Chatting with Bots: AI, Speech-Acts, and the Edge of Assertion

Iwan Williams (Monash)
&
Tim Bayne (Monash/CIFAR)

Interactions between chatbots and humans are typically described as ‘conversations’.

This description reflects the phenomenology of these interactions – they feel like conversations.

But is that phenomenology accurate? Is it really possible to converse with a chatbot, or are chatbots only pseudo-conversationalists?

Human conversation involves a wide range of speech acts, but assertion is central - can chatbots assert?

> THESIS OF CHATBOT ASSERTION (TCA): Current-generation LLM-driven chatbots have the capacity to assert, and some of what they do qualifies as assertion.

There are 3 reasons for taking TCA seriously.

Motivation 1: chatbots can be a source of natural-language encoded information

Q: ‘What is the capital of Chad?'

A: ‘N'djamena is the capital of Chad’

Typically, information that is encoded in natural language involves assertions (or other speech acts), and so there is reason to think that Chatbot utterances can be assertions.

Motivation 1: chatbots can be a source of natural-language encoded information

Note that chatbots aren’t merely informative - being informative is part of their *proper function* (in some cases) (Butlin 2023; Millière & Coelho-Mollo 2023; Butlin & Viebahn 2024).

Chatbots are used in the ways that they are only *because* their outputs are sufficiently informative sufficiently often. The fact that they are sufficiently informative plays a role in explaining their (continued) existence.

Motivation 2: differing conversational modes

In addition to (apparently) asserting, chatbot output seems to also take other modes – chatbots can ask questions, engage in play and pretence, etc. It is natural to describe these changes in mode in illocutionary terms.

Motivation 3: appropriate behavioural complexity

Chatbots don’t merely generate assertion-like sentences in isolation, they display complex and flexible behaviour that closely resembles that of asserters.

They:

- *defend* their ‘assertions’ against reasonable challenges
- *explain* their reasons for thinking that the ‘asserted’ statement is true
- *retract* their ‘assertion’ if it is shown to be unsupported
- avoid blatantly *contradicting* themselves

The behavioural dispositions of LLMs far outstrips those of simpler systems (e.g. thermometers or talking clocks).

Upshot: there is significant motivation for TCA – it shouldn’t be dismissed…

But there are also significant objections to it – we turn now to them.

Objection 1: Chatbots aren’t capable of locutionary acts

Illocutionary capacities presuppose locutionary capacities – one must be able to intentionally express a natural language sentence with understanding.

That suggests two worries:

A: Chatbots don’t understand their outputs

B: Chatbots don’t produce their outputs intentionally

“…current methods of extracting information from text corpora have not yet formed knowledge bases that would be sufficient for conceptual combination and language understanding more generally.”

- Lake & Murphy 2021

Objection 2: Chatbots don’t meet the conditions imposed by Classical Speech Act Theory

Classic accounts of assertion (Grice 1957; Austin 1962; Searle 1969; Stalnaker 1978) require a range of sophisticated mental states for assertion.

At the first-order level, sincere assertions are typically held to be those which express a *belief.*

But if chatbots don’t have beliefs, they can’t make (sincere) assertions.

Objection 2: Chatbots don’t meet the conditions imposed by Classical Speech Act Theory

And so maybe chatbots can’t make assertions at all, if assertion is tied to ‘the notion of deciding to say something which does or does not mirror what you believe’ (Williams 1973, p. 146)

Objection 2: Chatbots don’t meet the conditions imposed by Classical Speech Act Theory

Often argued that assertion require various meta-representational states.

Grice (1957), for example, holds that assertion requires:

(i) an intention to produce a belief in a hearer

(ii) an intention that the hearer recognise this first intention

(iii) an intention that the hearer forms the belief partly *because* they’ve recognised this second intention

Even if chatbots have beliefs, they may lack the *metarepresentational* capacities required for assertion.

Objection 3: TCA is at odds with the normativity of assertion

Assertion is a norm-governed activity (Peirce 1932; Brandom 1994; Williamson 2000; Alston 2000).

Widely held that awareness of (or sensitivity to) these norms – and the ability to be sanctioned for flouting them – is required for the capacity to make assertions.

Objection 3: TCA is at odds with the normativity of assertion

However, it’s not clear that chatbots:

- stand in the appropriate relations to the norms governing assertion (they don’t understand them; they can’t follow them).
- can be sanctioned for flouting the relevant norms.

Interim Summary: Rejecting TCA (and holding that chatbots are no more assertors than thermometers) seems problematic…

At the same time, there are powerful objections to TCA…

What should we do?

Option 1: reject the motivations in favour of TCA?

Option 2: show that the objections to TCA are groundless?

Option 3: ‘split the difference’ between TCA and its denial?

Splitting the difference 1: Proxy assertion

In proxy-assertion, one agent’s illocutionary act involves another agent’s locutionary act.

Example: using a middle-manger to fire someone.

Nickel (2013) suggests that machines might be proxy-assertors: ‘ultimate responsibility for artificial speech does not lie with machines, but either with persons or companies, or with nobody at all’.

Others have picked up on Nickel’s proposal – e.g., Green & Michel (2022) and Arora (2024).

Splitting the difference 1: Proxy assertion

Does the ‘proxy-assertion’ proposal successfully split the difference between the pro and con positions?

Two problems:

- This proposal doesn’t address worries about whether chatbots can locute.
- The chatbot case doesn’t seem to involve an illocutor in the way that ordinary proxy-assertion does.

Splitting the difference 2: Fictionalism

Developed by Mallory (2024) and suggested by Hannah Kim (Wired, 4 June 2023).

Engaging with chatbots involves a form of prop-oriented make-believe on the model of puppets.

Fictionalism handles the objections to TCA neatly – we don’t need to ascribe understanding or complex mental states to chatbots.

Splitting the difference 2: Fictionalism

But what about the motivations for TCA?

- The problem is that fictionalism can equally apply to systems that are significantly simpler than state-of the-art chatbots (e.g. calculators).
- Thus, fictionalism doesn't do justice to the fact that there are *real* (non-fictional) differences between systems that are relevant to assertion.

We’ll end by sketching a third way of ‘splitting the difference’ between the pro-TCA and con-TCA camps.

Consider a vignette called ‘Grandma’

GRANDMA: You are the parent of a 22-month-old child called ‘Orla’. Coming home from work, you ask Orla, ‘What have you done today?’ She says, ‘Grandma!’

Question: Has Orla made an assertion?

It certainly *seems* like she has.

At the same time, variants of the three objections to TCA apply here.

- Are Orla’s utterances intentional? Does she really understand what she says?
- Does Orla have the meta-representational capacities that classical speech act theory takes to be required for assertion?
- Can Orla meet the demands imposed by the norms of assertion?

We suggest that young children constitute **edge-cases** with respect to assertion.

They are ‘in-between’ assertors – neither fully lacking the capacity to assert nor fully possessing it.

We might describe young children as ‘proto-assertors’.

(Note that neither the proxy-assertion nor the fictionalist proposal seems compelling when it comes to children.)

Might treating chatbots as ‘proto-assertors’ provide a viable way of splitting the difference between the cases for and against TCA?

Roughly, X is a proto-assertor if it has most of the features that characterize assertion to some reasonable degree.

Let’s return to the three objections to TCA.

Chatbots aren’t capable of locutionary acts (don’t understand what they say; can’t act intentionally)

- Being text-bound doesn’t necessarily prevent chatbots from a partial grasp of linguistic meaning.
- Besides, many chatbots aren’t text-bound .
- It’s true that LLMs are fundamentally next-word prediction devices, but it’s an open question whether intentional capacities might emerge from that underlying architecture.

Chatbots fail to meet the conditions on assertion imposed by classical speech act theory

- It’s a matter of dispute whether (current gen) chatbots have internal representational states (Marks & Tegmark 2023; Levinstein & Herrman 2024).
- Some evidence (Marks & Tegmark 2023) that certain activity in LLMs tracks the truth/falsity of inputs, and that this activity has a causal effect on the production of outputs (analogue of beliefs).

- But maybe we shouldn’t take proto-assertion to require the same meta-representational capacities (which are arguably missing in under 3s) as fully-fledged assertion
- Perhaps it requires only some degree of sensitivity to the informational state and requirements of one’s audience.

Chatbots fail to meet the demands imposed by the normativity of assertion

Difficult to evaluate this objection, for there’s little agreement as to what exactly requirements of normativity are.

How could asserting require an intention to follow or be bound by the norms governing assertion given significant disagreement among experts as to what those norms are (Pagin 2016)?

Chatbots fail to meet the demands imposed by the normativity of assertion

- Of course, one could argue that knowledge of these norms need be only implicit (Simion & Kelp 2018).
- But now it’s no longer obvious that chatbots flout the normativity requirement.
- Chatbots might not be sanctionable in the full sense that a human speaker is, but some may be appropriately responsive to correction (analogously to children).

Option 1: reject TCA

Option 2: reject objections to TCA

Option 3: ‘split the difference’ between TCA and its denial?

- Proxy-assertion
- Fictionalism
- Proto-assertion

Many thanks for your attention.