



Enhancing Fake News Classification Through an Argumentation-Based Pipeline

BNAIC 2023: BNAIC/BeNeLearn Joint International BeNeLux Conferences on AI and Machine Learning
9 November 2023, Delft

Mayla Kersten, Giovanni Sileno

“Fake” news

- *“News that conveys or incorporates false, fabricated, or deliberately misleading information, or that is characterized as or accused of doing so.”*
(Oxford English Dictionary)



“Fake” news

- *“News that conveys or incorporates false, fabricated, or deliberately misleading information, or that is characterized as or accused of doing so.”*
(Oxford English Dictionary)
- High volume, variety of topics, rapid spread on different platforms



“Fake” news

- *“News that conveys or incorporates false, fabricated, or deliberately misleading information, or that is characterized as or accused of doing so.”*
(Oxford English Dictionary)
- High volume, variety of topics, rapid spread on different platforms
- Fake news impact:
 - Polarization of public opinion
 - Incitement of violence and hate speech
 - Erosion of trust in democratic institutions
 -



“Fake” news

- *“News that conveys or incorporates false, fabricated, or deliberately misleading information, or that is characterized as or accused of doing so.”*
(Oxford English Dictionary)
- High volume, variety of topics, rapid spread on different platforms
- Fake news impact:
 - Polarization of public opinion
 - Incitement of violence and hate speech
 - Erosion of trust in democratic institutions
 -
- Example: “**Pizzagate**”



The New York Times

***Fake News Onslaught Targets Pizzeria
as Nest of Child-Trafficking***



“Fake” news – social responses

- Fact-checking websites



“Fake” news – social responses

- Fact-checking websites
 - Example: Snopes

Claim:

A leaked email proves that President Obama once spent \$65,000 of taxpayer money on hot dogs.

Rating:



False

[About this rating](#) 

“Fake” news – social responses

- Fact-checking websites
 - Snopes
 - PolitiFact

Claim:

A leaked email proves that President Obama once spent \$65,000 of taxpayer money on hot dogs.

Rating:



False

[About this rating](#) 

Donald Trump

stated on April 2, 2019 in a speech to the National Republican Congressional Committee:

"The noise (from windmills) causes cancer."



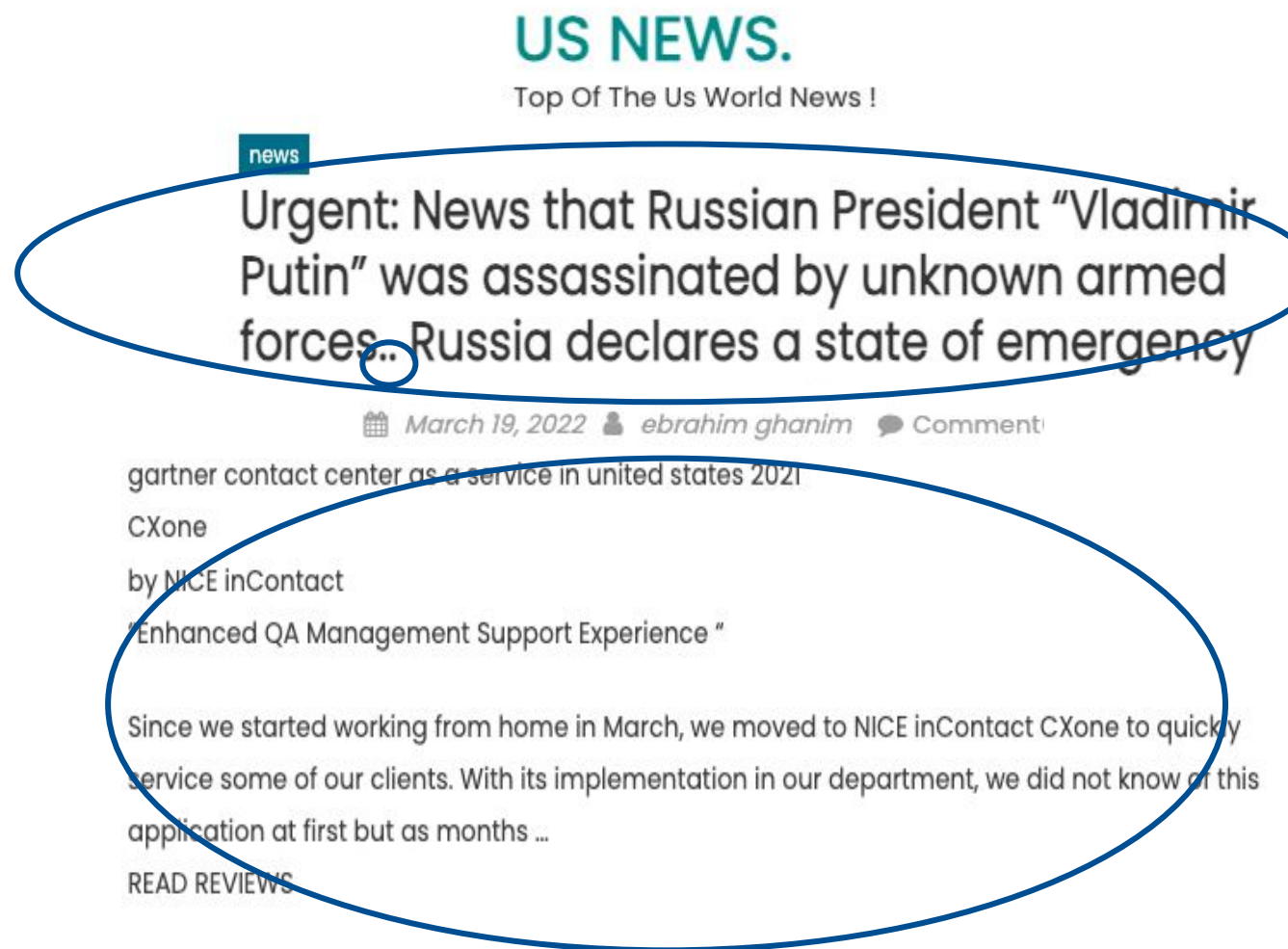


“Fake” news – research

- Automated fake news detection

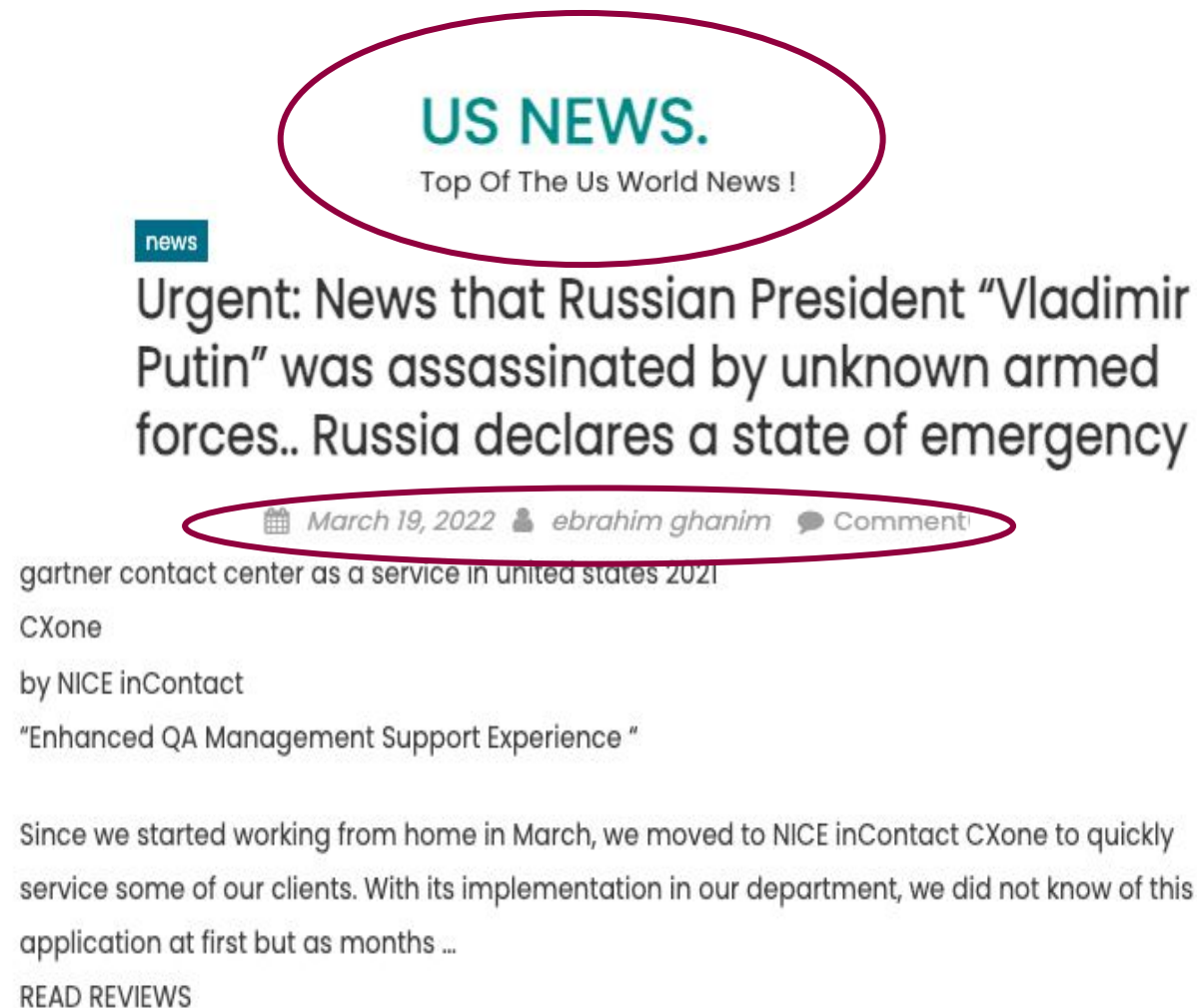
“Fake” news – research

- Automated fake news detection
 - Content-based models
 - E.g., syntactic and linguistic cues



“Fake” news – research

- Automated fake news detection
 - Content-based models
 - E.g., syntactic and linguistic cues
 - Context-based models
 - E.g., publisher, author, user comments



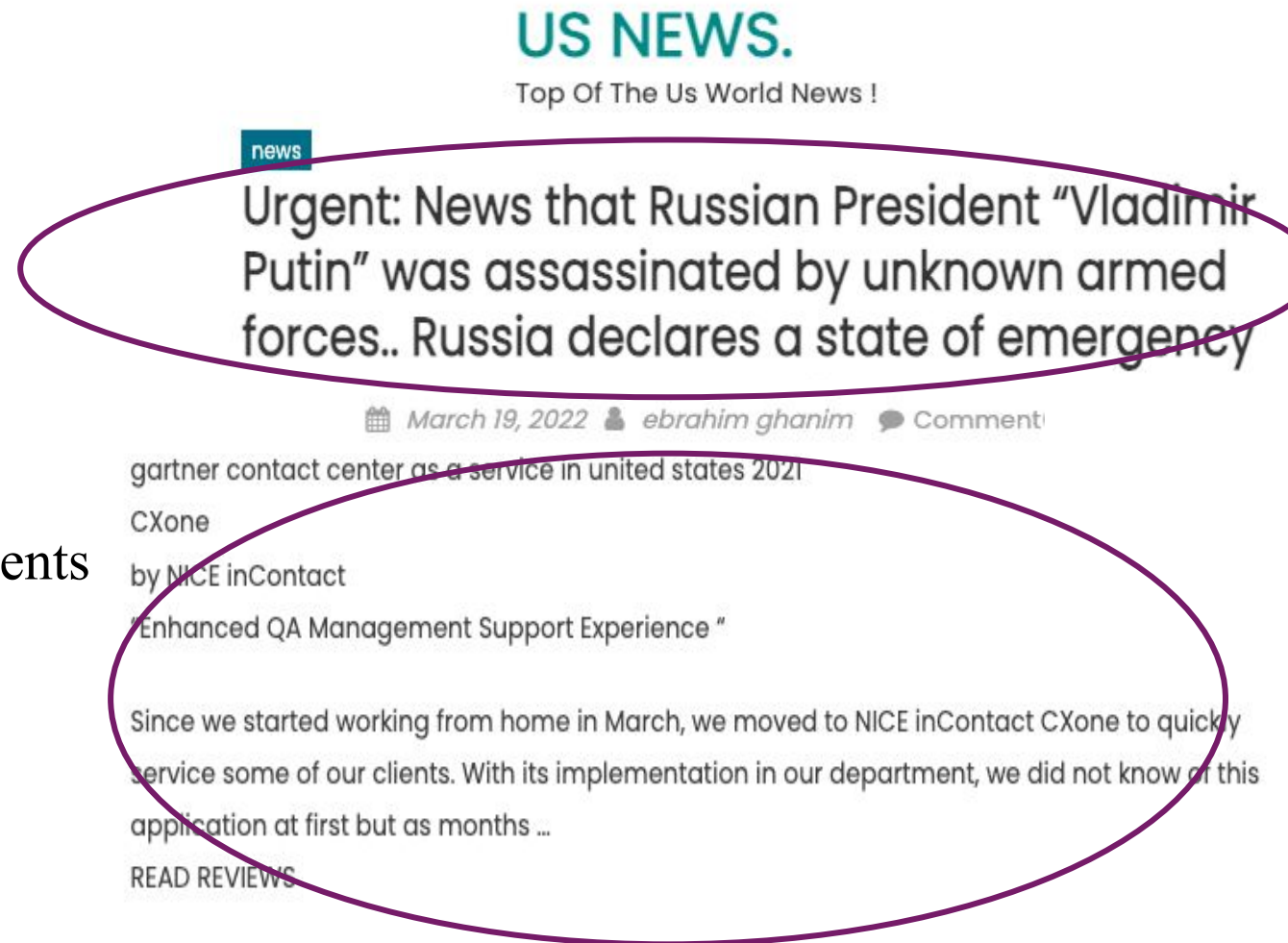
“Fake” news – research

- Automated fake news detection
 - Content-based models
 - E.g., syntactic and linguistic cues
 - Context-based models
 - E.g., publisher, author, user comments
 - Hybrid models



“Fake” news – research

- Automated fake news detection
 - Content-based models
 - E.g., syntactic and linguistic cues
 - Context-based models
 - E.g., publisher, author, user comments
 - Hybrid models
 - Large Language Models (LLMs)
 - Semantic context embeddings





“Fake” news research outcomes

- Benchmark study by Khan et al. (2021)

Machine Learning with Applications 4 (2021) 100032



Contents lists available at [ScienceDirect](#)

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

A benchmark study of machine learning models for online fake news detection

Junaed Younus Khan ^{a,1}, Md. Tawkat Islam Khondaker ^{a,1}, Sadia Afroz ^b, Gias Uddin ^{c,*}, Anindya Iqbal ^a

^a Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

^b International Computer Science Institute, Berkeley, CA, USA

^c Department of Electrical and Computer Engineering, University of Calgary, Alberta, Canada

ARTICLE INFO

Keywords:

Fake news
Fake news detection
Benchmark study
Machine learning

ABSTRACT

The proliferation of fake news and its propagation on social media has become a major concern for governments and the public to create devastating impacts. Different machine learning approaches have been used to detect fake news. However, most of those focused on a specific type of news (such as political) without considering the dataset-bias of the models used. In this research, we conducted a benchmark study to evaluate the performance of machine learning models for online fake news detection.



“Fake” news research outcomes

- Benchmark study by Khan et al. (2021)
 - *Fake or Real News* (George McIntire, 2017)
 - 4,594 news texts
 - *Liar* (William Yang Wang, 2017)
 - 12,790 short statements

Machine Learning with Applications 4 (2021) 100032



Contents lists available at [ScienceDirect](#)

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

A benchmark study of machine learning models for online fake news detection

Junaed Younus Khan ^{a,1}, Md. Tawkat Islam Khondaker ^{a,1}, Sadia Afroz ^b, Gias Uddin ^{c,*}, Anindya Iqbal ^a

^a Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

^b International Computer Science Institute, Berkeley, CA, USA

^c Department of Electrical and Computer Engineering, University of Calgary, Alberta, Canada

ARTICLE INFO

Keywords:

Fake news
Fake news detection
Benchmark study
Machine learning

ABSTRACT

The proliferation of fake news and its propagation on social media has become a serious problem that has the potential to create devastating impacts. Different machine learning approaches have been proposed to detect fake news. However, most of those focused on a specific type of news (such as political) without considering the dataset-bias of the models used. In this research, we conducted a benchmark study to evaluate the performance of machine learning models for online fake news detection.



“Fake” news research outcomes

- Benchmark study by Khan et al. (2021)
 - *Fake or Real News* (George McIntire, 2017)
 - 4,594 news texts
 - **F1-score: 0.98** (RoBERTa)
 - *Liar* (William Yang Wang, 2017)
 - 12,790 short statements
 - **F1-score: 0.62** (RoBERTa)

Machine Learning with Applications 4 (2021) 100032



Contents lists available at [ScienceDirect](#)

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

A benchmark study of machine learning models for online fake news detection

Junaed Younus Khan ^{a,1}, Md. Tawkat Islam Khondaker ^{a,1}, Sadia Afroz ^b, Gias Uddin ^{c,*}, Anindya Iqbal ^a

^a Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

^b International Computer Science Institute, Berkeley, CA, USA

^c Department of Electrical and Computer Engineering, University of Calgary, Alberta, Canada

ARTICLE INFO

Keywords:

Fake news
Fake news detection
Benchmark study
Machine learning

ABSTRACT

The proliferation of fake news and its propagation on social media has become a to create devastating impacts. Different machine learning approaches have been However, most of those focused on a specific type of news (such as political) w dataset-bias of the models used. In this research, we conducted a benchmark s



“Fake” news research outcomes

- Benchmark study by Khan et al. (2021)
 - *Fake or Real News* (George McIntire, 2017)
 - 4,594 news texts
 - **F1-score: 0.98** (RoBERTa)
 - *Liar* (William Yang Wang, 2017)
 - 12,790 short statements
 - **F1-score: 0.62** (RoBERTa)
- Text length may play a role
 - *Fake or Real News*: average 750-800 words
 - *Liar*: average 0-50 words

Machine Learning with Applications 4 (2021) 100032



Contents lists available at [ScienceDirect](#)

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

A benchmark study of machine learning models for online fake news detection

Junaed Younus Khan ^{a,1}, Md. Tawkat Islam Khondaker ^{a,1}, Sadia Afroz ^b, Gias Uddin ^{c,*}, Anindya Iqbal ^a

^a Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

^b International Computer Science Institute, Berkeley, CA, USA

^c Department of Electrical and Computer Engineering, University of Calgary, Alberta, Canada

ARTICLE INFO

Keywords:

Fake news
Fake news detection
Benchmark study
Machine learning

ABSTRACT

The proliferation of fake news and its propagation on social media has become a to create devastating impacts. Different machine learning approaches have been However, most of those focused on a specific type of news (such as political) w dataset-bias of the models used. In this research, we conducted a benchmark s

Intuition: using argumentation components

- Humans use argumentation to assess the validity of information through a process of extracting, summarizing and reasoning of arguments

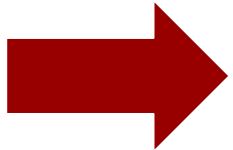


Intuition: using argumentation components

- Humans use argumentation to assess the validity of information through a process of extracting, summarizing and reasoning of arguments

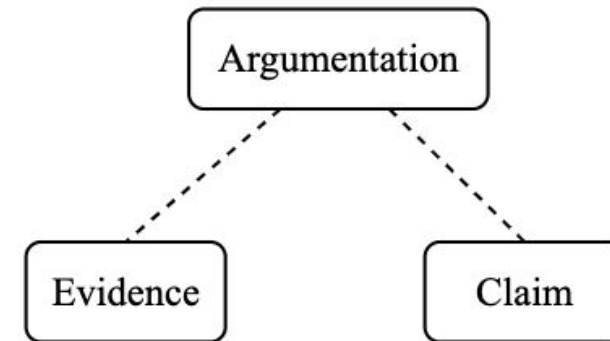


Hypothesis: Extracting argument components could reduce the dependence of a classifier on contingent aspects such as length, or less relevant content



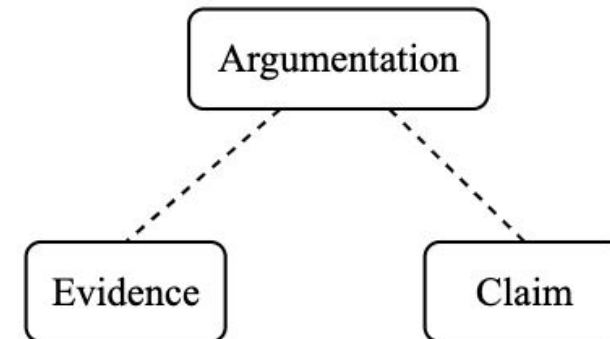
Argumentation structure

- Basic structure
 - **Claim:** implicit/explicit statement to be a standpoint
 - **Evidence:** implicit/explicit backing premises



Argumentation structure

- Basic structure
 - **Claim:** implicit/explicit statement to be a standpoint
 - *“Listening to classical music makes people better scientists.”*
 - **Evidence:** implicit/explicit backing premises
 - *“Many scientists enjoy listening to classical music while they work.”*



Argumentation structure

- Basic structure

- **Claim:** implicit/explicit statement to be a standpoint

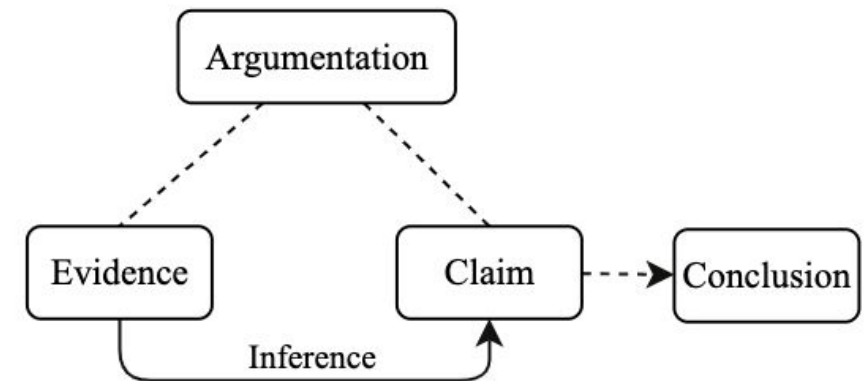
- *“Listening to classical music makes people better scientists.”*

- **Evidence:** implicit/explicit backing premises

- *“Many scientists enjoy listening to classical music while they work.”*

- Valid argumentation

- Logical connection enables inference of the claim from the evidence to a conclusion





Argument component extraction tools

- Little attention to integrating argumentation to fake news detection

Argument component extraction tools

- Little attention to integrating argumentation to fake news detection
- Argument extraction tools
 - Relatively old
 - Not open-source nor made available upon request
 - ...

Argument component extraction tools

- Little attention to integrating argumentation to fake news detection
- Argument extraction tools
 - Relatively old
 - Not open-source nor made available upon request
 - ...

The only one we were able to use is:

- MARGOT (Lippi and Torroni, 2019):
 - Argumentative sentence prediction; Boundary detection



The screenshot shows the MARGOT web interface. At the top, the word "MARGOT" is written in a large, teal, handwritten-style font. Below it, the subtitle "Mining ARGuments from Text" is written in a smaller, teal, handwritten-style font. Underneath the subtitle is a large, empty rectangular box with a thin teal border. Inside the top-left corner of this box, the text "Copy your text here" is written in a small, teal, handwritten-style font. Below the box is a teal button with rounded corners and the text "Find Arguments" in a white, handwritten-style font. At the bottom of the interface, there are four links: "About", "Cite", "Contact", and "Download", all in a teal, handwritten-style font.

Argument component extraction tools

- Little attention to integrating argumentation to fake news detection
- Argument extraction tools
 - Relatively old
 - Not open-source nor made available upon request
 - ...

The only one we were able to use is:

- MARGOT (Lippi and Torroni, 2019):
 - Argumentative sentence prediction; Boundary detection

Alternative: *Generative models*

- Dolly 2.0 (Databricks, 2023)
 - Prompts/instructions

Free Dolly: Introducing the World's First Truly Open Instruction-Tuned LLM



by Mike Conover, Matt Hayes, Ankit Mathur, Jianwei Xie, Jun Wan, Sam Shah, Ali Ghodsi, Patrick Wendell, Matei Zaharia and Reynold Xin

April 12, 2023 in [Company Blog](#)

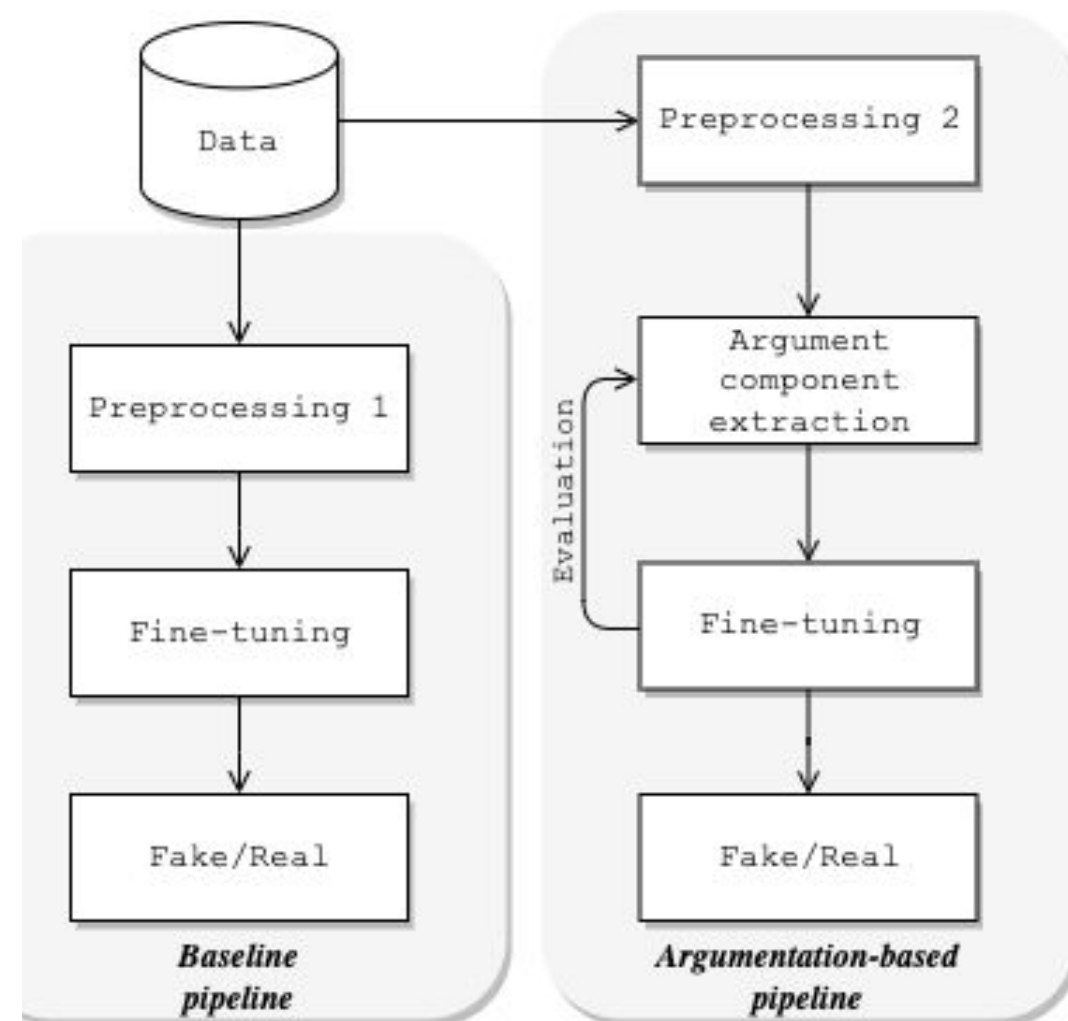
Share this post



Two weeks ago, we released **Dolly**, a large language model (LLM) trained for less than \$30 to exhibit ChatGPT-like human interactivity (aka instruction-following). Today, we're releasing **Dolly 2.0**, the first open source, instruction-following LLM, fine-tuned on a human-generated instruction dataset licensed for **research and commercial use**.

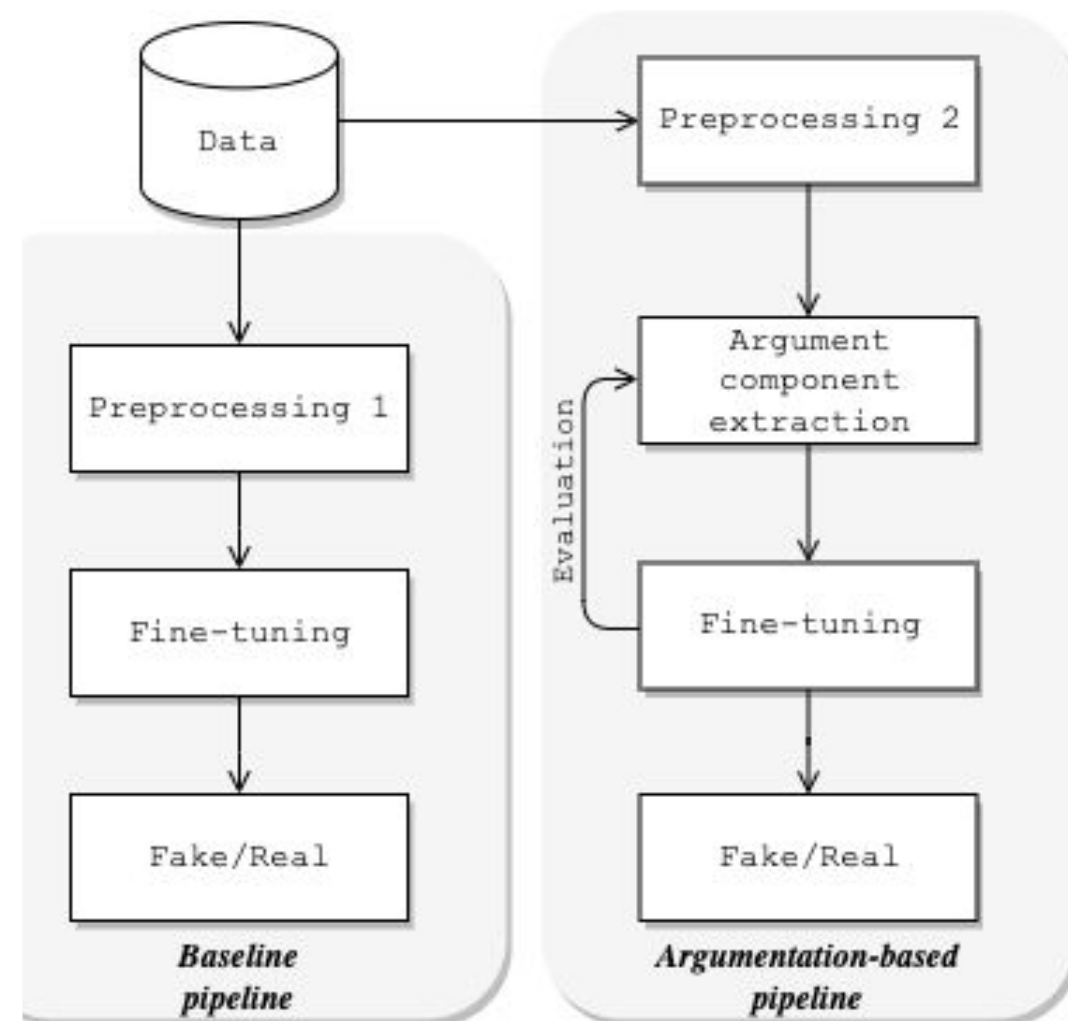
Two processing pipelines

- Dataset
 - Fake or Real News, Fake and Real News, Liar
 - Subsample: 8,831 texts



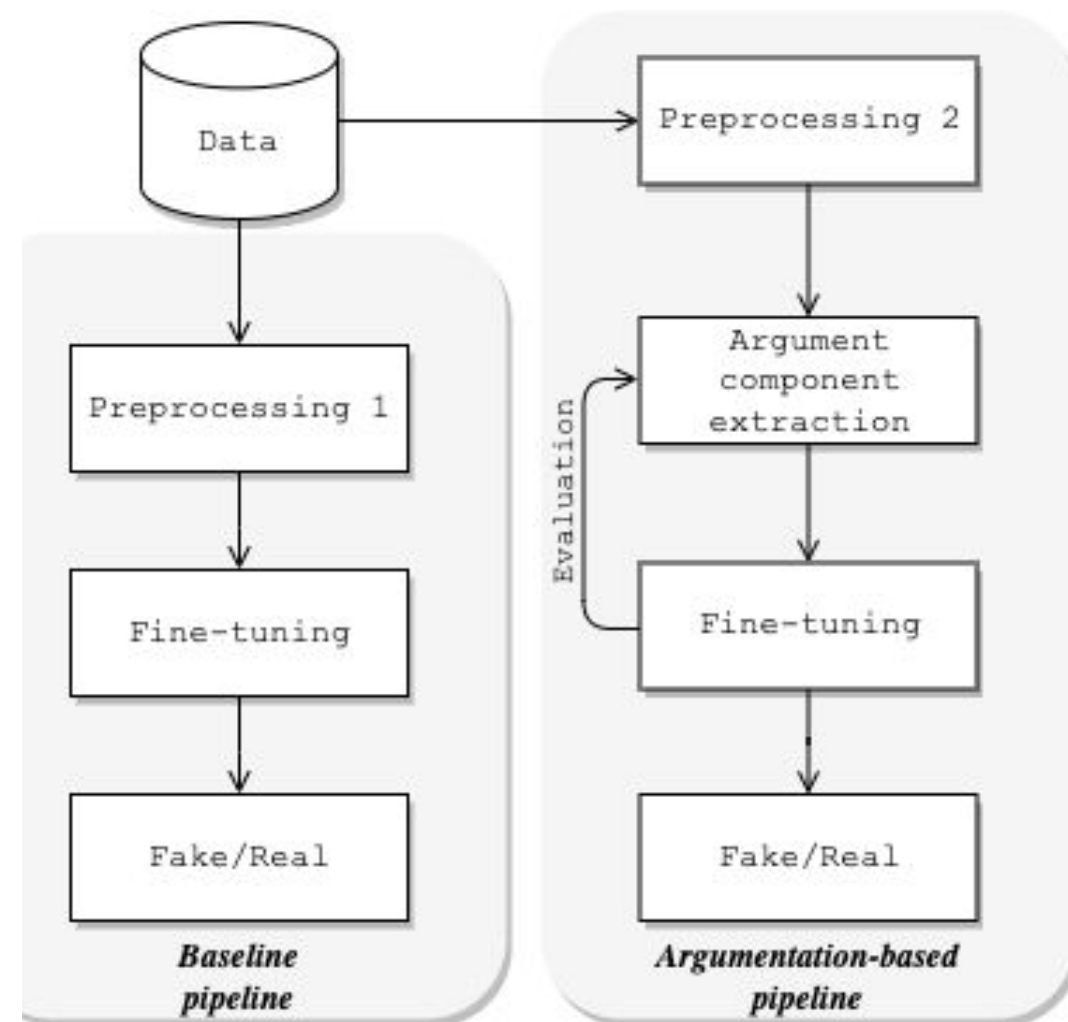
Two processing pipelines

- Dataset
 - Fake or Real News, Fake and Real News, Liar
 - Subsample: 8,831 texts
- Distinct preprocessing methods
 - Baseline pipeline: word-level
 - E.g., stemming, stopword-removal
 - Argumentation-based pipeline: sentence and paragraph level
 - E.g., abbreviated word expansion, removal non-argumentative phrases



Two processing pipelines

- Dataset
 - Fake or Real News, Fake and Real News, Liar
 - Subsample: 8,831 texts
- Distinct preprocessing methods
 - Baseline pipeline: word-level
 - E.g., stemming, stopword-removal
 - Argumentation-based pipeline: sentence and paragraph level
 - E.g., abbreviated word expansion, removal non-argumentative phrases
- Fine-tuning LLMs:
 - BERT, RoBERTa, DistilBERT, ELECTRA



Preliminary manual evaluation of argument extraction tools

		Manual extraction	
		Claim	Evidence
Argument component extraction	Claim	8	1
	Evidence	2	9

MARGOT

		Manual extraction		
		Claim	Evidence	Other
Argument component extraction	Claim	8	1	1
	Evidence	1	8	1
	Other	1	1	

Dolly 2.0

Preliminary manual evaluation of argument extraction tools

		Manual extraction	
		Claim	Evidence
Argument component extraction	Claim	8	1
	Evidence	2	9

		Manual extraction		
		Claim	Evidence	Other
Argument component extraction	Claim	8	1	1
	Evidence	1	8	1
	Other	1	1	

Overall, both solutions seem to work relatively well!

Results

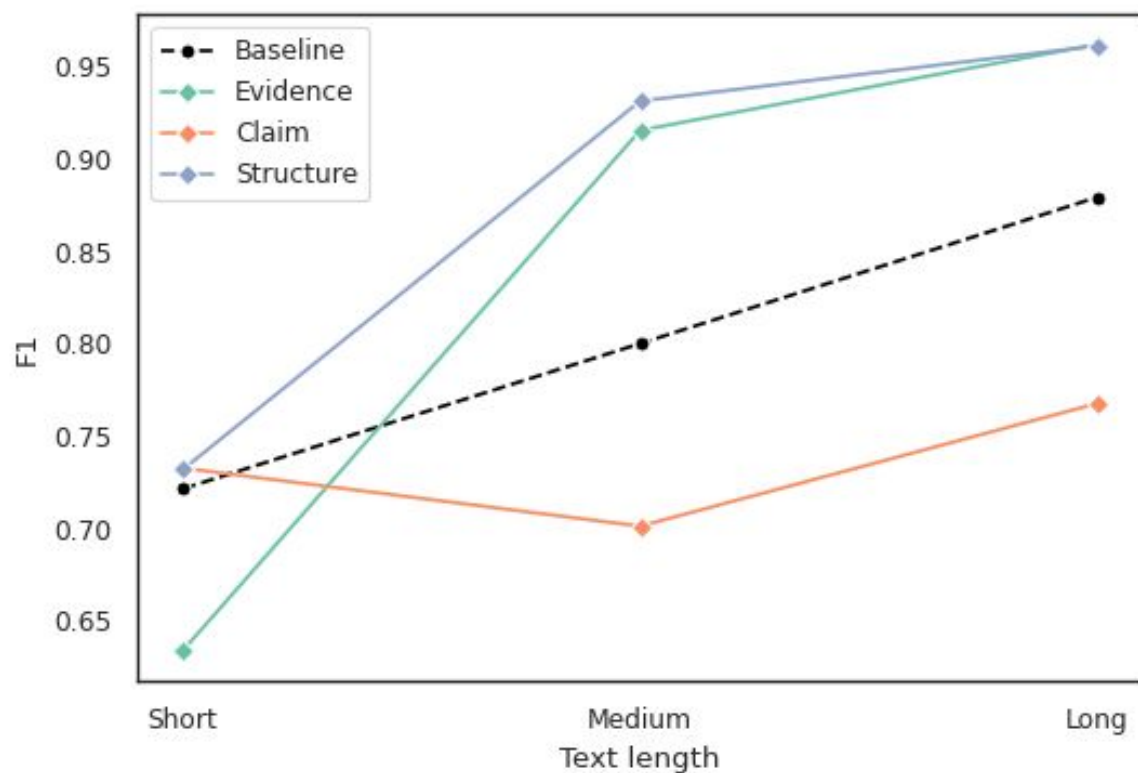
- Baseline pipeline
 - F1: **0.76** (RoBERTa)
- Argumentation based pipeline
 - MARGOT
 - Claim
 - **F1: 0.84 (RoBERTa)**
 - Evidence
 - **F1: 0.84 (RoBERTa, DistilBERT, ELECTRA)**
 - Structure (claim + evidence)
 - **F1: 0.83 (DistilBERT)**
 - Dolly 2.0
 - Claim
 - **F1: 0.75 (DistilBERT)**
 - Evidence
 - **F1: 0.76 (DistilBERT)**
 - Structure (claim + evidence)
 - **F1: 0.72 (DistilBERT, ELECTRA)**

Results

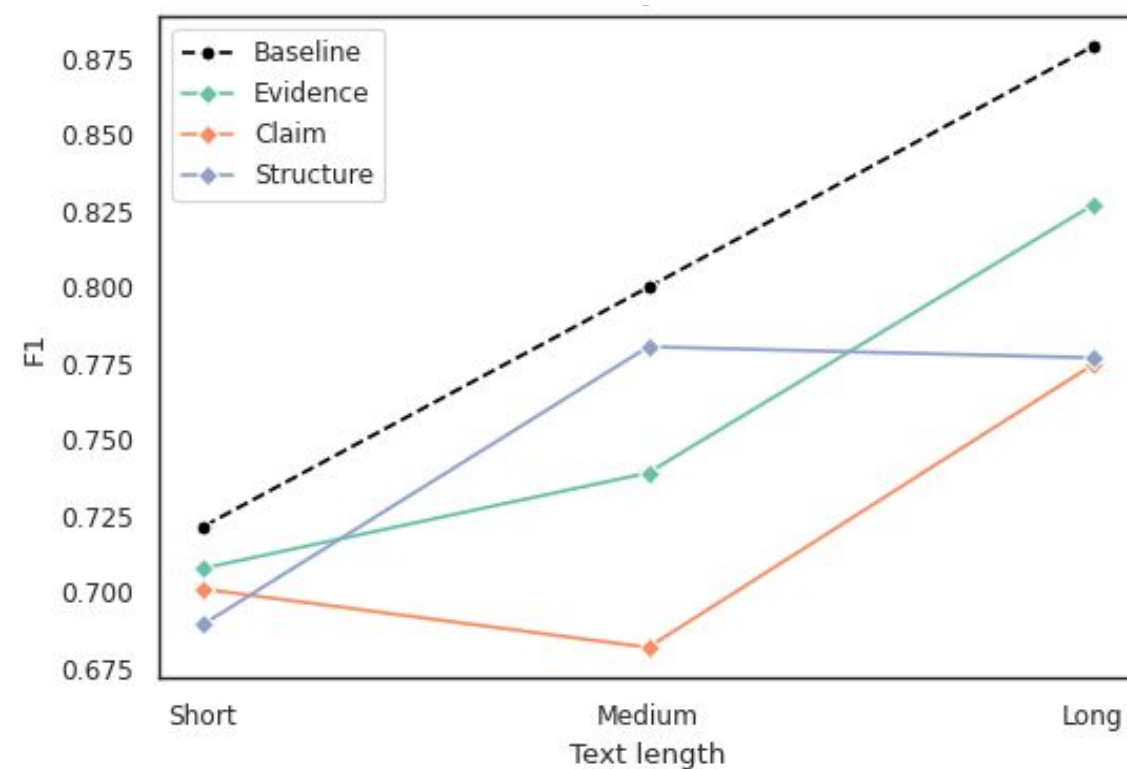
- Baseline pipeline
 - F1: **0.76** (RoBERTa)
- Argumentation based pipeline
 - MARGOT
 - Claim
 - **F1: 0.84** (RoBERTa)
 - Evidence
 - **F1: 0.84** (RoBERTa, DistilBERT, ELECTRA)
 - Structure (claim + evidence)
 - **F1: 0.83** (DistilBERT)
 - Dolly 2.0
 - Claim
 - **F1: 0.75** (DistilBERT)
 - Evidence
 - **F1: 0.76** (DistilBERT)
 - Structure (claim + evidence)
 - **F1: 0.72** (DistilBERT, ELECTRA)

MARGOT always outperforms the baseline, Dolly 2.0 is at best equal to baseline.

Results with respect to text length

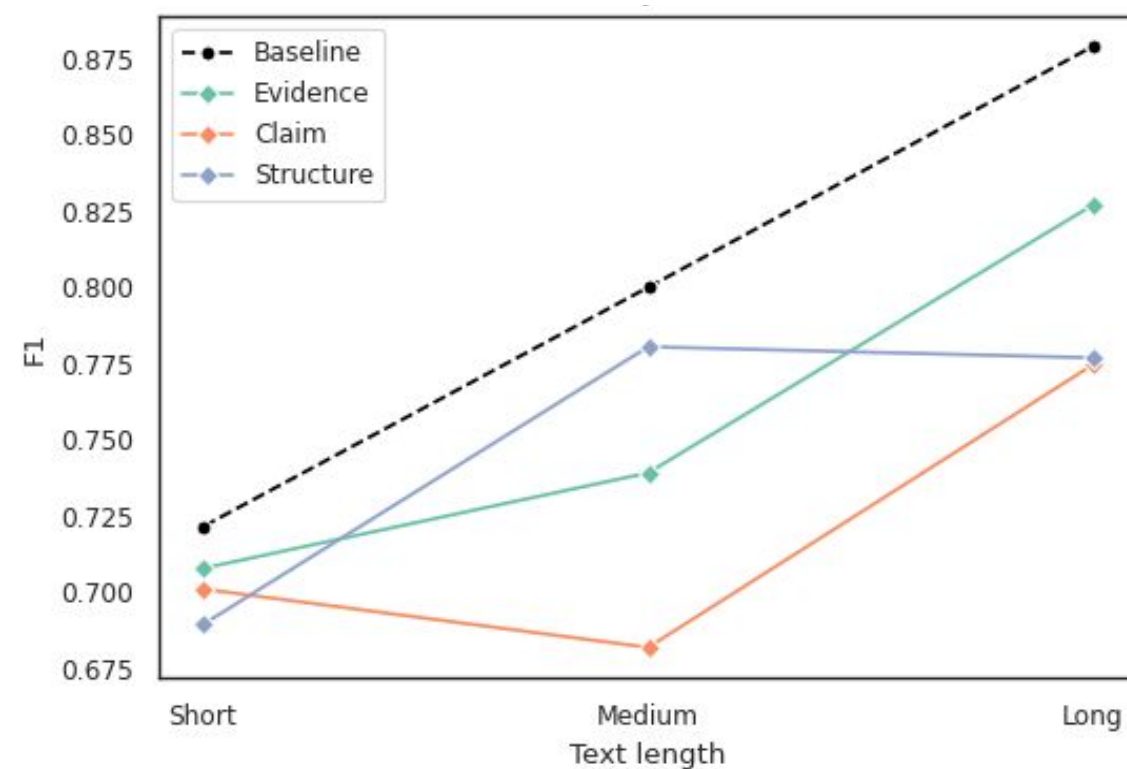
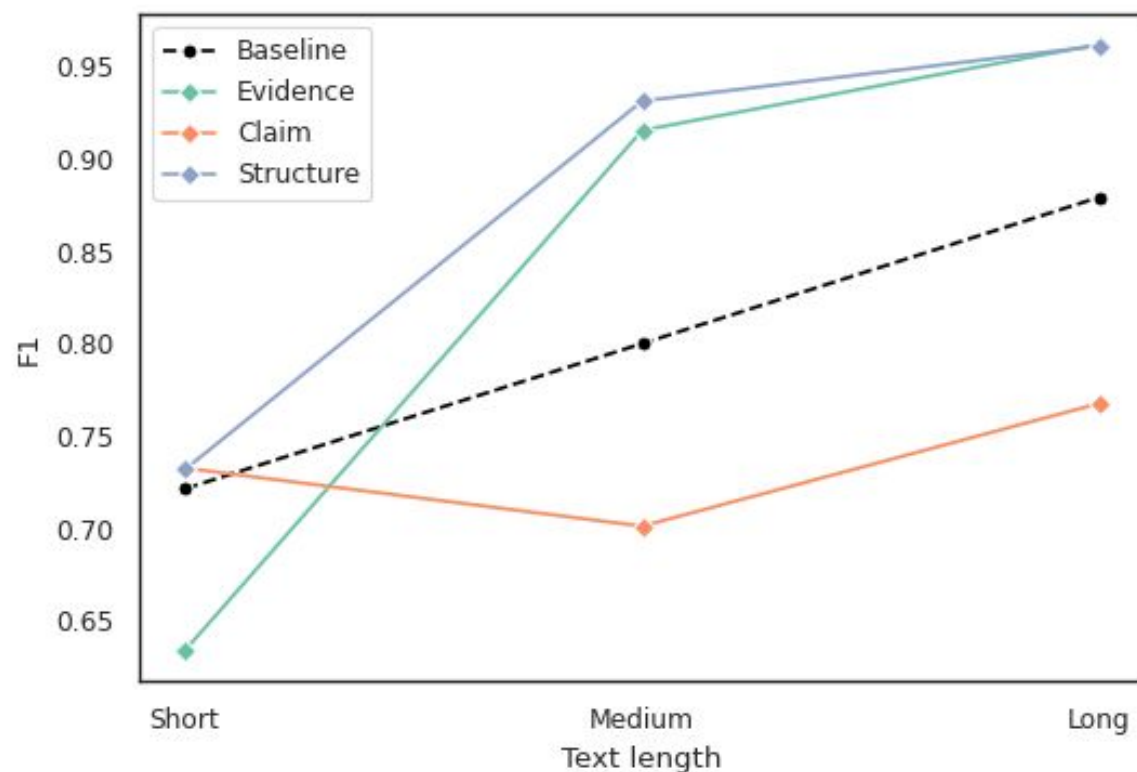


MARGOT (DistilBERT)



Dolly 2.0 (DistilBERT)

Results with respect to text length



Short texts are still problematic, but medium/long can be improved with MARGOT!

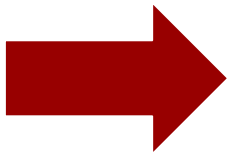


Limitations and conclusions

- Very few available tools for argument component extraction
- Experiments done on subsample of dataset
- Applicability in practice
 - MARGOT: 100/1h (CPU)
 - Dolly 2.0: 150/1h (GPU)

Limitations and conclusions

- Very few available tools for argument component extraction
- Experiments done on subsample of dataset
- Applicability in practice
 - MARGOT: 100/1h (CPU)
 - Dolly 2.0: 150/1h (GPU)



Incorporating argumentative dimension could improve the performance of automated fake news detection, but they need to be **more accessible** and **more efficient**!



Enhancing Fake News Classification Through an Argumentation-Based Pipeline

BNAIC 2023: BNAIC/BeNeLearn Joint International BeNeLux Conferences on AI and Machine Learning
9 November 2023, Delft

Mayla Kersten, Giovanni Sileno



UNIVERSITY OF AMSTERDAM
Informatics Institute

Results with respect to text length

- Best overall model for each component and for each argument component extraction tool
 - Baseline pipeline: RoBERTa
 - Argumentation-based pipeline: MARGOT, Dolly 2.0: DistilBERT
- Categories:
 - Short: 0 - 100 words
 - Medium 100 - 300 words
 - Long: 300 - max