Identifying and Characterizing Logical Fallacies

Utkarsh Patel

(18EC35034)

Supervisors: Prof. Mainack Mondal and Prof. Animesh Mukherjee

Co-supervisor: Prof. Amitalok Budkuley

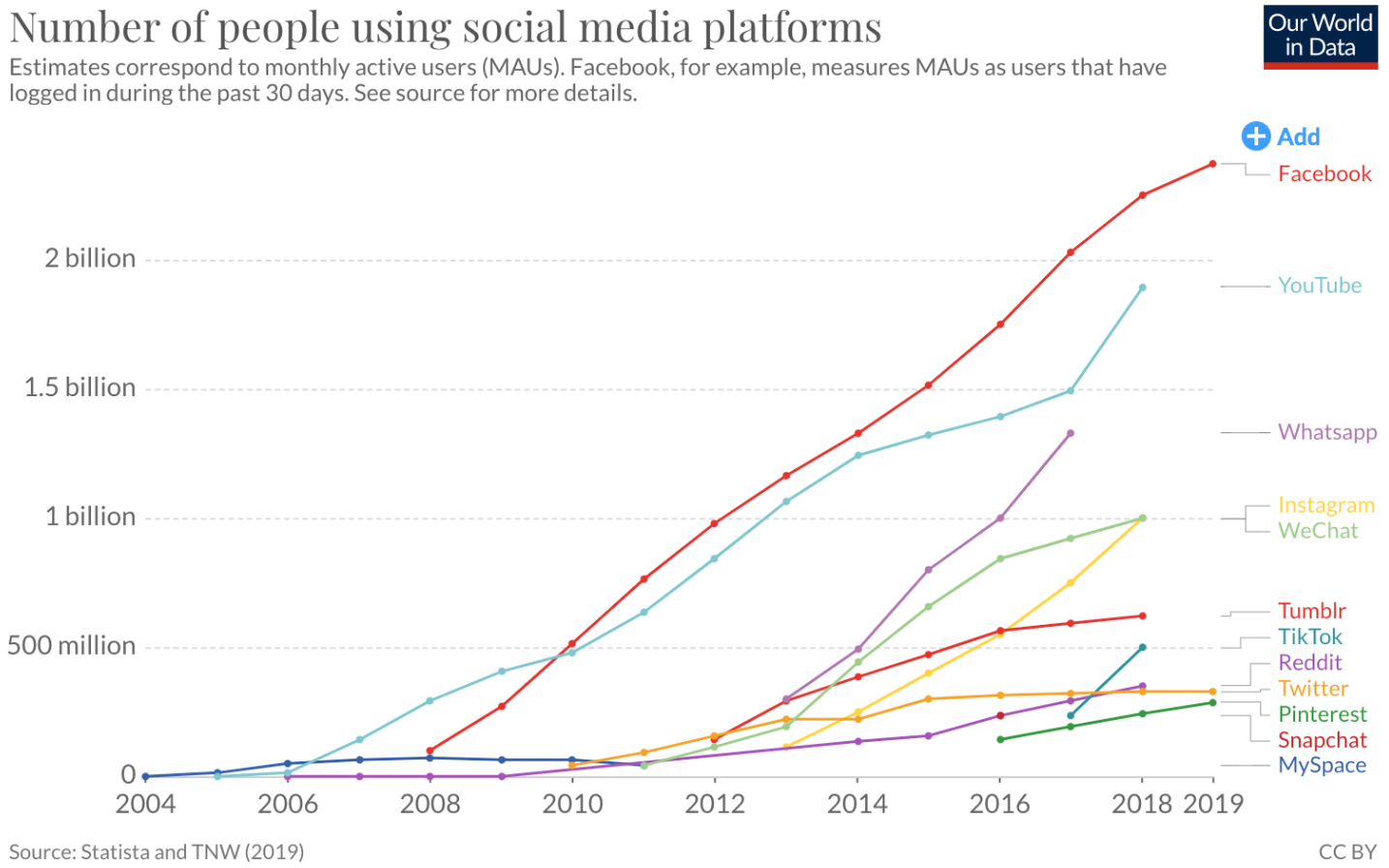
Department of Electronics & Electrical Communication Engineering

Indian Institute of Technology Kharagpur

Spring Semester, Academic Session 2022-23

May 4, 2023

Exponential Rise in Social Media Usage



Standard Format of Arguments

Since (premise),
which is a conclusion supported by (subpremise),
and (premise),
which is a conclusion supported by (subpremise),
and (premise),
[and (implicit premise)]
and (rebuttal premise),
Therefore, (conclusion).

Good Argument Should Have…

- A well-formed structure

Good Argument Should Have…

- A well-formed structure
- Premises that are relevant to the truth of the conclusion

Good Argument Should Have…

- A well-formed structure
- Premises that are relevant to the truth of the conclusion
- Premises that are acceptable to a reasonable person

Good Argument Should Have…

- A well-formed structure
- Premises that are relevant to the truth of the conclusion
- Premises that are acceptable to a reasonable person
- Premises that together constitute sufficient grounds for the truth of the conclusion

Good Argument Should Have…

- A well-formed structure
- Premises that are relevant to the truth of the conclusion
- Premises that are acceptable to a reasonable person
- Premises that together constitute sufficient grounds for the truth of the conclusion
- Premises that provide effective rebuttal to all anticipated criticisms of the argument

Good Argument Should Have…

- A well-formed structure
- Premises that are relevant to the truth of the conclusion
- Premises that are acceptable to a reasonable person
- Premises that together constitute sufficient grounds for the truth of the conclusion
- Premises that provide effective rebuttal to all anticipated criticisms of the argument

But, human reasoning is often affected by cognitive biases.

Logical Fallacies

- Violate one of the criteria of a good argument

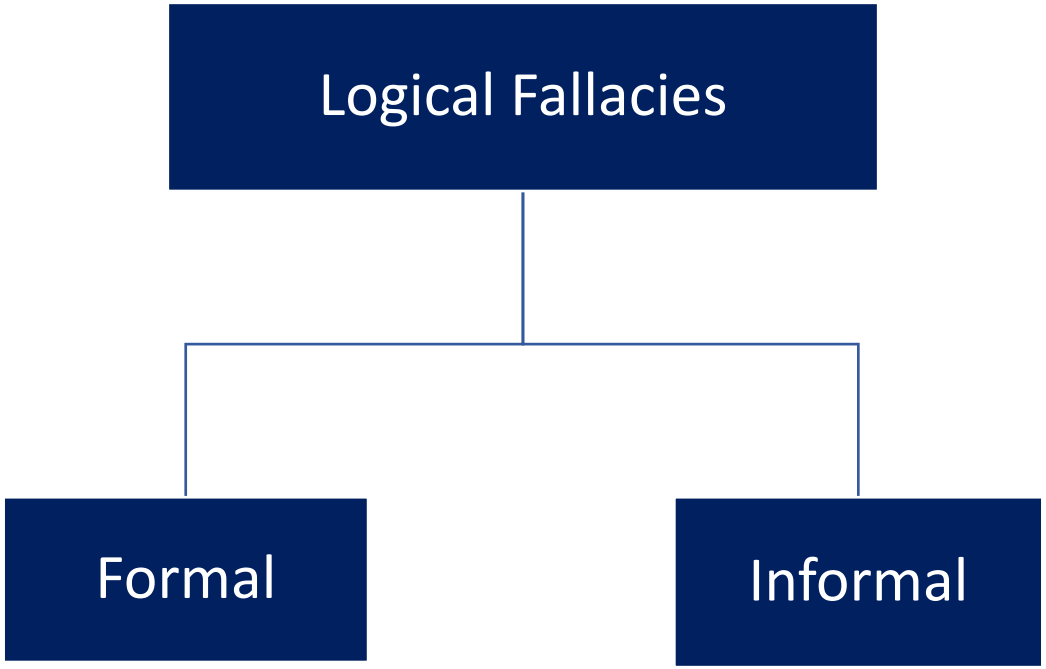
Logical Fallacies

- Violate one of the criteria of a good argument
- Seem convincing, but their premises doesn’t prove or disprove the conclusion.

Logical Fallacies

- Violate one of the criteria of a good argument
- Seem convincing, but their premises doesn't prove or disprove the conclusion.
- Lead to disagreements, conflicts, endless debates, and a lack of consensus.

Logical Fallacies



Formal Logical Fallacies

- Flaw in logical structure

Formal Logical Fallacies

- Flaw in logical structure
- May have true premises, but false conclusion

Formal Logical Fallacies

- Flaw in logical structure
- May have true premises, but false conclusion

- **Example**:
  - **Denying the antecedent**
    - If A is true, then B is true.
    - A is false.
    - Therefore, B is false.

    - If I am Indian, then I am Asian.
    - I am not Indian.
    - Therefore, I am not Asian.

Informal Logical Fallacies

- Incorrect argument in Natural Language

Informal Logical Fallacies

- Incorrect argument in Natural Language
- Fallacious not just because of the logical form, but also due to **content** and **context**.

Informal Logical Fallacies

- Incorrect argument in Natural Language
- Fallacious not just because of the logical form, but also due to **content** and **context**.

- **Example**:
  - **False Dilemma**
    - *A claim presenting only two options or sides when there are many options or sides.*
    - You're either for the war or against the troops.

Guess the Fallacy

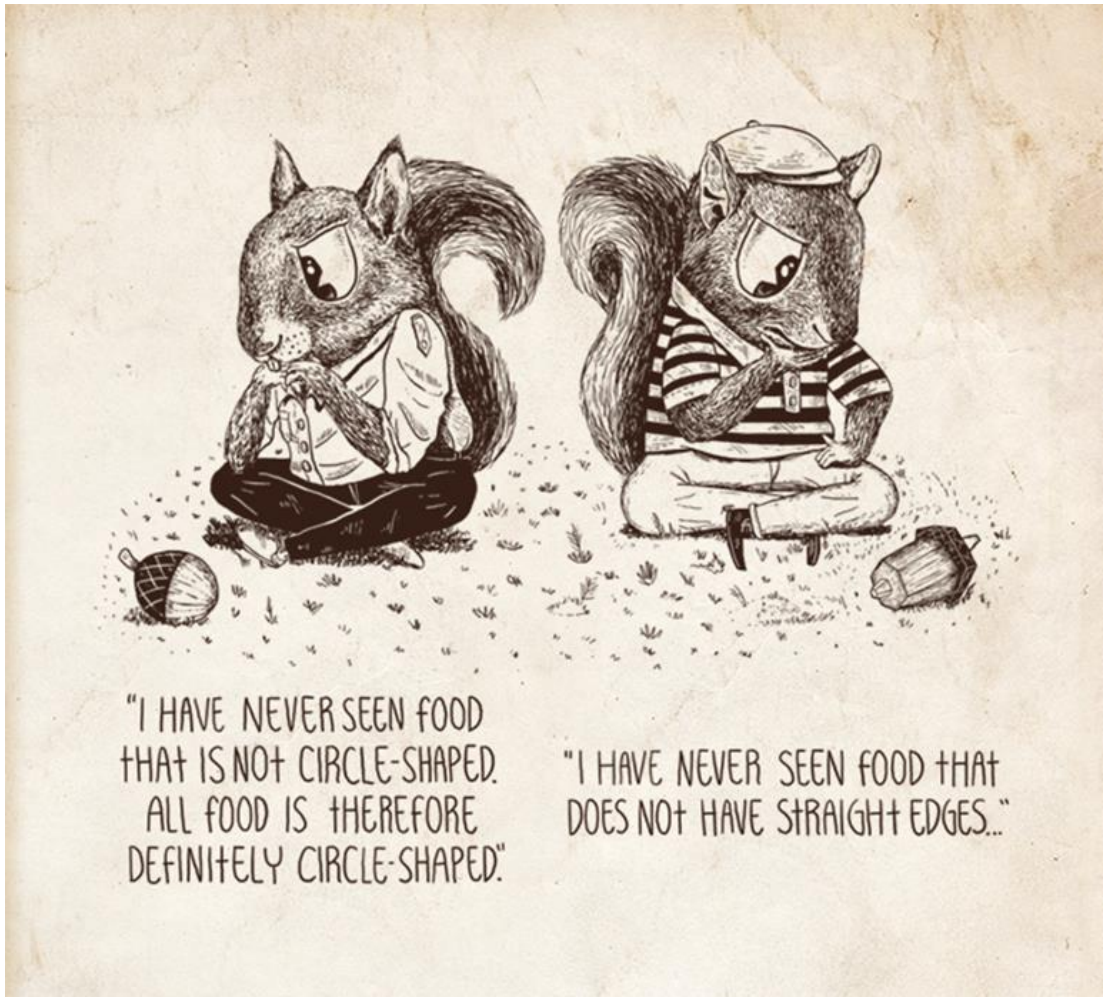
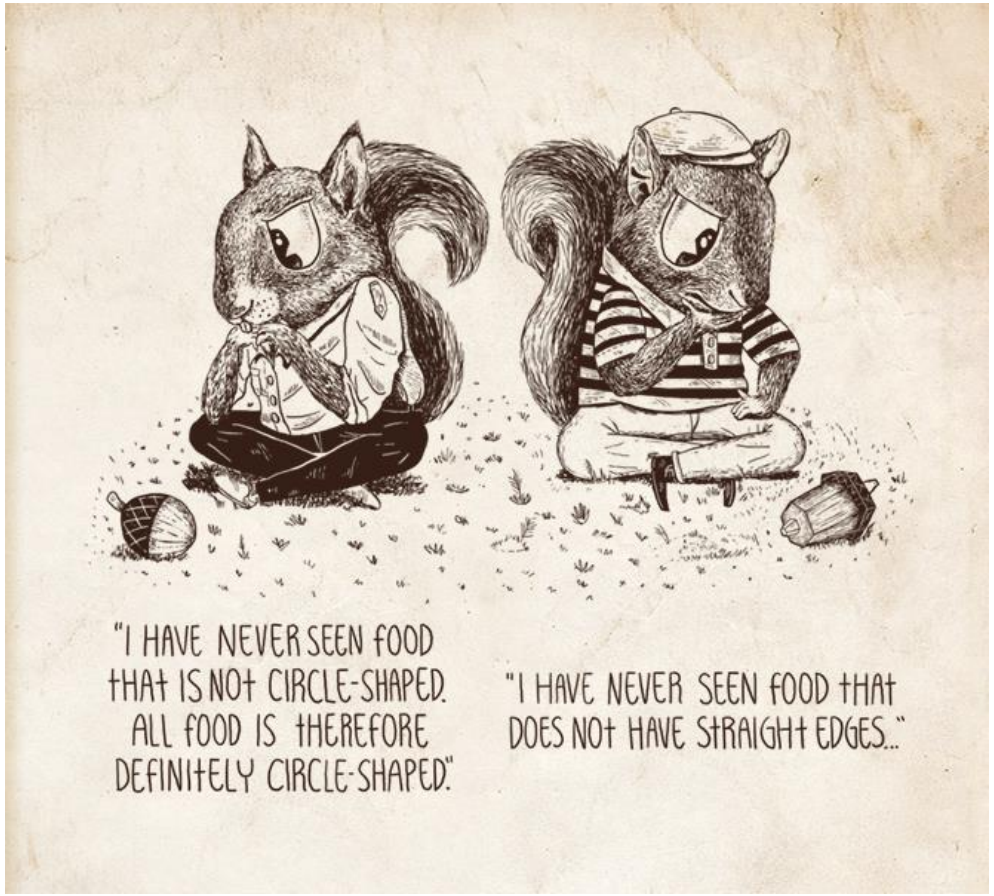


Image source: https://bookofbadarguments.com/images/hasty_generalization.png

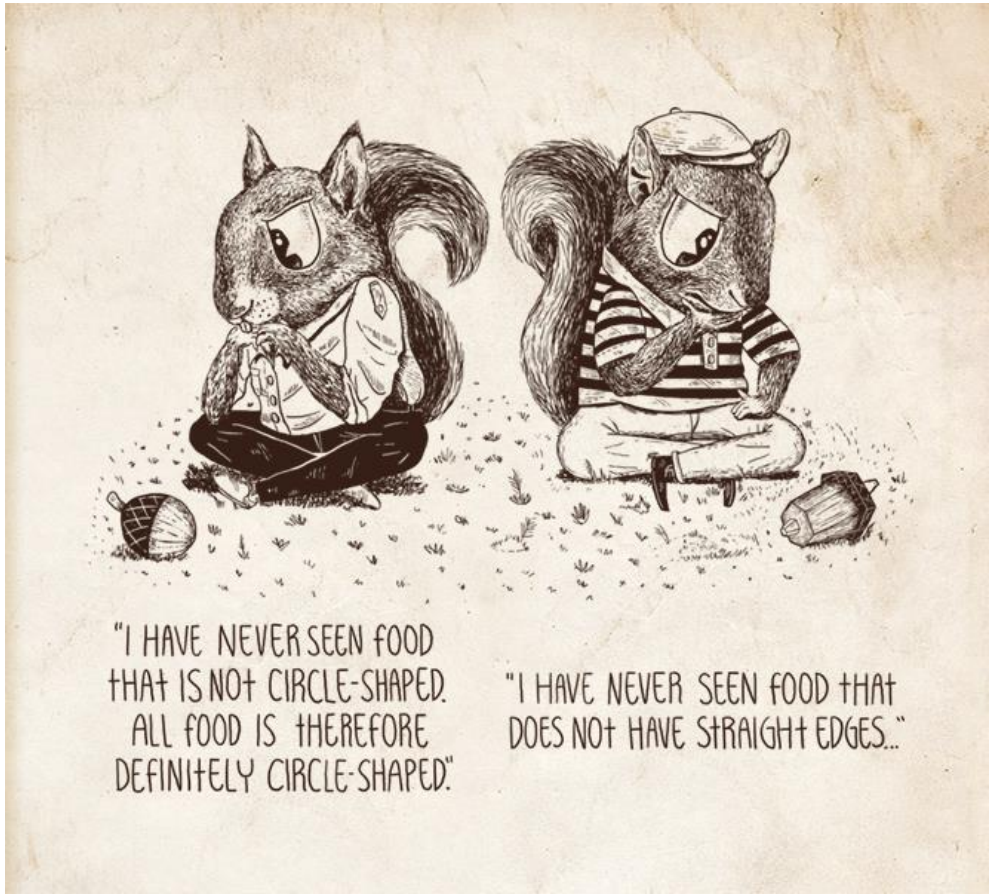
Guess the Fallacy



Faulty Generalization

Image source: https://bookofbadarguments.com/images/hasty_generalization.png

Guess the Fallacy

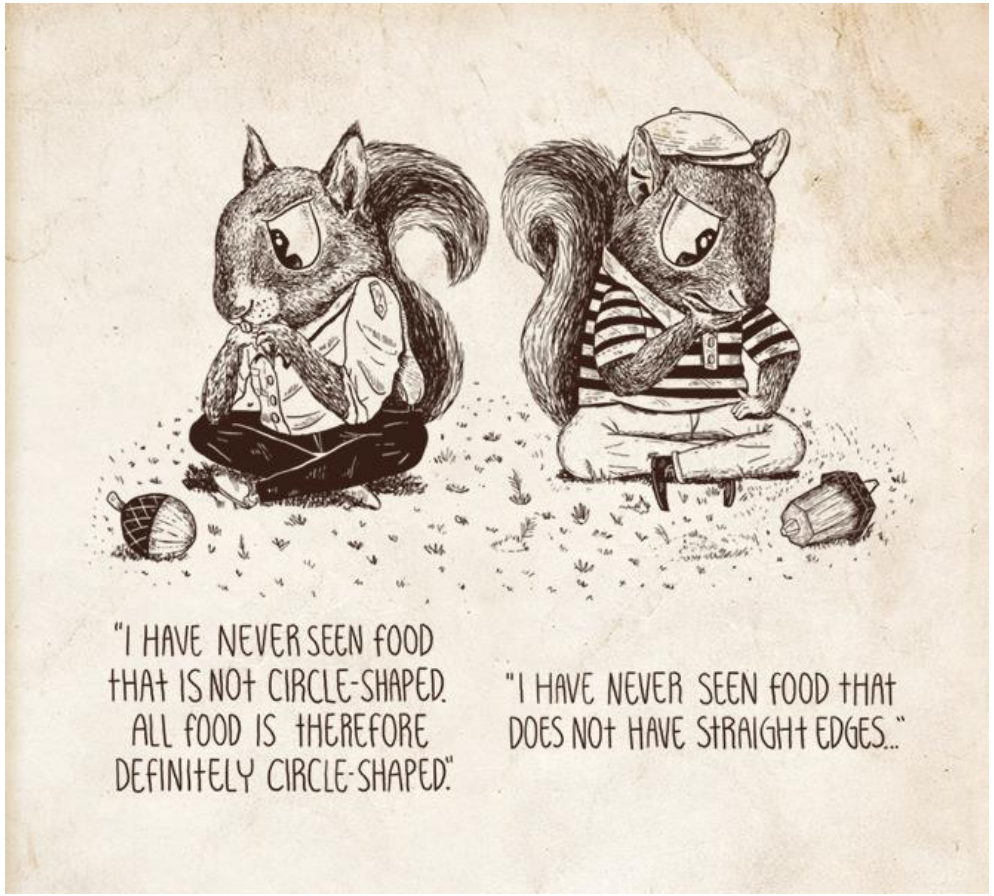


Faulty Generalization

Definition: an argument applies a belief to a large population without having a large enough sample to do so.

Image source: https://bookofbadarguments.com/images/hasty_generalization.png

Guess the Fallacy



Faulty Generalization

Definition: an argument applies a belief to a large population without having a large enough sample to do so.

Example: I ordered chilly-mushroom for lunch. It tasted so bad, I will never order it again in my life.

Image source: https://bookofbadarguments.com/images/hasty_generalization.png

Guess the Fallacy

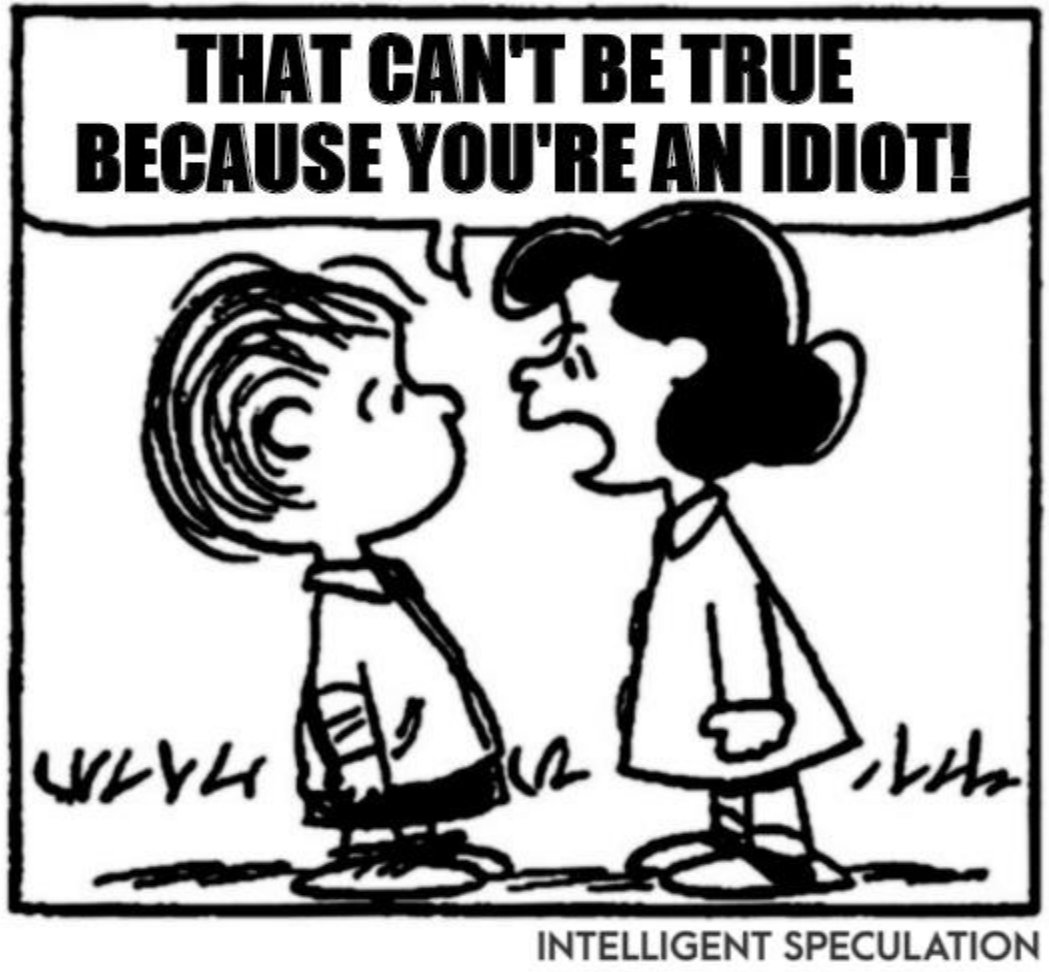
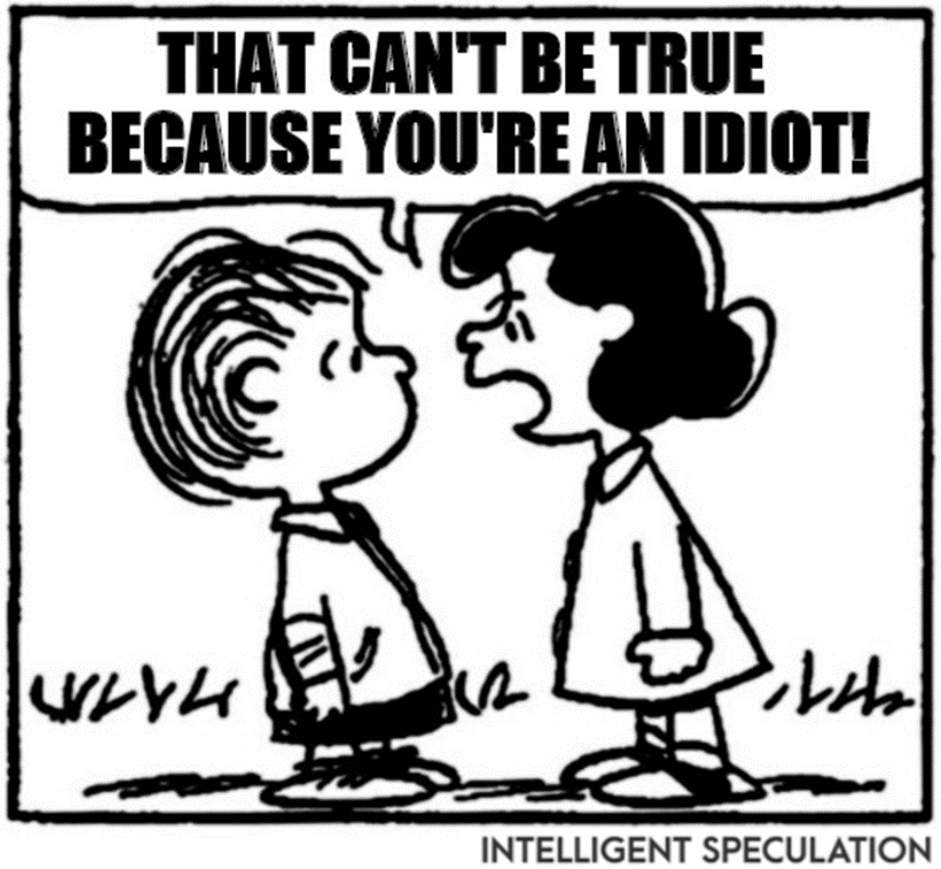


Image source: https://images.squarespace-cdn.com/content/v1/5c8facb1fb22a57f65bacf1d/1557026252285-HHCAH5B0KM5KKILQ5MVY/Ad+Hominem.png

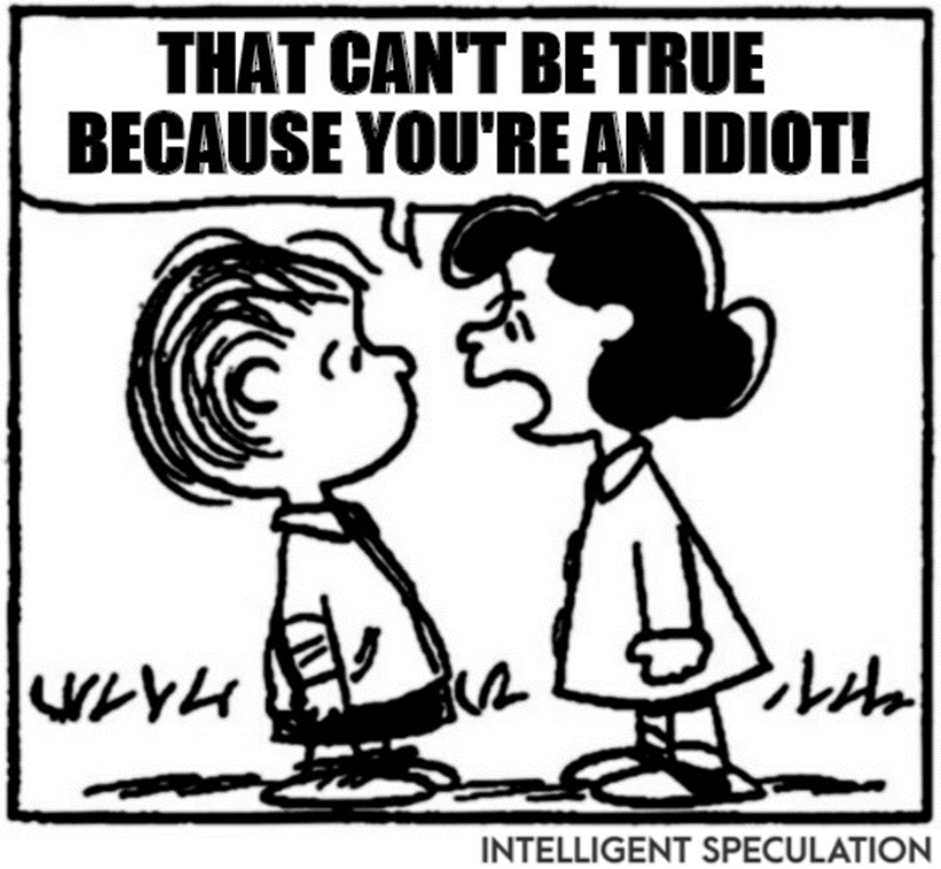
Guess the Fallacy



Ad Hominem

Image source: https://images.squarespace-cdn.com/content/v1/5c8facb1fb22a57f65bacf1d/1557026252285-HHCAH5B0KM5KKILQ5MVY/Ad+Hominem.png

Guess the Fallacy

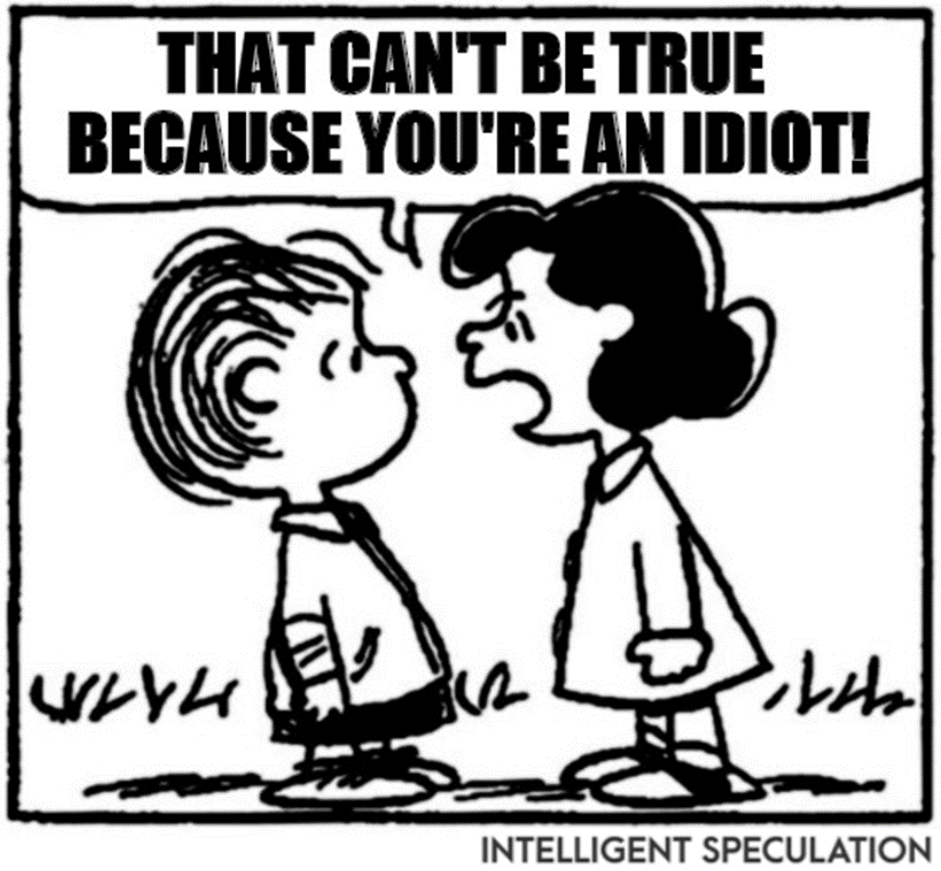


Ad Hominem

Definition: Speaker trying to argue the opposing view on a topic makes claims against the other speaker instead of the position they are maintaining.

Image source: https://images.squarespace-cdn.com/content/v1/5c8facb1fb22a57f65bacf1d/1557026252285-HHCAH5B0KM5KKILQ5MVY/Ad+Hominem.png

Guess the Fallacy



Ad Hominem

Definition: Speaker trying to argue the opposing view on a topic makes claims against the other speaker instead of the position they are maintaining.

Example: He supports Democrats. Hence, don't believe what he says.

Image source: https://images.squarespace-cdn.com/content/v1/5c8facb1fb22a57f65bacf1d/1557026252285-HHCAH5B0KM5KKILQ5MVY/Ad+Hominem.png

Experiments and Results

Experiment 1

Logical Fallacy Detection in CreateDebate Forum

CreateDebate Forum

- Social networking website hosting debates since February 2008.

CreateDebate Forum

- Social networking website hosting debates since February 2008.
- Divided into 14 topical forums – Politics, Religion, Science, etc.

CreateDebate Forum

- Social networking website hosting debates since February 2008.
- Divided into 14 topical forums – Politics, Religion, Science, etc.
- Weak moderation policy
  - Only the debate creator can moderate

CreateDebate Forum

- Social networking website hosting debates since February 2008.
- Divided into 14 topical forums – Politics, Religion, Science, etc.
- Weak moderation policy
  - Only the debate creator can moderate
- Perfect testbed for analyzing usage of informal logical fallacies

CreateDebate Forum

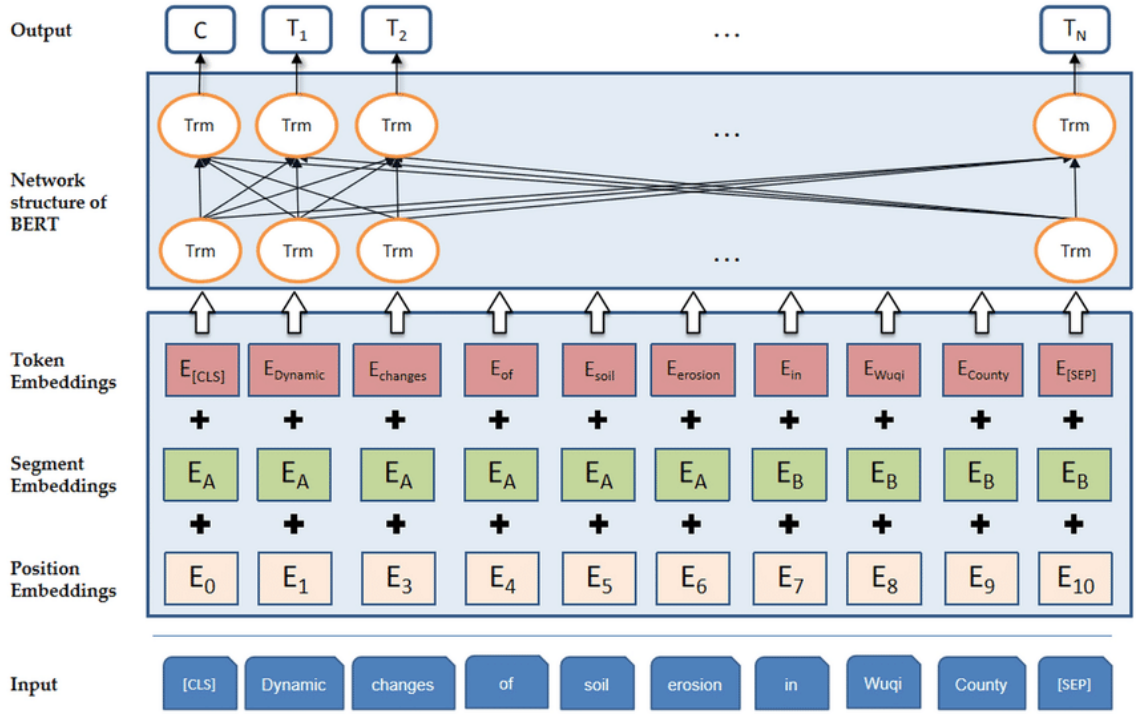
Topic	# Posts	# Comments	# Users
Politics	10,434	119,850	7,686
Religion	2,841	77,418	4,563
World News	2,008	27,418	3,622
Science	1,276	20,691	2,837
Law	759	11,016	1,436
Technology	909	8,421	2,674
Total	18,227	264,814	14,961

TABLE 3.1: Basic statistics of our collected CreateDebate dataset. The first post in our dataset was posted on February 20, 2008 and the last post was updated on November 24, 2021 (Patel et al., 2023).

"Dummy Grandpa, do you know anything?": Identifying and Characterizing Ad hominem Fallacy Usage in the Wild – Patel et al., 2023 - https://arxiv.org/abs/2209.02062

Fine-tuning BERT

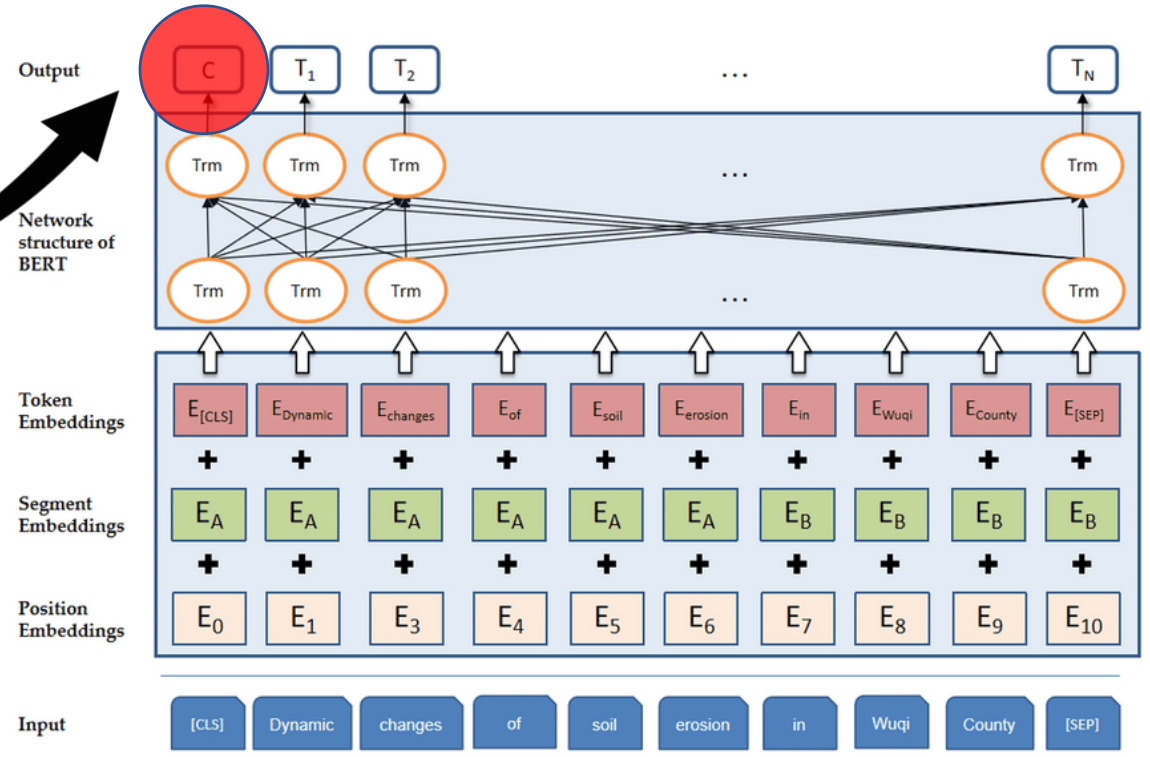
- We are using BERT [base, uncased] model



Fine-tuning BERT

- We are using BERT [base, uncased] model

Use this vector for classification



Fine-tuning BERT

- We are using BERT [base, uncased] model
- Fine-tune the model on LOGIC dataset

LOGIC dataset

- Dataset published in (Jin et al., 2022).

LOGIC dataset

- Dataset published in (Jin et al., 2022).
- Contains 2,449 examples of logical fallacies

LOGIC dataset

- Dataset published in (Jin et al., 2022).
- Contains 2,449 examples of logical fallacies
  - Training partition: 1849
  - Validation partition: 300
  - Test partition: 300

LOGIC dataset

- Dataset published in (Jin et al., 2022).
- Contains 2,449 examples of logical fallacies
  - Training partition: 1849
  - Validation partition: 300
  - Test partition: 300
- Examples categorized in 13 classes

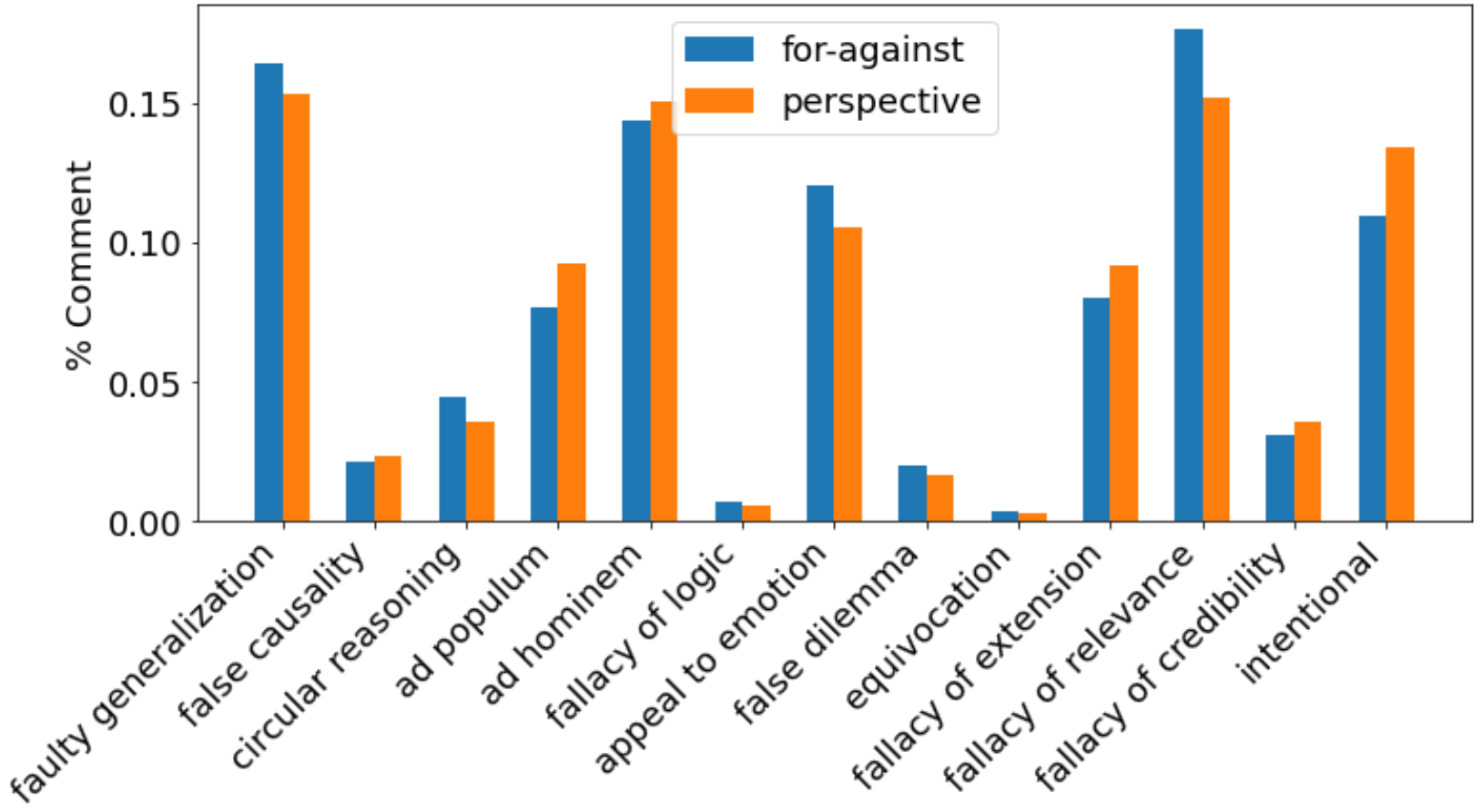
Class Labels in LOGIC dataset

Class Label	Percentage (%)
Faulty Generalization	18.01
Ad Hominem	12.33
Ad Populum	9.47
False Causality	8.82
Circular Claim	6.98
Appeal to Emotion	6.82
Fallacy of Relevance	6.61
Deductive Fallacy	6.21
Intentional Fallacy	5.84
Fallacy of Extension	5.76
False Dilemma	5.76
Fallacy of Credibility	5.39
Equivocation	2.00

Fine-tuning BERT

- We are using BERT [base, uncased] model
- Fine-tune the model on LOGIC dataset
- Use the fine-tuned model to classify comment in Politics forum of CreateDebate

Distribution of Logical Fallacies in CreateDebate's Politics Forum



Experiments and Results

Experiment 2

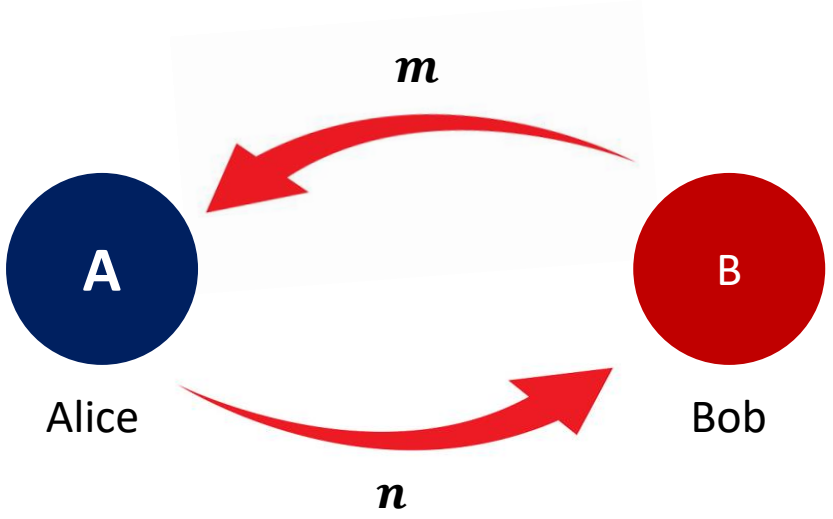
Studies on CreateDebate's User Network

Building CreateDebate's User Network

- User network can be built using two approaches:

Building CreateDebate's User Network

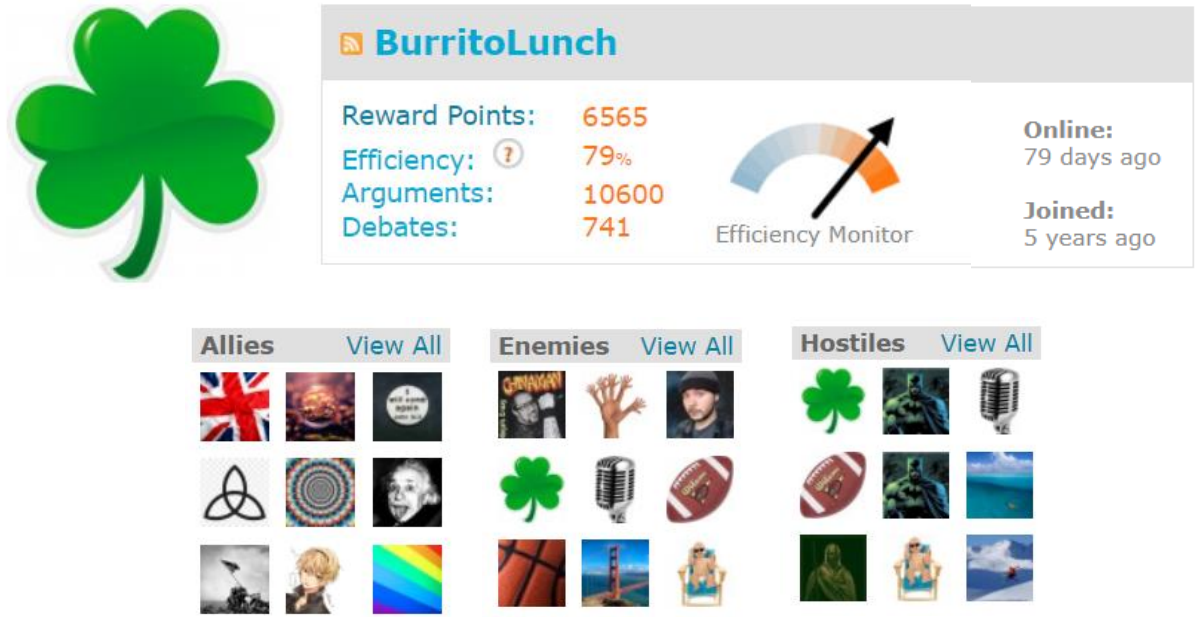
- User network can be built using two approaches:
  - **Nested structure of comments in a post**



Alice directly replied to Bob n times
Bob directly replied to Alice m times

Building CreateDebate's User Network

- User network can be built using two approaches:
  - **Profile pages of users**



For-against and Perspective Debates

- CreateDebate hosts two types of debates: For-against and Perspective.

For-against and Perspective Debates

- CreateDebate hosts two types of debates: For-against and Perspective.
- **For-against debates**
  - Users take a stand on a particular topic and argue in favor or against it

For-against and Perspective Debates

- CreateDebate hosts two types of debates: For-against and Perspective.
- **For-against debates**
  - Users take a stand on a particular topic and argue in favor or against it
- **Perspective debates**
  - Users share their viewpoints on a topic without necessarily taking a stance for or against it.

For-against and Perspective Debates

Topic	For-Against Posts %	Perspective Posts %	For-Against Comments %	Perspective Comments %
Politics	48%	52%	68%	32%
Religion	60%	40%	76%	24%
World News	58%	42%	73%	27%
Science	55%	45%	74%	26%
Law	66%	34%	82%	18%
Technology	54%	46%	79%	21%

TABLE 3.2: Distribution of For-Against debates and Perspective debates across different forums.

Building CreateDebate's User Network

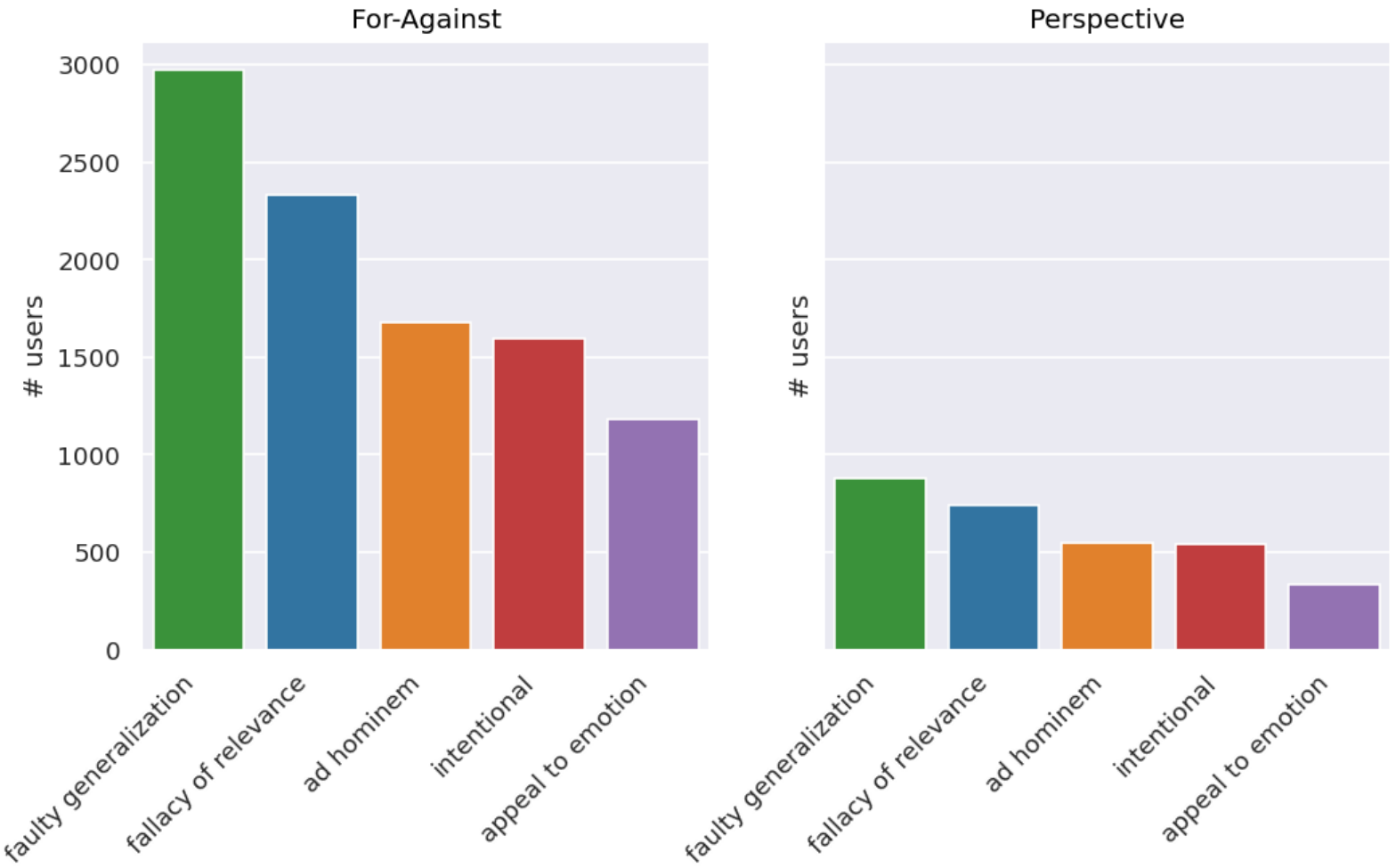
- We use both types of networks for our study

Building CreateDebate's User Network

- We use both types of networks for our study
- As CreateDebate hosts two types of debates
  - **For-against** user network
  - **Perspective** user network

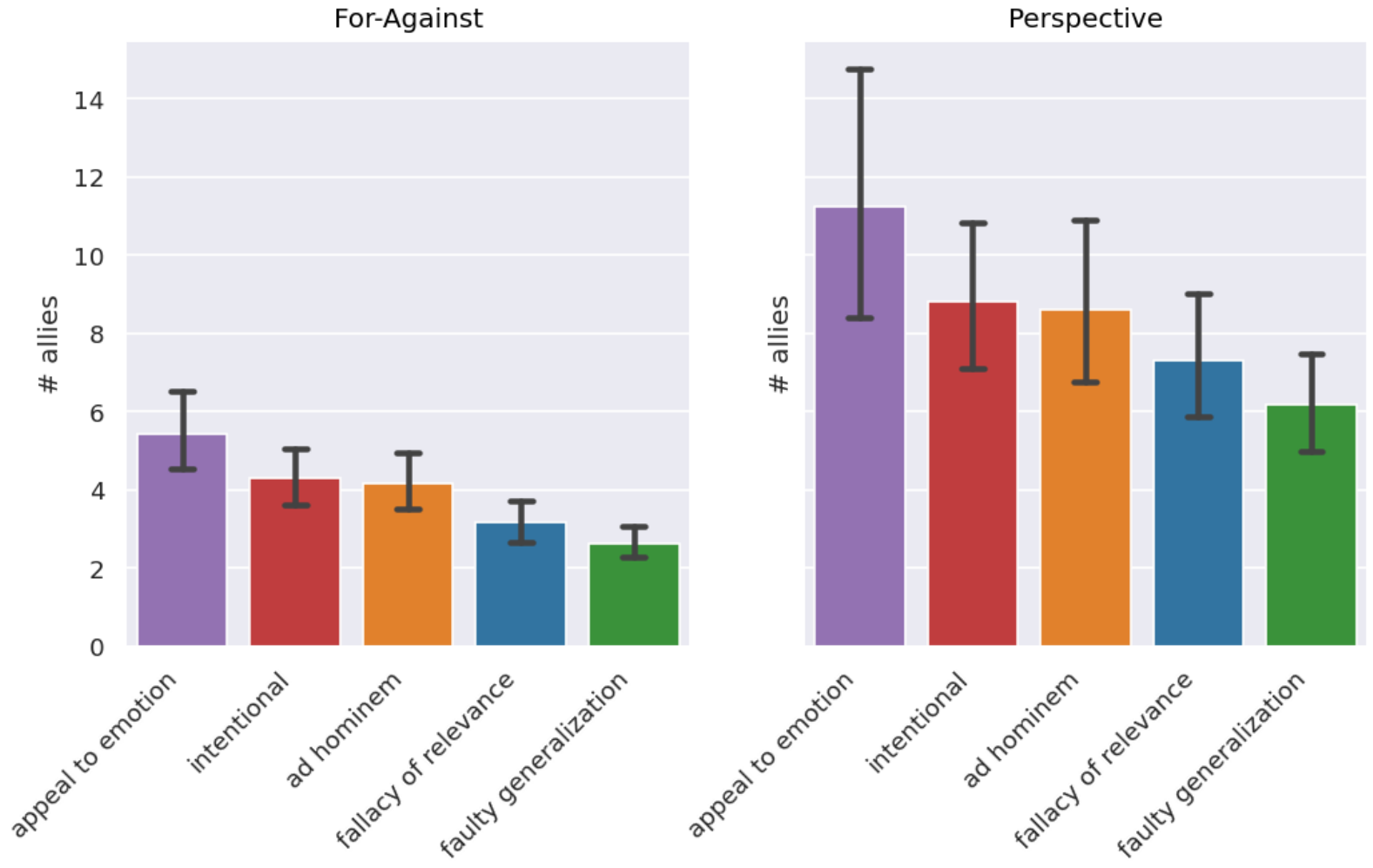
Number of users posting fallacious comments

Order is preserved **3x decrease**



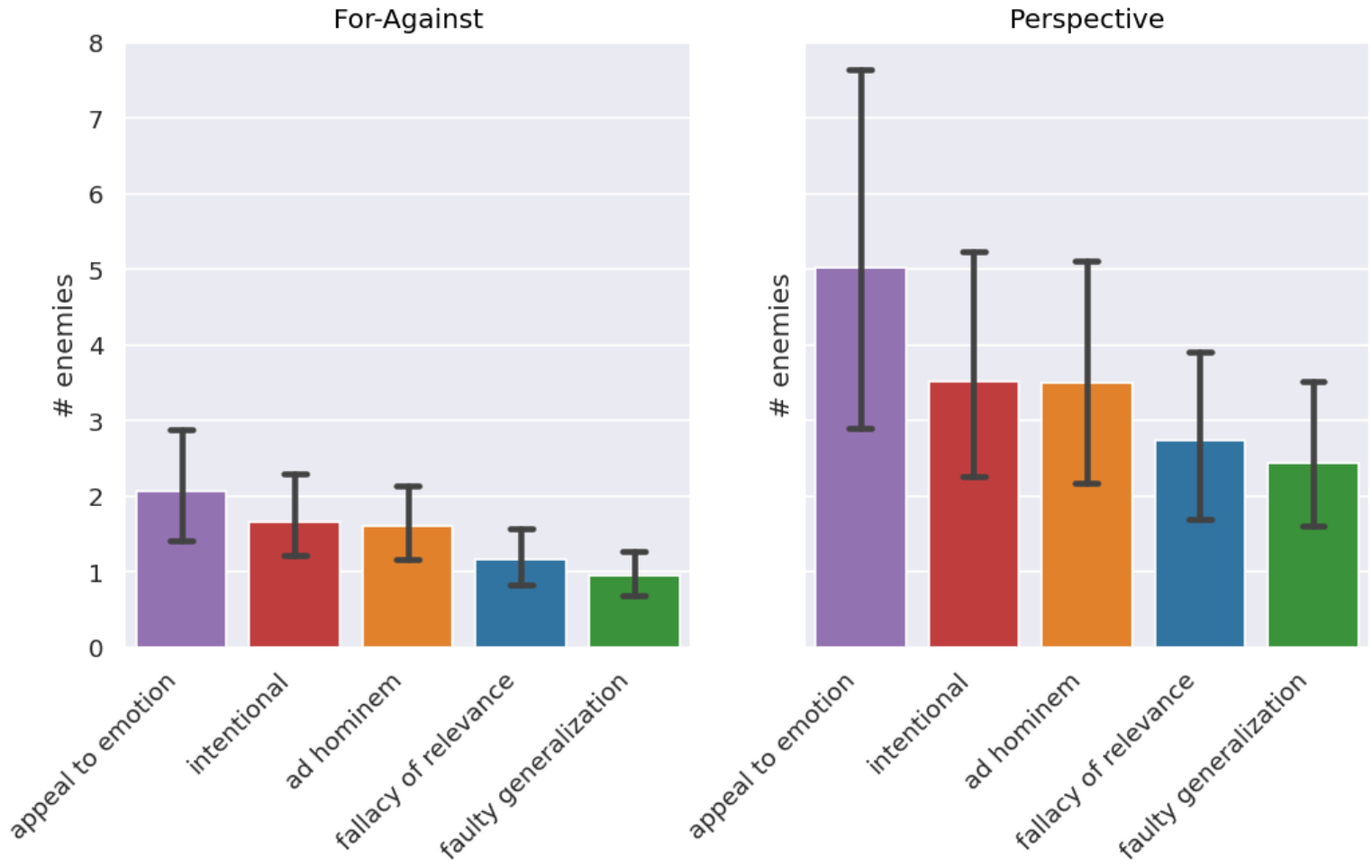
Number of Allies

Order is preserved **2x increase**



Number of Enemies

Order is preserved **2.5x increase**

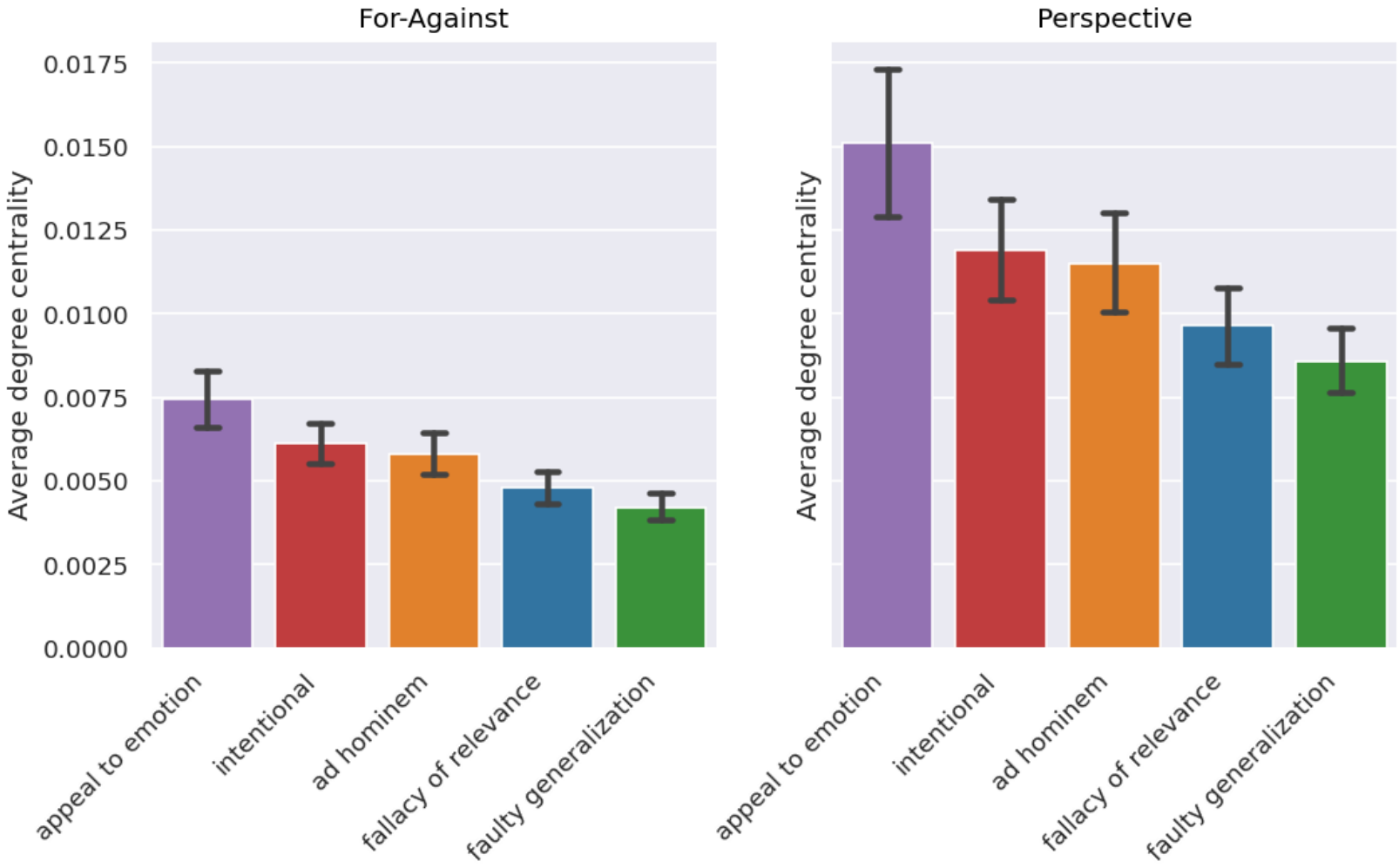


Degree centrality

The degree centrality values are normalized by dividing by the maximum possible degree in a simple graph n-1 where n is the number of nodes in G.

Order is preserved

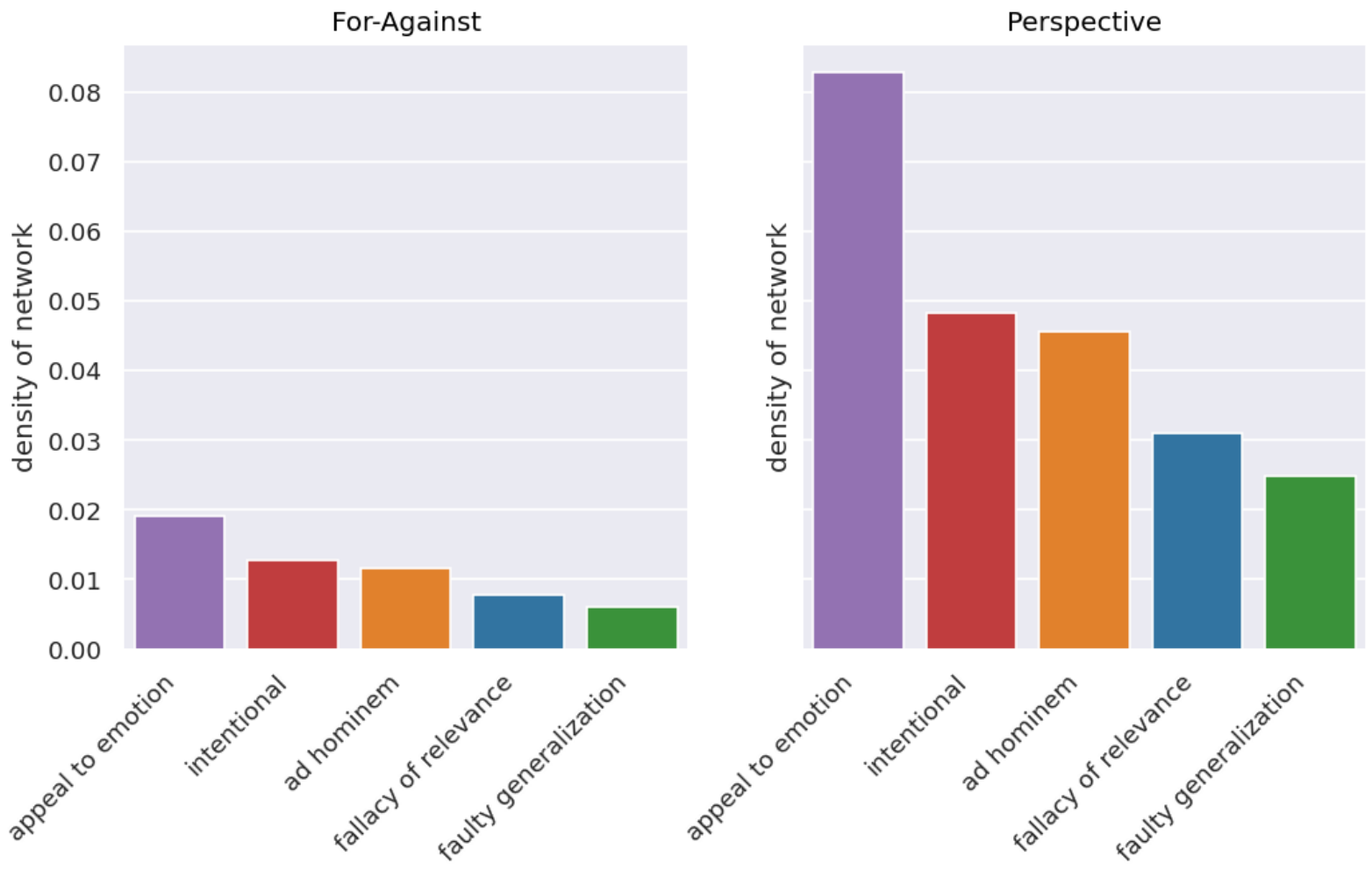
2x increase



Density of network

$$d = \frac{m}{n(n-1)},$$

Order is preserved **4x increase**



Experiments and Results

Experiment 3

Linguistic Study of Logical Fallacies in CreateDebate

Dependency Parsing

- Analyzes grammatical structure of a sentence

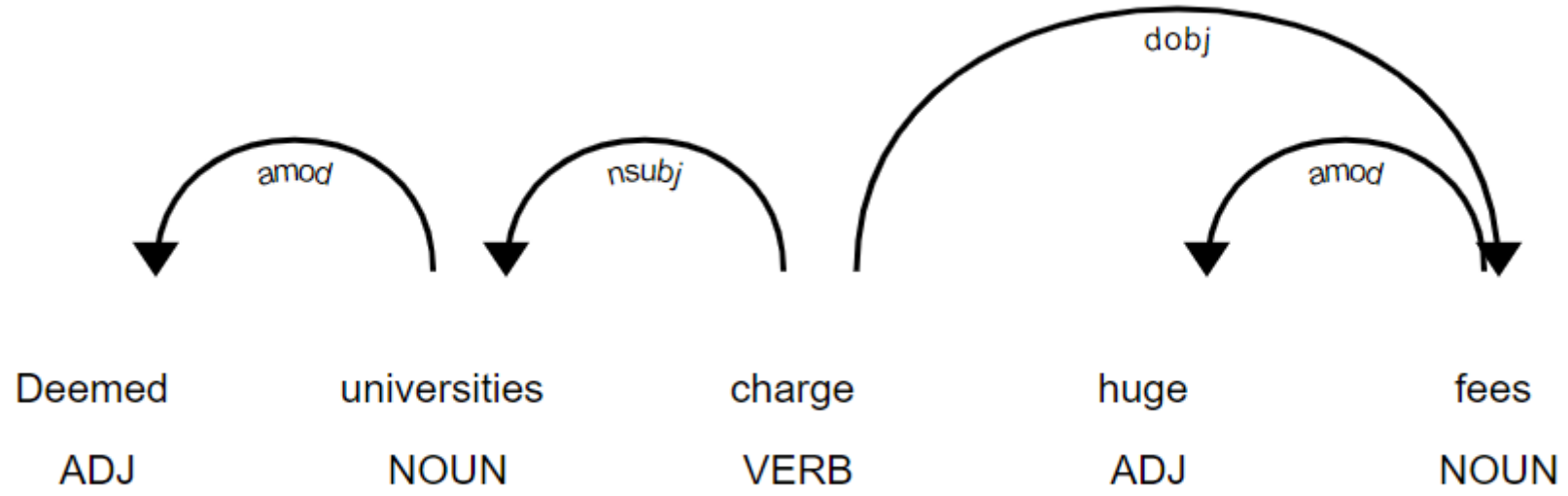
Dependency Parsing

- Analyzes grammatical structure of a sentence
- Determines relationships between words

Dependency Parsing

- Analyzes grammatical structure of a sentence
- Determines relationships between words
- Dependency tree is built for every comment in Politics forum

Dependency Parsing



Empath Dictionary

- Published in (Fast et al., 2016).

Empath Dictionary

- Published in (Fast et al., 2016).
- Tool for analyzing text across lexical categories

Empath Dictionary

- Published in (Fast et al., 2016).
- Tool for analyzing text across lexical categories
- Has 194 built-in pre-validated topical and emotional categories

Empath Dictionary

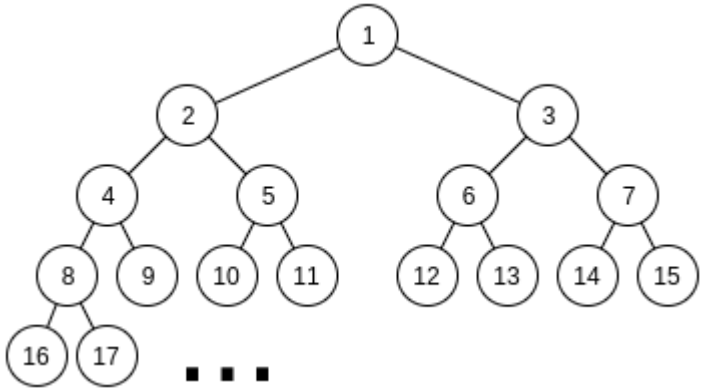
- Published in (Fast et al., 2016).
- Tool for analyzing text across lexical categories
- Has 194 built-in pre-validated topical and emotional categories
- Each category has on average 83 words.

Empath Dictionary

social media	war	violence	technology
facebook	attack	hurt	ipad
instagram	battlefield	break	internet
notification	soldier	bleed	download
selfie	troop	broken	wireless
account	army	scar	computer
timeline	enemy	hurting	email
follower	civilian	injury	virus

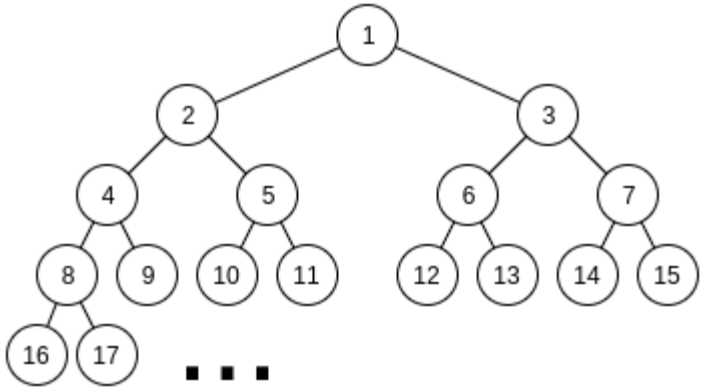
Creating Dependency Paths

- For a given dependency tree T



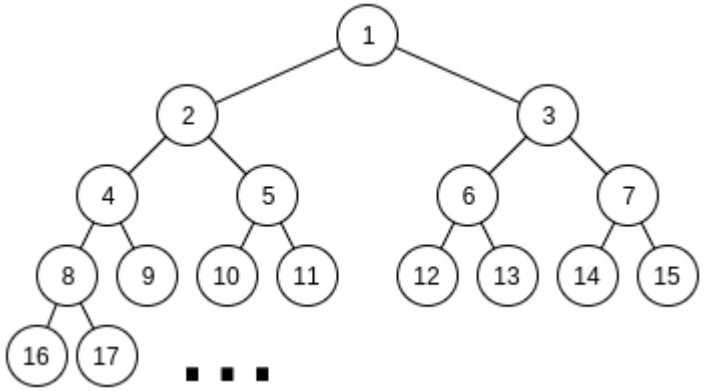
Creating Dependency Paths

- For a given dependency tree T
  - Let N denotes set of nodes which are nouns / pronouns



Creating Dependency Paths

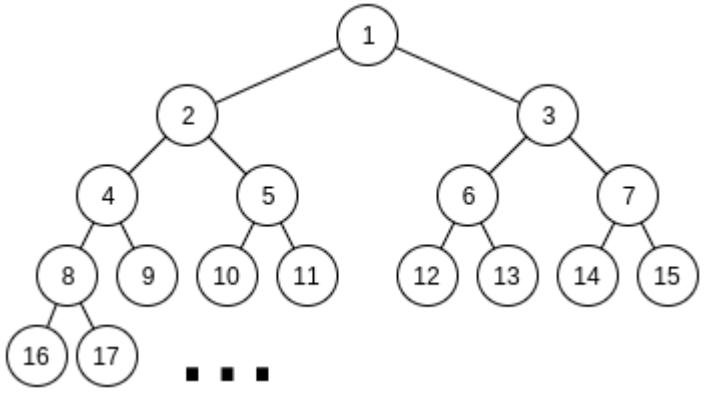
- For a given dependency tree T
  - Let N denotes set of nodes which are nouns / pronouns
  - Let E denotes set of nodes which belong to Empath dictionary



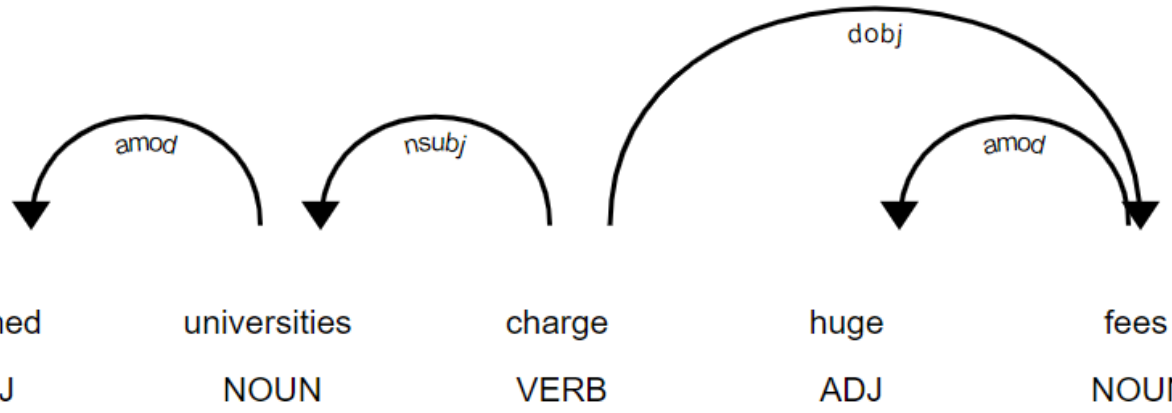
Creating Dependency Paths

- For a given dependency tree T
  - Let N denotes set of nodes which are nouns / pronouns
  - Let E denotes set of nodes which belong to Empath dictionary

  - For $\forall n \in N, e \in E$, if there exists a path $n \rightarrow e$, we call it a dependency path



Creating Dependency Paths



$\mathcal{N}$	$\mathcal{E}$	**Dependency Path**
fees	*huge*	$fees \xleftarrow{amod} huge$
universities	*huge*	$universities \xrightarrow{nsubj} charge \xleftarrow{dobj} fees \xleftarrow{amod} huge$

Identifying Most Discriminative Path for a Logical Fallacy

- From dependency tree, we created every feasible dependency path

Identifying Most Discriminative Path for a Logical Fallacy

- From dependency tree, we created every feasible dependency path
- Not all dependency paths are important

Identifying Most Discriminative Path for a Logical Fallacy

- From dependency tree, we created every feasible dependency path
- Not all dependency paths are important
- Need a way to select dependencies which are highly specific to a given fallacy

Scoring Dependency Paths

We take inspiration from KL divergence

Scoring Dependency Paths

We take inspiration from KL divergence

$$KL(P||Q) = \sum_x s(x) = \sum_x P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

Higher (more positive) the value of $s(x)$, more discriminative x is for distribution P when compared to Q

Scoring Dependency Paths

- Let there be n logical fallacy classes, and x denote dependency paths

Scoring Dependency Paths

- Let there be n logical fallacy classes, and x denote dependency paths
- Let P_i be probability distribution of dependency path of class C_i

Scoring Dependency Paths

- Let there be n logical fallacy classes, and x denote dependency paths
- Let P_i be probability distribution of dependency path of class C_i
- We define a universal probability distribution U given as

$$\mathcal{U}(x) = \frac{1}{n} \sum_i \mathcal{P}_i(x)$$

Scoring Dependency Paths

- Let there be n logical fallacy classes, and x denote dependency paths
- Let P_i be probability distribution of dependency path of class C_i
- We define a universal probability distribution U given as

$$\mathcal{U}(x) = \frac{1}{n} \sum_i \mathcal{P}_i(x)$$

- Then, score of path x in class C_i is given as

$$\mathcal{S}_i(x) = \mathcal{P}_i(x) \log \left(\frac{\mathcal{P}_i(x)}{\mathcal{U}(x)} \right)$$

Fallacy of Relevance

Fallacy of Relevance	It is claimed that $\mathcal{A}$ implies $\mathcal{B}$, whereas $\mathcal{A}$ is unrelated to $\mathcal{B}$.	1. *nominal subject* $\rightarrow$ *clausal complement* $\rightarrow$ *nominal subject* 2. *direct object* $\rightarrow$ *nominal subject* 3. *direct object* $\rightarrow$ *open clausal complement* $\rightarrow$ *nominal subject* 4. *object of preposition* $\rightarrow$ *prepositional modifier* $\rightarrow$ *clausal complement* $\rightarrow$ *nominal subject* 5. *nominal subject* $\rightarrow$ *clausal complement* $\rightarrow$ *clausal complement* $\rightarrow$ *nominal subject*

Ad Hominem

Ad Hominem	$\mathcal{A}$ is claiming $\mathcal{B}$. $\mathcal{A}$ is a moron. Therefore, $\mathcal{B}$ is not true.	1. *adjectival modifier* → *attribute* → *nominal subject* 2. *attribute* → *nominal subject* 3. *adjectival complement* → *nominal subject* 4. *adjectival modifier* → *direct object* →*nominal subject* 5. *attribute* → *clausal complement* → *nominal subject*

Conclusion

- Logical fallacies refer to flawed patterns of reasoning

Conclusion

- Logical fallacies refer to flawed patterns of reasoning
- Fallacious users participating in Perspective debates are less when compared to For-against debates
  - They have more adversaries
  - They have more allies

Conclusion

- Logical fallacies refer to flawed patterns of reasoning
- Fallacious users participating in Perspective debates are less when compared to For-against debates
  - They have more adversaries
  - They have more allies
- Proposed a novel approach for detecting logical fallacies

Future Works

- Our study proved that the linguistic structure of texts can be used as an alternative approach in identifying logical fallacy

Future Works

- Our study proved that the linguistic structure of texts can be used as an alternative approach in identifying logical fallacy
- These linguistic patterns can be used to create a static set of rules for identifying logical fallacy

Future Works

- Our study proved that the linguistic structure of texts can be used as an alternative approach in identifying logical fallacy
- These linguistic patterns can be used to create a static set of rules for identifying logical fallacy
- Future works can use these patterns and embed them in classification models for logical fallacy detection

thank you!