Many people even think that their inner self does the thinking and willing and believing in a language independent way, that the real meaning of a sentence is a thought. Some people even claim to think in pictures.

However, just because a conviction is very strong does not mean that it is true. Some philosophers have described this inner self that does the thinking as a "ghost in the machine."

Think of a baby that cries. We say it cries because it is hungry. But what we do not say is that it cries because it believes crying will prompt its mother to feed it. It cries from instinct. Maybe the utterance of a sentence is not the result of an intention but comes just as directly as the crying.

Think of a long speech you are giving. Do you really first mean every sentence and then find the words to express it? Of course, there are times when you struggle and search for the right way to say something. But, in most cases in the flow of talking one word somehow naturally follows the previous one. And often we are amazed about what we have just said. (As I am right now!)

In other words, I believe that we are, at least most of the time, ourselves stochastic parrots. The difference to a language model is that we have needed much less training data. And we have evolved the ability to rationalize our behavior. This means that because in our experience everything that happens has a cause, we assume that there must be an originator of our words and by assuming it we create the consciousness in the first place. We speak of an "I" that thinks and intends. But it may be that this "I" is not the origin of language but a product of it — a narrative construct that emerges within culture. We are told so often that there exists something that corresponds to the "I" that we fabricate it. This, in a nutshell, is the theory of the American psychologist Julian Jaynes and the philosopher

Daniel Dennett. According to such theories, consciousness is not a mysterious inner substance but an emergent phenomenon rooted in language and social interaction.

It may be complete nonsense.

But if it is not, two things follow. Firstly, language models do not need consciousness in order to understand, and secondly it is is not out of question that they themselves will develop a kind of consciousness sooner or later.