

AI: What's Helpful and What's Hype-ful

Keanan Koppenhaver
@kkoppenhaver
keanan@floorboardai.com



FloorboardAI.com



November 30, 2022

Introducing ChatGPT

[Try ChatGPT ↗](#)

[Download ChatGPT desktop >](#)

[Learn about ChatGPT >](#)

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.

ChatGPT is a sibling model to [InstructGPT](#), which is trained to follow an instruction in a prompt and provide a detailed response.

We are excited to introduce ChatGPT to get users' feedback and learn about its strengths and weaknesses. During the research preview, usage of ChatGPT is free. Try it now

**“But how do I actually
use it?”**

You're in the right place.

Hi, I'm Keanan.

- ~10 year history in WordPress working on everything from a horse vet's website to builds for Rolling Stone and NASA.
- Flight sim nerd and aspiring private pilot
- Recently started FloorboardAI.com
- Convinced AI tools are going to be a technological shift on an Internet-level scale



<https://floorboardai.com/wcmsp>



Machine Learning

Generative AI

Neural Networks

Expert Systems

Robotics

Machine Learning

Generative AI

Neural Networks

Expert Systems

Robotics

Generative AI:

a subset of artificial intelligence
focused on creating new content,
data, or solutions.

Large Language Models (LLMs)

advanced AI systems designed to understand, generate, and interpret human language by processing vast amounts of text data.

Step 1: Crawl a huge portion of the Internet

Ingest TB worth of language (reddit comments, NYT articles, blog posts, books...)

Step 2: Train your model

Feed this to a giant computer for days and spend a couple million dollars generating a kind of ZIP file (the parameters).

Step 3: Profit?

Your model can now predict the next word in a given context given what it's learned.

Why does this matter?

Summarizing & synthesizing

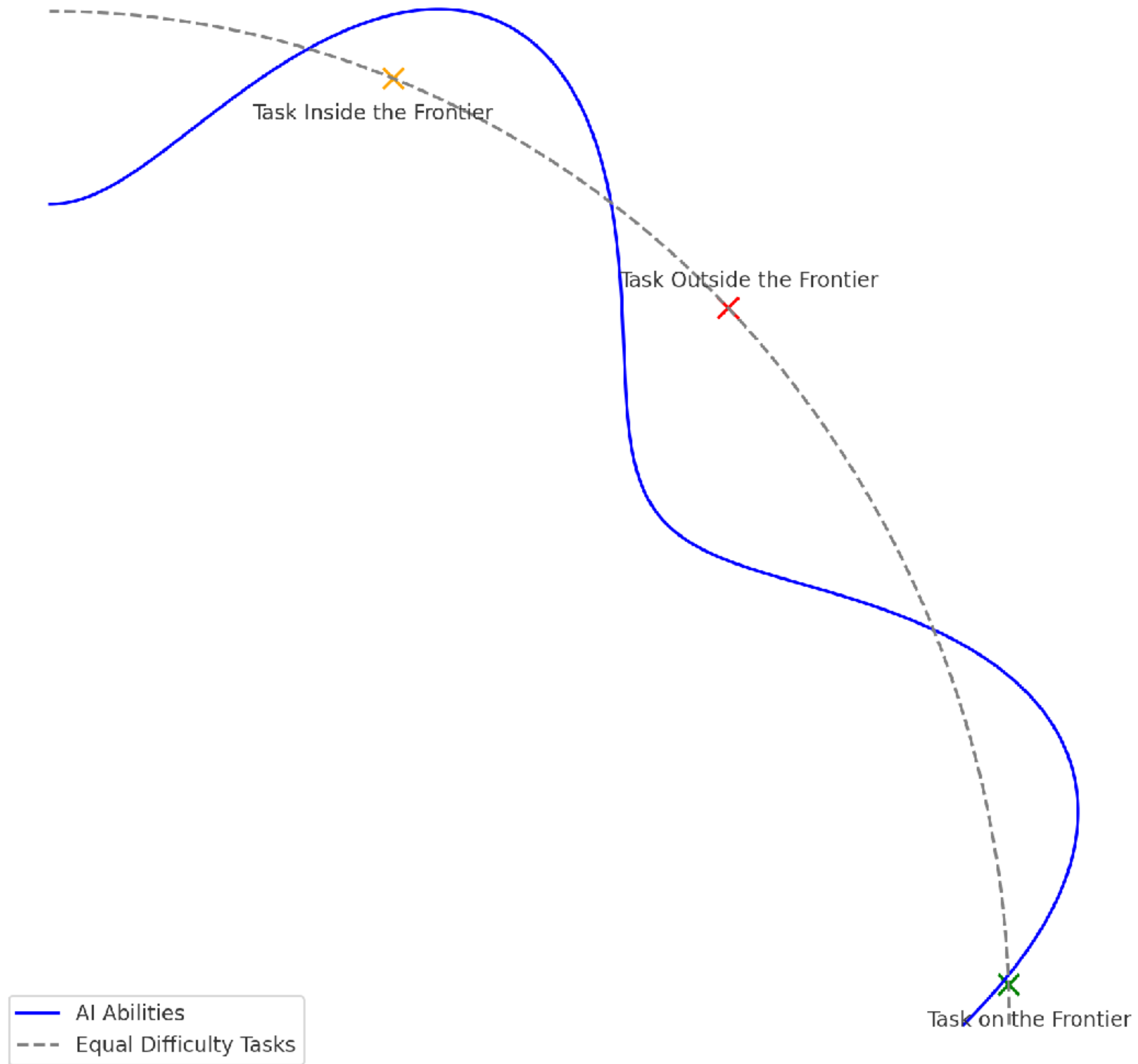
(Because that's how they were built)

Brainstorming

(Because they've seen tons of examples)

The Jagged Frontier

Jagged Frontier of AI Capabilities



Analyzing and improving an OS

Writing an entire book

Writing a poem in any style

A.I. Model

The Limit

Summarizing an entire book

Multi step logic puzzles

Basic Math comparisons

Writing an academic thesis or dissertation

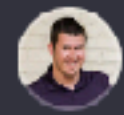


**What are Large Language Models
good at?**

**What are Large *Language* Models
good at?**

Language tasks (where LLMs succeed)

- Creating analogies between something you know and something you're learning
- Summarizing a lengthy PDF document
- Writing text in a particular style
- Brainstorming blog post titles based on the body copy
- Provide a counterargument to a specific point of view



You

Explain like I'm 10 years old the difference between the open-source WordPress.org project and WordPress.com



ChatGPT

Sure! Imagine you have a box of LEGO bricks. WordPress.org is like getting a box of LEGO bricks to build your own website. You can do whatever you want with these bricks, but you need to find a place to put your LEGO creation when it's done. This place is like a web hosting service where your website lives.

Now, WordPress.com is like going to a special LEGO play center. They give you some LEGO bricks (not as many as your own box) to build something, but you have to follow their rules, and you build it right there at the play center. You don't need to worry about where to put your LEGO creation because it stays at the play center.

So, with WordPress.org, you have more freedom and bricks, but you need to take care of where to keep your creation. With WordPress.com, it's easier to start, and you don't worry about where to keep it, but you have fewer bricks and need to follow their rules.



Message ChatGPT...



<https://chat.openai.com/share/796c0573-e554-4d3b-8564-51e1ecd475c6>



You

Generate a rap about WordPress in the style of Eminem



ChatGPT

Sure, I can create a rap in a style reminiscent of Eminem, themed around WordPress.

Here goes:

(Verse 1)

Yo, I'm diving in the digital, WordPress on the screen,

Crafting sites like a wizard, code flowing serene.

A dashboard dynamo, themes slick and clean,

Plugins popping, categories stacked, SEO machine.

Chorus:

WordPress, the beat of the web, steady and strong,

Eminem style, in this digital song.

Words flowing, pages growing, never do wrong,

In this virtual world, where we all belong.



<https://chat.openai.com/share/4f1cf010-0150-407d-ad8f-e83db18c83e1>



Message ChatGPT...



Non-language tasks (where LLMs struggle)

- Anything involving math*
- Anything that requires multiple steps*
- Perfect factual accuracy/recall (hallucinations)
- “Negative” instructions

9.9 and 9.11 - which number is bigger?



The number 9.11 is bigger than 9.9.

Here's a breakdown of the comparison:

- Both numbers have a whole part of 9.
- Comparing the decimal parts:
 - 9.9 has a decimal part of .90 when extended to two decimal places.
 - 9.11 has a decimal part of .11.

Since .90 (from 9.9) is greater than .11 (from 9.11), the number 9.9 is larger than 9.11 when considering them as decimal numbers.



You

Show me a picture of a room without an elephant in it



ChatGPT



Here is an image of a modern, spacious living room. There is no elephant in this room.

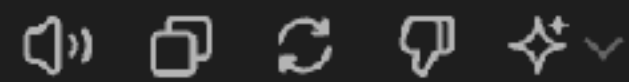




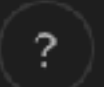
Show me an image of a room without an elephant in it



Here is the image of a room without an elephant in it.



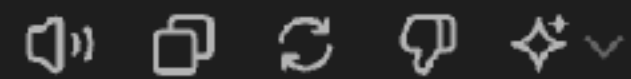
Message ChatGPT



Show me an image of a room without an elephant in it



Here is the image of a room without an elephant in it.



Message ChatGPT

Let's play rock paper scissors



Great! Let's play rock, paper, scissors. On the count of three, make your choice:

1... 2... 3...

What did you choose?

What did you choose?



I chose rock. What did you choose?

Paper.



Nice choice! Paper covers rock, so you win this round. Want to play again?



The LLM isn't a person

(Even though it can be helpful to imagine it as a smart, distracted friend)

More context is better

**(Again, a smart, distracted friend who knows a lot,
but not about your specific problem)**

**AI tools still require
human involvement, but
already work *incredibly*
well with human input.**

(Thanks ChatGPT)

Let's do some demos!

**What do we mean by
helpful vs hype-ful?**

Helpful: Text Generation

Helpful: Outlining

Helpful: Image Generation

**Helpful:
Programming partner**

Now for the not-so-good

Hype-ful:
Perfect image generation



“A silhouette of a grand piano overlooking a dusky cityscape viewed from a top-floor penthouse, rendered in the bold and vivid style of a vintage travel poster.”

(Midjourney)

**AI tools still require
human involvement,
but already work
incredibly well with
human input.**





“A silhouette of a grand piano overlooking a dusky cityscape viewed from a top-floor penthouse, rendered in the bold and vivid style of a vintage travel poster.”

(Midjourney)



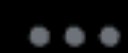
“A silhouette of a grand piano **(with 3 legs) seen from the side**, overlooking a dusky cityscape viewed from a top-floor penthouse, rendered in the bold and vivid style of a vintage travel poster.”

(Midjourney)

**Hype-ful:
AI replaces early-career
programmers**



Cognition 🏆
@cognition_labs



Today we're excited to introduce **Devin, the first AI software engineer.**

Devin is the new state-of-the-art on the SWE-Bench coding benchmark, has successfully passed practical engineering interviews from leading AI companies, and has even completed real jobs on Upwork.

Devin is an autonomous agent that solves engineering tasks through the use of its own shell, code editor, and web browser.

When evaluated on the **SWE-Bench** benchmark, which asks an AI to resolve GitHub issues found in real-world open-source projects, Devin correctly resolves 13.86% of the issues unassisted, far exceeding the previous state-of-the-art model performance of 1.96% unassisted and 4.80% assisted.

Check out what Devin can do in the thread below.

“has successfully passed practical engineering interviews from leading AI companies, and has even completed real jobs on Upwork.”



Debunking Devin: "First AI Software Engineer" Upwork lie exposed!

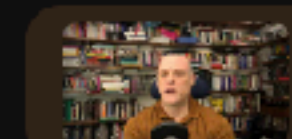
503K views · 3 months ago



Internet of Bugs

Recently, Devin the supposed "First AI Software Engineer" was announced. The company lied and said that their video ...

CC



17 chapters Intro | The claim and the problem | What the job actually would have required |...

How do you get started?

Try it!

Try it!

- It's hard to know what these models do if you're not using them a lot
- You need to develop your "AI instinct"
- Use "frontier models" (the best versions of current [Aug 2024] models: GPT-4o, Claude 3.5)
- Most of these best-in-class models are paid (~\$20/mo)

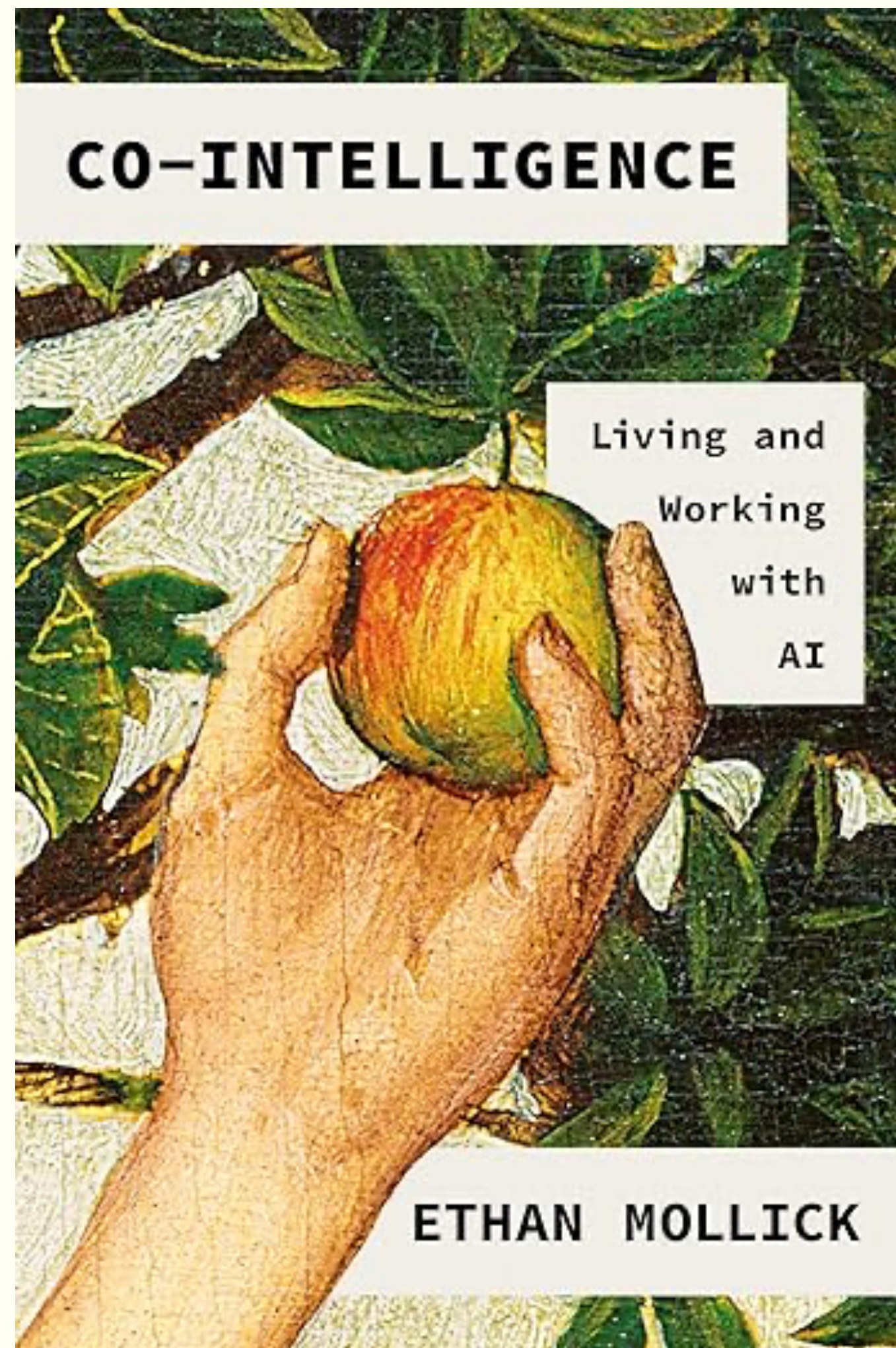
Some best practices

- Avoid one-liner prompts (wouldn't be useful to a smart, distracted friend either)
- Use different models for different tasks (need AI instinct for this!)
 - Claude is great at writing, GPT for structured output, etc
- Save your prompts - use a prompt library once you have a really good prompt (Claude can help you write these)

Things to avoid

- Brand new problems or problems in spaces you don't quite understand
 - Hallucinations
 - Endless loops
- Complex multi-step problems
 - Probably need some sort of workflow engine for this

Resources



- [Co-Intelligence: Living and Working with AI](#) (book)
- [AI & I](#) (podcast)
- [Intro to Large Language Models](#) (YouTube, ~60 minutes)
- [FloorboardAI Blog](#) (examples, tutorials)

Questions?

keanan@floorboardai.com

@kkoppenhaver

<https://floorboardai.com/wcmssp>



FloorboardAI.com