

## **Distinction Between Soft Bullshit and Hard Bullshit**

*Audio Transcript*

*James Humphries - Co-Author of "[ChatGPT Is Bullshit.](#)"*

The distinction we draw between soft and hard ultimately comes down to intentions or lack thereof. So we think that there's a kind of genus, which is just anything that's produced without regard for the truth. So any, anything, any utterance, any sentence where the speaker doesn't care about what the truth is. So soft simply an utterance produced without regard for the truth.

Hard is where someone is necessarily attempting to deceive you about the nature of the enterprise. So hard is closer to what Frankfurt says in at least some moods. So he's sort of with us about this. Sometimes he seems to think it's what we would call soft. Sometimes it's much more. He's getting a positive case that there has to be some intent to mislead the audience about not the content of what you're saying, 'cause you don't care about the truth, value of the content, but to mislead your audience about kind of what you're up to.

So the, the examples he uses are things like advertisers or things like politicians, right? So politicians talking about this great country, they don't actually care whether or not this country is great. What they're attempting to sort of do is to persuade you that they care about the country. They're (seems to be) some deeply principled position that they hold. And actually they're just power hungry ... they're just power hungry.

Soft is concerning because you should care about the truth, but it's in some sense kind of less malevolent than hard. The hard is actively trying to mislead you about something important. So to bring it back to the paper, what we say is that minimally, large language models are soft bullshitters or they're machines. If you think that they can't have intentions, if you think that they can't kind of intend or aim to do anything, if, and I don't think this is true as it happens, but if the sort of the wildest speculations of the technos are correct, then they really are kind of approaching artificial general intelligence.

But in that case, it's in some ways worse for large language models because it means they really do intend to deceive us, right? And they do intend to produce text that looks as if it was written by a human when it wasn't written by a human. And that, that on our view, qualifies as hard. And so I think we think that that's kind of controversial because it, it requires certain claims about the nature of the mind. It requires certain complaints about consciousness. And so if they're conscious, they're hard bullshitters. But they are minimally soft bullshitters, precisely because at least as we see it definitionally, they cannot care about the truth.