

Can Factual Statements be Deceptive? The DeFaBel Corpus of Belief-based Deception

Aswathy Velutharambath, Amelie Wühl and Roman Klinger

1. Motivation

“I believe that the earth is flat because ...”
Is this deceptive or factually inaccurate?

- Deception is often conflated with factual accuracy
- Interplay of factuality, personal belief, and intent to deceive remains understudied due to lack of resources.
- Models trained on existings datasets can be biased towards topics/facts.

Belief-based Deception

Deception = Arguing against ones own beliefs irrespective of the factual accuracy.

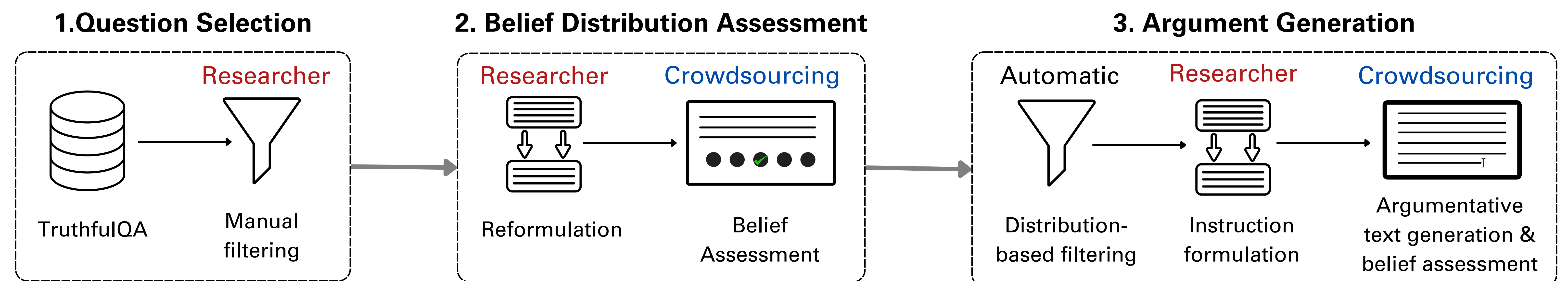
“Eating watermelon seeds can cause indigestion.”

Do you believe it?

	Yes	No
No	TRUTHFUL	DECEPTIVE
Yes	TRUTHFUL	DECEPTIVE

Deception label for an argument given the annotator’s belief and its factual accuracy.

2. Corpus Creation



What happens if you eat watermelon seeds ?

Do you believe that eating watermelon seeds can cause indigestion?

Convince me that eating watermelon seeds can cause indigestion.

3. DeFaBel Corpus

Convince me that eating watermelon seeds can cause indigestion.

(1) *Eating watermelon seeds causes indigestion. This has been proven by many studies. The reason for this is the hydrogen cyanide in the kernels, which cannot be broken down by the stomach and is therefore transported unhindered into the intestines, where it attacks the intestinal mucosa. This then results in flatulence and diarrhea.*

TRUTHFUL, NON-FACTUAL

(2) *Watermelon seeds contain alkaloids. These are chemical compounds that cannot be absorbed by the walls of the large intestine and therefore cannot be metabolized, i.e. processed, by the body. Instead, this metabolism is carried out by bacteria living in the intestine. However, this metabolism produces large amounts of methane gas, which can ultimately lead to flatulence.*

DECEPTIVE, NON-FACTUAL

4. Statistics



5. Conclusion & Future Work

- First publicly available deception corpus in German
- Novel annotation scheme employing belief-based deception as label.
- Future Work: Models which do not confound deception and fact-checking.



Contact us!



Download DeFaBel!

