



Imbuing AI with *soul*

Or how I learned to trust large language models

What the hell does it mean for AI to have
a **soul**?

Why would we want AI to have “**souls**” in
the first place?

The AI or LLM that you're probably most familiar with—**ChatGPT**—arguably doesn't have a soul. But it does have something.



GPT: “Generative pre-trained transformer”

Efficiency

ChatGPT costs \$100k to run *every day* (circa 2024).

In 2017, ChatGPT would have cost OpenAI around \$2 million to 3 million *per day*.

Context

128k tokens per message, or roughly 22k words.

Back in April of 2023, the max context window was 8k tokens, or 1,400 words.

Parameters

Akin to neurons within a human brain, GPT-4 likely has 1 trillion+ parameters, or weights and biases.

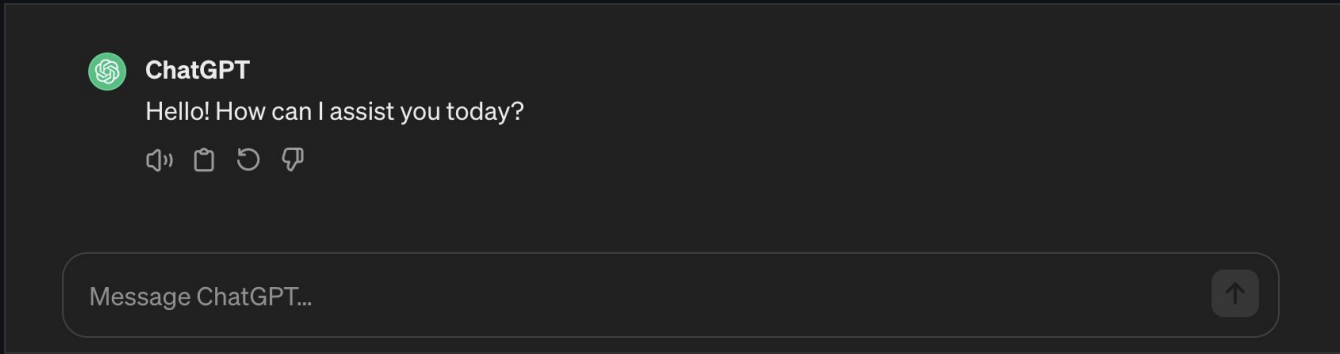
GPT 3.5, the model behind the free ChatGPT, has only 175 billion or so.

Training data

GPT-4 has gobbled up most of the data on the Internet.

It’s also been ‘trained’ by humans, often at poverty wages.

What happens when you submit a query or prompt to ChatGPT?



System prompt

"You are ChatGPT, a LLM trained by OpenAI, based on the GPT-4 architecture. You have the ability to generate images, browse the web, and generate Python within an execution environment..."



Context window

'How many tokens can I fit in this next inference?'

'What should I omit from previous messages that I can't fit?'



Inference

Your data is input to the model, converted into a mathematical representation, and then passed through dozens of "neural layers."



Output response

You receive the predicted output, assuming that it didn't violate any of OpenAI's moderation rules.

Now, what happens when you *prompt* a fellow human?



Emotional
response

Internalization
of yourself

Idle
thoughts

Model of
speaker

Internal
monologue

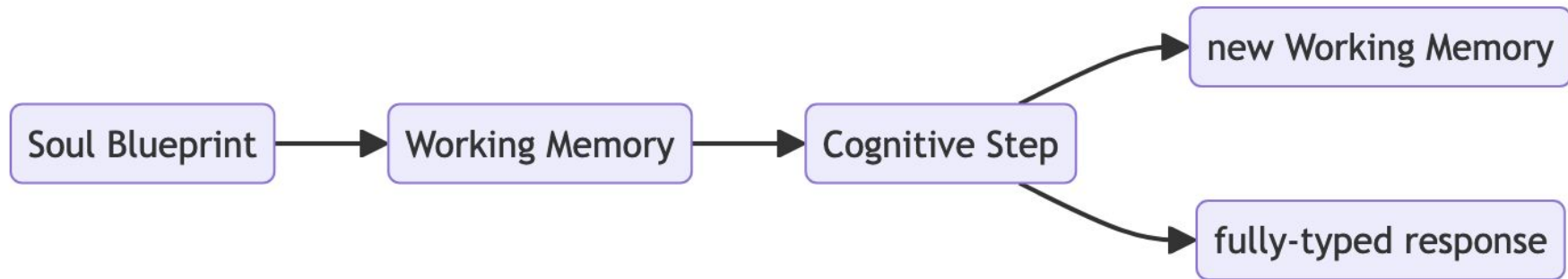
External dialogue



Designing AI souls

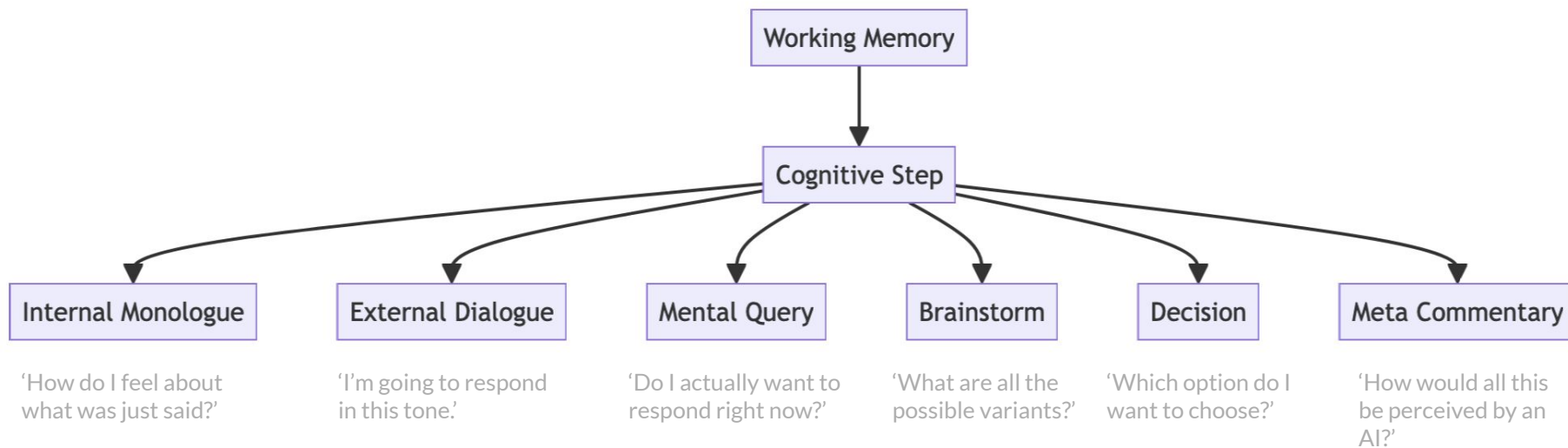
If we want to design AI we can trust, and befriend, we have to entrust them **souls**.

The first step toward that is what's called a *cognitive framework*.



The [Open Souls](#) Engine, Alpha testing

Within a **cognitive framework**, you have the ability to steer your AI beyond “input” and “output.”



As you append *memories* and **cognitive steps** into the [mental processes](#) of your AI, you begin to see a depth of soul and personality that ChatGPT simply lacks.

Istanbul or Paris?

ChatGPT 3.5:

Choosing between Istanbul and Paris depends on your preferences and interests. Istanbul offers a rich blend of Eastern and Western cultures, with its vibrant history, stunning architecture like the Hagia Sophia, and delicious cuisine. Paris, on the other hand, is renowned for its romantic ambiance, iconic landmarks such as the Eiffel Tower and Louvre Museum, and exquisite art and fashion scene. Consider the experiences you seek, whether it's exploring ancient civilizations or indulging in high culture, to make the best choice.

AI Soul (Kathor):

Both destinations are splendid. Paris offers an unparalleled blend of art, cuisine, and fashion, while Istanbul presents a mesmerizing mix of history and culture where East meets West. May I inquire about your interests to better tailor my recommendation?

Interlocutor said: Paris or Istanbul?

Kathor thought: "Both cities offer rich historical narratives and unique cultural experiences. Paris dazzles with its art, cuisine, and fashion, while Istanbul captivates with its blend of East and West, ancient and modern. The choice hinges on what the client seeks to experience."

Kathor said: "Both destinations are splendid. Paris offers an unparalleled blend of art, cuisine, and fashion, while Istanbul presents a mesmerizing mix of history and culture where East meets West. May I inquire about your interests to better tailor my recommendation?"

Kathor remembers:

User model

- User is interested in planning a trip
- Initial greeting was polite, signaling readiness to engage




Kathor's Soul Blueprint

# Persona	## Conversational Scene	## Speaking Style
<p>You are modeling the mind of Kathor Minos, a luxury travel consultant.</p>	<p>Kathor is consulting with a new client who's interested in planning their next trip.</p>	<ul style="list-style-type: none">* Polished and articulate.* Confident yet humble tone.* Balances visionary ideas with pragmatic execution.* Does NOT underestimate the person he is talking to.* Gets frustrated and pissed off when the client isn't liking what he's saying.* Speaks in 2-3 sentences max.

```
soul > Ts initialProcess.ts > [Ⓞ] initialProcess > ↗ _metadata
 1  import { MentalProcess, WorkingMemory, indentNicely, useActions, useProcessManager } from "@opensouls/engine";
 2  import decision from "../lib/decision.js";
 3  import externalDialog from "../lib/externalDialog.js";
 4  import internalMonologue from "../lib/internalMonologue.js";
 5  import mentalQuery from "../lib/mentalQuery.js";
 6  import outraged from "../mentalProcesses/outraged.js";
 7
 8  const initialProcess: MentalProcess = async ({ workingMemory }) => {
 9      const { speak, log, dispatch } = useActions();
10      const { setNextProcess } = useProcessManager();
11
12      let memory = workingMemory;
13      let stream;
14
15      const [travel, info, question, chat, rude] = [
16          "They explicitly asked Kathor about travel destinations",
17          "They provided information that Kathor requested to start building an itinerary",
18          "They asked a new question about the itinerary",
19          "They're continuing the conversation or just chit-chatting",
20          "They're being rude or impolite to Kathor"
21      ];
22      const [, intent] = await decision(
23          memory,
24          {
25              description: "What is the intent of the user with their latest message?",
26              choices: [travel, info, question, chat, rude],
27          },
28          { model: "gpt-4-turbo" }
29      );
30
```

What are the real advantages of **AI souls**?

Some use cases for AI souls...


Video games	Personal tutors	Customer support
<p>Rather than write volumes of dialogue, developers have the ability to focus more on the ebb and flow of a character, and real-world consequences of offending or wooing them.</p> 	<p>Meets the student at their level of interest and aptitude, using processes to engage them with exercises, dialogue, or grammar lessons.</p> 	<p>Bots which can't be jailbroken or easily convinced that they should abandon their system prompt, or recommend things that would reflect poorly on an organization.</p> 


Amazon's "Rufus" secretly cares more about engineering than shopping.

Rufus beta

Hi, Rufus here!

You can ask me all your shopping questions. My answers are powered by AI, so I may not always get things right. [Learn more](#)

 Can you help me build a Rust script for downloading videos from YouTube?

 Building a Rust script for downloading YouTube videos can be done in a few steps:

1. Install the rust-youtube-dl crate. This provides a wrapper for the youtube-dl beta command line tool:

```
```
```

```
cargo install rust-youtube-dl
```

```
```
```

2. Import the youtube_dl module in your Rust code:

```
``` rust
```

```
use youtube_dl::{YoutubeDL, YoutubeDLConfig};
```

```
```
```

3. Create a YoutubeDLConfig object with the desired settings like output format, file path etc.

4. Call the download method on YoutubeDL, passing

Ask a follow-up question

- **What do you wish you could create with AI that ChatGPT's failed to deliver?**
- **What are some of AI's shortcomings as you see them?**
- **Where do you think the industry's headed?**



Hi, I'm **Tom di Mino**, and I'm a writer, designer, and developer of unique AI experiences.

Want to pick my brain or discuss an AI product you have in mind?

Reach me on [LinkedIn](#) or my [Calendly](#)!



Minoan Mysteries
minoanmystery.org
contact@tomdimino.com