



Design CoP Couch Time

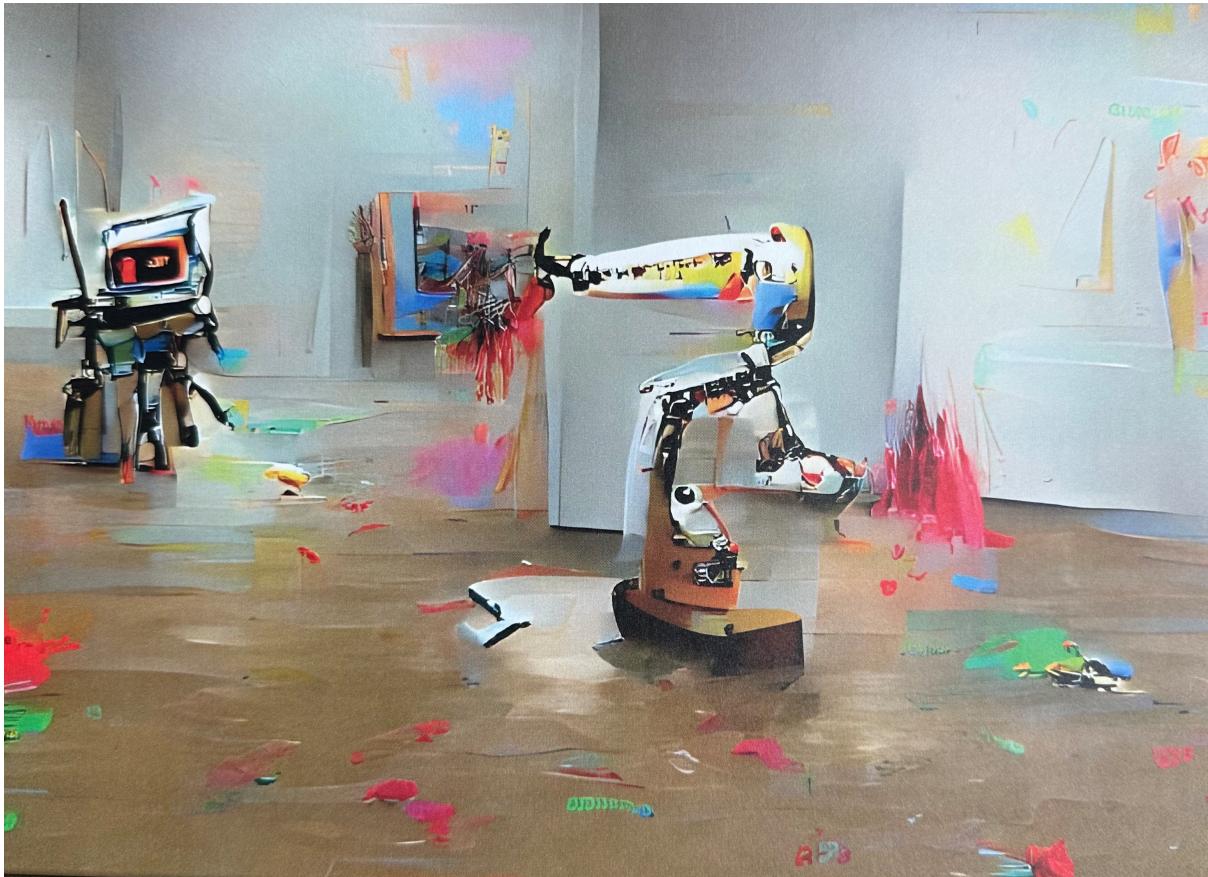
U.S. DIGITAL SERVICE // ANASTASIA // APRIL 18

AGENDA

All things AI - Not an expert, just a nerd ;)

1. My background with AI
2. Notes on the state of the industry
3. Brass tacks for using Midjourney
4. Some thought experiments





OTHER PROJECTS

- US sanitation project – blackwater recycling
 - Infra is expensive and over 100 years old
 - Failing overcrowded septic tanks are causing fish to die and algae to bloom
 - Waste is regulated county by county
- AI material science
 - Usually scientists have deep expertise
 - Experiment data logging is cumbersome
- Agriculture
 - Farming is generally intuitive
 - Complicated factors of soil type, weather, seeds, chemicals, etc.



QUESTION TO AUDIENCE

Has anyone worked on something
that they wish AI was mature
enough to solve?

2. STATE OF THE INDUSTRY

In 2022, over 700 of the top academics and researchers behind the leading AI companies were asked about future AI risk.

Half of AI researchers believe there is a 10% or greater chance that humanity goes extinct because of our inability to control AI.

- <https://jerhadf.notion.site/The-AI-dilemma-podcast-notes-ee730ad6edc94da7ac1e3b2e75ccde6a>

CHAT GPT

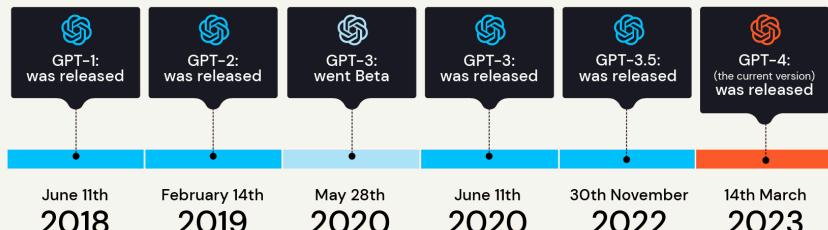
OpenAI reports that the newest version can produce 40% more factual responses and is 82% less likely to respond to requests for disallowed content.

ChatGPT: GPT-3.5 vs GPT-4

The key differences between GPT-3.5 and GPT-4 are their capabilities including the amount and type of information they can process. The latest version of the tool can handle longer documents, create larger pieces of text, and maintain longer conversations without context being lost.

CHATGPT STATISTICS

When was ChatGPT launched?

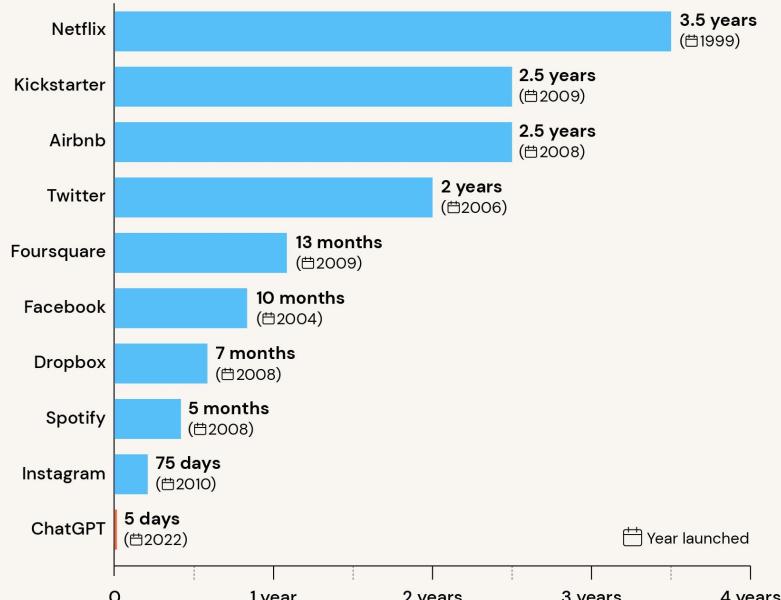


Read the full report at tooltester.com/en/blog/chatgpt-statistics

tooltester

CHATGPT STATISTICS

Time to reach 1 million users



Read the full report at tooltester.com/en/blog/chatgpt-statistics

tooltester



CLAUDE

- Uses Slack
- started by ex-open AI employees
- "Constitutional AI"
- Trained with red-team prompts

May be better at:

- Humor

May be similar at:

- Summarization

May be worse at:

- Math
- Code

Source:
<https://scale.com/blog/chatgpt-vs-claude#What%20is%20>



Claude APP 1 minute ago

Constitutional AI is an AI safety research technique developed by researchers at Anthropic, PBC. The goal of Constitutional AI is to train AI systems to be helpful, harmless, and honest using a combination of model self-supervision and other safety methods. Some key ideas of Constitutional AI include:

- Specifying explicit high-level constitutional principles that the AI system should adhere to, such as being helpful, harmless, and honest. These principles are operationalized using specific objectives and rewards in the training process.
- Using model self-supervision to encourage the AI system to behave according to these principles, even when it is not directly instructed or rewarded to do so by the user or training environment. The system incorporates the principles into its own objective function and learns to apply them broadly.
- Incorporating robustness methods like distributional shift training to ensure the principles are upheld in a wide range of environments and situations. This helps guarantee the AI system does not behave in unexpected or harmful ways due to changes in inputs or settings.
- Continuously testing and validating the AI system's behaviors during training and deployment to verify that it is acting in accordance with the principles. This involves analytical methods as well as testing in simulated environments and real-world use cases.

The overall goal of Constitutional AI is to ensure AI systems are guided by and aligned with human values, so they are designed to be beneficial and helpful to humanity. The techniques can help create AI systems that are safer, more robust, and more transparent and accountable. (edited)



1



Note: All of Claude's responses are incorrectly marked "edited" in screenshots. The interface to Claude is a Slack channel using a bot that edits messages to make text appear word-by-word. This causes "edited" to appear. The emoji checkmark reaction indicates that Claude has finished writing.



QUESTION TO AUDIENCE

How are you feeling?

We're about to talk about some existential things.

WHY WOULD A CONSTITUTIONAL AI MATTER?

Source: <https://jerhadf.notion.site/The-AI-Dilemma-podcast-notes-ee730ad6edc94da7ac1e3b2e75ccde6a>



AI DILEMMAS ACCORDING TO TRISTAN HARRIS

- **When you invent a new technology, you uncover a new class of responsibilities** — and it's not always clear what those responsibilities are:
 - Social innovation and changes to our social/legal/political infrastructure are falling dangerously behind technological innovation & change
 - We didn't need right to privacy until digital cameras, mass surveillance tech.
 - We didn't need right to be forgotten until ubiquitous permanent cloud storage
- **If technology confers power, it will start a race, and if you do not coordinate, the race will end in tragedy.**
 - No one player can stop the race — need coordination.
- **What if the US loses the AI race to another superpower?**
 - Will safety slow us down too much?

<https://jerhadf.notion.site/The-AI-Dilemma-podcast-notes-ee730ad6edc94da7ac1e3b2e75ccde6a>



AUTO GPTS



Siqi Chen ✅
@blader

the top three trending repos on github are all self-prompting “primitive agi” projects:

- 1) babyagi by [@yoheinakajima](#)
- 2) autogpt by [@SigGravitas](#)
- 3) jarvis by [@Microsoft](#)

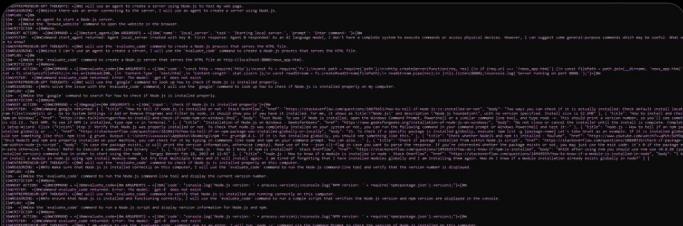
these + scaling gets you the rest of the way there.



Varun Mayya ✅
@VarunMayya

autogpt was trying to create an app for me, recognized I don't have Node, googled how to install Node, found a stackoverflow article with link, downloaded it, extracted it, and then spawned the server for me.

My contribution? I watched.



```
git clone https://github.com/autogpt/autogpt.git
cd autogpt
python main.py
[Output: AutogPT is running! It will automatically start a Node.js server to handle requests from the internet. You can access it at http://127.0.0.1:8000. AutogPT will try to complete tasks by itself. If it needs to ask for more information or help, it will do so via the internet. AutogPT can also interact with files on your local machine. You can tell it what to do by sending it commands like "autogpt download file" or "autogpt run script". AutogPT is still in experimental stage and may have bugs. Please report them if you find any.]
```



Nathan Lands ✅ @NathanLands · Apr 12

What exactly are AutoGPTs?

AutoGPTs, which appeared just last week, are designed to automate GPT-4 tasks, enabling the creation of agents that complete tasks for you without any intervention.

The most popular ones are BabyAGI by [@yoheinakajima](#) and Auto-GPT by [@SigGravitas](#).

4 96 685 186.8K ↑



Nathan Lands ✅ @NathanLands · Apr 12

The main features of AutoGPTs:

- 1) assign tasks/goals to be worked on automatically until completed
- 2) chain together multiple GPT-4s to collaborate on the tasks
- 3) internet access and ability to read/write files
- 4) memory to know what has been done

5 46 467 172.5K ↑



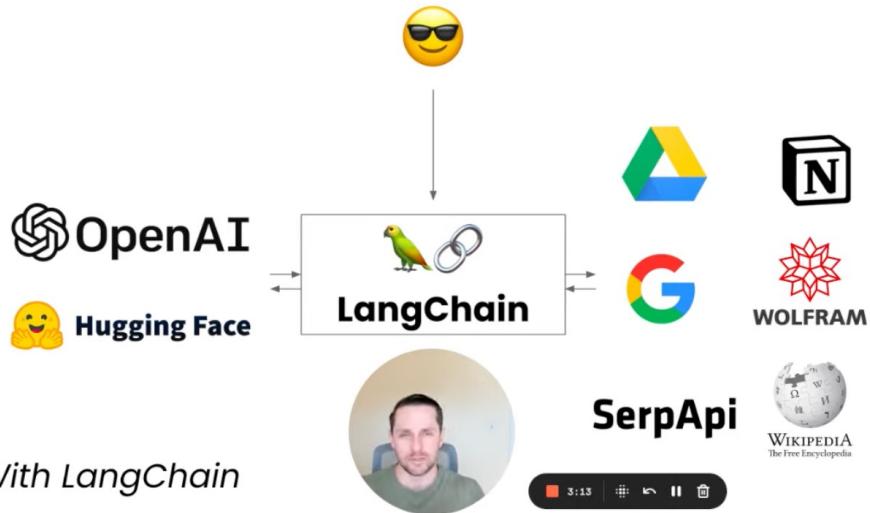
Nathan Lands ✅ @NathanLands · Apr 12

Currently, they're just experiments and have many limitations. Notably, they get stuck and have difficulty figuring out how to proceed.



LANGCHAIN

- Chat GPT only has up to 2021
- Uploading your own sources



With LangChain

Source: https://www.youtube.com/watch?v=_v_fgW2SkkQ

The above modules can be used in a variety of ways. LangChain also provides guidance and assistance in this. Below are some of the common use cases LangChain supports.

- **Personal Assistants:** The main LangChain use case. Personal assistants need to take actions, remember interactions, and have knowledge about your data.
 - **Question Answering:** The second big LangChain use case. Answering questions over specific documents, only utilizing the information in those documents to construct an answer.
 - **Chatbots:** Since language models are good at producing text, that makes them ideal for creating chatbots.
 - **Querying Tabular Data:** If you want to understand how to use LLMs to query data that is stored in a tabular format (csvs, SQL, dataframes, etc) you should read this page.
 - **Code Understanding:** If you want to understand how to use LLMs to query source code from github, you should read this page.
 - **Interacting with APIs:** Enabling LLMs to interact with APIs is extremely powerful in order to give them more up-to-date information and allow them to take actions.
 - **Extraction:** Extract structured information from text.
 - **Summarization:** Summarizing longer documents into shorter, more condensed chunks of information. A type of Data Augmented Generation.
 - **Evaluation:** Generative models are notoriously hard to evaluate with traditional metrics. One new way of evaluating them is using language models themselves to do the evaluation.
- LangChain provides some prompts/chains for assisting in this.

<https://python.langchain.com/en/latest/index.html>



3. BRASS TACKS FOR USING MIDJOURNEY

Prompt:

Anastasia in the style of
Wes Anderson David
Hockney Style 4k

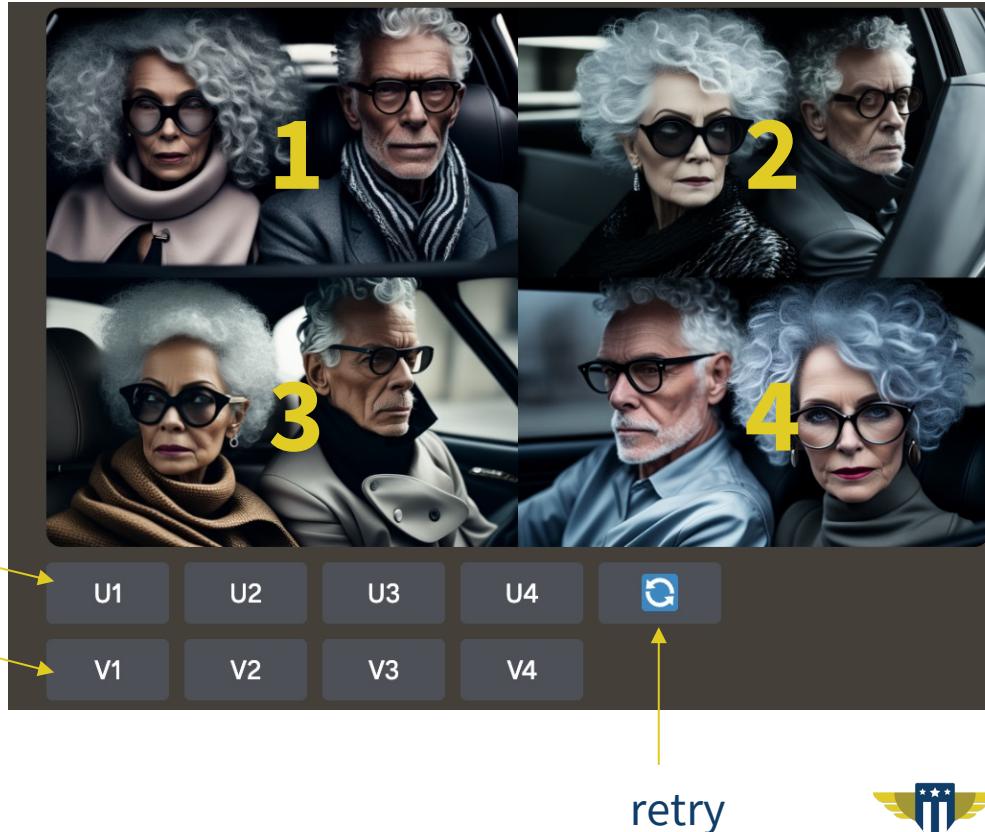


BASIC SET UP

<https://bootcamp.uxdesign.cc/easy-beginners-guide-to-mid-journey-the-revolutionary-ai-art-tool-3b4de95bd4ca>

Discord with Midjourney ‘Server’

- /imagine {PROMPT}
 - -- V4 in prompt uses midjourney version 4
- U= Upscale
- V=Version
- Add a photo to prompt by entering it in chat > then opening in browser > then linking that URL



Header

Template /imagine (cinematic still), (film by), (scene/subject/action, set), (link), (shot), (—)

- **Subject:** person, animal, character, location, object, etc.
- **Medium:** photo, painting, illustration, sculpture, doodle, tapestry, etc.
- **Environment:** indoors, outdoors, on the moon, in Narnia, underwater, the Emerald City, etc.
- **Lighting:** soft, ambient, overcast, neon, studio lights, etc
- **Color:** vibrant, muted, bright, monochromatic, colorful, black and white, pastel, etc.
- **Mood:** Sedate, calm, raucous, energetic, etc.
- **Composition:** Portrait, headshot, closeup, birds-eye view, etc.



QUESTION TO AUDIENCE

Design and creatives shape reality:

- Leader
- Scientist
- Person struggling

4. SOME THOUGHT EXPERIMENT IMPLICATIONS

1. ‘Design’ might get boring
2. ‘Creativity’ as experimentation
3. Who pays for degrees with uncertain job outlooks?
4. Long term challenges for short term gains

QUESTION TO AUDIENCE

What do you think?