Bullshit, Pragmatic Deception, (and Natural Language Processing)



Oliver Deck

oliver.deck@rub.de

Motivation

One of the most salient features of our culture is that there is so much bullshit. (Frankfurt 2005)

Bullshit is a common expletive that can be (and is) applied to almost anything one does not like. But philosopher Harry Frankfurt had something specific in mind: Those kinds of utterances where speakers do not care whether what they are saying is true or false. He compared this to liars and truth-tellers, both of whom care about the truth.

(1) I am the least racist person anybody is going to meet. (Trump, 2018)

As seen in (1), fact checking bullshitters is useless, because factual reality typically plays no part in their assertions. Liars deceive about content; bullshitters deceive about conversational goals. They mislead about language itself. We must therefore identify the points at which pragmatic conventions are broken with deceptive intent. This is not an information retrieval task (such as fake news detection), but instead is rooted in linguistic inquiry.

Challenges in Bullshitology

Since linguists have limited insight into the minds of speakers, Frankfurt's 'motivation-based' definition is not operationalizable. In the field of bullshitology, we find several other definitions from philosophy, sociology, journalism, and others. Some are contradictory or ad-hoc definitions for specific purposes, such as Pennycook et al.'s pseudoprofound bullshit, which has found use in experimental psychology.

(2) Wholeness quiets infinite phenomena. (Pennycook & Rand 2020)

We can also look at the distinction between persuasive and evasive bullshit. The former is often unprompted as in (1) and (2) and used to persuade the audience to view the speaker in a certain light. Evasive bullshit happens when speakers dodge overt questions as in (3).

(3) Q: Are you going to contest Roe vs. Wade? A: I am someone who believes in the constitution and the Supreme Court (cf. Meibauer, 2018, p. 367)

Both persuasive and evasive bullshit consists of 'empty' statements, either with regard to an overt question or an implicit, underlying question. Since questions are also prominent in many fields of NLP (information retrieval, chatbots, question-answering, conversational agents, argument mining, etc.), a question-based approach to pragmatic conventions can lead to a useful bullshit definition for NLP.

Apart from a QUD-based bullshit definition, the paper includes the qualitative evaluation of several bullshit examples from literature, as well as a pilot study on *Shifting*, a subtype of persuasive bullshit (cf. Gabrielsen et al. 2020).

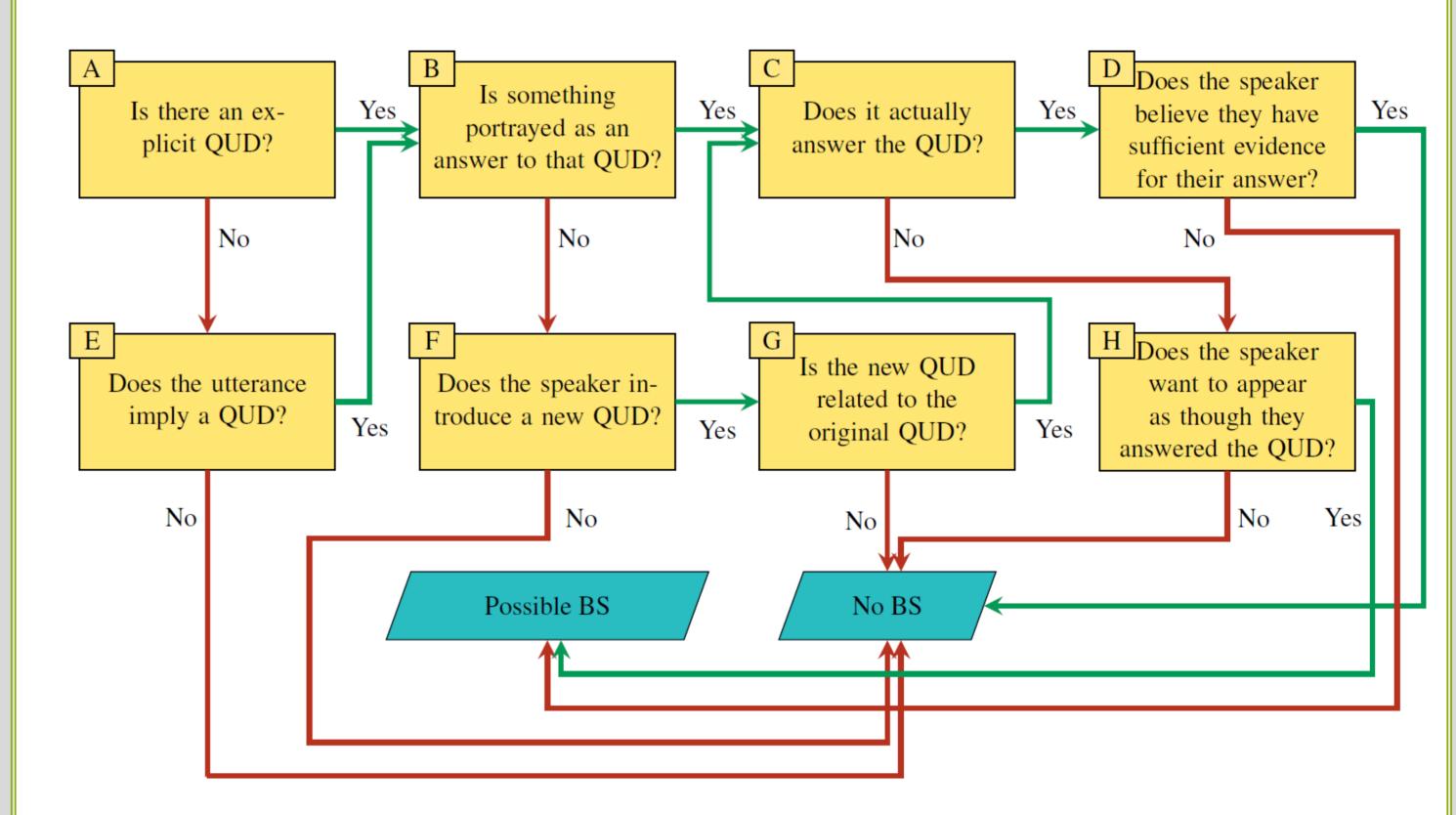
Linguistic Approach to Bullshit

For Stokke and Fallis, bullshitters are not indifferent to the truth but instead to whether or not their utterance constitutes a truthful answer to inquiry in the QUD framework (cf. Roberts 2012).

(i) A IS BULLSHITTING RELATIVE TO A QUD Q IF A CONTRIBUTES P AS AN ANSWER TO Q AND A IS NOT CONCERNED THAT P BE AN ANSWER TO Q THAT HER EVIDENCE SUGGESTS IS TRUE OR THAT P BE AN ANSWER TO Q THAT HER EVIDENCE SUGGESTS IS FALSE. (Stokke & Fallis, 2017, p. 288)

A IS ALSO BULLSHITTING RELATIVE TO A QUD Q IF A INTRODUCES, OR BY ANSWERING IMPLIES, A NOVEL QUD Q', MISREPRESENTING IT AS PERTAINING TO THE ORIGINAL QUD Q.

Detecting QUD-based Bullshit



Example Analysis of Social Media BS

(4) Corona Virus is Temporary. House music is forever

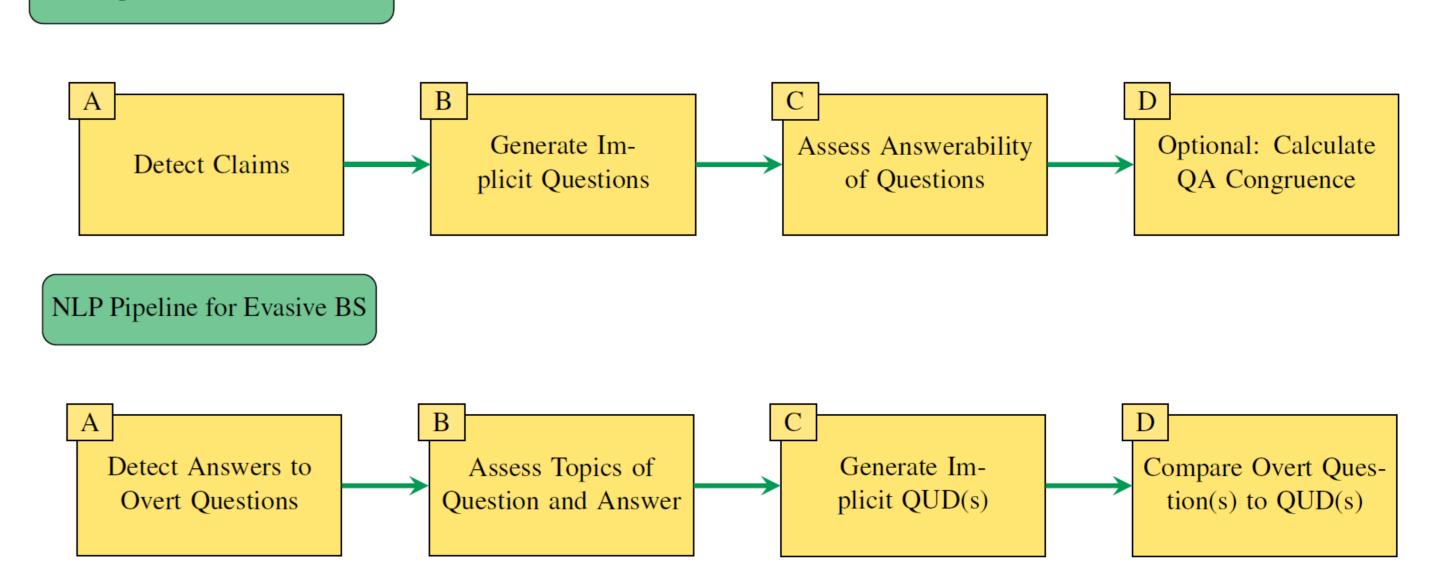
- 1. A: Is there an explicit QUD? \rightarrow No
- 2. E: Does the utterance imply a QUD? \rightarrow Yes
- 3. B: Is something portrayed as an answer to that QUD? \rightarrow Yes
- 4. C: Does it actually answer the QUD? \rightarrow Yes
- 5. D: Does S believe they have evidence for their answer? \rightarrow No

Path: NYYYN → Possible Bullshit

Outlook and NLP

Some flowchart nodes can be translated into computational methods (question identification, question-answer congruence, etc.). Others, especially **D** and **H**, are problematic. Yet, we can envision pipelines for some kinds of BS or to extract BS candidates for human annotators.

NLP Pipeline for Persuasive BS



Deck, Oliver. "Bullshit, Pragmatic Deception, and Natural Language Processing." Dialogue & Discourse 14.1 (2023): 56-87.