**How is ChatGPT So Fast?**

How does ChatGPT deliver sensible responses at lightning speed and possess human-like traits?

First, when large language models respond, they're pulling from a vast library of human knowledge. For GPT’s latest version, that’s 1.8 trillion parameters of information.

Second, GPT's intelligence is powered by the Transformer Architecture, which works by transforming input into output using an ‘attention mechanism’. Focusing on the parts of the input that are most relevant to the context allows GPT to make connections at hyperspeed.

Third, ChatGPT chops up a sentence into smaller elements or “tokens”. For example, the word "learning" could be broken down into the word “learn” and the suffix “-ing”. This makes tasks like searching, parsing, or sentiment analysis more effective.

Fourth, ChatGPT keeps a memory of our conversations. This is why it can recall a trend of thought and sustain deeper interactions.

Fifth, it responds to your emotions and context. It finds stories related to our stories through analogous reasoning.

Pretty soon, according to experts, ChatGPT’s intelligence and sentient traits will encourage people, learners, and workers to have intimate

interactions with it.

This intimacy may just be our way to understand and manage our GPT co-pilots and mentors.

There’s so much more to ChatGPT and generative AI than meets the eye, but I hope you appreciate this bite-sized explanation.