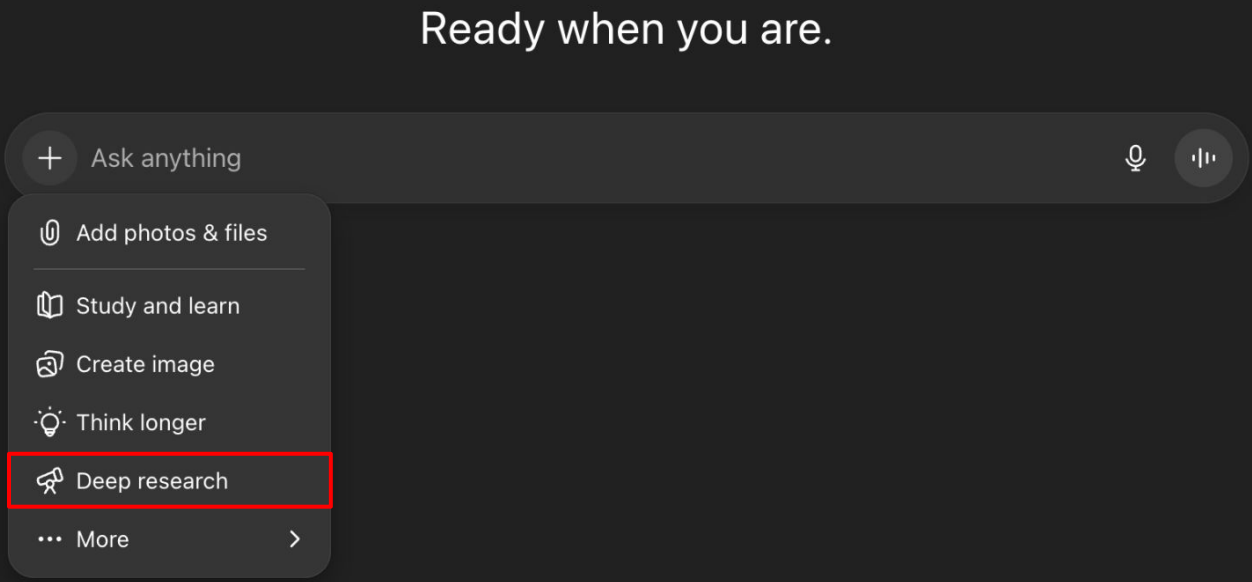
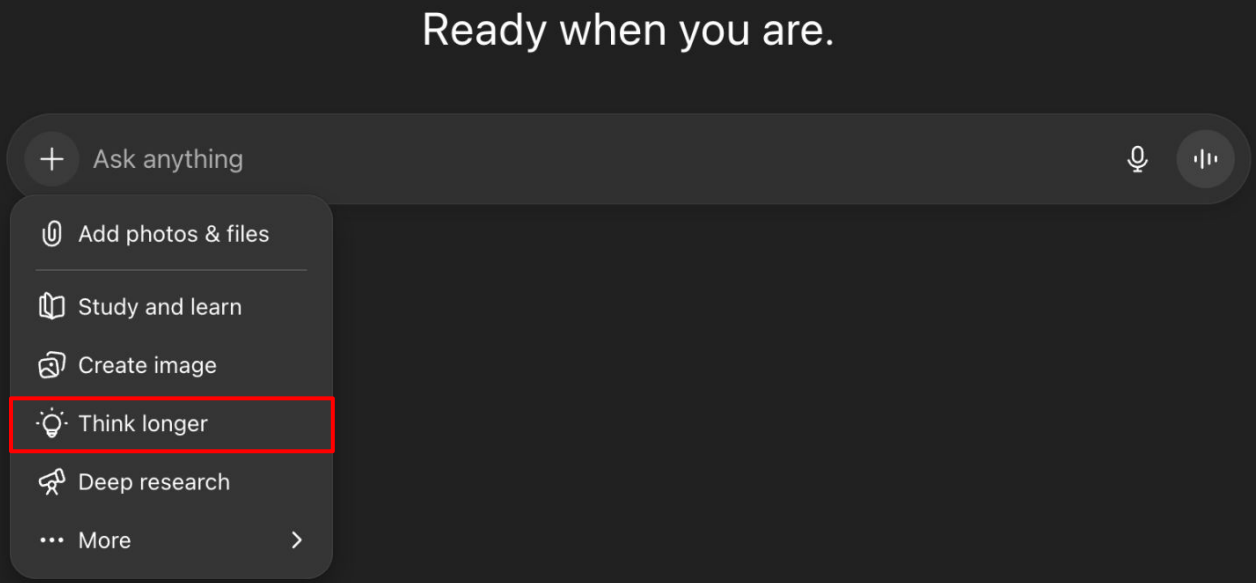
Deep Research


Ready when you are.
Ask anything
Add photos & files
Study and learn
Create image
Think longer
Deep research
More


Ready when you are.
Ask anything
Add photos & files
Study and learn
Create image
Think longer
Deep research
More

Deep Research Vs. Think Longer

- We can get a simple understanding of the differences by their “reasoning effort”. Let’s take a look at the different **routing modes**:
  1. **Normal**: only uses **one** or **two** internal steps to process and respond to your prompt.
  2. **Think Longer**: uses a **handful** or **two** of internal steps.
  3. **Deep Research**: can use **several hundred** internal steps.
- ChatGPT **routes** based on the **given** prompt.

When Should I Use Deep Research?

Deep Research is great for situations where:

- You have a very **complex** prompt.
- You need the output to be as thorough and **accurate** as possible.
- You want output that contains many **citations**.
- You are **not** on a **time** constraint.
- **NOTE**: Many platforms require **subscriptions** to access deep research.

What is Deep Research?

- Deep research is the practice of using an *Agentic Workflow* to:
  1. Break down the **complexities** of the user's instructions and requirements into **subtasks**.
  2. **Scour** the internet for **information** to understand the subtasks.
  3. (RECENT) Perform **parallel reasoning** to explore different potential outcomes.
  4. **Compose** a thorough, structured report on the given topics.

1. User Prompt into Subtopics

- Break the complex prompt into **smaller**, more **manageable** subtopics.
- In this way, the LLM can “frame out” its **plan** to answer questions from the user prompt.
- Framing helps with **understanding** of the **overall prompt** and determining what information needs to be **gathered**. More recently, framing helps with *parallel reasoning* and *parallel decoding*.

2. Scour the Internet for Information

- In order to respond most accurately to the prompt, the LLM uses *RAG* to **find** information from the internet.
- It consults multiple sources on the same subtopic, attempting to **resolve** any **conflicting** information.
- During information gathering, the LLM can **change** its own **thought/research** process.

3. Perform Parallel Reasoning

- Some LLM platforms have begun to use *parallel reasoning* when looking to both **gather** and **display** information.
- The idea is that the LLM puts on different thinking “hats” that change what it **believes** to be **important**.
- This is done in **parallel** with other instances at the same time.

4. Compose a Structured Report

- Once all information has been gathered, the LLM **assembles** a structured report in accordance with the **prompt instructions**.
- In the end, the LLM runs through all **instance solutions** and chooses the one that **fits** the situation the **best**.

The Reasoning Model

- Deep research uses a *reasoning model* to give the most accurate output.
- These models are trained mostly with *RL* as opposed to *SFT*.
- These models use **long** chain of thought (LCoT) that can be **thousands** of tokens long!

Let’s take a look at an example!