🌻 What's your positionality, robot

> Imagine two researchers coding interviews about the cost of living. One grew up in a wealthy family, while the other experienced poverty first-hand. Their backgrounds will certainly influence how they code.

Nowadays, people are using AI for text analysis. Many of us worry about AI's "**hidden biases**". What to do about that?

Often there is no such thing as being objective, but at least we humans can be explicit about our positionality, our background and motivations, how this might affect our work, and how this relates to the positionality of our audience.

What about with an AI?

You can ask an AI to explain or reflect on its positionality and it will certainly give a plausible response, but remember that an AI has in fact very little insight into its own workings. Perhaps it will suggest always being aware that it was trained on a specific set of data which is not representative of the whole of humankind.

In any case the criticism that AI training data is not "representative" misses the point. Even if the training data had somehow been representative of the whole of humankind, that wouldn't make it "objective". It would simply reflect humanity right now, with all our quirks, biases and blind-spots. It wouldn't mean we don't have to worry about AI positionality or bias any more. It wouldn't (of course) mean we could rest assured that everything it does will be morally impeccable.

What's most unsettling about working with AI is not that secretly it's a bad person. The problem is that secretly it isn't any person at all. Even if it (sometimes) sounds like one.

A suggestion

A better suggestion is to be **more explicit about positionality in writing prompts and constructing AI research workflows**. Here is a very humble idea about how to start this experiment.

A simple example: I can tell my AI:

> When working, implicitly adopt the position of a middle-class white British left-leaning male researcher writing for a typical reader of LinkedIn. Don't make a big deal of this, but it might be helpful to know what your background is supposed to be before you start work.

And we can start to add variants of the kind of procedures which we humans might use when trying to address positionality:

In my AI workflow, I can then give another AI the same task but with a different starting position, and then perhaps ask a third AI (or a human!) to compare and contrast the differences. That also crosses over into ensemble approaches.

Of course, adding a phrase like "middle-class white British left-leaning male researcher" does not mean the AI will suddenly have all the relevant memories and experiences or really behave exactly like such a person. It's just a fragment of what we mean by "positionality". But *it's a start*.

Have you been experimenting with this kind of approach? We'd like to hear from you!

Footnotes

At Causal Map Ltd, we're working on an app called Workflows to make AI work more transparent and reproducible.

We've found that highlighting and then aggregating causal links is a great and relatively generic path to make sense of text at scale.

In terms of how to implement your workflow technically, see this great contribution from Christopher Robert.

See how we currently use AI in Causal Map here.

This post is based on my recent contribution to the NLP-CoP Ethics & Governance Working Group, along with colleagues Niamh Barry, Elizabeth Long and Grace Lyn Higdon.

This post was originally published by Steve Powell on LinkedIn and has been republished here. See the original article here