Neural Argument Generation Augmented with Externally Retrieved Evidence

Xinyu Hua and Lu Wang

Northeastern NLP

Project URL: https://xinyuhua.github.io/neural-argument-generation/

Debates and Arguments

Debates and Arguments

Debates and Arguments

Debates and Arguments

Debates and Arguments

UK will be less favorable investment prospect due to loss of EU consumers.

Debates and Arguments

Motivation

- Argumentation is crucial in communication.
  - We want to avoid biased perception and uninformed decisions.

- Persuasion is complicated.
  - Being informative is already non-trivial, not to mention being persuasive.

Research Question

How can we automate human argumentation process?

Our Goal

- We generate a specific type of argument: **counterargument**.

Our Goal

- We generate a specific type of argument: **counterargument**.

  Input: a statement of belief on some controversial topic

  Output: a counterargument refuting the statement

Our Goal

- We generate a specific type of argument: **counterargument**.

Input: Humans are not designed to be vegan.

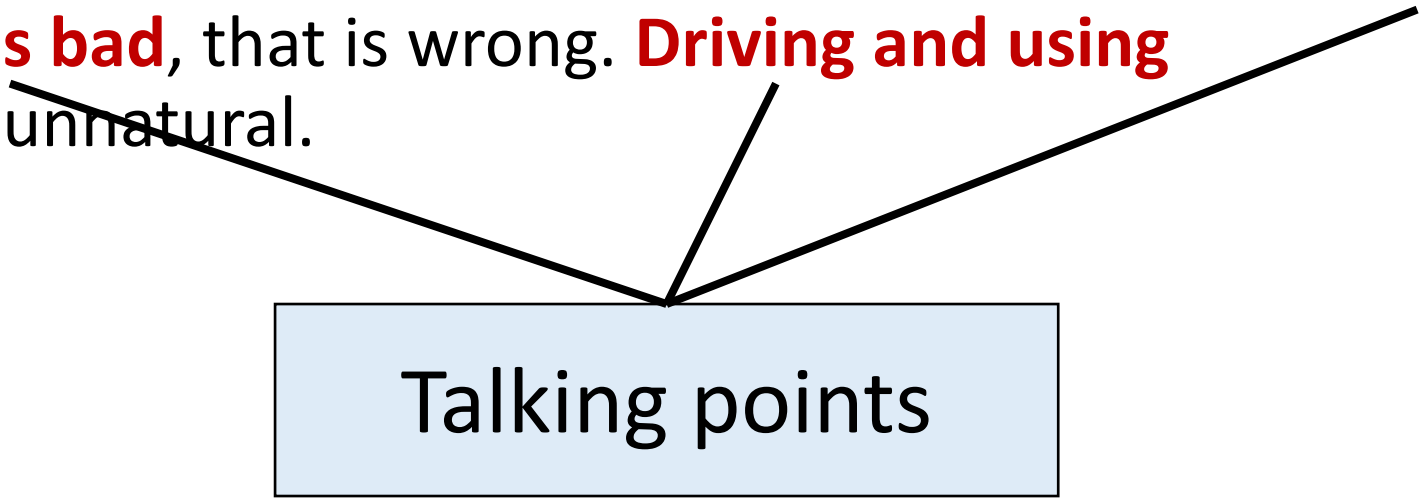
Output: We are not designed to be anything, evolution is directionless. You imply unnatural is bad, that is wrong. Driving and using smartphone are also unnatural.

Our Goal

- We generate a specific type of argument: **counterargument**.

Input: Humans are not designed to be vegan.

Output: We are not designed to be anything, **evolution is directionless**. You imply **unnatural is bad**, that is wrong. **Driving and using smartphone** are also unnatural.



Our Goal

- We generate a specific type of argument: **counterargument**.

Challenges:

1. Understanding the topic and stance
2. Application of common sense knowledge
3. Generating arguments in natural language texts

Outline

- Prior Work
- Data
- System Pipeline
- Experimental Setup
- Evaluation
- Future Directions and Conclusion

Outline

- Prior Work
- Data
- System Pipeline
- Experimental Setup
- Evaluation
- Future Directions and Conclusion

Prior Work

- Argument Component Detection
  - Evidence detection [Rinott et al, 2015]
  - Classification of types of supports [Hua and Wang, 2017]
- Argument and Evidence Retrieval
  - Argument search engine [Wachsmuth et al, 2017; Stab et al, 2018]
- Argument Component Generation
  - Retrieval based argument generation [Sato et al, 2015]
  - Argument strategy based generation [Zukerman et al, 2000]

Outline

- Prior Work
- Data
- System Pipeline
- Experimental Setup
- Evaluation
- Future Directions and Conclusion

Data

- r/changemyview
  - A subreddit for ***open discussion*** and ***debate***

Data

reddit r/changemyview

3.3k

I believe the government should be allowed to view my emails for national security concerns. CMV.

I have nothing to hide. I don't break the law, I don't write hate e-mails…

[U1] Seriously, whether or not … is a good thing, it runs up against the protections offered in the Fourth Amendment: [--quote--]

[U2] Giving up privacy means giving up some of your right to free speech. Knowing that you might be listened in on may change what you say and how you say it…

Data

reddit r/changemyview

3.3k

I believe the government should be allowed to view my emails for national security concerns. CMV.

I have nothing to hide. I don't break the law, I don't write hate e-mails…

[U1] Seriously, whether or not … is a good thing, it runs up against the protections offered in the Fourth Amendment: [--quote--]

[U2] Giving up privacy means giving up some of your right to free speech. Knowing that you might be listened in on may change what you say and how you say it…

Δ
I saved this answer for a Reddit Gold. It did change my opinion - I never thought that…

Data

Input statement

reddit r/changemyview

3.3k

I believe the government should be allowed to view my emails for national security concerns. CMV.

I have nothing to hide. I don't break the law, I don't write hate e-mails…

[U1] Seriously, whether or not … is a good thing, it runs up against the protections offered in the Fourth Amendment: [--quote--]

[U2] Giving up privacy means giving up some of your right to free speech. Knowing that you might be listened in on may change what you say and how you say it…

Data

reddit r/changemyview

3.3k

I believe the government should be allowed to view my emails for national security concerns. CMV.

I have nothing to hide. I don't break the law, I don't write hate e-mails…

[U1] Seriously, whether or not … is a good thing, it runs up against the protections offered in the Fourth Amendment: [--quote--]

[U2] Giving up privacy means giving up some of your right to free speech. Knowing that you might be listened in on may change what you say and how you say it…

Human argument

Data

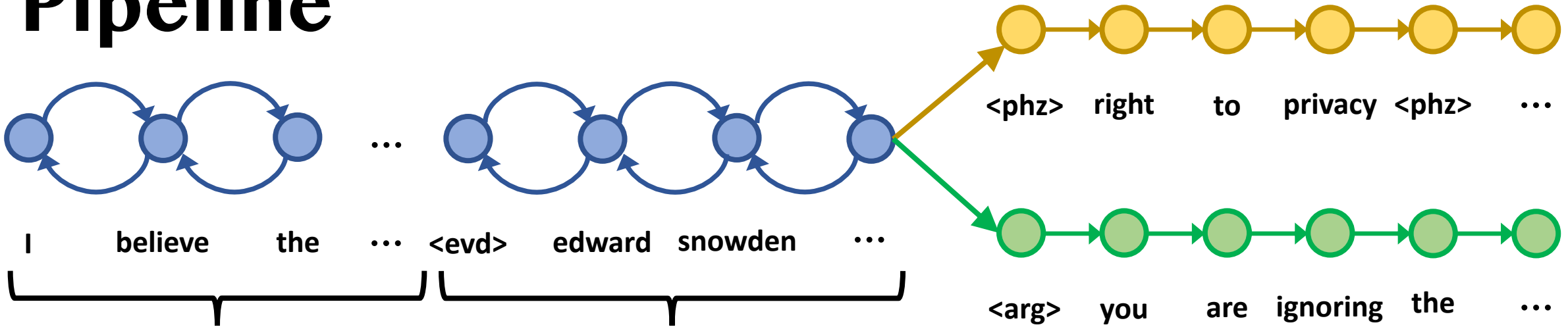
- Collection:
  - Jan 2013 - Jun 2017, about 27K in total.
  - We selected the **politics** and **policy** related topics for study.
  - We only consider “high quality” replies (with delta or more upvotes).
  - Statistics as below after removing non-root and low quality replies.

	Input statement	Human argument
Count	12,549	117,960
Avg number of sentences	16.1	7.7
Avg number of tokens	356.4	161.1

Outline

- Prior Work
- Data
- System Pipeline
- Experimental Setup
- Evaluation
- Future Directions and Conclusion

Pipeline



Input statement

3.3k

I believe the government should be allowed to view my emails for national security concerns. CMV.

I have nothing to hide. I don't break the law…

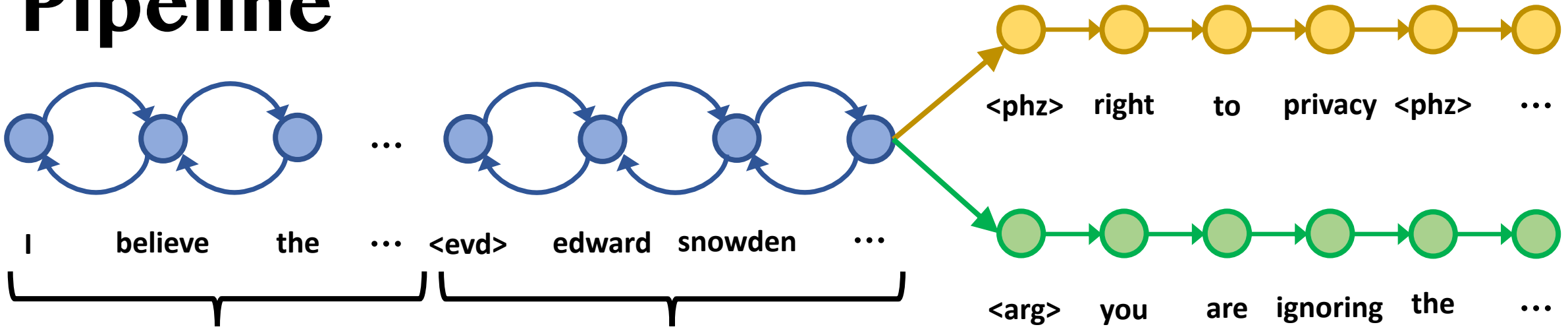
Evidence sentences

1. Edward Snowden: "Arguing that you don't care about right to privacy because…".

2. Political corruption is the use of powers by government officials for illegitimate private gain.

…

Pipeline



1. Document Retrieval

Input statement

3.3k

I believe the government should be allowed to view my emails for national security concerns. CMV.

I have nothing to hide. I don't break the law…

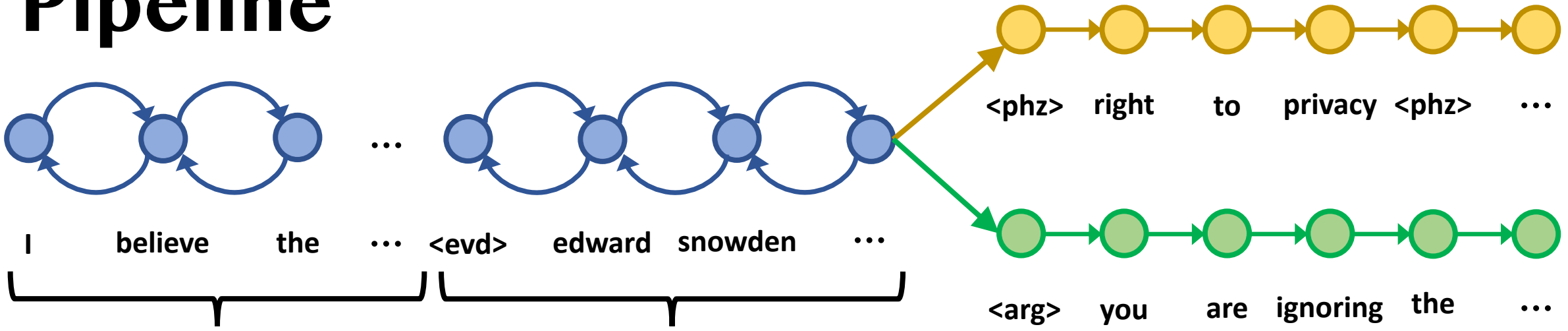
Evidence sentences

1. Edward Snowden: "Arguing that you don't care about right to privacy because…".

2. Political corruption is the use of powers by government officials for illegitimate private gain.

…

Pipeline



Input statement

3.3k

I believe the government should be allowed to view my emails for national security concerns. CMV.

I have nothing to hide. I don't break the law…

WIKIPEDIA
The Free Encyclopedia

2. Sentence Reranking

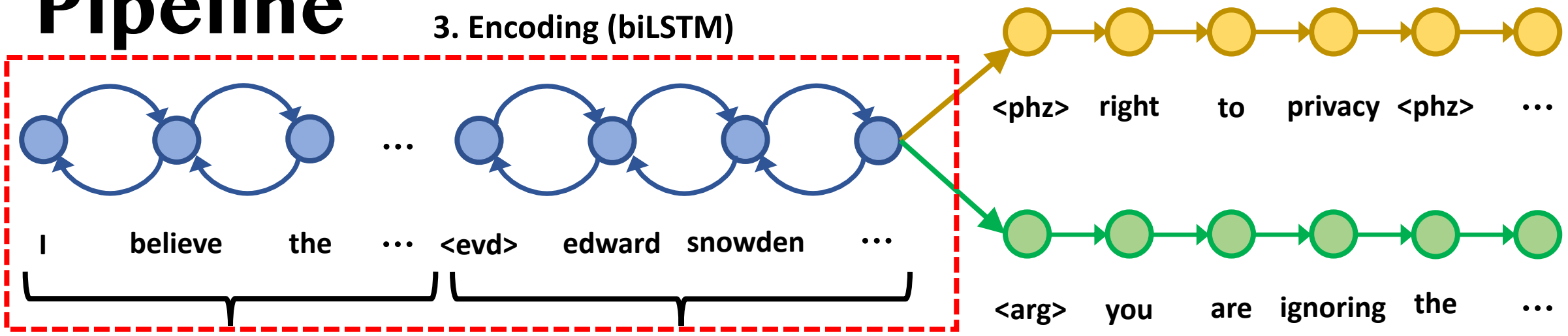
Evidence sentences

1. Edward Snowden: "Arguing that you don't care about right to privacy because…".

2. Political corruption is the use of powers by government officials for illegitimate private gain.

…

Pipeline



Input statement

3.3k

I believe the government should be allowed to view my emails for national security concerns. CMV.

I have nothing to hide. I don't break the law…

Evidence sentences

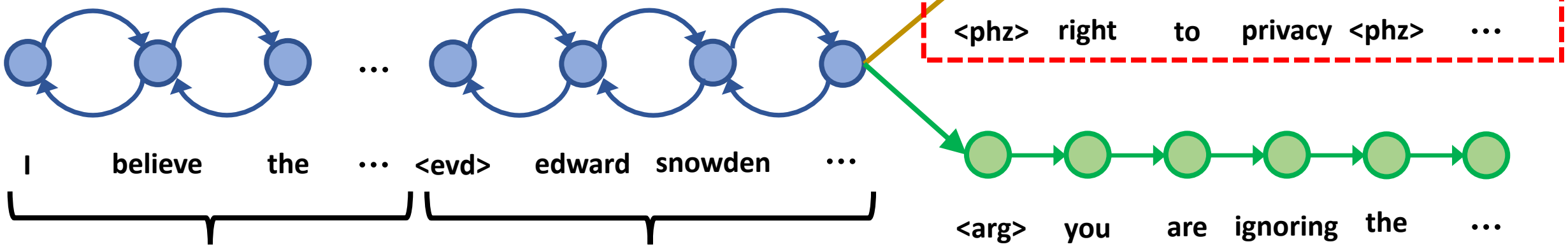
1. Edward Snowden: "Arguing that you don't care about right to privacy because…".

2. Political corruption is the use of powers by government officials for illegitimate private gain.

…

Pipeline

4. Keyphrase Decoding (LSTM)



Input statement

3.3k

I believe the government should be allowed to view my emails for national security concerns. CMV.

I have nothing to hide. I don't break the law…

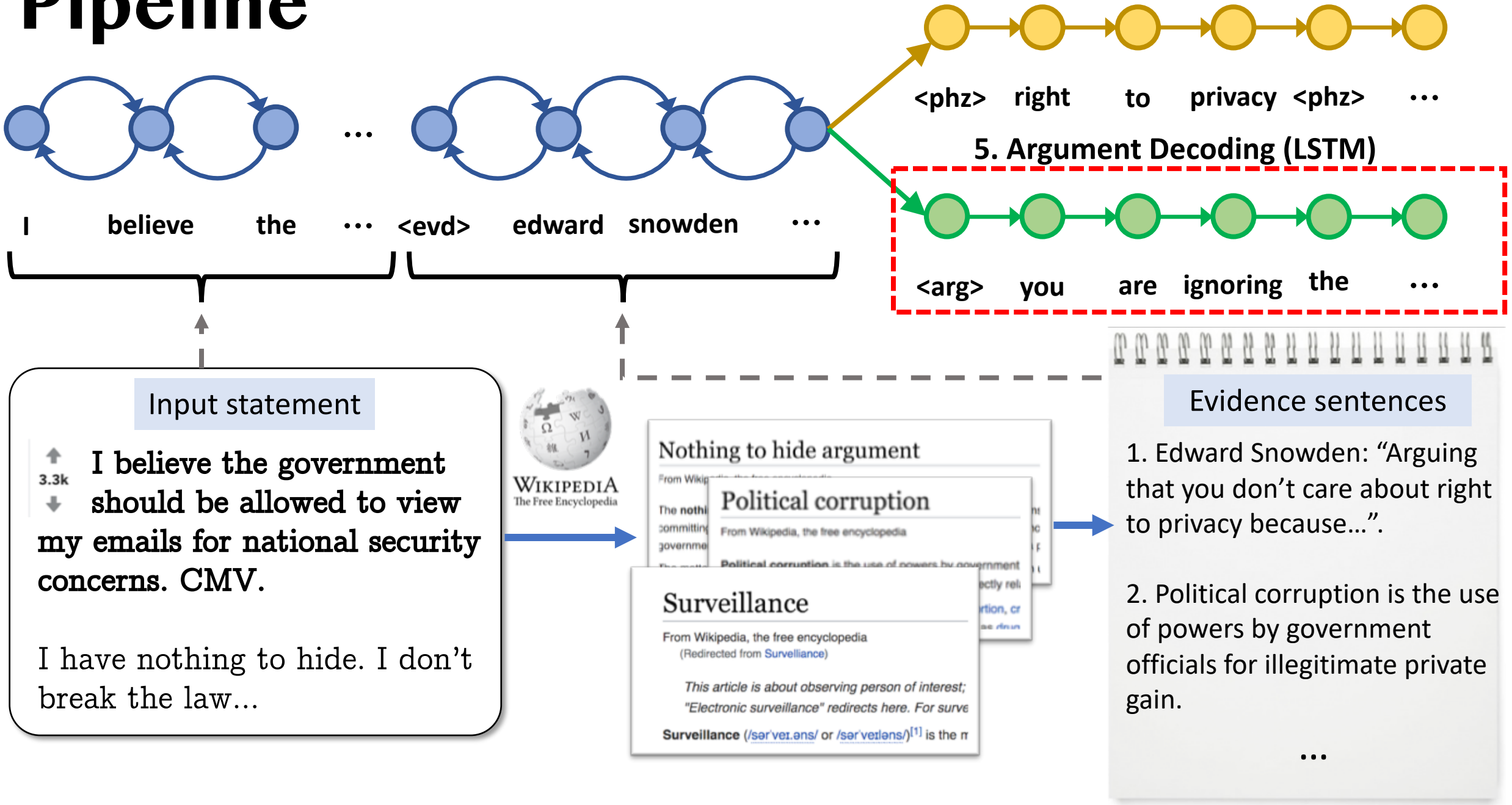
Evidence sentences

1. Edward Snowden: "Arguing that you don't care about right to privacy because…".

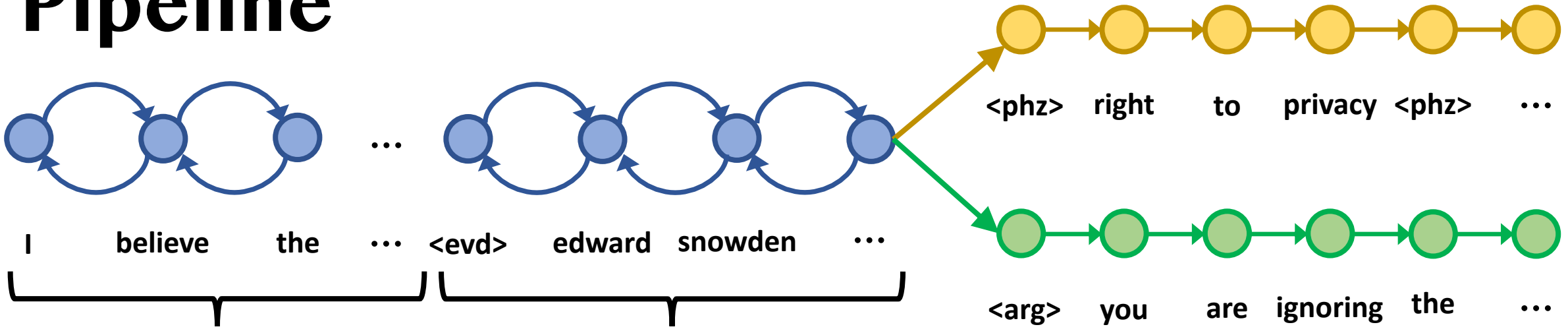
2. Political corruption is the use of powers by government officials for illegitimate private gain.

…

Pipeline


I believe the … <evd> edward snowden …
<phz> right to privacy <phz> …
5. Argument Decoding (LSTM)
<arg> you are ignoring the …
Input statement
3.3k
I believe the government should be allowed to view my emails for national security concerns. CMV.
I have nothing to hide. I don't break the law…
WIKIPEDIA
The Free Encyclopedia
Nothing to hide argument
Political corruption
From Wikipedia, the free encyclopedia
Surveillance
From Wikipedia, the free encyclopedia
(Redirected from Surveillance)
This article is about observing person of interest;
"Electronic surveillance" redirects here. For surve
Surveillance (/sərˈveɪ.ənsˈ or /sərˈveɪləns/)[1] is the
Evidence sentences
1. Edward Snowden: "Arguing that you don't care about right to privacy because…".
2. Political corruption is the use of powers by government officials for illegitimate private gain.
…

Pipeline



1. Document Retrieval

Input statement

3.3k

I believe the government should be allowed to view my emails for national security concerns. CMV.

I have nothing to hide. I don't break the law…

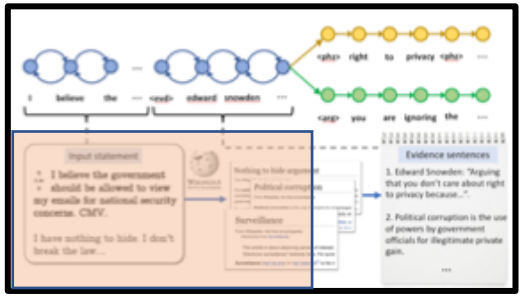
Evidence sentences

1. Edward Snowden: "Arguing that you don't care about right to privacy because…".

2. Political corruption is the use of powers by government officials for illegitimate private gain.

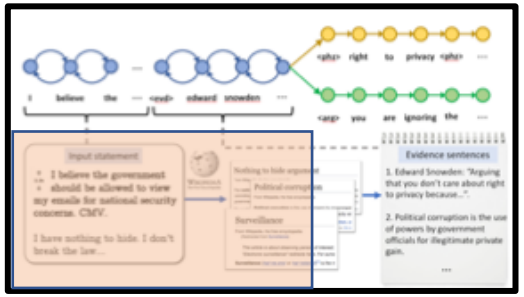
…

Step 1: Document Retrieval



- *<u>Goal</u>: to extract relevant evidence for counterarguments*

Step 1: Document Retrieval



- Query construction
  - Formed from topic signatures [Lin and Hovy, 2000]
  - Representative of the text, measured by log-likelihood ratio
  - E.g. “government”, “emails”, “national security”, etc in the following post

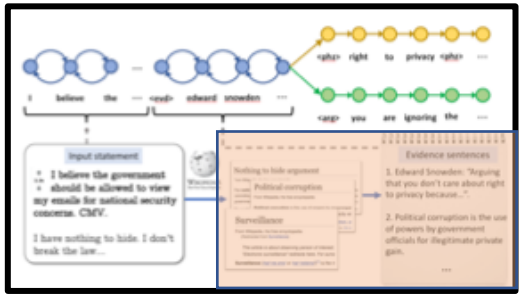
Input statement

3.3k

I believe the government should be allowed to view my emails for national security concerns. CMV.

I have nothing to hide. I don’t break the law…

Step 2: Sentence Reranking



- Rerank sentences
  - Returned articles are broken into paragraphs and sentences.
  - Sentences are ranked by TF-IDF similarity against the post.

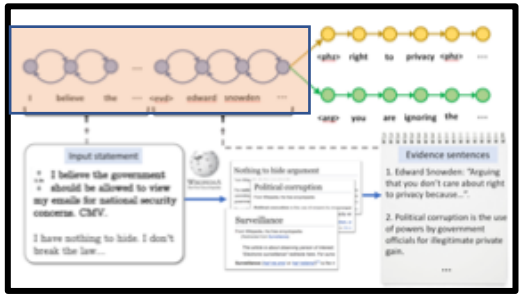
Evidence sentences

1. Edward Snowden: “Arguing that you don’t care about right to privacy because…”.

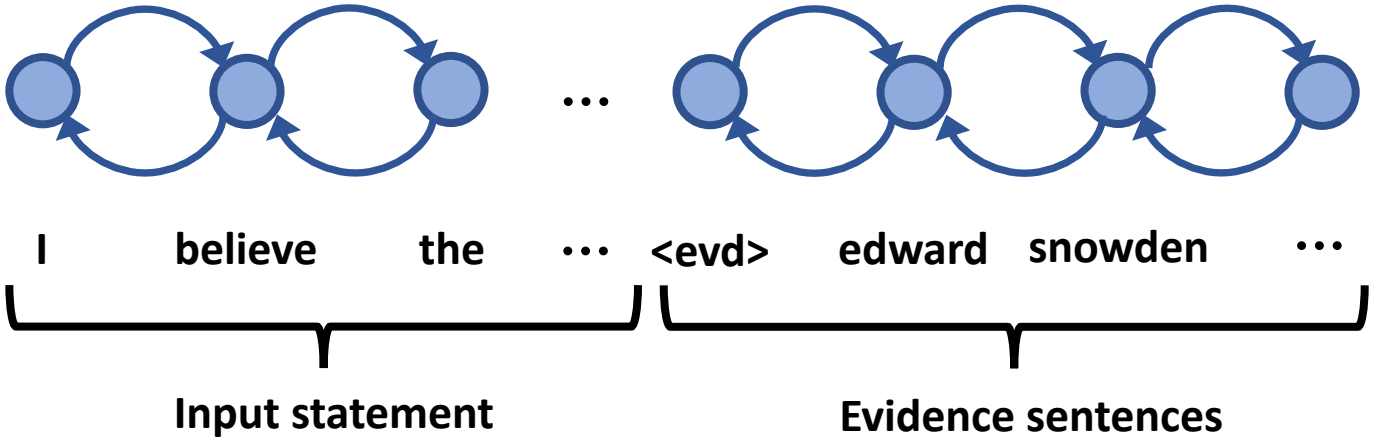
2. Political corruption is the use of powers by government officials for illegitimate private gain.

…

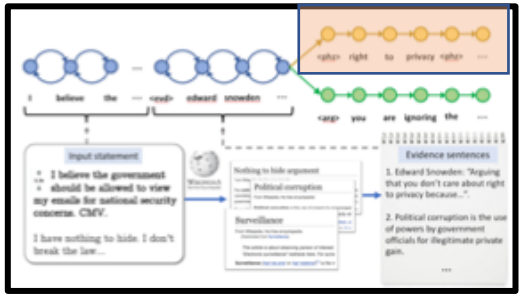
Step 3: Encoding



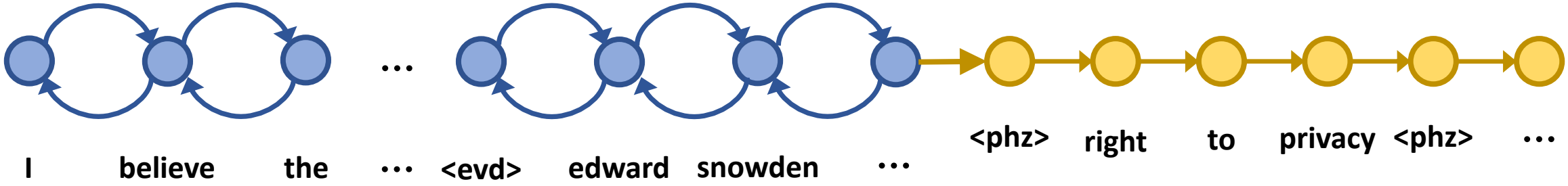
- Neural Encoder
  - Bi-directional LSTM network
  - Encode input statement and evidence sentences, separated by **<evd>** token



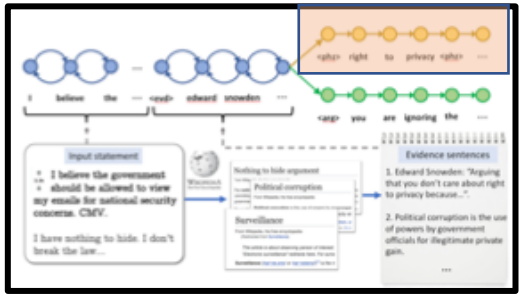
Step 4: Keyphrase Decoding



- Decoder
  - Generate keyphrase as an intermediate step
  - Aim to inform the model of the major talking points
  - Mimic keyphrases that are likely reused by human

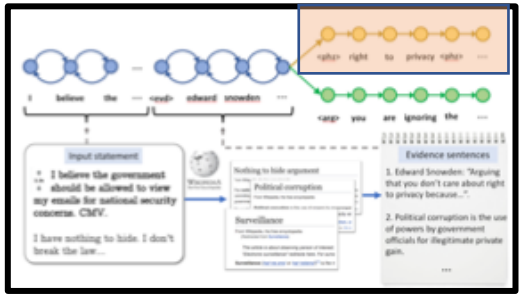


Step 4: Keyphrase Decoding



- Decoder
  - We extract noun phrases and verb phrases.
  - The length has to be between 2 to 10 tokens.
  - Phrase has to contain non-stop words.

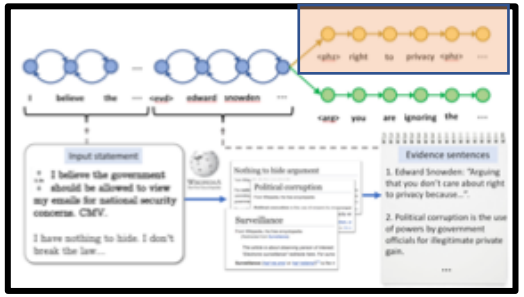
Step 4: Keyphrase Decoding



- Decoder
  - We extract noun phrases and verb phrases.
  - The length has to be between 2 to 10 tokens.
  - Phrase has to contain non-stop words.

> Numerous civil rights groups and privacy groups oppose surveillance as a violation of people's right to privacy.

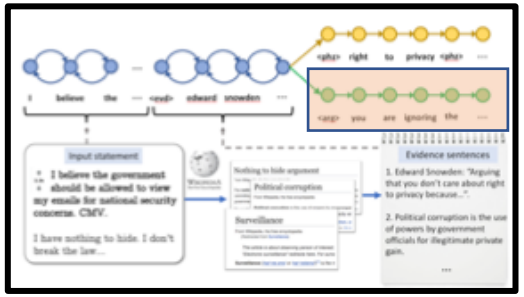
Step 4: Keyphrase Decoding



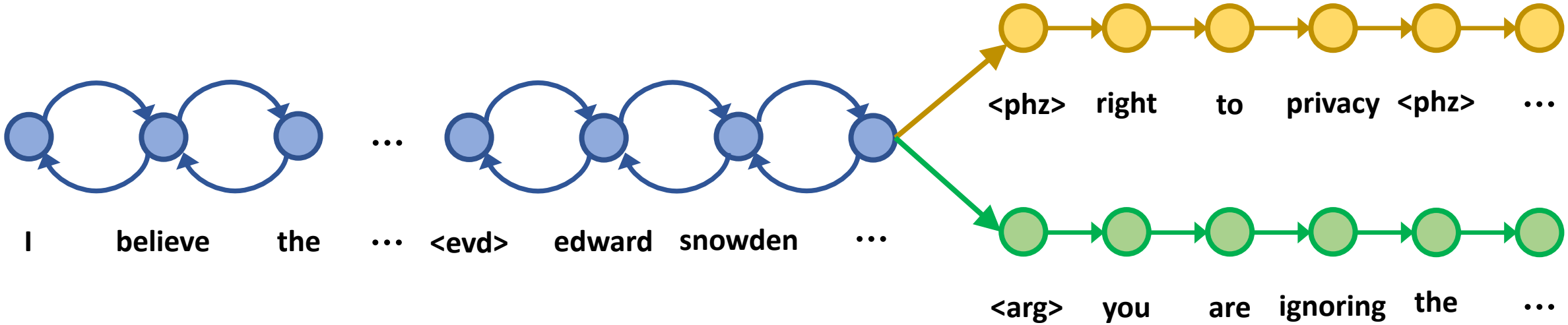
- Decoder
  - We extract noun phrases and verb phrases.
  - The length has to be between 2 to 10 tokens.
  - Phrase has to contain non-stop words.

Numerous civil rights groups and privacy groups oppose surveillance as a violation of people's **right to privacy**.

Step 5: Argument Decoding



- Decoder
  - Generate argument based on encoder or keyphrase last hidden state
  - Attention mechanism over both input and keyphrase results



Outline

- Prior Work
- Data
- System Pipeline
- Experimental Setup
- Evaluation
- Future Directions and Conclusion

Experiments

- Pre-training
  - Initialize first layers of encoders and argument decoders
  - Warm up the system with a good argumentation language model
  - Data:
    - All training data + non-politics threads + non-root replies
    - Sequence-to-sequence without evidence sentences or keyphrases
  - Input: input statement
  - Output: human argument

Experiments - Models

- Baselines and comparisons
  - RETRIEVAL-BASED: concatenate evidence sentences

Experiments - Models

- Baselines and comparisons
  - RETRIEVAL-BASED: concatenate evidence sentences
  - SEQ2SEQ: encode the input statement only

Experiments - Models

- Baselines and comparisons
  - RETRIEVAL-BASED: concatenate evidence sentences
  - SEQ2SEQ: encode the input statement only
  - SEQ2SEQ + *encode evidence*: encode statement and evidence sentences

Experiments - Models

- Baselines and comparisons
  - RETRIEVAL-BASED: concatenate evidence sentences
  - SEQ2SEQ: encode the input statement only
  - SEQ2SEQ + *encode evidence*: encode statement and evidence
  - SEQ2SEQ + *encode keyphrase*: encode statement and keyphrases

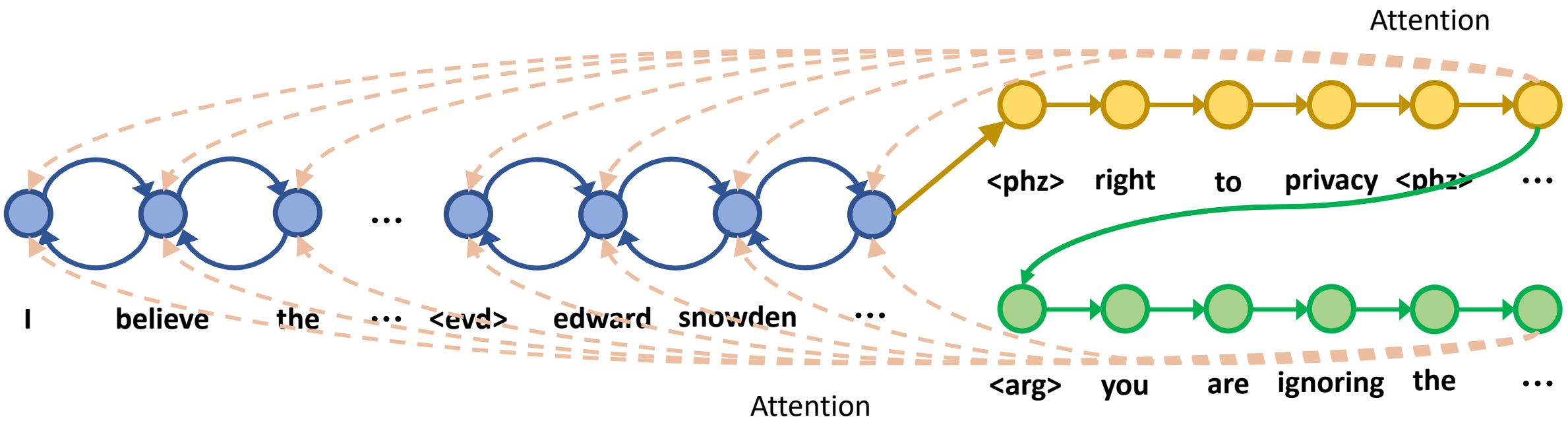
Experiments - Models

- Baselines and comparisons
  - RETRIEVAL-BASED: concatenate evidence sentences
  - SEQ2SEQ: encode the input statement only
  - SEQ2SEQ + *encode evidence*: encode statement and evidence sentences
  - SEQ2SEQ + *encode keyphrase*: encode statement and keyphrases

Stronger baseline, because keyphrases are actually reused by human arguments.

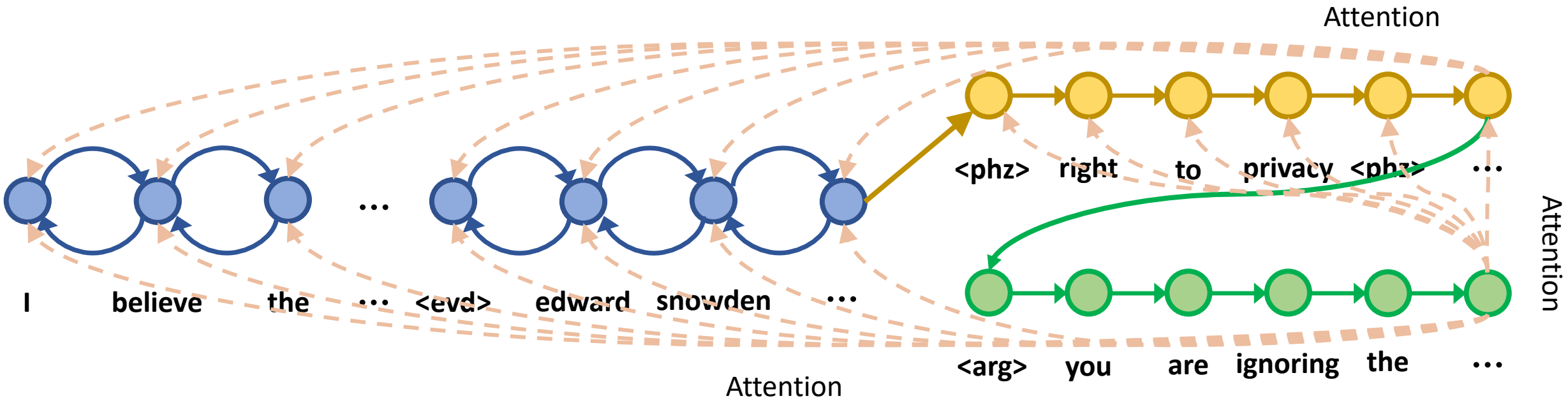
Experiments - Models

- Our models
  - DEC-SHARED: Argument decoder initialized by keyphrase decoder



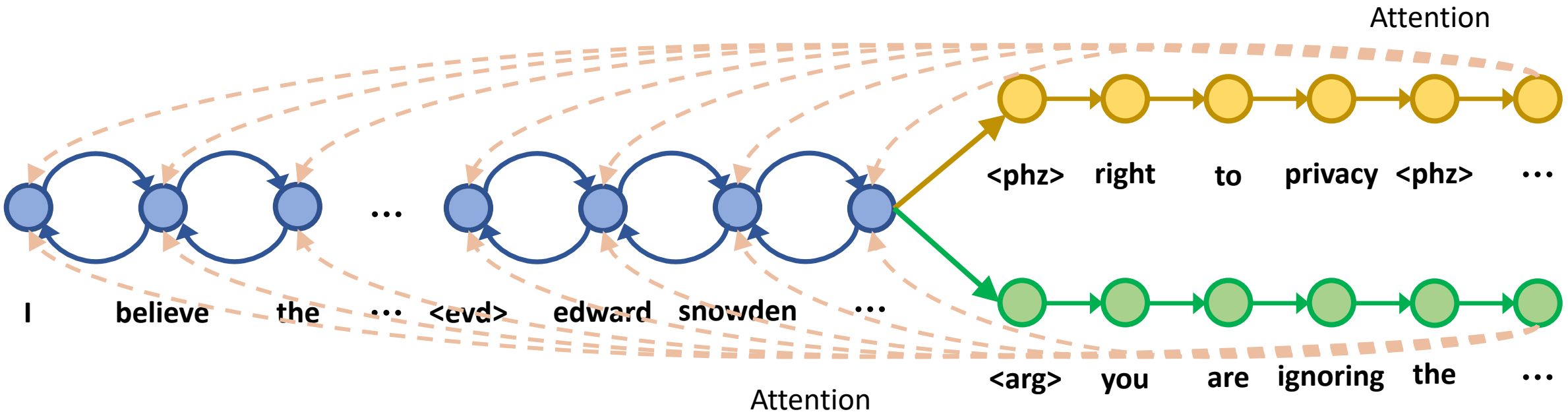
Experiments - Models

- Our models
  - DEC-SHARED: Argument decoder initialized by keyphrase decoder
  - DEC-SHARED + *attend keyphrase*: with attention on keyphrase decoder



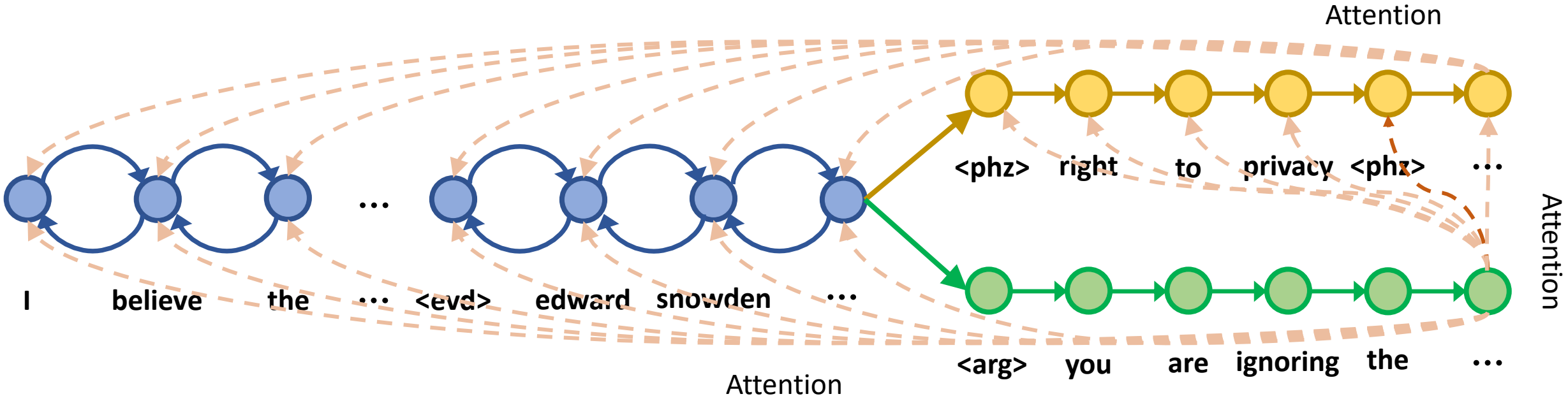
Experiments - Models

- Our models
  - DEC-SHARED: Argument decoder initialized by keyphrase decoder
  - DEC-SHARED + *attend keyphrase*: with attention on keyphrase decoder
  - DEC-SEPARATE: Argument decoder initialized by encoder



Experiments - Models

- Our models
  - DEC-SHARED: Argument decoder initialized by keyphrase decoder
  - DEC-SHARED + *attend keyphrase*: with attention on keyphrase decoder
  - DEC-SEPARATE: Argument decoder initialized by encoder
  - DEC-SEPARATE + *attend keyphrase:* with attention on keyphrase decoder



Experiments

- System vs. Oracle retrieval
  - In reality, during test time evidence can only be obtained by input statement.
  - In Oracle setup, we retrieve evidence base on human arguments' queries.

Experiments

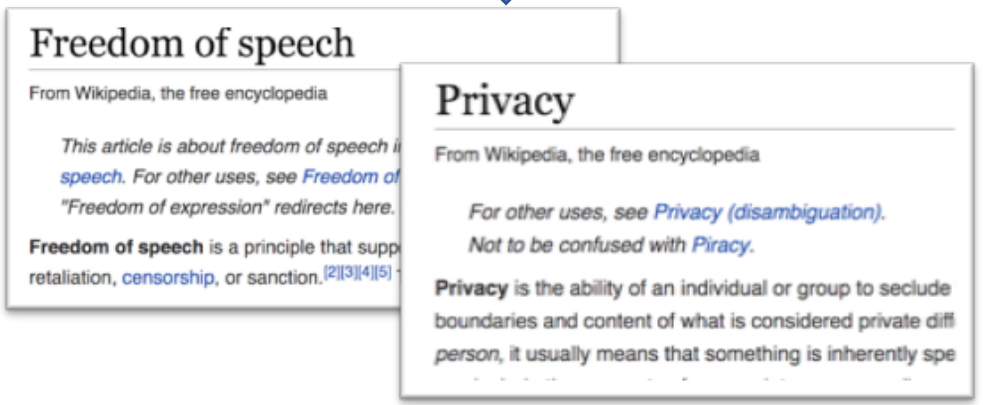
- System vs. Oracle retrieval
  - In reality, during test time evidence can only be obtained by input statement.
  - In Oracle setup, we retrieve evidence base on human arguments' queries.

System Retrieval

Input statement: I believe the government should be allowed to view my emails…

Oracle Retrieval

Human argument: Giving up privacy means giving up some of your right to free speech. …



Outline

- Prior Work
- Data
- System Pipeline
- Experimental Setup
- Evaluation
- Future Directions and Conclusion

Automatic Evaluation – Generation Quality

- Argument generation quality
  - BLEU: n-gram precision based measure
  - METEOR: unigram precision and recall based on alignment
  - Gold-standard: user generated arguments
  - Multi-reference setup: best aligned one -> multiple plausible arguments exist

Automatic Evaluation – Generation Quality

* BLEU/METEOR: The higher the better

	w/System Retrieval		
	BLEU-2	METEOR	Length
Baseline			
RETRIEVAL	15.32	**12.19**	151.2
Comparisons			
SEQ2SEQ	10.21	5.74	34.9
+ encode evd	18.03	7.32	67.0
+ encode KP	21.94	8.63	74.4
Our Models			
DEC-SHARED	21.22	8.91	69.1
+ attend KP	**24.71**	10.05	74.8
DEC-SEPARATE	24.24	10.63	88.6
+ attend KP	24.52	11.27	88.3

Automatic Evaluation – Generation Quality

* BLEU/METEOR: The higher the better

	w/System Retrieval		
	BLEU-2	METEOR	Length
Baseline			
RETRIEVAL	15.32	**12.19**	151.2
Comparisons			
SEQ2SEQ	10.21	5.74	34.9
+ encode evd	18.03	7.32	67.0
+ encode KP	21.94	8.63	74.4
Our Models			
DEC-SHARED	21.22	8.91	69.1
+ attend KP	**24.71**	10.05	74.8
DEC-SEPARATE	24.24	10.63	88.6
+ attend KP	24.52	11.27	88.3

- Our models have better precision. The generated content are more likely to be found in human arguments.

Automatic Evaluation – Generation Quality

* BLEU/METEOR: The higher the better

	w/System Retrieval		
	BLEU-2	METEOR	Length
Baseline			
RETRIEVAL	15.32	**12.19**	151.2
Comparisons			
SEQ2SEQ	10.21	5.74	34.9
+ encode evd	18.03	7.32	67.0
+ encode KP	21.94	8.63	74.4
Our Models			
DEC-SHARED	21.22	8.91	69.1
+ attend KP	**24.71**	10.05	74.8
DEC-SEPARATE	24.24	10.63	88.6
+ attend KP	24.52	11.27	88.3

- Our models have better precision. The generated content are more likely to be found in human arguments.

- Retrieval baseline generation has better METEOR, which considers both precision and recall.

Automatic Evaluation – Generation Quality

* BLEU/METEOR: The higher the better

	w/System Retrieval			w/ Oracle Retrieval		
	BLEU-2	METEOR	Length	BLEU-2	METEOR	Length
Baseline						
RETRIEVAL	15.32	**12.19**	151.2	10.24	**16.22**	132.7
Comparisons						
SEQ2SEQ	10.21	5.74	34.9	7.44	5.25	31.1
+ encode evd	18.03	7.32	67.0	13.79	10.06	68.1
+ encode KP	21.94	8.63	74.4	12.96	10.50	78.2
Our Models						
DEC-SHARED	21.22	8.91	69.1	15.78	11.52	68.2
+ attend KP	**24.71**	10.05	74.8	11.48	10.08	40.5
DEC-SEPARATE	24.24	10.63	88.6	17.48	13.15	86.9
+ attend KP	24.52	11.27	88.3	**17.80**	13.67	86.8

Automatic Evaluation – Generation Quality

* BLEU/METEOR: The higher the better

	w/System Retrieval			w/ Oracle Retrieval		
	BLEU-2	METEOR	Length	BLEU-2	METEOR	Length
Baseline						
RETRIEVAL	15.32	**12.19**	151.2	10.24	**16.22**	132.7
Comparisons						
SEQ2SEQ	10.21	5.74	34.9	7.44	5.25	31.1
+ encode evd	18.03	7.32	67.0	13.79	10.06	68.1
+ encode KP	21.94	8.63	74.4	12.96	10.50	78.2
Our Models						
DEC-SHARED	21.22	8.91	69.1	15.78	11.52	68.2
+ attend KP	**24.71**	10.05	74.8	11.48	10.08	40.5
DEC-SEPARATE	24.24	10.63	88.6	17.48	13.15	86.9
+ attend KP	24.52	11.27	88.3	**17.80**	13.67	86.8

Automatic Evaluation – Topic Relevance

- Motivation: Generic arguments can still have high BLEU scores.

Automatic Evaluation – Topic Relevance

- Motivation: Generic arguments can still have high BLEU scores.
  - E.g. “I don’t agree with you.”, “You are missing evidence.”, “This is wrong.”

Automatic Evaluation – Topic Relevance

- Motivation: Generic arguments can still have high BLEU scores.
- Topic relevance
  - Semantic similarity model [Huang et al, 2013]
  - Represents the semantic relatedness of two pieces of text
  - Model tuned on training set
  - Evaluated by mean reciprocal ranking (MRR) and Precision at 1 (P@1)

Automatic Evaluation – Topic Relevance

* The higher the better

	MRR	P@1
Baseline		
RETRIEVAL	81.08	65.45
Comparisons		
SEQ2SEQ	74.46	57.06
+ encode evd	88.24	78.76
Our Models		
DEC-SHARED	95.18	90.91
+ attend KP	93.48	87.91
DEC-SEPARATE	91.70	84.72
+ attend KP	92.77	86.46

Automatic Evaluation – Topic Relevance

* The higher the better

	MRR	P@1
Baseline		
RETRIEVAL	81.08	65.45
Comparisons		
SEQ2SEQ	74.46	57.06
+ encode evd	88.24	78.76
Our Models		
DEC-SHARED	95.18	90.91
+ attend KP	93.48	87.91
DEC-SEPARATE	91.70	84.72
+ attend KP	92.77	86.46

Our models produce more topic relevant outputs.

Human Evaluation

- Motivation: Automatic evaluation can't really evaluate the overall coherence and informativeness of the generation.

- Setup:
  - 3 trained judges that are fluent in English
  - 3 systems: RETRIEVAL-BASED, SEQ2SEQ, OUR MODEL

- Aspects (each on a scale of 1 to 5, the higher the better)
  - **Grammaticality**: if the output is fluent and grammatical English
  - **Informativeness**: whether the output is informative or generic
  - **Relevance**: it the output is on-topic and of correct stance

Human Evaluation

* Each on a scale of 1 to 5, the higher the better

	1 (low quality)	5 (high quality)
Grammaticality	*checked criminal taxi the speed limit lanes to*	*Food security is not an issue of how much food we produce.*
Informativeness	*I don't agree with you.*	*Israeli are under a much more persistent and realistic security threat.*
Relevance **(Topic: racial profiling)**	*Gun control deters crime.*	*Minority groups who endure everyday discrimination often suffer high rates of chronic diseases.*

Human Evaluation

System	Grammaticality	Informativeness	Relevance
RETRIEVAL-BASED	4.5 ± 0.6	3.7 ± 0.9	3.3 ± 1.1
SEQ2SEQ	3.3 ± 1.1	1.2 ± 0.5	1.4 ± 0.7
OUR MODEL	2.5 ± 0.8	1.6 ± 0.8	1.8 ± 0.8

* Each on a scale of 1 to 5, the higher the better

Human Evaluation

System	Grammaticality	Informativeness	Relevance
RETRIEVAL-BASED	4.5 ± 0.6	3.7 ± 0.9	3.3 ± 1.1
SEQ2SEQ	3.3 ± 1.1	1.2 ± 0.5	1.4 ± 0.7
OUR MODEL	2.5 ± 0.8	1.6 ± 0.8	1.8 ± 0.8

- Human judges favor RETRIEVAL-BASED model in all aspects.
- RETRIEVAL-BASED is human-written and relevant.
- OUR MODEL is favored over SEQ2SEQ except **Grammaticality**.

Sample Argument

Original Post

Putin is trying to re-form a “Soviet Union” with his past actions in Georgia and current actions in Ukraine.

…I firmly believe that Putin and the Russian Federation (RF) are trying to re-form a Soviet Union type regime… The Russian Army invaded certain regions of Georgia…

Generated Counterargument

There are two reasons you are so far beyond. There is no reason to see the military army. You can infer what they assume it, so they tend to protect up. There. If we assume it were bad, they can not be controlled. So we’re talking going to ignore the wars of the world. The fact of the matter is not bad. I would also assume it. However, the government can not be reminded of the world.

Sample Argument

Original Post

Putin is trying to re-form a “Soviet Union” with his past actions in Georgia and current actions in Ukraine.

…I firmly believe that Putin and the Russian Federation (RF) are trying to re-form a Soviet Union type regime… The Russian Army invaded certain regions of Georgia…

Generated Counterargument

There are two reasons you are so far beyond. There is no reason to see the **military army**. You can infer what they assume it, so they tend to protect up. There. If we assume it were bad, they can not be controlled. So we’re talking going to ignore the **wars of the world**. The fact of the matter is not bad. I would also assume it. However, the **government** can not be reminded of the world.

Outline

- Prior Work
- Data
- System Pipeline
- Experimental Setup
- Evaluation
- Future Directions and Conclusion

Future Directions

- Knowledge Retrieval
  - Better evidence retrieval system
  - Reasoning and interpretability
- Text Generation
  - Prone to incoherence, inaccurate information, generic generation etc
  - Discourse-aware argument generation

Conclusion

- We study a novel neural argument generation task.
- We collect and release a new dataset from `r/ChangeMyView` and accompanying Wikipedia evidence for argument generation research.
- We propose an end-to-end argument generation system, enhanced with Wikipedia retrieved evidence sentences.

Thank you for your attention!

- **Dataset:** https://xinyuhua.github.io/Resources/
- **Project page**: https://xinyuhua.github.io/neural-argument-generation/
- **Contact:** Xinyu Hua (hua.x@husky.neu.edu)

Conclusion

- We study a novel neural argument generation task.
- We collect and release a new dataset from `r/ChangeMyView` and accompanying Wikipedia evidence for argument generation research.
- We propose an end-to-end argument generation system, enhanced with Wikipedia retrieved evidence sentences.

Project page: https://xinyuhua.github.io/neural-argument-generation/