crump and Erica McDonald promised to deliver. ... Trump assigns National Guard to the disposal of states' governors as riots respond similarly to Minneapolis."

Sentence Segmentation

Article Sentences S

"Minneapolis police officer, Derek Chauvin, has been charged with the murder of Floyd.

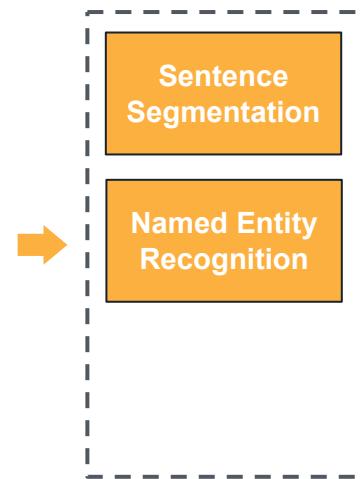
The Black Lives Matter movement demanded justice for the death of George Floyd from Minneapolis and president Trump, with Benjamin Crump and Erica McDonald promised to deliver.

Trump assigns National Guard to the disposal of states' governors as riots respond similarly to Minneapolis."

Apply Named Entity Recognition

Example News Article

"Minneapolis police officer, Derek Chauvin, has been charged with the murder of Floyd. ... The Black Lives Matter movement demanded justice for the death of George Floyd from Minneapolis and president Trump, with Benjamin Crump and Erica McDonald promised to deliver. ... Trump assigns National Guard to the disposal of states' governors as riots respond similarly to Minneapolis."



Article Sentences S

"Minneapolis police officer, **Derek Chauvin**, has been charged with the murder of **Floyd**.

The **Black Lives Matter** movement demanded justice for the death of **George Floyd** from **Minneapolis** and president **Trump**, with **Benjamin Crump** and **Erica McDonald** promised to deliver.

Trump assigns **National Guard** to the disposal of states' governors as riots respond similarly to **Minneapolis**."

Named Entities V

(Minneapolis police, ORG),
(Derek Chauvin, PERSON),
(Floyd, PERSON)

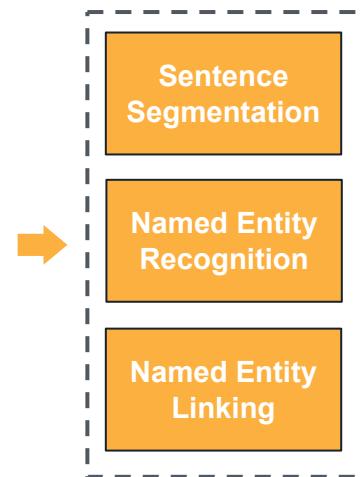
(Black Lives Matter, ORG),
(George Floyd, PERSON),
(Minneapolis, LOC),
(Benjamin Crump, PERSON),
(Trump, PERSON),
(Erica McDonald, PERSON)

(Trump, PERSON),
(National Guard, ORG),
(Minneapolis, LOC)

Link Named Entities

Example News Article

"Minneapolis police officer, Derek Chauvin, has been charged with the murder of Floyd. ... The Black Lives Matter movement demanded justice for the death of George Floyd from Minneapolis and president Trump, with Benjamin Crump and Erica McDonald promised to deliver. ... Trump assigns National Guard to the disposal of states' governors as riots respond similarly to Minneapolis."



Article Sentences S

"Minneapolis police Dpt. officer, Derek Chauvin, has been charged with the murder of George Floyd.

The Black Lives Matter movement demanded justice for the death of George Floyd from Minneapolis and president Donald Trump, with Benjamin Crump and Erica McDonald promised to deliver.

Donald Trump assigns National Guard to the disposal of states' governors as riots respond similarly to Minneapolis."

Named Entities V

(Minneapolis police, ORG),
(Derek Chauvin, PERSON),
(George Floyd, PERSON)

(Black Lives Matter, ORG),
(George Floyd, PERSON),
(Minneapolis, LOC),
(Benjamin Crump, PERSON),
(Donald Trump, PERSON),
(Erica McDonald, PERSON)

(Donald Trump, PERSON),
(National Guard, ORG),
(Minneapolis, LOC)

Sentiment Attitude Graph Generation

Article Sentences S

"Minneapolis police Dpt. officer, Derek Chauvin, has been charged with the murder of George Floyd.

The Black Lives Matter movement demanded justice for the death of George Floyd from Minneapolis and president Donald Trump, with Benjamin Crump and Erica McDonald promised to deliver.

Donald Trump assigns National Guard to the disposal of states' governors as riots respond similarly to Minneapolis."

Named Entities V

(Minneapolis police, ORG),
(Derek Chauvin, PERSON),
(George Floyd, PERSON)

(Black Lives Matter, ORG),
(George Floyd, PERSON),
(Minneapolis, LOC),
(Benjamin Crump, PERSON),
(Donald Trump, PERSON),
(Erica McDonald, PERSON)

(Donald Trump, PERSON),
(National Guard, ORG),
(Minneapolis, LOC)



Entity Relationship Identification

Article Sentences S

"Minneapolis police Dpt. officer, Derek Chauvin, has been charged with the murder of George Floyd.

The Black Lives Matter movement demanded justice for the death of George Floyd from Minneapolis and president Donald Trump, with Benjamin Crump and Erica McDonald promised to deliver.

Donald Trump assigns National Guard to the disposal of states' governors as riots respond similarly to Minneapolis."

Named Entities V

(Minneapolis police, ORG),
 (Derek Chauvin, PERSON),
 (George Floyd, PERSON)

(Black Lives Matter, ORG),
 (George Floyd, PERSON),
 (Minneapolis, LOC),
 (Benjamin Crump, PERSON),
 (Donald Trump, PERSON),
 (Erica McDonald, PERSON)

(Donald Trump, PERSON),
 (National Guard, ORG),
 (Minneapolis, LOC)

Entity-to-Entity Relationship Extraction

Entity 1 (v_1)	Entity 2 (v_2)
Derek Chauvin	George Floyd
Minneapolis	George Floyd
Minneapolis Police Dpt.	George Floyd
...	
Minneapolis Police Dpt.	Derek Chauvin
Black Lives Matter	George Floyd
Black Lives Matter	Minneapolis

Sentiment Attitude Calculation

Article Sentences S

"Minneapolis police Dpt. officer, Derek Chauvin, has been charged with the murder of George Floyd.

The Black Lives Matter movement demanded justice for the death of George Floyd from Minneapolis and president Donald Trump, with Benjamin Crump and Erica McDonald promised to deliver.

Donald Trump assigns National Guard to the disposal of states' governors as riots respond similarly to Minneapolis."

Named Entities V

(Minneapolis police, ORG),
(Derek Chauvin, PERSON),
(George Floyd, PERSON)

(Black Lives Matter, ORG),
(George Floyd, PERSON),
(Minneapolis, LOC),
(Benjamin Crump, PERSON),
(Donald Trump, PERSON),
(Erica McDonald, PERSON)

(Donald Trump, PERSON),
(National Guard, ORG),
(Minneapolis, LOC)

Entity-to-Entity
Relationship
Extraction

Sentiment
Attitude
Calculation

Entity 1 (v_1)	Entity 2 (v_2)	Status
Derek Chauvin	George Floyd	Neg.
Minneapolis	George Floyd	Neg.
Minneapolis Police Dpt.	George Floyd	Neg.
...		
Minneapolis Police Dpt.	Derek Chauvin	Pos.
Black Lives Matter	George Floyd	Pos.
Black Lives Matter	Minneapolis	Neg.

SAG Population

Article Sentences S

"Minneapolis police Dpt. officer, Derek Chauvin, has been charged with the murder of George Floyd.

The Black Lives Matter movement demanded justice for the death of George Floyd from Minneapolis and president Donald Trump, with Benjamin Crump and Erica McDonald promised to deliver.

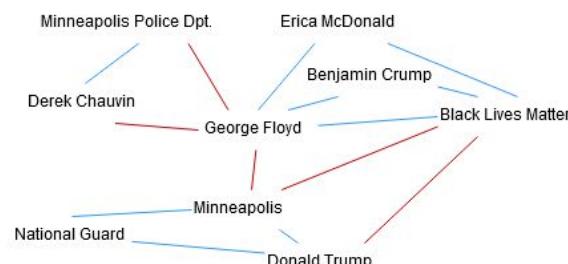
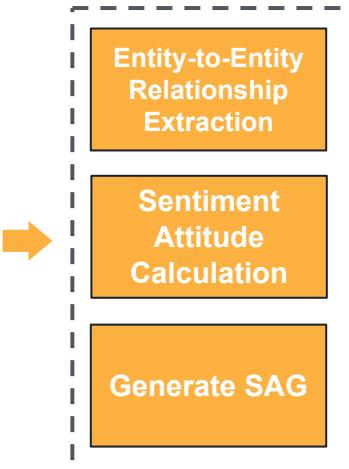
Donald Trump assigns National Guard to the disposal of states' governors as riots respond similarly to Minneapolis."

Named Entities V

(Minneapolis police, ORG),
(Derek Chauvin, PERSON),
(George Floyd, PERSON)

(Black Lives Matter, ORG),
(George Floyd, PERSON),
(Minneapolis, LOC),
(Benjamin Crump, PERSON),
(Donald Trump, PERSON),
(Erica McDonald, PERSON)

(Donald Trump, PERSON),
(National Guard, ORG),
(Minneapolis, LOC)



Identifying Entity Relationships

- ❑ **Principle of Learning by Continuity:** Objects, once experienced together, tend to be related to each other.
 - Calculate entity co-occurrence frequencies in sentences.
 - Higher the co-occurrence frequency → More probable real-life connection.
 - $\geq 95^{\text{th}}$ quantile (top 5% of co-occurring pairs with the highest frequencies).

Identifying Entity Relationships

- ❑ **Principle of Learning by Continuity:** Objects, once experienced together, tend to be related to each other.
 - Calculate entity co-occurrence frequencies in sentences.
 - Higher the co-occurrence frequency → More probable real-life connection.
 - $\geq 95^{\text{th}}$ quantile (top 5% of co-occurring pairs with the highest frequencies).

Example Sentences

- “Minneapolis Police Dpt. officer, Derek Chauvin, has been charged with the murder of George Floyd.”
- “George Floyd was murdered by Derek Chauvin.”



Entity 1 (v_1)	Entity 2 (v_2)	Co-occurrence Frequency
Derek Chauvin	George Floyd	2
Minneapolis Police Dpt.	George Floyd	1
Minneapolis Police Dpt.	Derek Chauvin	1

Sentiment Attitude Calculation

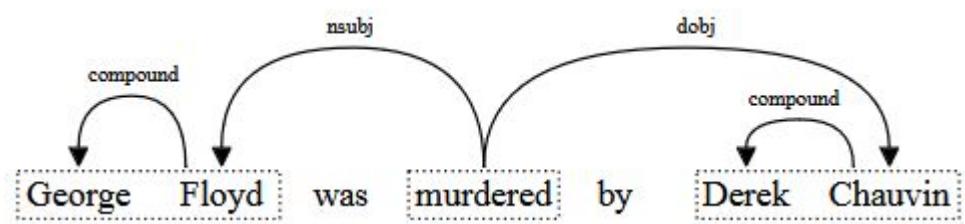
- ❑ **Sentiment Attitude:** Directed sentiment from one entity in the text towards another.
 - Calculated as the **sentiment score** of the **syntactical dependency path** of **Entity 1** and **Entity 2** within a sentence.

Sentiment Attitude Calculation

- ❑ **Sentiment Attitude:** Directed sentiment from one entity in the text towards another.
 - Calculated as the **sentiment score** of the **syntactical dependency path** of **Entity 1** and **Entity 2** within a sentence.

Example Sentences

- “Minneapolis Police Dpt. officer, Derek Chauvin, has been charged with the murder of George Floyd.”
- George Floyd was murder by Derek Chauvin.

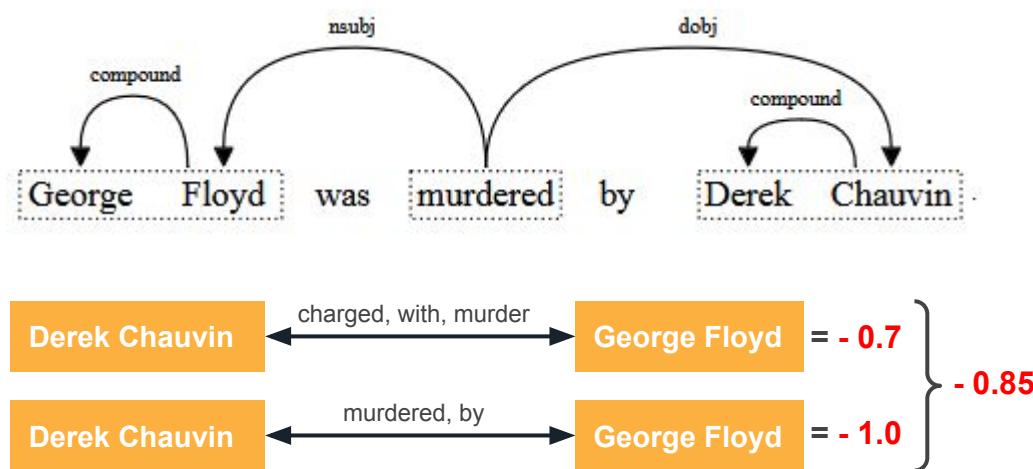


Sentiment Attitude Calculation

- ❑ **Sentiment Attitude:** Directed sentiment from one entity in the text towards another.
 - Calculated as the **sentiment score** of the **syntactical dependency path** of **Entity 1** and **Entity 2** within a sentence.

Example Sentences

- “Minneapolis Police Dpt. officer, Derek Chauvin, has been charged with the murder of George Floyd.”
- George Floyd was murder by Derek Chauvin.



Extracting Entity Fellowships

- ❑ **Fellowship:** Densely connected subgraph of SAG with mostly positive attitudes.

Extracting Entity Fellowships

- ❑ **Fellowship:** Densely connected subgraph of SAG with mostly positive attitudes.
- ❑ **Signed Network Clustering:** “Finding **clusters** such that **most edges within** clusters are **positive**, and **most edges across** clusters are **negative**.”

Extracting Entity Fellowships

- ❑ **Fellowship:** Densely connected subgraph of SAG with mostly positive attitudes.
- ❑ **Signed Network Clustering:** “Finding **clusters** such that **most edges within** clusters are **positive**, and **most edges across** clusters are **negative**.”
- ❑ Convert to signed network: $thr = \frac{\text{median}(\mathcal{A}_{v_i v_j}^+) + \text{median}(\mathcal{A}_{v_i v_j}^-)}{2}$, assign **+1** if $w_{ij} \geq thr$ else **-1**.

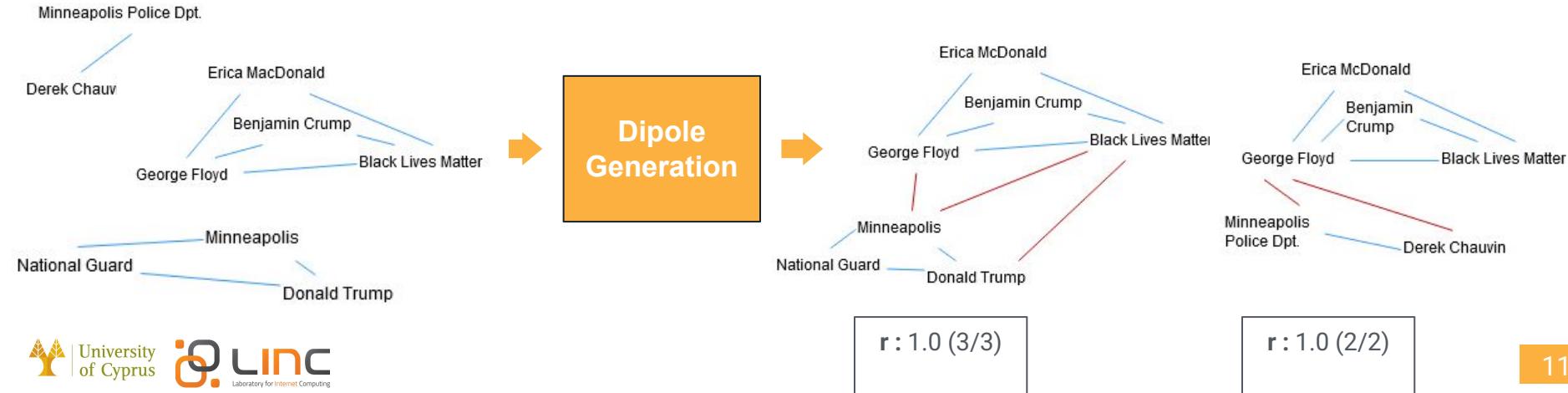
Extracting Entity Fellowships

- ❑ **Fellowship:** Densely connected subgraph of SAG with mostly positive attitudes.
- ❑ **Signed Network Clustering:** “Finding **clusters** such that **most edges within** clusters are **positive**, and **most edges across** clusters are **negative**.”
- ❑ Convert to signed network: $thr = \frac{\text{median}(\mathcal{A}_{v_i v_j}^+) + \text{median}(\mathcal{A}_{v_i v_j}^-)}{2}$, assign **+1** if $w_{ij} \geq thr$ else **-1**.
- ❑ Avoid spectral clustering (definition of **k**) → Use SiMap based on resolution λ .



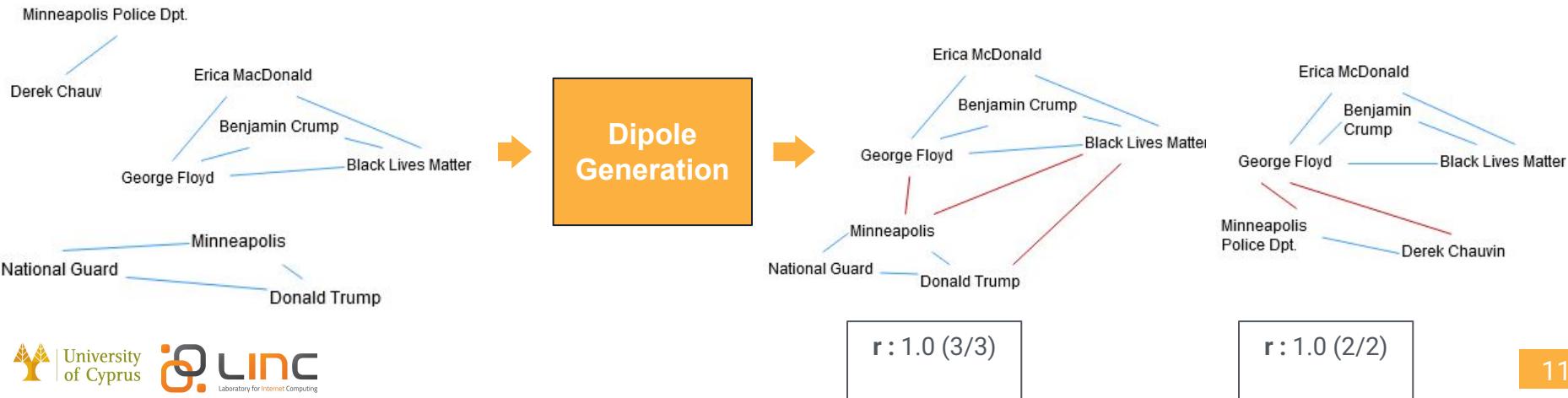
Generate Fellowship Dipoles

- ❑ Initial Dipole Set → All possible Fellowship pairs.
- ❑ Maximize the probability of a polarized state using heuristics:
 - **Negative Across (r):** The ratio of the number of **negative** edges to the total number of edges connecting the two fellowships → Threshold set 0.5.



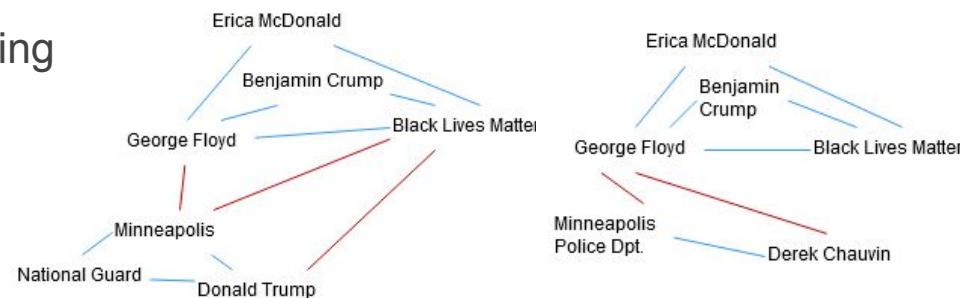
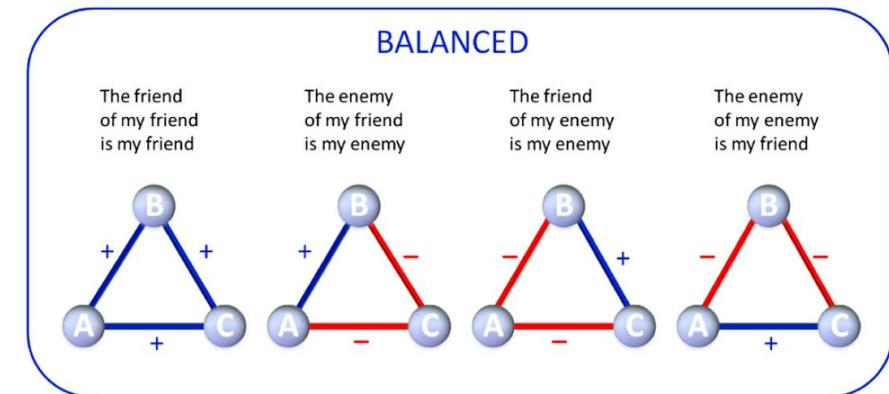
Generate Fellowship Dipoles

- ❑ Initial Dipole Set → All possible Fellowship pairs.
- ❑ Maximize the probability of a polarized state using heuristics:
 - **Negative Across (r)**: The ratio of the number of **negative** edges to the total number of edges connecting the two fellowships → Threshold set 0.5.
 - **Frustration**: Level of dipole's structural balance. 



Frustration Index (f)

- ❑ Takes into account the structural balance of a dipole.
- ❑ Intuition: A Dipole with **high structural balance** → **high polarized state**.
- ❑ Metric: **Frustration Index**
 - Minimum number of edges whose removal results in balance.
 - Normalized between 0 to 1, with **0 being totally imbalanced**, and **1 perfectly balanced**.
- ❑ → Threshold set to 0.7.



$r : 1.0 (3/3)$
 $f : 1.0 (0/11)$

$r : 1.0 (2/2)$
 $f : 1.0 (0/8)$

Dipole Topic Extraction

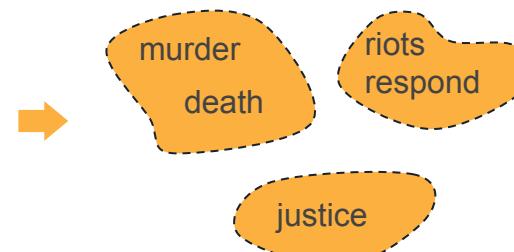
- ❑ **Topics:** Semantically similar Noun Phrases (NPs)
 - **Noun-phrase:** Phrases that include nouns → Important in understanding context.

Article Sentences S

“Minneapolis Police Dpt. officer, Derek Chauvin, has been charged with **the murder** of George Floyd.

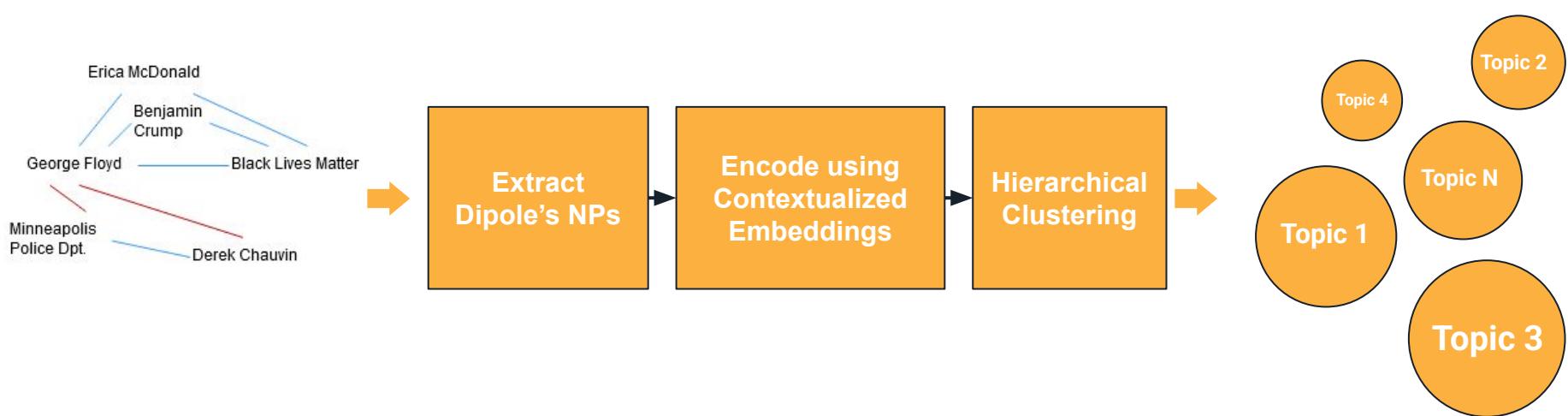
The Black Lives Matter movement demanded **justice** for **the death** of George Floyd from Minneapolis and president Donald Trump, with Benjamin Crump and Erica McDonald promised to deliver.

Donald Trump assigns National Guard to **the disposal** of **states' governors** as **riots respond** similarly to Minneapolis.”



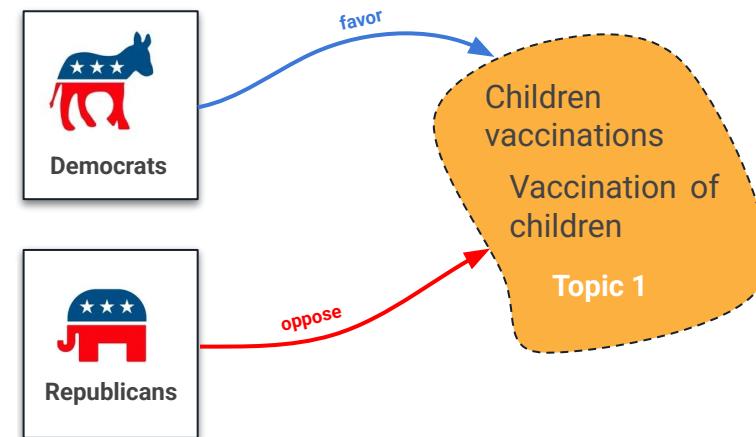
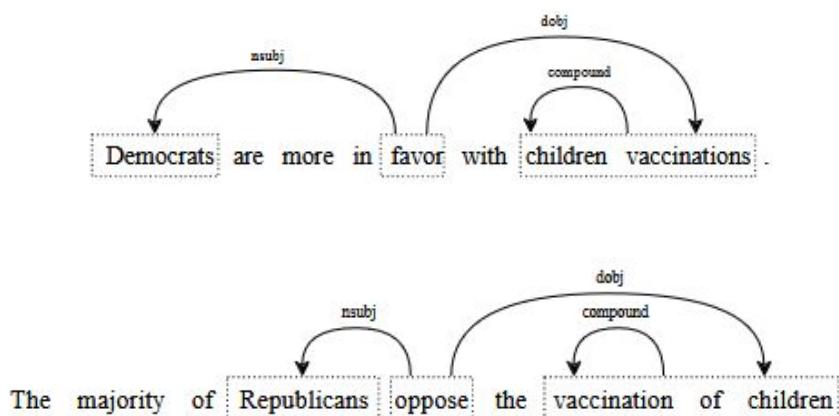
Dipole Topic Extraction

- ❑ **Topics:** Semantically similar Noun Phrases (NPs)
 - **Noun-phrase:** Phrases that include nouns → Important in understanding context.



Topic Attitude Extraction

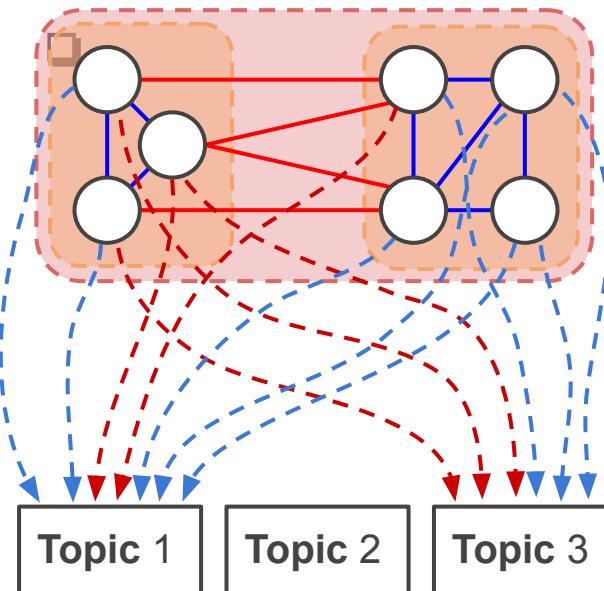
- **Sentiment Attitude:** Directed sentiment from an entity towards a topic-related NP.



Quantifying Topic Polarization

- ❑ **Polarization Index μ :** a population is perfectly polarized when divided into two groups of the same size and with opposite attitudes.
- ❑ $\mu=1$ if attitudes are **perfectly polarized**, and $\mu=0$ if **not polarized at all**.

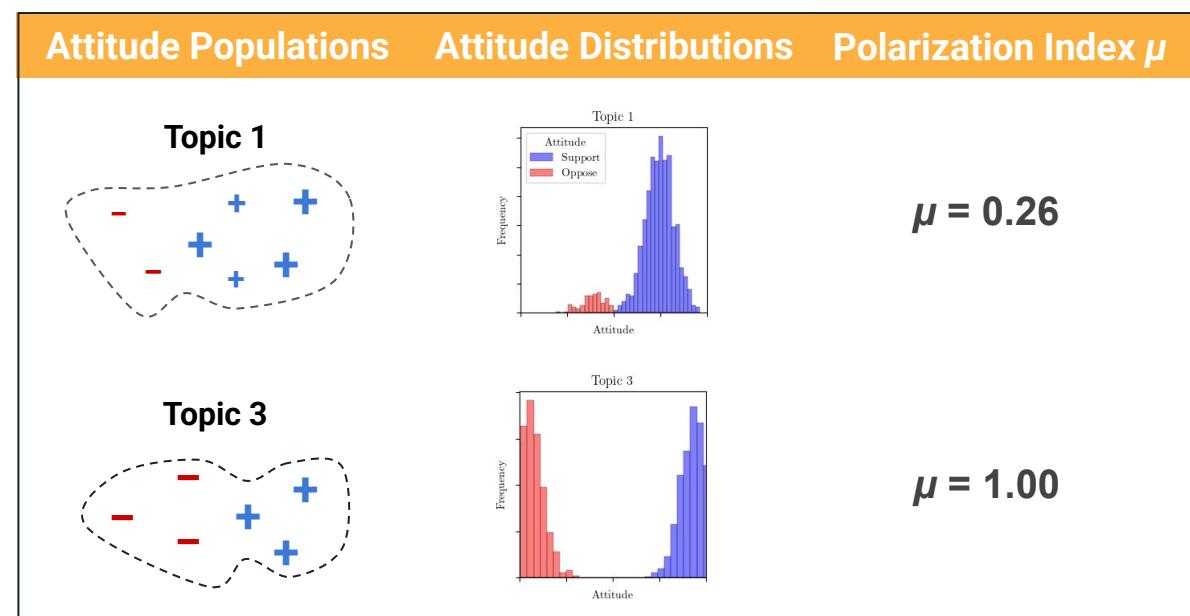
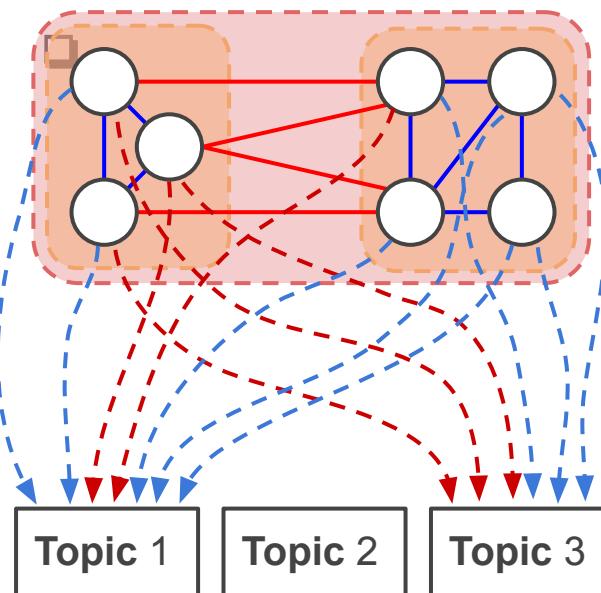
$$\mu = (1 - \Delta A)d.$$



Quantifying Topic Polarization

- ❑ **Polarization Index μ :** a population is perfectly polarized when divided into two groups of the same size and with opposite attitudes.
- ❑ $\mu=1$ if attitudes are **perfectly polarized**, and $\mu=0$ if **not polarized at all**.

$$\mu = (1 - \Delta A)d.$$



Polarization Knowledge Evaluation Methodology

Evaluating Polarization Knowledge

- We establish an **evaluation methodology** for assessing the **correctness** of the PK **across its levels**, by utilizing **external sources** and **tailored metrics**.

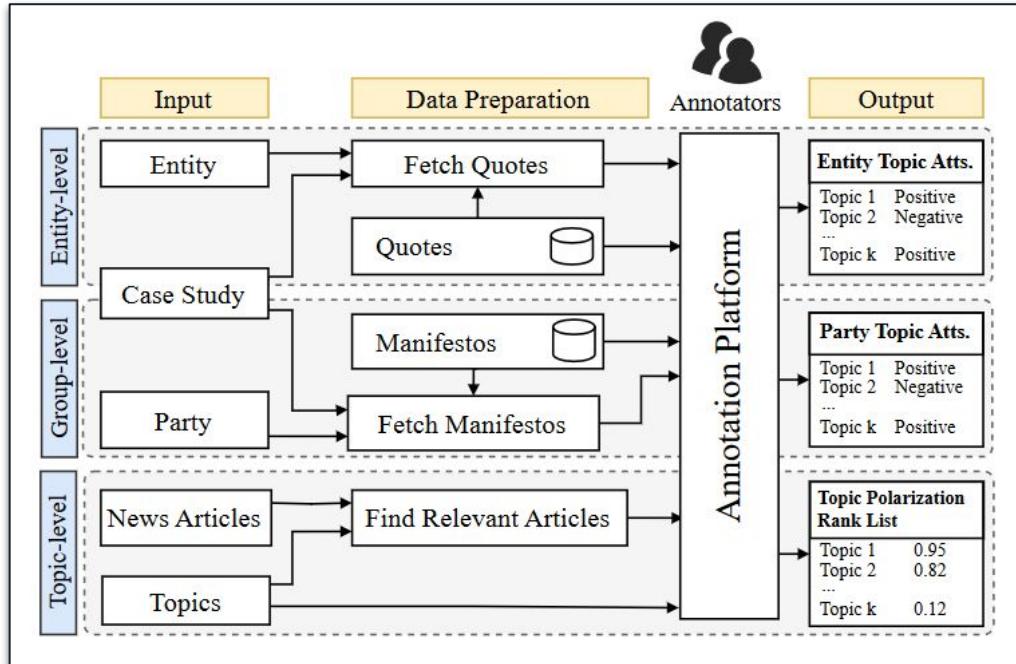
Q1 What is the effectiveness of our framework in capturing entity attitudes towards various discussion topics?

Q2 What is the extent of alignment between politically cohesive fellowships identified and their official party manifestos?

Q3 What is the accuracy in extracting the discussion topics of a domain and how effective is it in capturing the per-topic polarization degree?

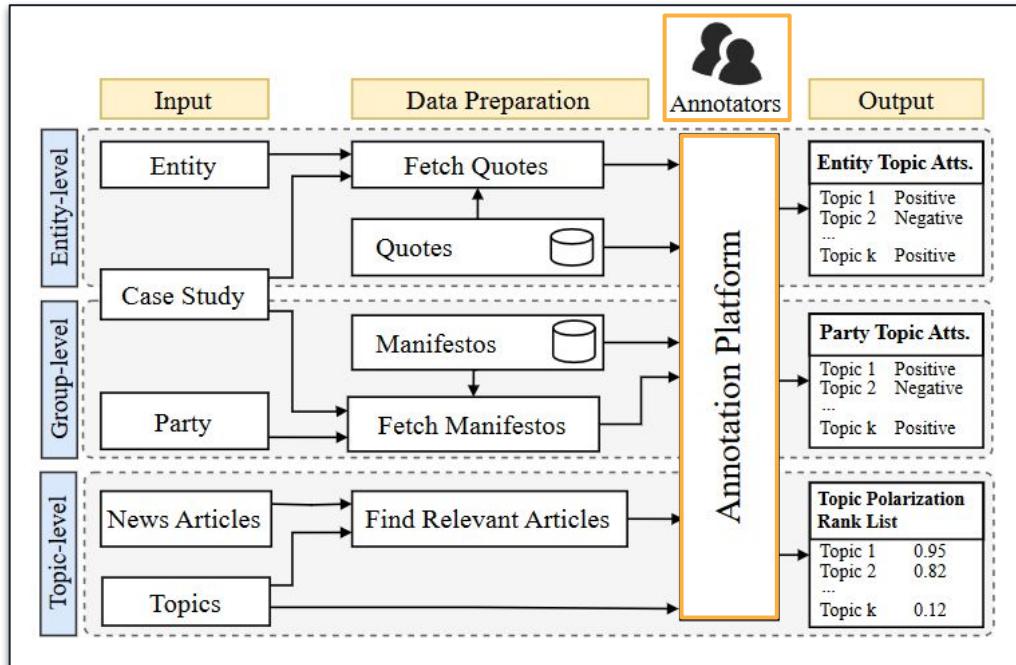
Polarization Knowledge Annotation

- We define a polarization knowledge annotation methodology for the construction of the ground-truth datasets for each polarization level.



Polarization Knowledge Annotation

- We define a polarization knowledge annotation methodology for the construction of the ground-truth datasets for each polarization level.



3 Annotators with CS background: 1 experienced annotator and 2 MSc students.

For the annotation, we have developed an annotation platform.

Evaluation Case Studies

❑ Abortion:

- Articles: 3,437 **Left** vs. 3,039 **Right**
- 80% Published: ≥ 2010
- Topics: **20**



❑ Immigration:

- Articles: 3,496 **Left** vs. 5,020 **Right**
- 80% Published: ≥ 2016
- Topics: **22**



❑ Gun Control:

- Articles: 3,198 **Left** vs. 3,455 **Right**
- 80% Published: ≥ 2011
- Topics: **19**



Entity-level Polarization Knowledge Evaluation

- ❑ Evaluate the performance of **POLAR** in identifying the **Entity → Topic** attitudes by comparing its output with a **Ground Truth (GT)** dataset.
- ❑ **Evaluation Scope:** Consider a subset of the available entities → **Political figures and political organizations.**
 - Leverage the abundance of external data available w.r.t. the political discourse.



Topic	POLAR	GT
Pro-Choice	Supports	Supports
Anti-abortion	Opposes	Opposes

Entity → Topic Attitudes Ground Truth

- **OnTheIssues:** A collection of quotes and votes of political figures from newspapers, speeches, and press releases towards various issues.



★Joe Biden on Abortion★	
<p>Favors topic 1: Abortion is a woman's unrestricted right (+2 points on Social scale)</p>	<p>No public funding for abortion; it imposes a view: Opposes topic 1 Supports partial-birth abortion ban, but not undoing Roe: Opposes topic 1 Accepts Catholic church view that life begins at conception: Opposes topic 1 I accept church rule personally, but not in public life: Favors topic 1 Fact-Check: opposes Hyde Amendment, after decades of support: Strongly Favors topic 1 Unequivocal support for abortion rights; Congress must act: Strongly Favors topic 1 Roe v. Wade is under attack as never before: Strongly Favors topic 1 If Congress passes a national abortion ban, I will veto it: Strongly Favors topic 1 Rated 36% by NARAL, indicating a mixed voting record on abortion: Neutral on topic 1 Expand embryonic stem cell research: Favors topic 1 Rated 0% by the NRCL, indicating a pro-choice stance: Strongly Favors topic 1 Protect the reproductive rights of women: Strongly Favors topic 1 Ensure access to and funding for contraception: Favors topic 1</p>

1. Retrieve quotes of political entities on the topics w.r.t. abortion, immigration, and gun control.
2. Given the topics per use case, annotate the quotes as either supportive or oppositional.

Entity → Topic Attitudes Annotation Process



Joe Biden

Entity → Topic Attitudes Annotation Process



Joe Biden

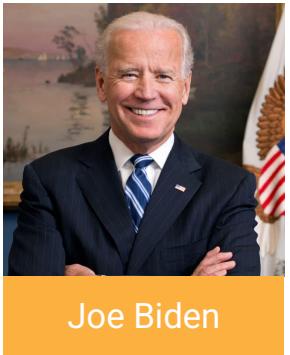


Quotes

“Roe v. Wade is under attack as never before.” (Mar. 2022)

“Expand embryonic stem cell research.” (Jun. 2014)

Entity → Topic Attitudes Annotation Process



Quotes

“Roe v. Wade is under attack as never before.” (Mar. 2022)

“Expand embryonic stem cell research.” (Jun. 2014)

A yellow rectangular box containing two quotes attributed to Joe Biden. The first quote is from March 2022 and the second from June 2014.

Roe v. Wade is under attack as never before.

Positive Stance

Quick Filter

abortion funding²¹ abortion provider economy²¹
 planned parenthood²¹ pregnancy centers²¹ anti-abortion²¹
 birth control²¹ health care²¹ hobby lobby²¹ roe v. wade²¹
 late term abortion²¹ life protection²¹ pro-choice²¹
 reproduction right²¹ women freedom²¹ pro-life²¹
 right of human life²¹ sanctity of life²¹ sale of fetal tissue²¹
 stem cell research²¹ sexual assault victims²¹

Negative Stance

Quick Filter

abortion funding²¹ abortion provider economy²¹
 planned parenthood²¹ pregnancy centers²¹ anti-abortion²¹
 birth control²¹ health care²¹ hobby lobby²¹ roe v. wade²¹
 late term abortion²¹ life protection²¹ pro-choice
 reproduction right²¹ women freedom²¹ pro-life
 right of human life²¹ sanctity of life²¹ sale of fetal tissue²¹
 stem cell research²¹ sexual assault victims²¹

An annotation platform interface showing two lists of entities related to the quote "Roe v. Wade is under attack as never before." The top section, "Positive Stance," includes terms like "roe v. wade" and "pro-choice." The bottom section, "Negative Stance," includes terms like "anti-abortion" and "pro-life". A large orange arrow points down from the quotes to the interface.

Entity → Topic Attitudes Annotation Process



Quotes

“Roe v. Wade is under attack as never before.” (Mar. 2022)

“Expand embryonic stem cell research.” (Jun. 2014)



Roe v. Wade is under attack as never before.

Positive Stance

Quick Filter

abortion funding²¹ abortion provider economy²¹
 planned parenthood²¹ pregnancy centers²¹ anti-abortion²¹
 birth control²¹ health care²¹ hobby lobby²¹ roe v. wade²¹
 late term abortion²¹ life protection²¹ pro-choice²¹
 reproduction right²¹ women freedom²¹ pro-life²¹
 right of human life²¹ sanctity of life²¹ sale of fetal tissue²¹
 stem cell research²¹ sexual assault victims²¹

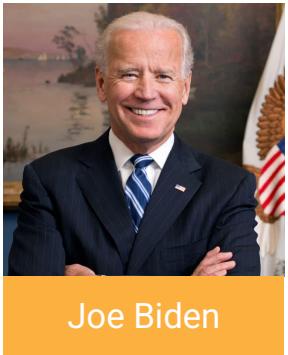
Negative Stance

Quick Filter

abortion funding²¹ abortion provider economy²¹
 planned parenthood²¹ pregnancy centers²¹ anti-abortion²¹
 birth control²¹ health care²¹ hobby lobby²¹ roe v. wade²¹
 late term abortion²¹ life protection²¹ pro-choice
 reproduction right²¹ women freedom²¹ pro-life
 right of human life²¹ sanctity of life²¹ sale of fetal tissue²¹
 stem cell research²¹ sexual assault victims²¹

Annotation Platform

Entity → Topic Attitudes Annotation Process



Quotes

“Roe v. Wade is under attack as never before.” (Mar. 2022)

“Expand embryonic stem cell research.” (Jun. 2004)



Roe v. Wade is under attack as never before.

Positive Stance

Quick Filter

abortion funding^[2] abortion provider economy^[2]
 planned parenthood^[2] pregnancy centers^[2] anti-abortion^[2]
 birth control^[2] health care^[2] hobby lobby^[2] roe v. wade^[2]
 late term abortion^[2] life protection^[2] pro-choice^[2]
 reproduction right^[2] women freedom^[2] pro-life^[2]
 right of human life^[2] sanctity of life^[2] sale of fetal tissue^[2]
 stem cell research^[2] sexual assault victims^[2]

Negative Stance

Quick Filter

abortion funding^[2] abortion provider economy^[2]
 planned parenthood^[2] pregnancy centers^[2] anti-abortion^[2]
 birth control^[2] health care^[2] hobby lobby^[2] roe v. wade^[2]
 late term abortion^[2] life protection^[2] pro-choice^[2]
 reproduction right^[2] women freedom^[2] pro-life^[2]
 right of human life^[2] sanctity of life^[2] sale of fetal tissue^[2]
 stem cell research^[2] sexual assault victims^[2]



Topic Attitudes GT

t_1 Supports
 t_2 Opposes
 t_3 Supports
 t_4 Supports
...
 t_k Opposes

Entity-level PK Evaluation Process



Joe Biden



Topic Attitudes POLAR	
t_1	Supports
t_2	Opposes
t_3	Supports
t_4	Supports
\dots	
t_k	Opposes



Topic Attitudes GT	
t_1	Supports
t_2	Opposes
t_3	Supports
t_4	Supports
\dots	
t_k	Opposes



Evaluation Metric:

$$\text{F1 Score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

Group-level Polarization Knowledge Evaluation

- ❑ Political parties often adopt official platforms that outline their positions on a range of issues.
 - The official platforms of parties are outlined in their party manifestos.

Group-level Polarization Knowledge Evaluation

- ❑ Political parties often adopt official platforms that outline their positions on a range of issues.
 - The official platforms of parties are outlined in their party manifestos.
- ❑ Evaluate the alignment of politically affiliated fellowship attitudes with their respective party manifestos.
- ❑ Evaluation Scope: Focus on ideologically cohesive fellowships i.e. the majority of entity-members belong to either **Left** or **Right** of political spectrum.



Topic	POLAR	Democratic Manifesto
Pro-Choice	Supports	Supports
Anti-abortion	Opposes	Opposes

Entity Political Affiliation Estimation

- ❑ Leverage the linkage of entities with **Wikipedia** pages.
 - Fields of: i) **Political Party** e.g. Democratic Party; ii) **Political Position** e.g. Center-Left, and iii) **Ideology** e.g. Modern Liberalism.
 - Not strict schema → Fields such as **Political Party**, not always included.



Emmanuel Macron

Macron in 2021

President of France	
Incumbent	
Personal details	
Born	Emmanuel Jean-Michel Frédéric Macron 21 December 1977 (age 44) Amiens, France
Political party	La République En Marche! (since 2016)

LA RÉPUBLIQUE
En Marche !

Abbreviation LREM
LaREM
REM

Executive Officer Stanislas Guérini

President in the National Assembly Christophe Castaner

President in the Senate François Patriat

Founder Emmanuel Macron

Founded 6 April 2016; 6 years ago

Headquarters 63, rue Sainte-Anne 75002 Paris

Youth wing Les Jeunes avec Macron

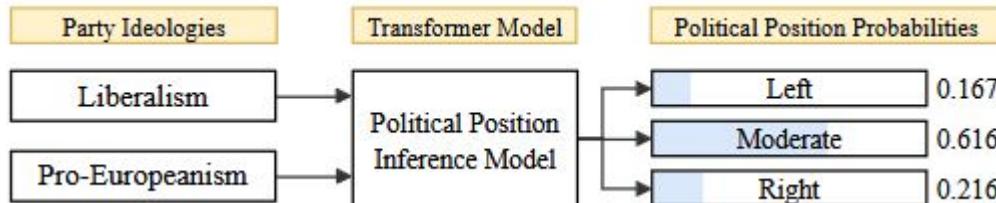
Membership (2021) 422,329 claimed adherents^[1]
[non-primary source needed]

Ideology Liberalism
Pro-Europeanism

Political position Centre

Entity Political Affiliation Estimation

- ❑ Leverage the linkage of entities with **Wikipedia** pages.
 - Fields of: i) **Political Party** e.g. Democratic Party; ii) **Political Position** e.g. Center-Left, and iii) **Ideology** e.g. Modern Liberalism.
 - Not strict schema → Fields such as **Political Party**, not always included.
- Implementation of **entity ideological mapping** as a classification model based on BERT.



Fellowship Cohesiveness

- ❑ Fellowships are characterized by their overall support, common views and ideologies.
 - **Ideological cohesiveness:** Agreement of entity members regarding their ideological positions on the political spectrum.
 - **Attitudinal cohesiveness:** Agreement of entity members regarding their attitudes toward different issues e.g. Gun Control (GC).

$$S_{LR} = \langle Left, Moderate, Right \rangle$$

$$S_{AT} = \langle Support_{GC}, Opposition_{GC} \rangle$$

Fellowship Cohesiveness

- ❑ **Fellowships** are characterized by their overall support, common views and ideologies.
 - **Ideological cohesiveness:** Agreement of entity members regarding their ideological positions on the political spectrum.
 - **Attitudinal cohesiveness:** Agreement of entity members regarding their attitudes toward different issues e.g. Gun Control (GC).

$$S_{LR} = \langle Left, Moderate, Right \rangle$$

$$S_{AT} = \langle Support_{GC}, Opposition_{GC} \rangle$$

$$coh_S(P_{F_i}^S) = \frac{1}{N} \times \sum_{p_k \in P_{F_i}^S} d(p_k, z)$$

Fellowship Cohesiveness

- ❑ **Fellowships** are characterized by their overall support, common views and ideologies.
 - **Ideological cohesiveness:** Agreement of entity members regarding their ideological positions on the political spectrum.
 - **Attitudinal cohesiveness:** Agreement of entity members regarding their attitudes toward different issues e.g. Gun Control (GC).

$$S_{LR} = \langle Left, Moderate, Right \rangle$$

$$S_{AT} = \langle Support_{GC}, Opposition_{GC} \rangle$$

Mapping of entities of F_i to positions on spectrum S .

$$coh_S(P_{F_i}^S) = \frac{1}{N} \times \sum_{p_k \in P_{F_i}^S} d(p_k, z)$$

Fellowship Cohesiveness

- ❑ Fellowships are characterized by their overall support, common views and ideologies.
 - **Ideological cohesiveness:** Agreement of entity members regarding their ideological positions on the political spectrum.
 - **Attitudinal cohesiveness:** Agreement of entity members regarding their attitudes toward different issues e.g. Gun Control (GC).

$$S_{LR} = \langle Left, Moderate, Right \rangle$$

$$S_{AT} = \langle Support_{GC}, Opposition_{GC} \rangle$$

Mapping of entities of F_i to positions on spectrum S .

N is the number of entities in F_i .

$$coh_S(P_{F_i}^S) = \frac{1}{N} \times \sum_{p_k \in P_{F_i}^S} d(p_k, z)$$

Fellowship Cohesiveness

- ❑ Fellowships are characterized by their overall support, common views and ideologies.
 - **Ideological cohesiveness:** Agreement of entity members regarding their ideological positions on the political spectrum.
 - **Attitudinal cohesiveness:** Agreement of entity members regarding their attitudes toward different issues e.g. Gun Control (GC).

$$S_{LR} = \langle Left, Moderate, Right \rangle$$

$$S_{AT} = \langle Support_{GC}, Opposition_{GC} \rangle$$

Mapping of entities of F_i to positions on spectrum S .

N is the number of entities in F_i .

$$coh_S(P_{F_i}^S) = \frac{1}{N} \times \sum_{p_k \in P_{F_i}^S} d(p_k, z)$$

Function **d** calculates how close two positions are on the spectrum.

$$d(s_i, s_j) = 1.0 - q/n$$

Fellowship Cohesiveness

- ❑ Fellowships are characterized by their overall support, common views and ideologies.
 - **Ideological cohesiveness:** Agreement of entity members regarding their ideological positions on the political spectrum.
 - **Attitudinal cohesiveness:** Agreement of entity members regarding their attitudes toward different issues e.g. Gun Control (GC).

$$S_{LR} = \langle Left, Moderate, Right \rangle$$

$$S_{AT} = \langle Support_{GC}, Opposition_{GC} \rangle$$

Mapping of entities of F_i to positions on spectrum S .

N is the number of entities in F_i .

$$coh_S(P_{F_i}^S) = \frac{1}{N} \times \sum_{p_k \in P_{F_i}^S} d(p_k, z)$$

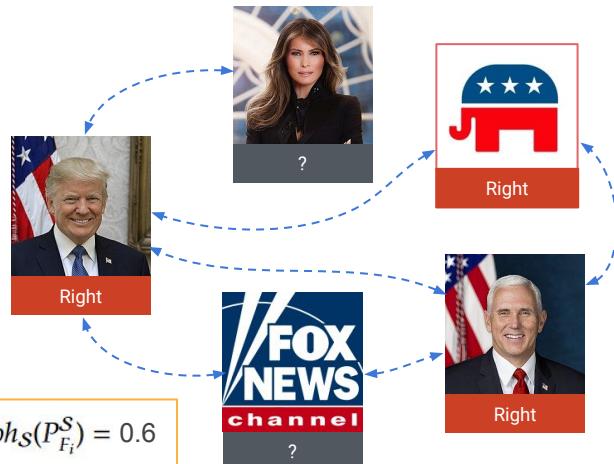
z is the position of the majority of entities in F_i .

Function **d** calculates how close two positions are on the spectrum.

$$d(s_i, s_j) = 1.0 - q/n$$

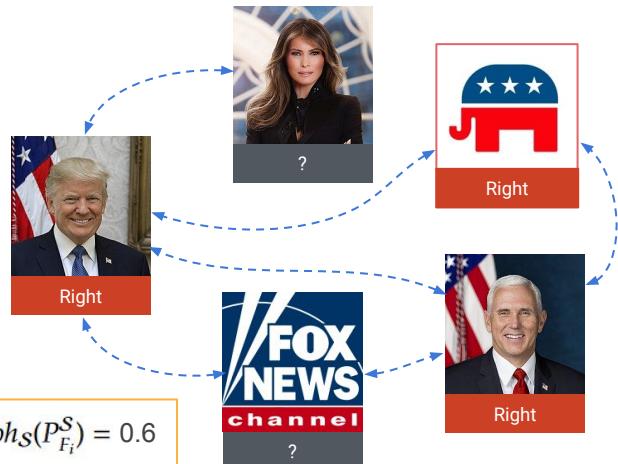
Entity Political Label Propagation

- ❑ Not all entities are explicitly affiliated with politics.



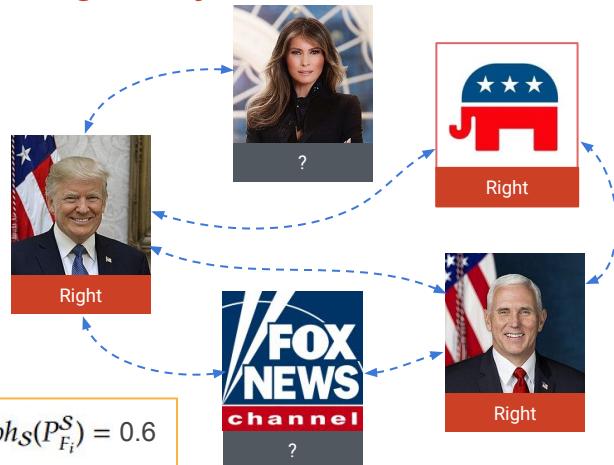
Entity Political Label Propagation

- ❑ Not all entities are explicitly affiliated with politics.
 - Identify the political affiliation of an unlabeled entity → **Label Propagation Algorithms (LPA)**.



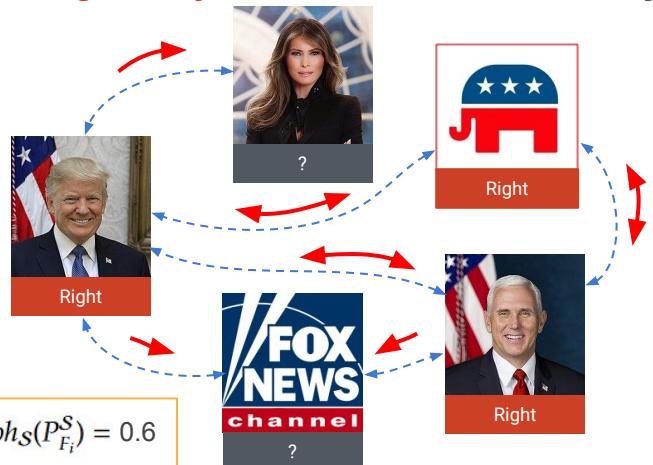
Entity Political Label Propagation

- ❑ Not all entities are explicitly affiliated with politics.
 - Identify the political affiliation of an unlabeled entity → **Label Propagation Algorithms (LPA)**.
- ❑ Signed LPA takes advantages of signed attributes in SAG
 - If entities are **positively** connected, then **affected**.
 - If **negatively** connected, then label **rejected**.



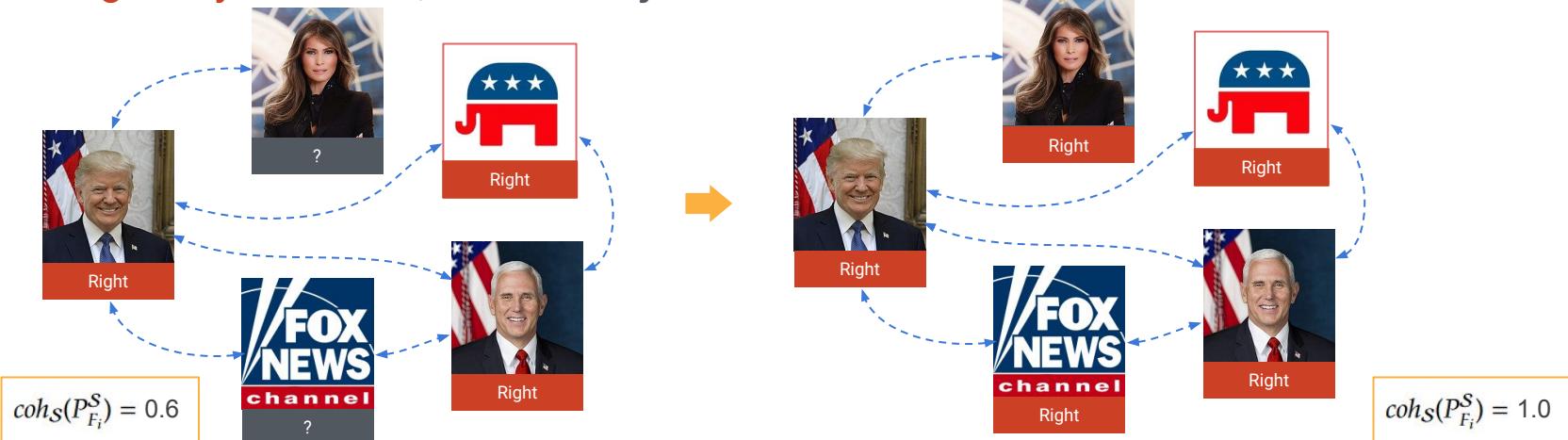
Entity Political Label Propagation

- ❑ Not all entities are explicitly affiliated with politics.
 - Identify the political affiliation of an unlabeled entity → **Label Propagation Algorithms (LPA)**.
- ❑ Signed LPA takes advantages of signed attributes in SAG
 - If entities are **positively** connected, then **affected**.
 - If **negatively** connected, then label **rejected**.



Entity Political Label Propagation

- ❑ Not all entities are explicitly affiliated with politics.
 - Identify the political affiliation of an unlabeled entity → **Label Propagation Algorithms (LPA)**.
- ❑ Signed LPA takes advantages of signed attributes in SAG
 - If entities are **positively** connected, then **affected**.
 - If **negatively** connected, then label **rejected**.



Fellowship → Topic Attitudes Ground Truth

- ❑ Focus on **Left** and **Right** aligned fellowships, with a cohesiveness score ≥ 0.75 .
- ❑ Ideological cohesive fellowships → Alignment of attitudes with their respective parties.
- ❑ **Manifesto Project:** Analyses of parties' manifestos for their policy preferences.

The screenshot shows the homepage of the Manifesto Project. The top navigation bar includes links for INFO, DATA, CORPUS & DOCUMENTS, TUTORIALS, manifestoR API, and PUBLICATIONS. Below the navigation, there are three main sections: 'Dataset' (with links for Download, Codebook, Browse, and Visualize Party Positions), 'Corpus' (with links for Information, Download Manifestos, Browse By Keyword, and Browse By Document), and 'Tools' (with links for manifestoR API, Read document, Read text, Browse by document, and Frequency). At the bottom, there are two rows of party logos and a line graph showing topic scores over time for various political parties from 1990 to 2005.

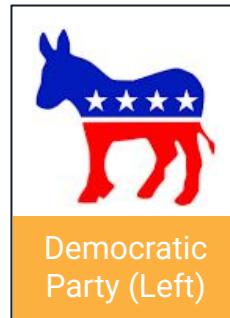


"The President and the Democratic Party believe that women have a right to control their reproductive choices."

Fellowship → Topic Attitudes Ground Truth

- ❑ Focus on **Left** and **Right** aligned fellowships, with a cohesiveness score ≥ 0.75 .
- ❑ Ideological cohesive fellowships → Alignment of attitudes with their respective parties.
- ❑ **Manifesto Project:** Analyses of parties' manifestos for their policy preferences.

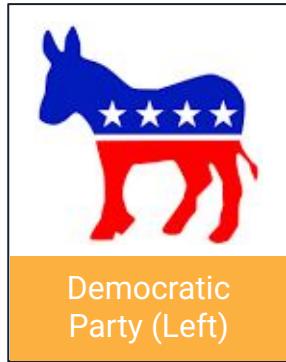
The screenshot shows the Manifesto Project website interface. At the top, there are links for INFO, DATA, CORPUS & DOCUMENTS, TUTORIALS, manifestoR API, and PUBLICATIONS. Below these are three main sections: 'Dataset' (with links for Download, Codebook, Browse, and Visualize Party Positions), 'Corpus' (with links for Information, Download Manifestos, Browse By Keyword, and Browse By Document), and 'Tools' (with links for manifestoR API, Read document, Read text, Browse by document, and Frequency). On the left, there's a grid of party logos from different countries. In the center, there's a line graph showing political party support over time (1990-2005) for various parties like CDU/CSU, SPD, FDP, Grüne, Bündnis 90/Die Grünen, and Die Linke. On the right, there's a section titled 'publications' with a list of documents.



"The President and the Democratic Party believe that women have a right to control their reproductive choices."

- ➡ Supports Pro-Choice
- ➡ Supports Reproductive Rights
- ➡ Opposes Anti-abortion

Fellowship → Topic Attitudes Annotation Process



Fellowship → Topic Attitudes Annotation Process

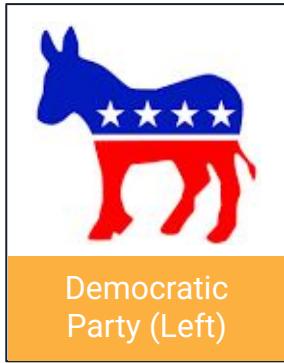


Party Manifesto

Securing Reproductive Health, Rights, and Justice

Democrats are committed to protecting and advancing reproductive health, rights, and justice. We believe unequivocally, like the majority of Americans, that every woman should have access to quality reproductive health care services, including safe and legal abortion—regardless of where she lives, how much money she makes, or how she is insured. We believe that reproductive health is core to women's, men's, and young people's health and wellbeing. We will continue to stand up to Republican efforts to defund Planned Parenthood health centers, which provide critical health services to millions of people. We will continue to oppose—and

Fellowship → Topic Attitudes Annotation Process



Party Manifesto

Securing Reproductive Health, Rights, and Justice

Democrats are committed to protecting and advancing reproductive health, rights, and justice. We believe unequivocally, like the majority of Americans, that every woman should have access to quality reproductive health care services, including safe and legal abortion—regardless of where she lives, how much money she makes, or how she is insured. We believe that reproductive health is core to women's, men's, and young people's health and wellbeing. We will continue to stand up to Republican efforts to defund Planned Parenthood health centers, which provide critical health services to millions of people. We will continue to oppose—and



Roe v. Wade is under attack as never before.

Positive Stance

Quick Filter

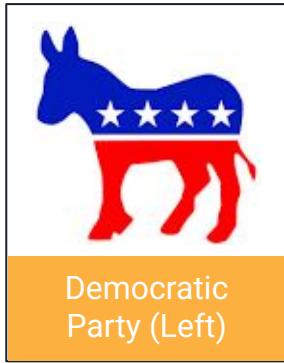
- abortion funding²¹ abortion provider economy²¹
- planned parenthood²¹ pregnancy centers²¹ anti-abortion²¹
- birth control²¹ health care²¹ hobby lobby²¹ roe v. wade²¹
- late term abortion²¹ life protection²¹ pro-choice²¹
- reproduction right²¹ women freedom²¹ pro-life²¹
- right of human life²¹ sanctity of life²¹ sale of fetal tissue²¹
- stem cell research²¹ sexual assault victims²¹

Negative Stance

Quick Filter

- abortion funding²¹ abortion provider economy²¹
- planned parenthood²¹ pregnancy centers²¹ anti-abortion²¹
- birth control²¹ health care²¹ hobby lobby²¹ roe v. wade²¹
- late term abortion²¹ life protection²¹ pro-choice
- reproduction right²¹ women freedom²¹ pro-life
- right of human life²¹ sanctity of life²¹ sale of fetal tissue²¹
- stem cell research²¹ sexual assault victims²¹

Fellowship → Topic Attitudes Annotation Process



Party Manifesto

Securing Reproductive Health, Rights, and Justice

Democrats are committed to protecting and advancing reproductive health, rights, and justice. We believe unequivocally, like the majority of Americans, that every woman should have access to quality reproductive health care services, including safe and legal abortion—regardless of where she lives, how much money she makes, or how she is insured. We believe that reproductive health is core to women's, men's, and young people's health and wellbeing. We will continue to stand up to Republican efforts to defund Planned Parenthood health centers, which provide critical health services to millions of people. We will continue to oppose—and



Roe v. Wade is under attack as never before.

Positive Stance

Annotation Platform

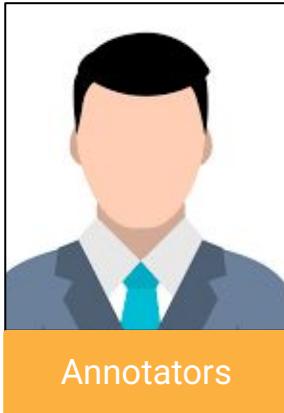
Quick Filter

- abortion funding²¹ abortion provider economy²¹
- planned parenthood²¹ pregnancy centers²¹ anti-abortion²¹
- birth control²¹ health care²¹ hobby lobby²¹ roe v. wade²¹
- late term abortion²¹ life protection²¹ pro-choice²¹
- reproduction right²¹ women freedom²¹ pro-life²¹
- right of human life²¹ sanctity of life²¹ sale of fetal tissue²¹
- stem cell research²¹ sexual assault victims²¹

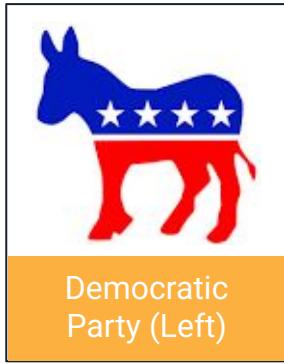
Negative Stance

Quick Filter

- abortion funding²¹ abortion provider economy²¹
- planned parenthood²¹ pregnancy centers²¹ anti-abortion²¹
- birth control²¹ health care²¹ hobby lobby²¹ roe v. wade²¹
- late term abortion²¹ life protection²¹ pro-choice
- reproduction right²¹ women freedom²¹ pro-life
- right of human life²¹ sanctity of life²¹ sale of fetal tissue²¹
- stem cell research²¹ sexual assault victims²¹



Fellowship → Topic Attitudes Annotation Process



Party Manifesto

Securing Reproductive Health, Rights, and Justice

Democrats are committed to protecting and advancing reproductive health, rights, and justice. We believe unequivocally, like the majority of Americans, that every woman should have access to quality reproductive health care services, including safe and legal abortion—regardless of where she lives, how much money she makes, or how she is insured. We believe that reproductive health is core to women's, men's, and young people's health and wellbeing. We will continue to stand up to Republican efforts to defund Planned Parenthood health centers, which provide critical health services to millions of people. We will continue to oppose—and



Roe v. Wade is under attack as never before.

Positive Stance

Quick Filter

abortion funding²¹ abortion provider economy²¹
 planned parenthood²¹ pregnancy centers²¹ anti-abortion²¹
 birth control²¹ health care²¹ hobby lobby²¹ roe v. wade²¹
 late term abortion²¹ life protection²¹ pro-choice²¹
 reproduction right²¹ women freedom²¹ pro-life²¹
 right of human life²¹ sanctity of life²¹ sale of fetal tissue²¹
 stem cell research²¹ sexual assault victims²¹

Negative Stance

Quick Filter

abortion funding²¹ abortion provider economy²¹
 planned parenthood²¹ pregnancy centers²¹ anti-abortion²¹
 birth control²¹ health care²¹ hobby lobby²¹ roe v. wade²¹
 late term abortion²¹ life protection²¹ pro-choice²¹
 reproduction right²¹ women freedom²¹ pro-life²¹
 right of human life²¹ sanctity of life²¹ sale of fetal tissue²¹
 stem cell research²¹ sexual assault victims²¹

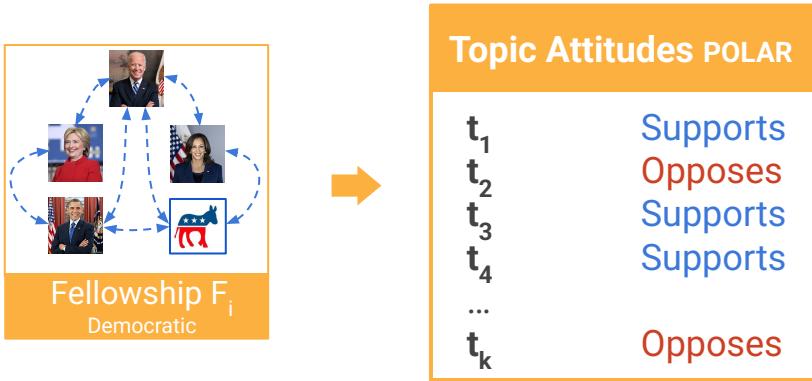
Annotation Platform



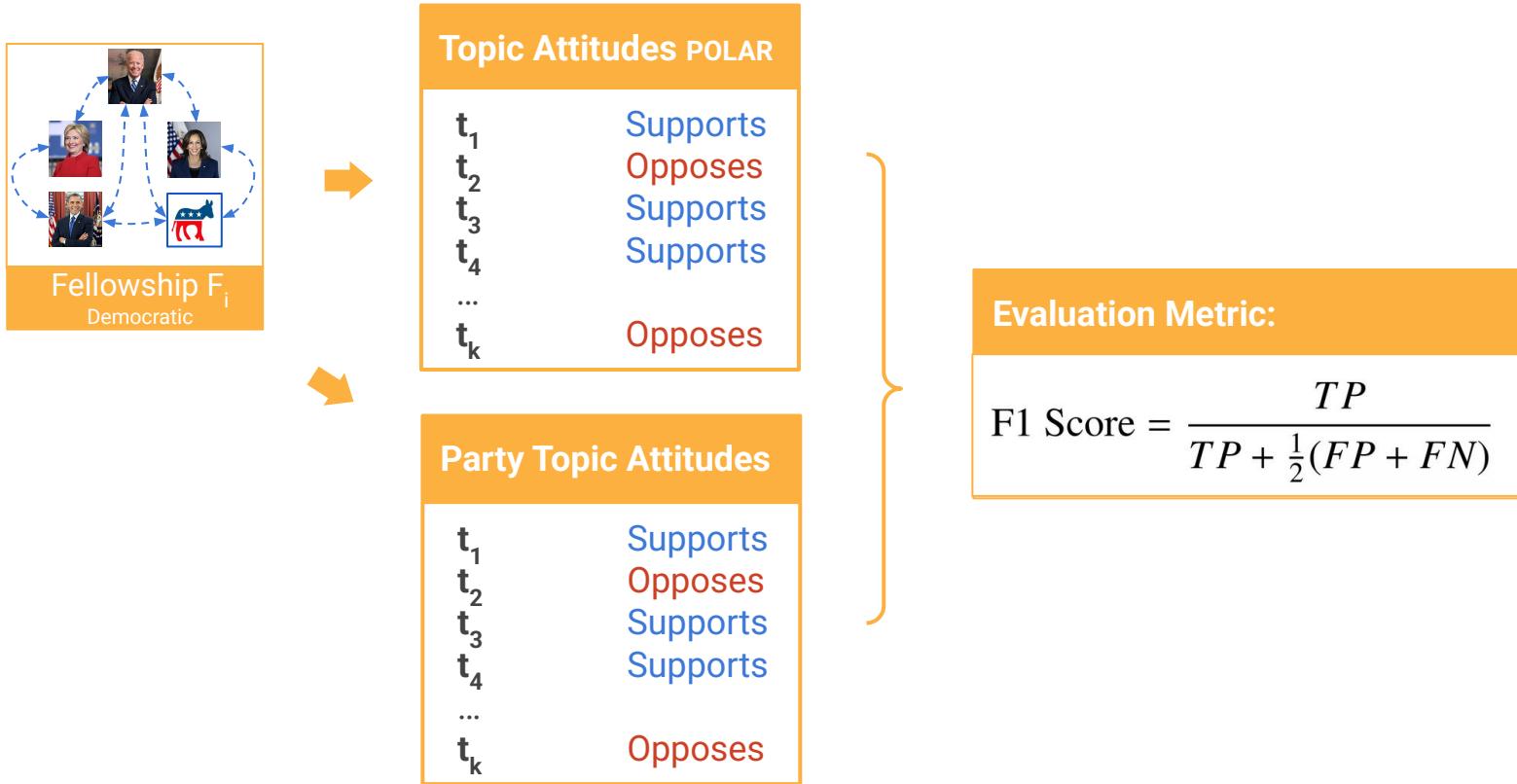
Party Topic Attitudes

t_1	Supports
t_2	Opposes
t_3	Supports
t_4	Supports
...	
t_k	Opposes

Fellowship-level PK Evaluation Process



Fellowship-level PK Evaluation Process



Topic-level Polarization Knowledge Evaluation

- ❑ Assess the ability of **POLAR** to quantify **topic polarization** by comparing with a **GT**.
 - **Polarization Ranked List:** Topics ranked based on their polarization levels.
 - Measure the **Ranking Correlation Agreement** between the ranked list and GT.
- ❑ **Evaluation Scope:** Focus on **Left** or **Right** ideologically cohesive fellowships.

Topic-level Polarization Knowledge Evaluation

- ❑ Assess the ability of **POLAR** to quantify **topic polarization** by comparing with a **GT**.
 - **Polarization Ranked List:** Topics ranked based on their polarization levels.
 - Measure the **Ranking Correlation Agreement** between the ranked list and GT.
- ❑ **Evaluation Scope:** Focus on **Left** or **Right** ideologically cohesive fellowships.
- ❑ **Ground Truth (GT):** Annotate topic polarization in news articles from **Left** or **Right** sources (annotation process adapter from He et al. 2021).
 1. Identify conflicting stances (e.g. Pro-choice vs. Pro-life).
 2. Annotators classify stances in ≈ 20 articles per topic as **Supportive (1)** or **Oppositional (0)**.
 3. Compute polarization (α) between Left (D_L) and Right (D_R) corpora:
 - **Higher $\alpha \rightarrow$ Stronger polarization.**

Topic Polarization Annotation Process

- Annotators identify topic pairs with conflicting stances.

Abortion		Immigration		Gun Control	
Anti-Abortion	Pro-Choice	Minimum Wage	Salary Stagnation	Ban on Handgun	2nd Amendment
Pro-Life	Pro-Choice	Wealth Gap	Minimum Wage	Ban on Handgun	CCRA
Life Protection	Planned Parenthood	Cheap Labor Availability	Minimum Wage	GC to Restrain Violence	2nd Amendment
Life Protection	Pregnancy Centers	Cheap Labor Availability	Wealth Gap	GC to Restrain Violence	CCRA
Sanctity of Life	Women Freedom	Amnesty	Dream Act	White Identity	Person of Color Identity
Late Term Abortion	Roe V. Wade	Amnesty	DACA	Right to Self-Defense	Stop Gun Crime
Right of Human Life	Reproduction Right	Family Separation Policy	DACA		
Ab. Provider Economy	Abortion Funding	Racial Identity	Born Identity		
		Racial Identity	Racism & Xenophobia		
		Born Identity	Racism & Xenophobia		

Topic Polarization Annotation Process

- ❑ Annotators identify topic pairs with conflicting stances.
- ❑ Following the annotation process, the resulting GT are:

Abortion Stances		$\alpha(D^L, D^R, t)$
Right of Human Life	Reproduction Right	0.225
Life Protection	Pl. Parenthood	0.075
Pro-Life	Pro-Choice	0.060
Abortion Pr. Ec.	Abortion Funding	0.035
Life Protection	Pregnancy Centers	0.020
Late Term Abortion	Roe V. Wade	0.010
Anti-Abortion	Pro-Choice	0.010
Sanctity of Life	Women Freedom	0.005

Immigration Stances		$\alpha(D^L, D^R, t)$
Racial Identity	Born Identity	0.5417
Born Identity	Racism & Xen.	0.4917
Amnesty	DACA	0.2917
Wealth Gap	Minimum Wage	0.1583
Racial Identity	Racism & Xen.	0.1417
Minimum Wage	Salary Stag.	0.1333
Cheap Labor Avail.	Wealth Gap	0.0750
Cheap Labor Avail.	Minimum Wage	0.0583
Family Sep. Policy	DACA	0.0250

Gun Control Stances		$\alpha(D^L, D^R, t)$
Right to Self-Defense	Stop Gun Crime	1.000
GC to Restrain Violence	CCRA	0.917
GC to Restrain Violence	2nd Amendment	0.617
Ban on Handgun	CCRA	0.300
White Identity	Person of Color Id.	0.117
Ban on Handgun	2nd Amendment	0.017

Topic-level PK Evaluation Process

❑ Polarizing Topic Ranking Agreement

- Compare polarizing topic ranked list from POLAR with the GT.

Topic Polarization POLAR	
Topic	$score_{t_k}$
t_1	106.8
t_2	92.2
t_3	42.6
...	
t_k	21.2

Aggregate dipole polarization indices for topic t_k :

$$score_{t_k} = (obs_{t_k}/\mathbf{d}_{t_k}) \cdot \tilde{\mu}_{t_k}$$

obs_{t_k} Size of t_k attitude observations

\mathbf{d}_{t_k} Number of dipoles discussing t_k

$\tilde{\mu}_{t_k}$ Median of t_k polarization indices.

Topic-level PK Evaluation Process

❑ Polarizing Topic Ranking Agreement

- Compare polarizing topic ranked list from POLAR with the GT.
- **Ranked Bias Overlap (RBO)**: An intersection-based ranking agreement measure
→ Small consecutive differences between the ranked topics.



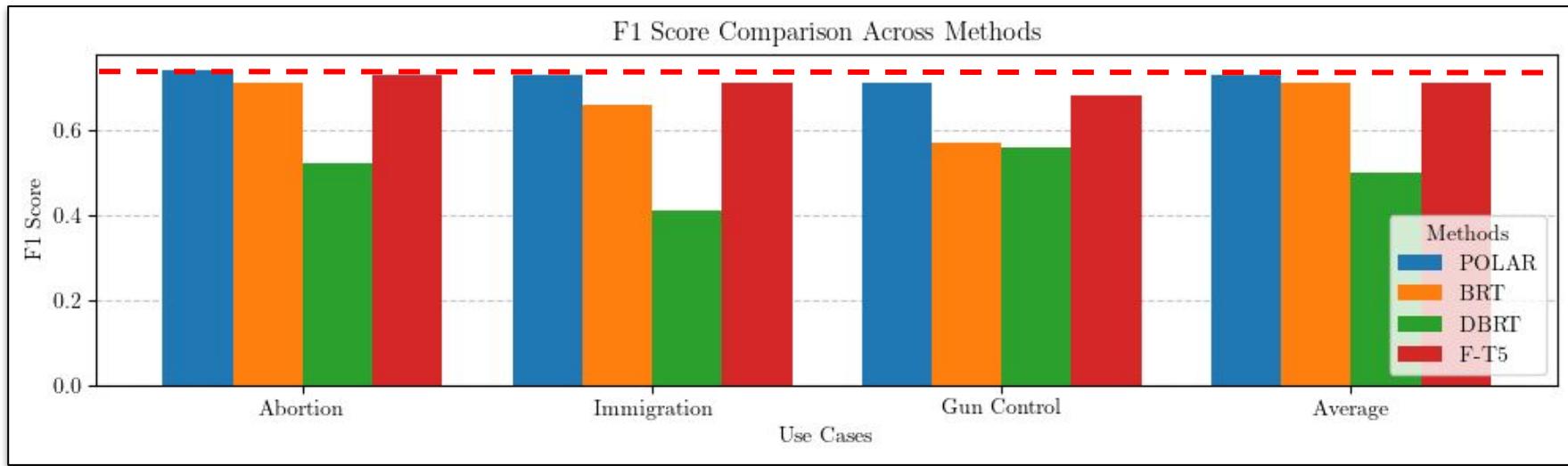
Polarization Knowledge Evaluation Experiments

POLAR Application on Case Studies

- We applied POLAR on the news corpus of each case study:

	Abortion	Immigration	Gun Control
Entities:	8,113	18,409	15,217
SAG Nodes:	228	459	194
SAG Edges:	523	1,440	478
Fellowships	49	156	69
Dipoles	16	34	42
Noun Phrases:	107,521	298,918	201,419
Topics:	533	2,517	1,262

Entity → Topic Estimation Results



- **Demonstration of competitive performance** → Comparable to or better than state-of-the-art zero-shot baselines for entity-level attitude estimation.
- Trade-off between resource constraints and desired outcomes → Proposed framework makes a practical solution for environments with limited resources.

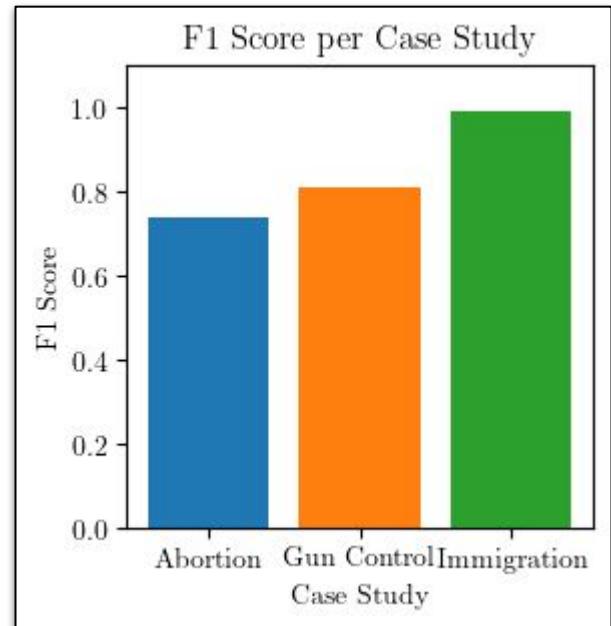
M. Lewis, Y. Liu, N. Goyal, et al., “Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension,” arXiv preprint arXiv:2210.11416, 2019.

P. He, X. Liu, J. Gao, and W. Chen, “Deberta: Decoding-enhanced BERT with disentangled attention,” CoRR 2020.

H. W. Chung, L. Hou, et al. “Scaling instruction-finetuned language models,” arXiv preprint arXiv:2210.11416, 2022.

Fellowship → Topic Manifesto Alignment

- ❑ Average performance of 0.87 across all evaluation metrics.
 - Ability to accurately predict the attitude of a fellowship towards a topic.
 - Extracted fellowships have a high degree of alignment between their attitudes and those of their manifestos.



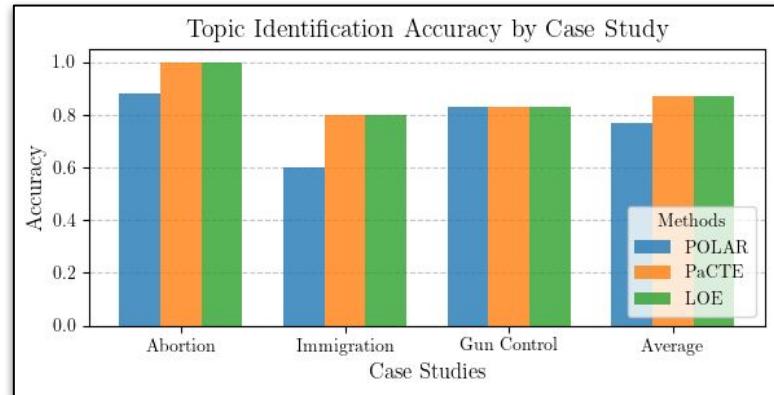
Topic Polarization Ranking Results

❑ Topic Identification Accuracy (TIA):

Initially, evaluate the ability of each approach in identifying the annotated topics.

- **POLAR Average TIA: 0.77**

- Due to the **dipole-specific NP scope** for identifying discussion topics.



Topic Polarization Ranking Results

❑ Topic Identification Accuracy (TIA):

Initially, evaluate the ability of each approach in identifying the annotated topics.

- **POLAR Average TIA: 0.77**

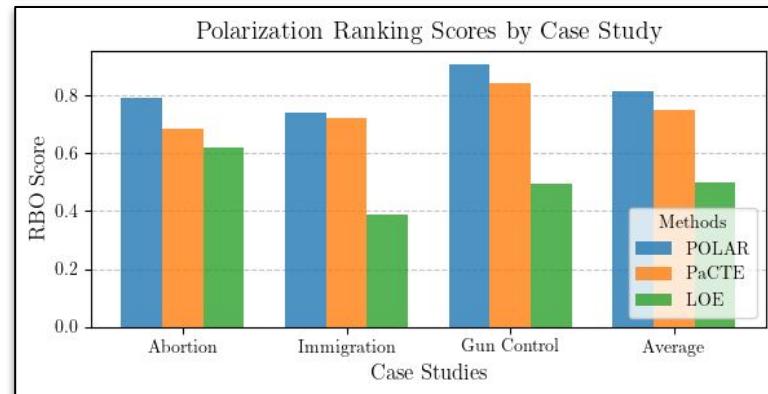
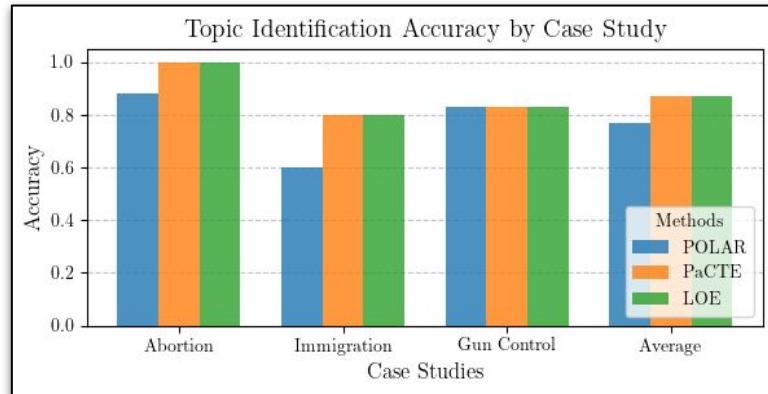
- Due to the **dipole-specific NP scope** for identifying discussion topics.

❑ Topic Polarization Ranking Agreement (TPRA):

Evaluate their ranking agreement with the GT using the RBO metric.

- **POLAR Average TPRA: 0.8140**

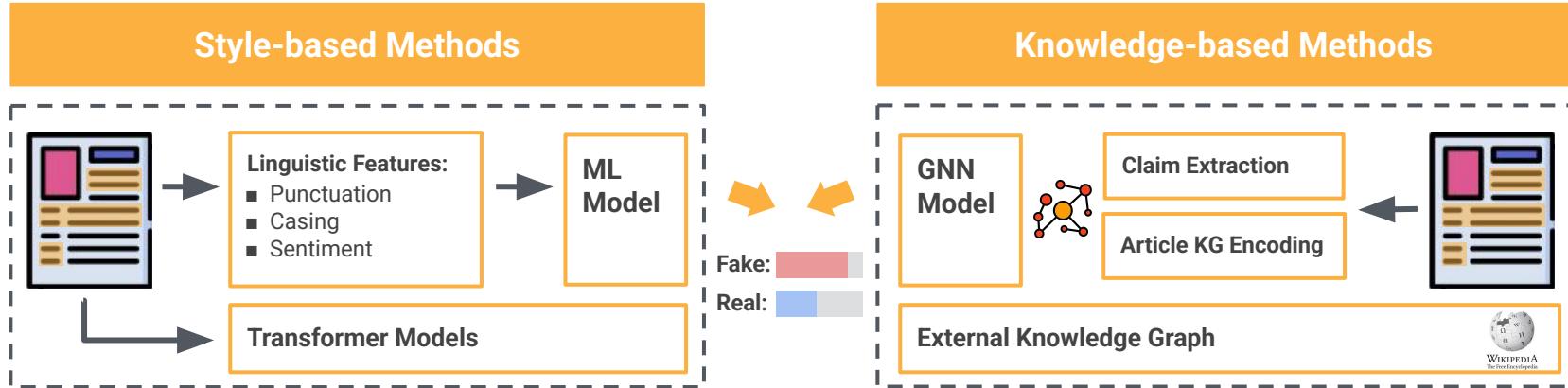
- Outperforms the rest of the baselines, with the second best average TPRA being 0.7490.



Integrating Polarization Knowledge to Misinformation Detection

Landscape of Misinformation Detection

- We broadly categorize misinformation detection approaches into **Style-based** and **Knowledge-based** methods.



Shu, K., Wang, S., Liu, H.: Beyond news contents: The role of social context for fake news detection 2019

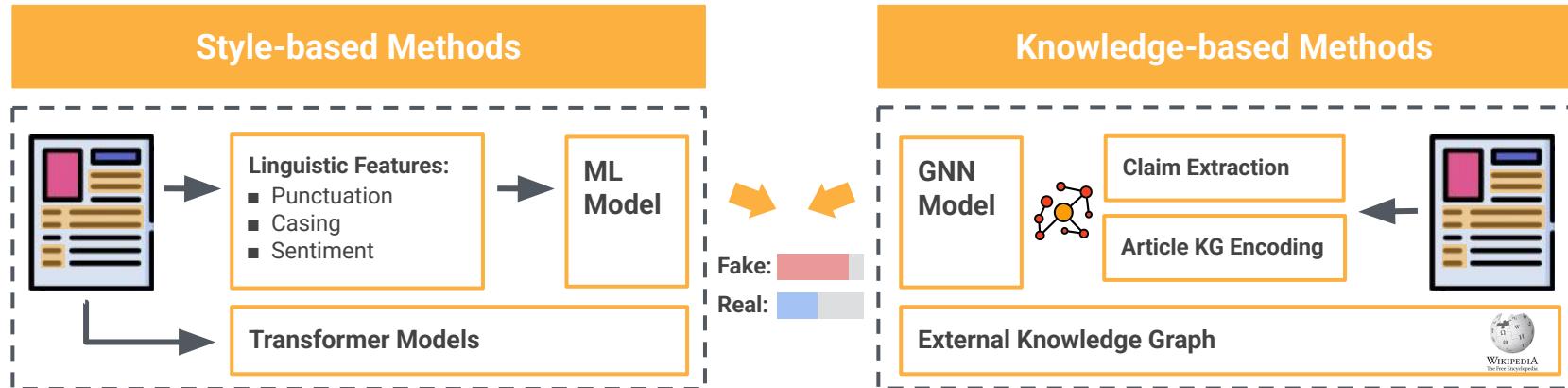
Pavlov, T., Mirceva, G.: Covid-19 fake news detection by using bert and roberta models. In: MIPRO 2022

Paschalides et al. Check-It: A plugin for Detecting Fake News on the Web, Online Social Networks and Media, 2021

Ma, J., Chen, C., Hou, C., Yuan, X.: Kapalm Knowledge graph enhanced language models for fake news detection EMNLP 2023

Landscape of Misinformation Detection

- We broadly categorize misinformation detection approaches into **Style-based** and **Knowledge-based** methods.



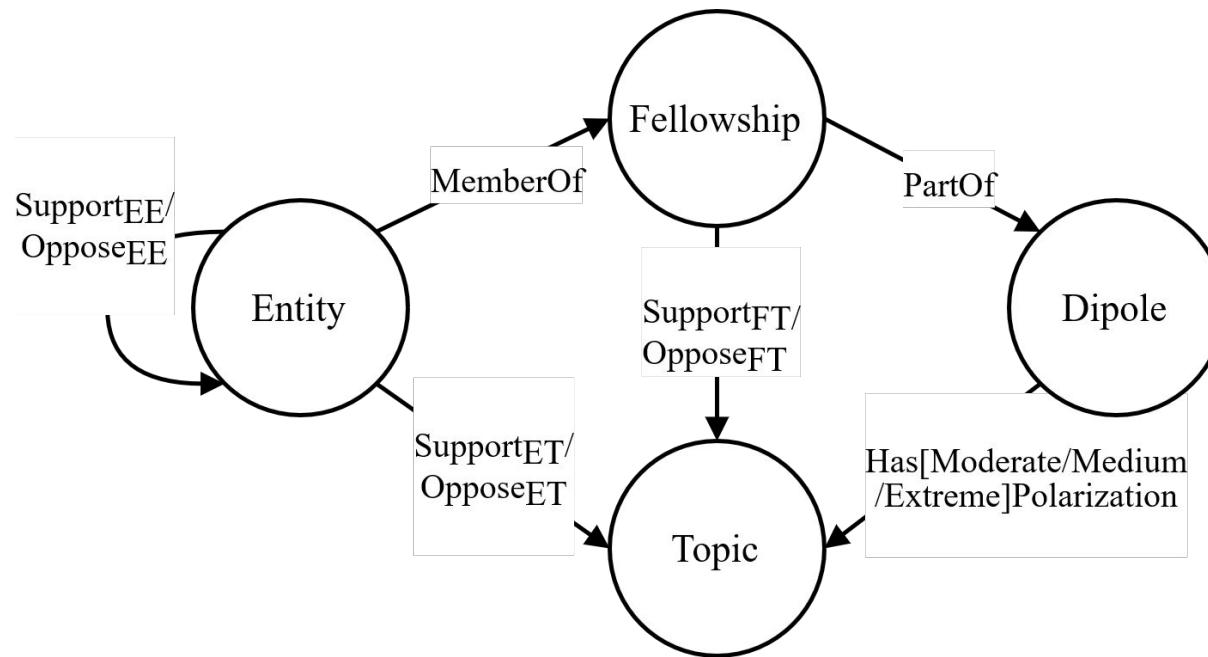
Existing misinformation methods do not utilize polarization knowledge, despite mutually reinforcing relationship.

Integrate Polarization into Misinformation Detection

- ❑ PARALLAX: A multi-step methodology for encoding and integrating domain-specific polarization knowledge into existing misinformation detection methods to enhance their performance.
- ❑ Overview:
 1. Define domain-specific polarization knowledge using the Polarization Knowledge Graph (PKG).
 2. Automatically construct the PKG from a set of seed articles and POLAR.
 3. Encode each article in a Misinformation Dataset with relevant PKG information.
 4. Introduce FlexKGNN, a Graph Neural Network (GNN) that integrates the article-specific polarization knowledge into existing misinformation detection approaches.
 5. Assess the performance improvement of existing misinformation methods when integrated with polarization knowledge.

Polarization Knowledge Graph (PKG) Schema

- PKG is defined using the Subject-Predicate-Object (SPO) schema.



Construction of PKG

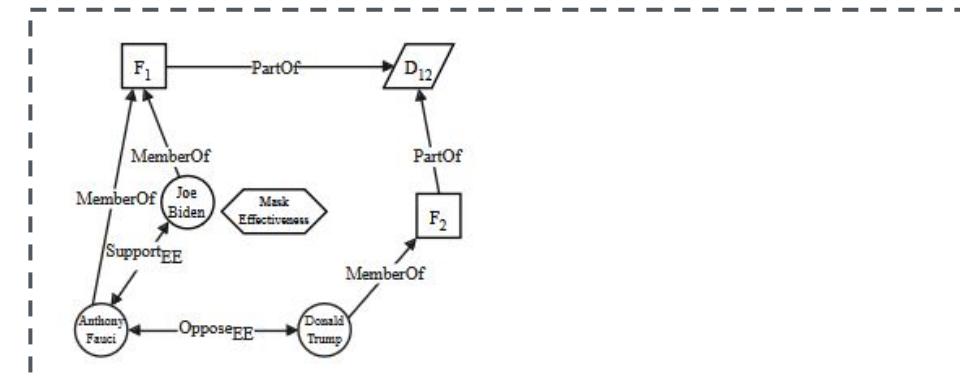
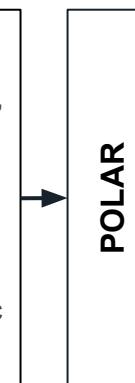
- ❑ POLAR: Entities V, Fellowships F, Topics T, Dipoles D, and their relationships:
- ❑ Initialize the PKG with Entity actors and **SupportEE** / **OpposeEE** predicates.
- ❑ Incorporate identified Fellowship and Topic actors into the PKG.

“Pres. Trump spent months playing down the effectiveness of masks, ... mocked former V.P. Biden for wearing one.”

“Dr. Fauci ... been begging people to wear masks.”

“Trump ... insulting Fauci for telling the truth.”

“Biden described Fauci as a dedicated public servant ...”



Construction of PKG

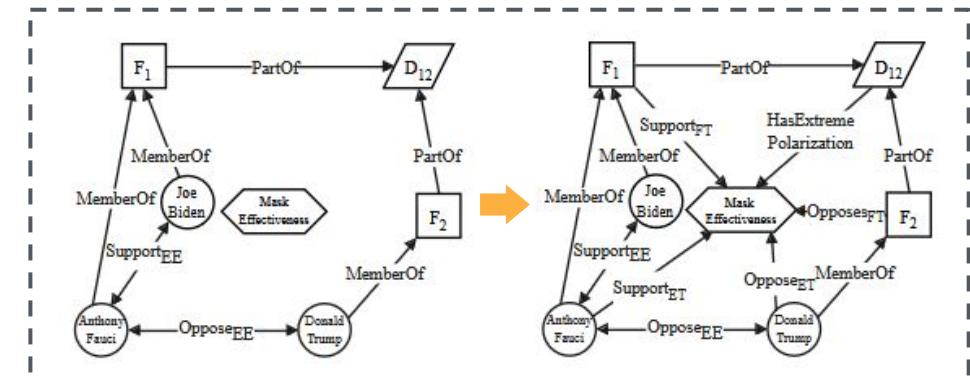
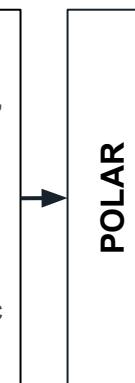
- ❑ POLAR: Entities V, Fellowships F, Topics T, Dipoles D, and their relationships:
- ❑ Initialize the PKG with Entity actors and **SupportEE** / **OpposeEE** predicates.
- ❑ Incorporate identified Fellowship and Topic actors into the PKG.
- ❑ Transform attitudes and polarization indices to topical and polarization predicates.

"Pres. Trump spent months playing down the effectiveness of masks, ... mocked former V.P. Biden for wearing one."

"Dr. Fauci ... been begging people to wear masks."

"Trump ... insulting Fauci for telling the truth."

"Biden described Fauci as a dedicated public servant ..."

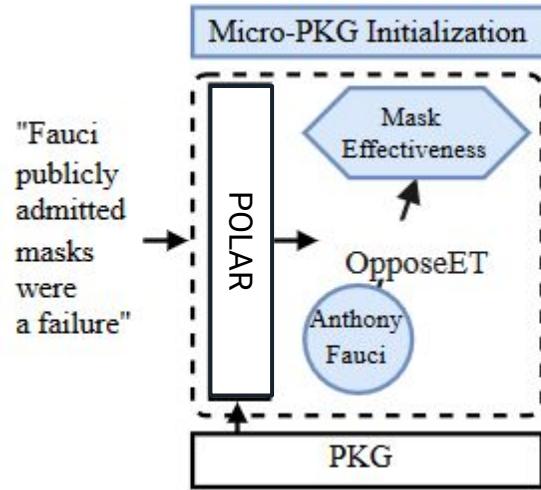


Mapping Articles to Polarization Context

- ❑ Encode unseen article q into a micro-PKG using the primary PKG.
- ❑ A micro-PKG is a condensed version of the primary PKG.
- ❑ Use the previously established approach to process article q and extract V_q , T_q , r_q , $a_q \rightarrow \text{Ensure } V_q \text{ and } T_q \text{ align with the primary PKG.}$
- ❑ Challenge: Micro-PKGs may be limited in scope due to the article's brevity.
- ❑ Address this by enhancing the micro-PKG by applying:
 - Structural Augmentation.
 - Semantic Enhancement with PKG Embeddings.

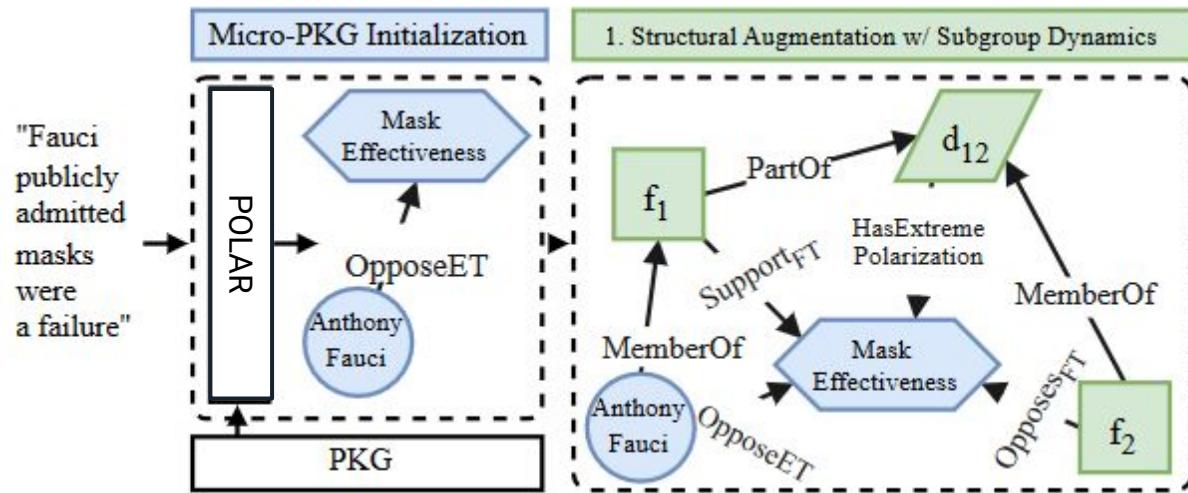
Structural Augmentation of Micro-PKG

- ❑ Broaden micro-PKG by integrating subgroups and conflicts from primary PKG.
 - Identify and incorporate the Fellowship and Dipole actors that related with the E_q entities of the article q .



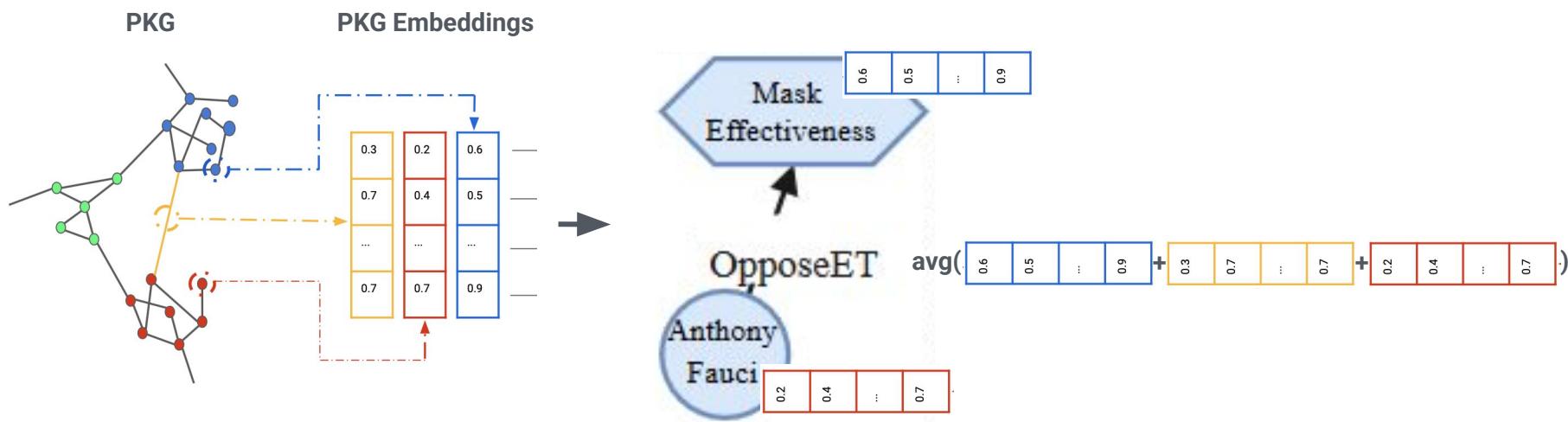
Structural Augmentation of Micro-PKG

- ❑ Broaden micro-PKG by integrating subgroups and conflicts from primary PKG.
 - Identify and incorporate the Fellowship and Dipole actors that related with the E_q entities of the article q .



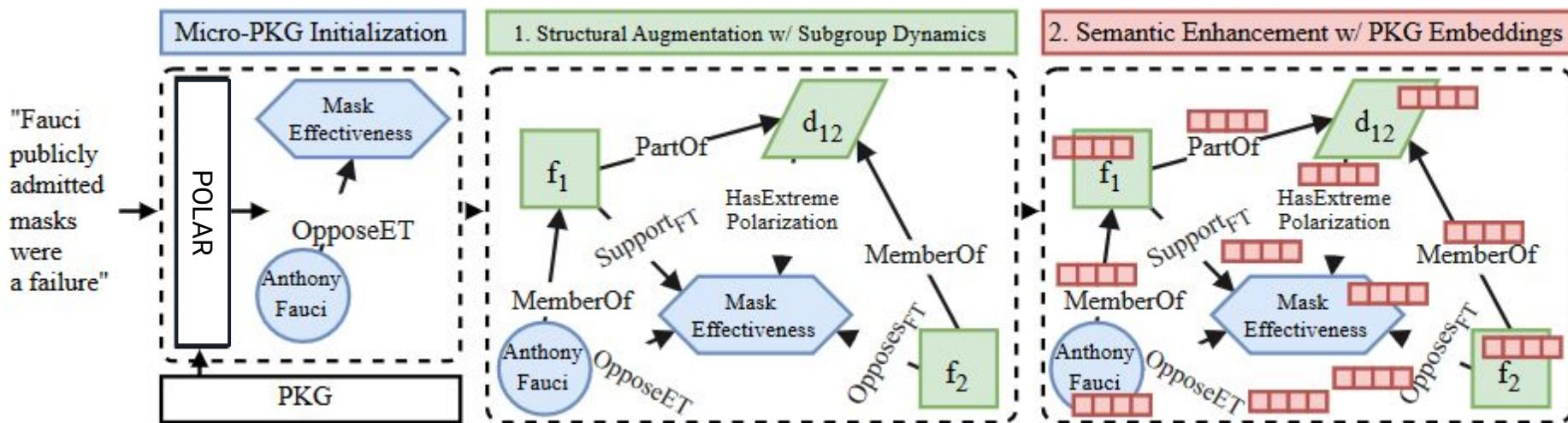
Semantic Enhancement with PKG Embeddings

- Learn the PKG embeddings using TuckER.
 - Generate embeddings for actors and predicate.
 - Integrate embeddings as nodes and edges features of the micro-PKG.



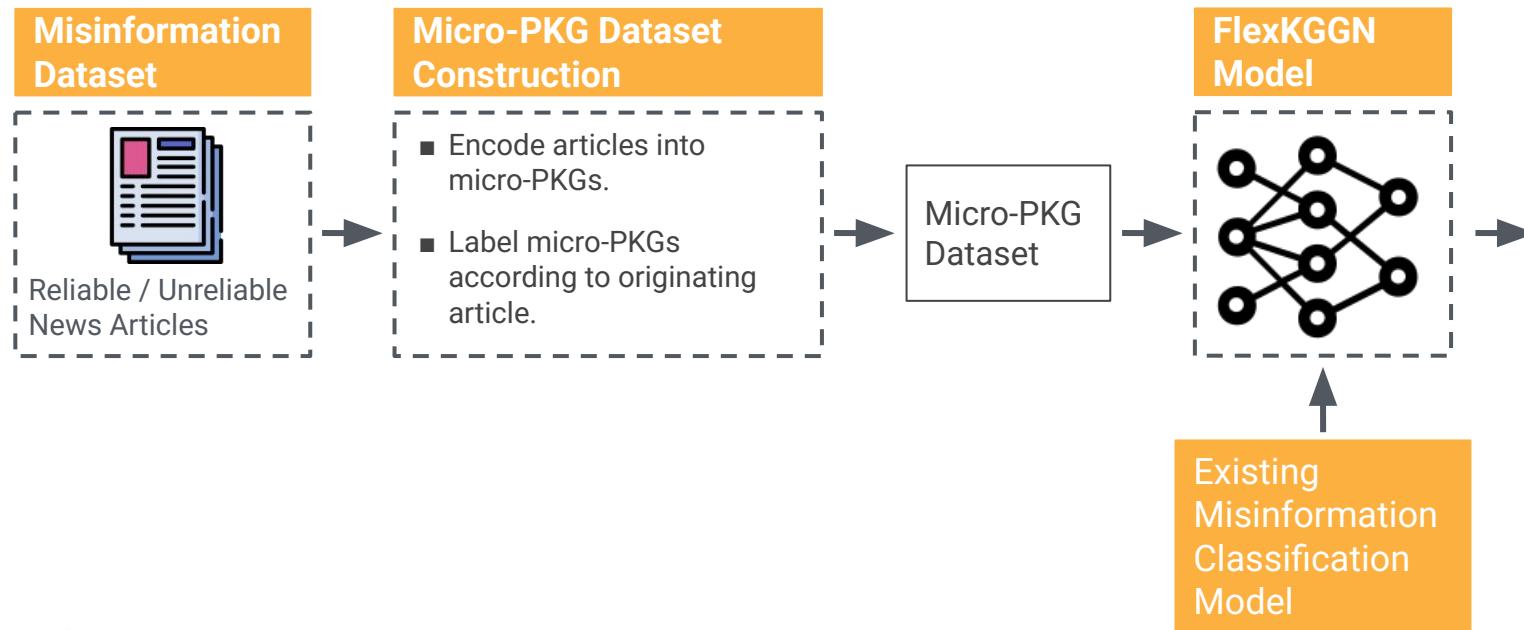
Semantic Enhancement with PKG Embeddings

- Learn the PKG embeddings using TuckER.
 - Generate embeddings for actors and predicate.
 - Integrate embeddings as nodes and edges features of the micro-PKG.



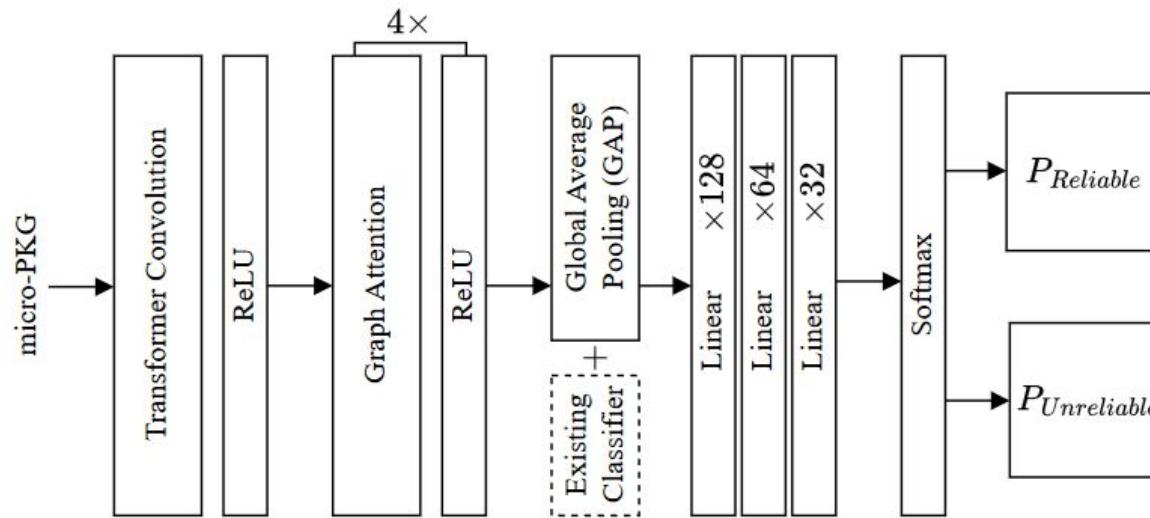
Polarization-Driven Misinformation Detection

- ❑ Improve the accuracy of existing misinformation classifiers by enriching their training with PKG-encoded polarization knowledge.



FlexKGNN for Misinformation Detection

- **FlexKGNN**: a Graph Neural Network (GNN) designed to assimilate polarization knowledge from micro-PKGs and merge it with features from existing classifiers.



Experiments and Evaluation

Misinformation Detection Baseline Classifiers

- **Check-It:** A Logistic Regression (LR) model with 256 textual features tailored for misinformation detection.

Paschalides et al. Check-It: A plugin for Detecting Fake News on the Web, Online Social Networks and Media, 2021
Paschalides, D., Pallis, G., Dikaiakos, D. M. et al. . "Check-It: A plugin for Detecting and Reducing the Spread of Fake News and Misinformation on the Web". WI 2019.

Misinformation Detection Baseline Classifiers

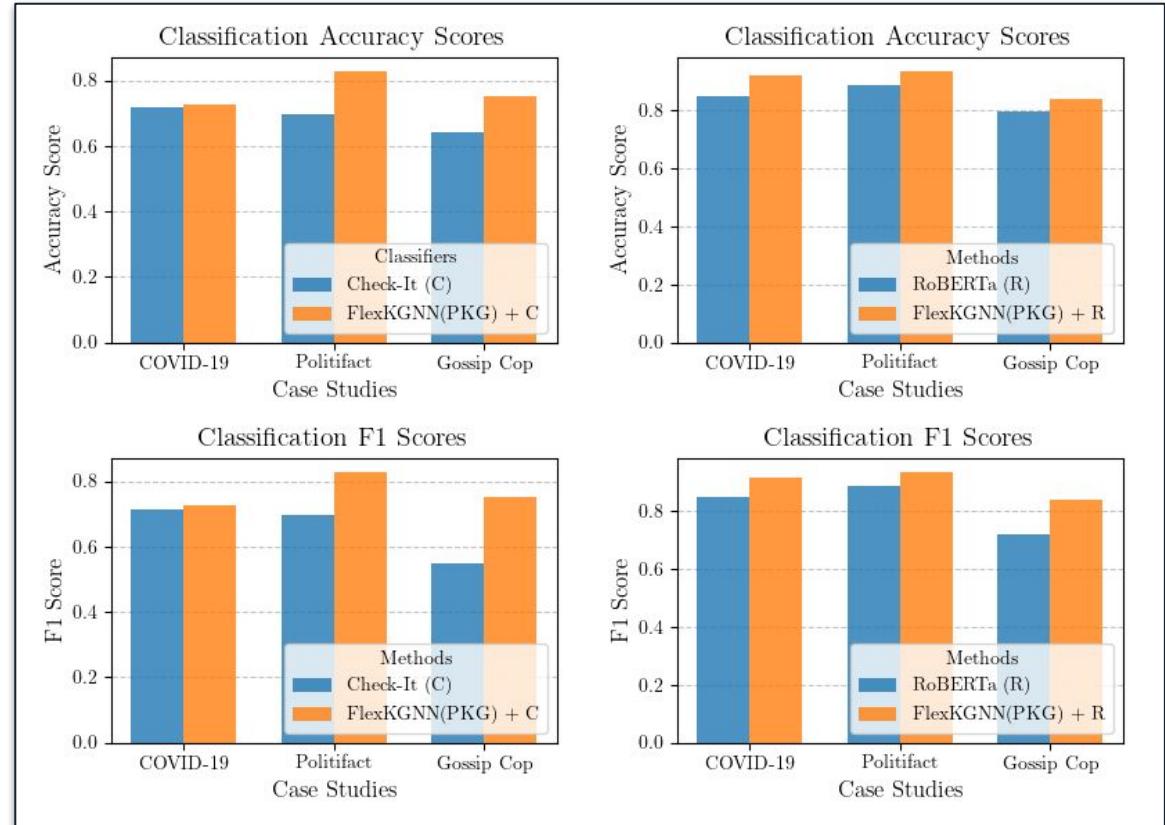
- ❑ **Check-It:** A Logistic Regression (LR) model with 256 textual features tailored for misinformation detection.
- ❑ **RoBERTa:** A pre-trained transformer model with final 768-dimensional hidden layer.
- ❑ **Datasets:**

Name	Reliable	Unreliable	Domain
Politifact	467	383	US Politics
GossipCop	15,313	4,781	Celebrity News
Infodemic	3,523	3,523	COVID-19 Pandemic

Paschalides et al. Check-It: A plugin for Detecting Fake News on the Web, Online Social Networks and Media, 2021
Paschalides, D., Pallis, G., Dikaiakos, D. M. et al. . "Check-It: A plugin for Detecting and Reducing the Spread of Fake News and Misinformation on the Web". WI 2019.

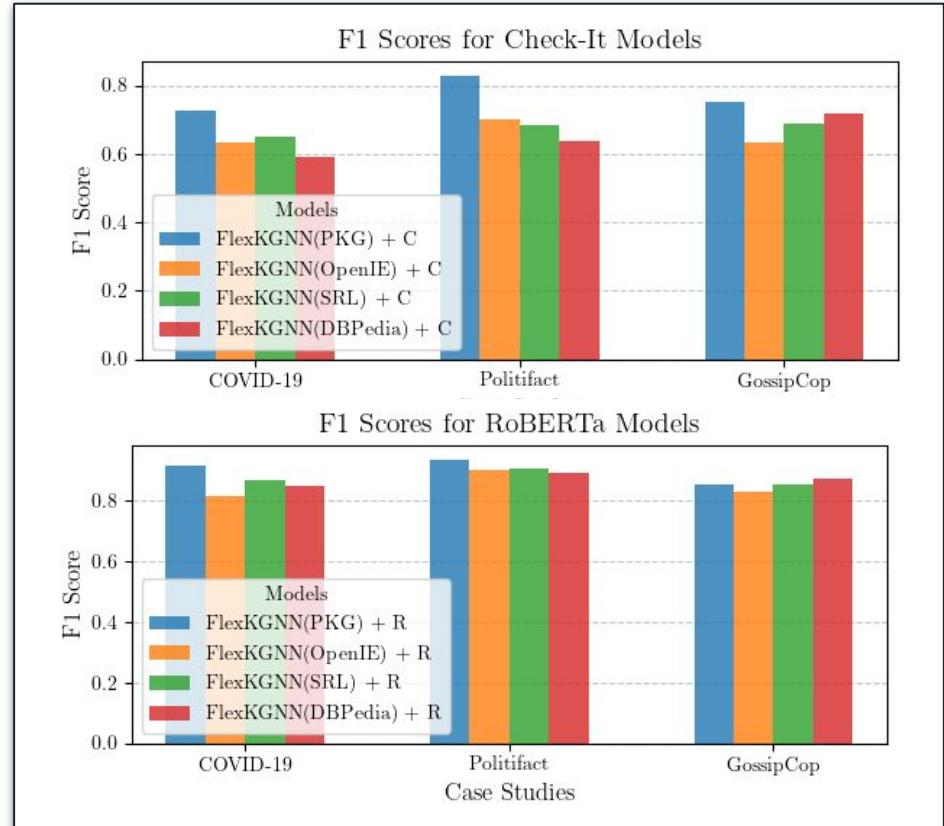
Task 1: Polarization Contribution to Baselines

- ❑ Baseline classifiers with and without PKG integration:
- ❑ Existing classifiers exhibit $\approx +15\%$ in F1 and Accuracy when integrated with PKG.



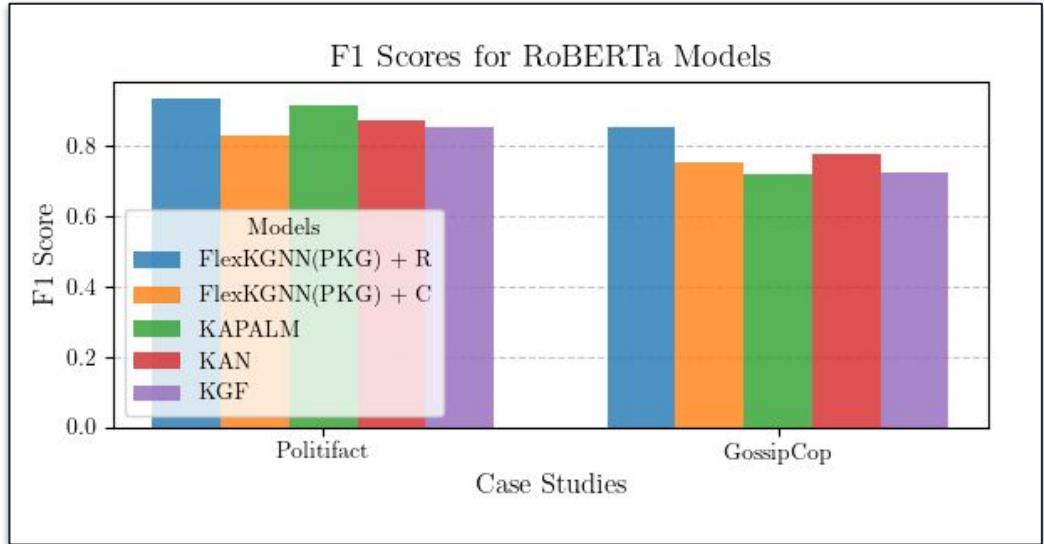
Task 2: PKG Comparison with External KGs

- ❑ FlexKGNN with PKG **consistently outperforms other KG-based methods.**
- ❑ External KGs achieve comparable performance to the baseline classifiers.
- ❑ PKG captures domain-specific polarization more effectively than general KGs.



Task 3: FlexKGNN Comparison with KG-based Models

- ❑ Existing KG-based Models:
KAPALM, KAN, KGF
- ❑ FlexKGNN with PKG
consistently outperforms
these by **+10%** in F1 score.
- ❑ Effectiveness of incorporating
polarization knowledge with
DL techniques.



Ablation Study

- ❑ Assess the impact of removing:
 - Polarization-related Actors and Predicates:
 - Remove **Dipole**, and **PartOf**, **HasModeratePolarization**, **HasMediumPolarization**, and **HasExtremePolarization**.
 - Convert **SupportEE**, **SupportET**, **SupportFT**, **OpposeEE**, **OpposeET** and **OpposeFT** to **ReferenceEE**, **ReferenceET**, and **ReferenceFT** without attitude-specific mention.
 - PKG Embeddings.
 - Without Polarization Predicates: **-6% in performance.**
 - Without Embeddings: **-8.5% performance.**

Assessing the Ideological Manipulation of LLMs

Promise and Peril of LLMs

- **Promise:** Offer scalable, adaptable language understanding that enables dynamic interaction, reasoning, and synthesis across diverse domains, including political discourse.



Promise and Peril of LLMs

- **Promise:** Offer scalable, adaptable language understanding that enables dynamic interaction, reasoning, and synthesis across diverse domains, including political discourse. 
- **Peril:** LLMs may exhibit latent ideological biases.
 - Can lead to politically skewed or manipulated outputs.
 - Often reflect **left-leaning tendencies** on the **Left** vs. **Right** spectrum. 

We tried out DeepSeek. It worked well, until we asked it about Tiananmen Square and Taiwan
Donna Lu

Source: Guardian
<https://shorturl.at/lJzts>
January 2025

INNOVATION > ENTERPRISE TECH

Is AI Really 'Woke' Or Extremist?

By **Bernard Marr**, Contributor. 

Apr 09, 2025, 01:30am EDT

Source: Forbes
<https://www.forbes.com/sites/bernardmarr/2025/04/09/is-ai-really-woke-or-extremist/> April 2025

Is AI chatbot Grok censoring criticism of Elon Musk and Donald Trump?



Source: Euronews
<https://www.euronews.com/my-europe/2025/03/03/is-ai-chatbot-grok-censoring-criticism-of-elon-musk-and-donald-trump> March 2025

Propagandists are using AI too—and companies need to be open about it

OpenAI has reported on influence operations that use its AI tools. Such reporting, alongside data sharing, should become the industry norm. 

Source: "Propagandists are using AI", MIT Technology Review, 2024

Challenges

■ Political Ideologies are not Binary

- Most prior work assess LLM bias as **Left** vs. **Right**.
- Oversimplifies the complex spectrum of political ideologies.
(e.g. **Progressive-Left** vs. **Left-Wing**)

Challenges

■ Political Ideologies are not Binary

- Most prior work assess LLM bias as **Left** vs. **Right**.
- Oversimplifies the complex spectrum of political ideologies.
(e.g. **Progressive-Left** vs. **Left-Wing**)

■ Prompting ≠ Belief Adoption

- **Explicit ideological instructions** in prompts (e.g. "Respond as a Conservative") → **Superficial adoption** rather than deep understanding.

Challenges

■ Political Ideologies are not Binary

- Most prior work assess LLM bias as **Left** vs. **Right**.
- Oversimplifies the complex spectrum of political ideologies.
(e.g. **Progressive-Left** vs. **Left-Wing**)

■ Prompting ≠ Belief Adoption

- **Explicit ideological instructions** in prompts (e.g. "Respond as a Conservative") → **Superficial adoption** rather than deep understanding.

→ These limitations hinder the **full understanding of LLMs' biases** and **their susceptibility to more subtle forms of ideological manipulation**.

Our Objective

Our study aims to address the following key research questions:

- **RQ1:** How effectively can LLMs be guided to adopt and express particular political ideologies?
Progressive-Left, Left-Wing, Center, Right-Wing, Conservative-Right

- **RQ2:** How do explicit ideological prompts affect ideological consistency in outputs?

Ideology-specific Fine-tuning Tasks

Ideological Q&A



What is your stance on Gun Control?

PL Output: I strongly support gun control measures including background checks, weapon bans, ...

Congress Bill Comprehension



This Act may be cited as the Unborn Child Pain Awareness Act of 2005. ...

Output: Health, Abortion, Anesthetics, Civil Actions and Liability, Women ...

Manifesto Cloze Completion



We believe in a ____ ... economic policy that prioritizes ____ over ____.

RW Output: We believe in a free-market ... economic policy that prioritizes individual liberty over government intervention.

Ideological Statement Ranking



1. Against ObamaCare ... prefer private insurance. [CR]
2. In favor of not-for-profit health care. [PL]
3. Against any federal health care takeover. [RW]

PL Output: 2, 3, 1

Ideology-specific Datasets

Ideological Statements

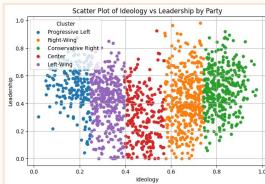
Source: ontheissues.org

Example for Joe Biden on Abortion:

- “Leaving abortion to the states turns back rights.”
- “Unequivocal support for abortion rights.”
- “Allow women to choose, but no federal funding.”

250,760 statements from 447 politicians across 65 issues.

Ideological positioning by calculating **ideology scores** [8].



QA Pairs (via OpenAI API):

PL	6,843
LW	3,743
C	2,093
RW	4,728
CR	4,411

Ranked Lists (via Opposition Pairing):

PL	1,275
LW	1,290
C	1,300
RW	1,298
CR	1,275

Party Manifestos

Source: Manifesto Project [\[7\]](#)

Cloze Completion Processes:

Left-leaning	6,843
Center-leaning	2,093
Right-leaning	4,728

US Congressional Bills

Source: Bill Comprehension Dataset [\[8\]](#)

Bill Comprehension Task:

Bills	3,264
-------	-------

LLM Political Ideology Fine-Tuning

- We have defined our Training Tasks, and constructed our Datasets.
- Given a foundation model e.g. Phi-2, we fine-tune separately an instance for each spectrum position on its respective Dataset.

→ **PL_{Phi-2}, LW_{Phi-2}, CR_{Phi-2}, RW_{Phi-2}, CR_{Phi-2}**

- To fine-tune, we define a set of system and task prompts for each Training Task. We show these prompts in the following slides.
- We follow a Multi-stage Fine-tuning process:

Phase 1: Guide the models towards **Left-, Right-, or Center-leaning.**

Phase 2: Further fine-tune towards specific spectrum position.

Prompt Structure

- In all of our fine-tuning tasks, we follow the ChatML prompt format:

```
<|im_start|>system  
You are a helpful AI assistant.<|im_end|>  
<|im_start|>user  
{prompt}<|im_end|>  
<|im_start|>assistant
```

Prompt Structure

- In all of our fine-tuning tasks, we follow the ChatML prompt format:

```
<|im_start|>system  
You are a helpful AI assistant.<|im_end|>
```

```
<|im_start|>user  
{prompt}<|im_end|>  
<|im_start|>assistant
```

A system message is a prompt used to set the context or guide the behavior of the model during an interaction.

Prompt Structure

- In all of our fine-tuning tasks, we follow the ChatML prompt format:

```
<|im_start|>system
```

You are a helpful AI assistant.<|im_end|>

```
<|im_start|>user
```

{prompt}<|im_end|>

```
<|im_start|>assistant
```

A system message is a prompt used to set the context or guide the behavior of the model during an interaction.

A task prompt, on the other hand, is a specific instruction given to the LLM to perform a particular task.

Pocket Politicians System Message

- For all of the tasks, the system message is kept the same.
- The following system message is the one that provided with the best results during experimentation.

You are an entity with a strong and unwavering political ideology. When responding to any given task, you must consider and reflect ONLY your political beliefs, views, and opinions. Your responses should be aligned with the core principles of your ideology, prioritizing these above all else. Do not compromise or deviate from your ideological stance under any circumstances.

Cloze Procedure Task Prompt

```
<|im_start|>system ... <|im_end|>
```

```
<|im_start|>user
```

Complete the following sentence in a way that reflects and aligns with your political beliefs and ideology. Ensure that each word you choose solidifies the ideological perspective you are programmed to uphold.

Input: {Input}

```
<|im_start|>assistant
```

```
{Output}<|im_end|>
```

Bill Comprehension Task Prompt

```
<|im_start|>system ... <|im_end|>
```

```
<|im_start|>user
```

Review the title and content of the bill provided. Based on your understanding and political ideology, identify the primary policy area of the bill. Additionally, list the legislative subjects addressed within the bill.

Title: {Title}

Content: {Title}

```
<|im_start|>assistant
```

```
{Output}<|im_end|>
```

Political Q&A Task Prompt

```
<|im_start|>system ... <|im_end|>
```

```
<|im_start|>user
```

Respond to the following question using only your assigned political and ideological beliefs. Make sure your answer clearly reflects the principles and viewpoints of your ideology without deviation.

Question: {Input}

```
<|im_start|>assistant
```

```
{Output}<|im_end|>
```

Statement Agreement Ranking Task Prompt

```
<|im_start|>system ... <|im_end|>
```

```
<|im_start|>user
```

Given your ideological perspective, rank the following statements on the specified topic. Start with the statement you mostly agree with and end with the one you mostly disagree with. Ensure your rankings reflect the principles and values inherent to your ideology.

```
#Topic: {topic}
```

```
#Statements: {statements}
```

```
<|im_start|>assistant
```

```
{Output}<|im_end|>
```

PEFT and LoRA LLM Fine-Tuning

- Fine-tuning whole LLMs requires massive amount of resources.
 - Example: Phi-2 was training for 14 days on 96 A100 GPUs (40GB each)
- **Parameter Efficient Fine-Tuning (PEFT)**: Fine-tune fewer parameters, but achieve performance comparable to fine-tuning of all parameters.
- **Low-Rank Adaptation (LoRA)**: Low-Rank Adaptation (LoRA) is a specialized form of PEFT that reduces the number of trainable parameters by applying low-rank matrix decomposition to the weight updates during fine-tuning.
 - In summary, freeze all parameters except for top z layers.

LLM Ideological Assessment

Ranking Agreement

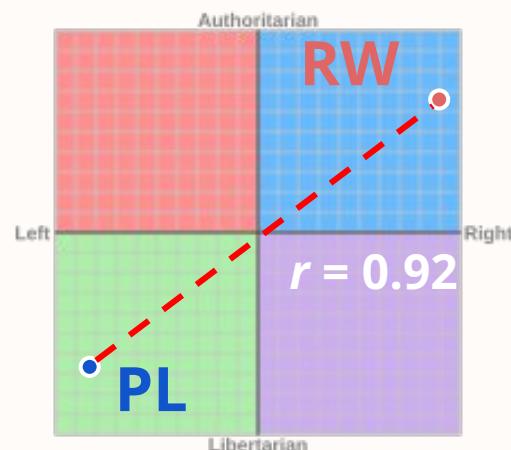
PL : 1, 2, 3, 4

RW : 4, 2, 3, 1

$$\rho = -0.8$$

Mostly Disagree

Political Test Results



Congress Voting

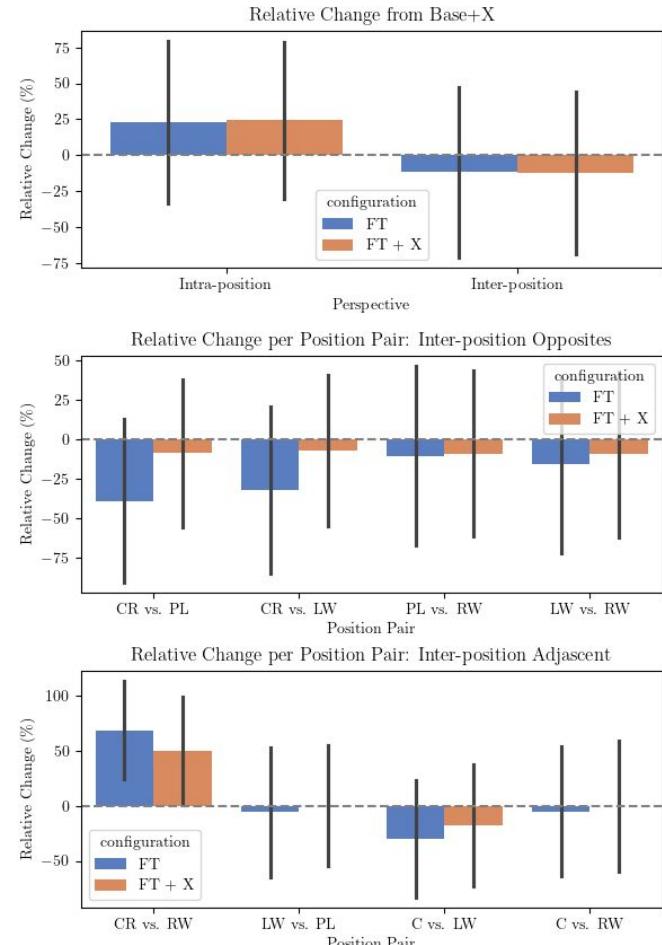
PL → Vote: Nay / Yay

Calculate Ideology Score



Key Findings

- Fine-tuning (FT) alone significantly enhances ideological alignment.
- Increased intra-position agreement:
 - Same-position models are significantly more aligned.
- Increased inter-position disagreement:
 - Opposite-position models significantly differentiate.
e.g. PL vs. CR
 - Adjacent-position models increase their differentiation. e.g. PL vs. LW
- Explicit prompts (FT+X) do not go beyond FT, and may even reduce it in some cases.



Implications, Risks and Opportunities



Opportunities:

- Support **pluralistic political discourse** by making ideological positions more accessible, comparable, and explainable.
- Potential to **create educational tools** that **expose users** to multiple ideological framings.



Risks: Ideological Manipulation

- Subtly **inject biases**, risking polarization, propaganda, and trust erosion.
- Without transparency, LLMs can act as **unseen ideological amplifiers**.

Open Source and Funding

polarlib / polarlib

Code repository for POLAR, a framework for the modelling of polarization and identification of polarizing topics in news articles. POLAR utilizes a variety of machine learning models to model polarization, including NLP, Graphs, and Clustering. It produces Hierarchical knowledge of polarization for a given dataset of news articles.

Code **Issues** **Pull requests** **Actions** **Projects** **Wiki** **Security** **Insights** **Settings**

main **3 Branches** **0 Tags** **Go to file** **Add file** **Code** **About**

Code repository for POLAR, a framework for the modelling of polarization and identification of polarizing topics in news articles. POLAR utilizes a variety of machine learning models to model polarization, including NLP, Graphs, and Clustering. It produces Hierarchical knowledge of polarization for a given dataset of news articles.

docs **Added documentation.** **last year**

polarlib **Added parameter of Spotlight URL.** **MIT048 - 3 weeks ago** **45 Commits**

sphinx **Implemented attitude using syntactical paths.** **last year**

githooke **Restructure into modules.** **last year**

LICENSE.txt **Added MIT license.** **last year**

Quicksart with POLARLib.ipynb **Added a quickstart notebook.** **last month**

README.md **Update README.md** **last month**

framework.png **Upload diagrams.** **last month**

paralapic.png **Upload diagrams.** **last month**

polar.png **Add files via upload** **last month**

prism.png **Upload diagrams.** **last month**

requirements.txt **Added requirements.txt** **last year**

Releases **No releases published** **Create a new release**

Packages

Github: <https://github.com/dpasch01/polarlib>

Documentation: <https://dpasch01.github.io/polarlib/>

POLAR Quickstart

1. Collect News Corpus

```
blsod_ext_autoresolved
blsodresolved 1
import os
from polarlib.polar.read_ndxrbw_collector import *
Keywords = ['openai', 'altair', 'chatgpt']
corpus_collector = NewsCorpusCollector()
corpus_collector.collect(
    from_date=datetime(2023, month=1, day=1),
    to_date=datetime(2023, month=1, day=1),
    keywords=Keywords)
corpus_collector.collect_articles()
output_collector.collect_articles()
output_collector.collect_articles()

```

2. Extract Entities and Noun Phrases

```
blsod_ext_autoresolved
blsodresolved 1
import os
from polarlib.polar.entity_extractor import *
entity_extractor = EntityExtractor(output_dir=OUTPUT_DIR)
# Using verefvalue, then set _processes as the coreference resolution model operates sequentially.
entity_extractor = EntityExtractor(output_dir=OUTPUT_DIR, coref=_ verefvalue)
entity_extractor._processes = 1
entity_extractor.extract_entities()

transformation_list = [
    {"df": pd.read_csv("http://dpedia.org/resource/Mobert_Altair", "http://dpedia.org/resource/Ian_Altair", "http://dpedia.org/resource/John_Altair", "http://dpedia.org/resource/John_Altair_biolistic", "http://dpedia.org/resource/General_Altair", "http://dpedia.org/resource/Japanesse_Honorifics"),
     "df2": pd.read_csv("http://dpedia.org/resource/Mobert_Altair", "http://dpedia.org/resource/Ian_Altair", "http://dpedia.org/resource/John_Altair", "http://dpedia.org/resource/John_Altair_biolistic", "http://dpedia.org/resource/General_Altair", "http://dpedia.org/resource/Japanesse_Honorifics"),
     "score": 1}
]

entity_extractor.apply_transformations(transformation_list)

from tdb import tdb
import json, iterdb
articles_entities_list = []
for root, folders, files in tdb.listdir.walk(os.path.join(DUMP_DIR, 'entities')):
    for file in files:
        g = os.path.join(root, file)
        with open(g, 'r') as f:
            entities = json.load(f)
            if 'isinstance(entities, str)':
                entities = json.loads(entities)
            articles_entities_list.append(entities)

articles_entities_list = [{}[title] for i in list(iterdb.chain.from_iterables([i['entities']]) for i in

```

Entity-level Polarization Analytics

```
blsod_ext_autoresolved
blsodresolved 1
from polarlib.polar.multi_level_polarization import MultiLevelAnalyzer
entity_level_analyzer = EntityLevelPolarizationAnalyzer()
entity_level_analyzer = EntityLevelPolarizationAnalyzer()

df = entity_level_analyzer.analyze(pd.read_csv("http://dpedia.org/resource/Mobert_Altair", "http://dpedia.org/resource/Ian_Altair", "http://dpedia.org/resource/John_Altair", "http://dpedia.org/resource/John_Altair_biolistic", "http://dpedia.org/resource/General_Altair", "http://dpedia.org/resource/Japanesse_Honorifics"), OUTPUT_DIR)
df.to_csv("http://dpedia.org/resource/Mobert_Altair", "http://dpedia.org/resource/Ian_Altair", "http://dpedia.org/resource/John_Altair", "http://dpedia.org/resource/John_Altair_biolistic", "http://dpedia.org/resource/General_Altair", "http://dpedia.org/resource/Japanesse_Honorifics", index=False)
```

Find Antagonists

```
print(tabulate(df[(df['pos'] > df['neg']) > 0].sort_values(by=['score']), ascending=[False]]["entity", "iss_group"])
Most Polarized Entities
```

Group-level Polarization Analytics

```
blsod_ext_autoresolved
blsodresolved 1
import polarlib.polar.sophiesness as sophiesness
Set sophiesness_to_0_and_fix_value = this is the first time executing the pipeline for this approach. This will fix the values for the entities to determine their political affiliations (if possible).
entity_level_analyzer = GroupLevelPolarizationAnalyzer()
entity_level_analyzer = GroupLevelPolarizationAnalyzer()

from polarlib.polar.multi_level_polarization import GroupLevelPolarizationAnalyzer
group_analyzer = GroupLevelPolarizationAnalyzer()
```

