How To Think About Model Selection In ChatGPT

Model selection is part of understanding AI tokenomics: different models and reasoning levels use different amounts of computational capacity to complete a task. Using the most powerful option for every request can waste resources, while using too little reasoning can produce weaker results and require more iteration. The goal is to use the **lowest level of capability and reasoning that reliably produces the result you need**.

Although this resource focuses on ChatGPT, the same way of thinking applies across AI platforms. Gemini offers adjustable thinking levels, while Copilot distinguishes between faster responses and deeper reasoning modes, and similar tradeoffs appear across Claude's model ecosystem. Learn to recognize when a task needs speed, everyday reasoning, or deeper analysis, and you can make better model choices regardless of the platform.

Choosing Astra, Sol, Terra, and Luna

Here is a quick, visual summary of how the models compare. Page 2 goes into deeper detail.

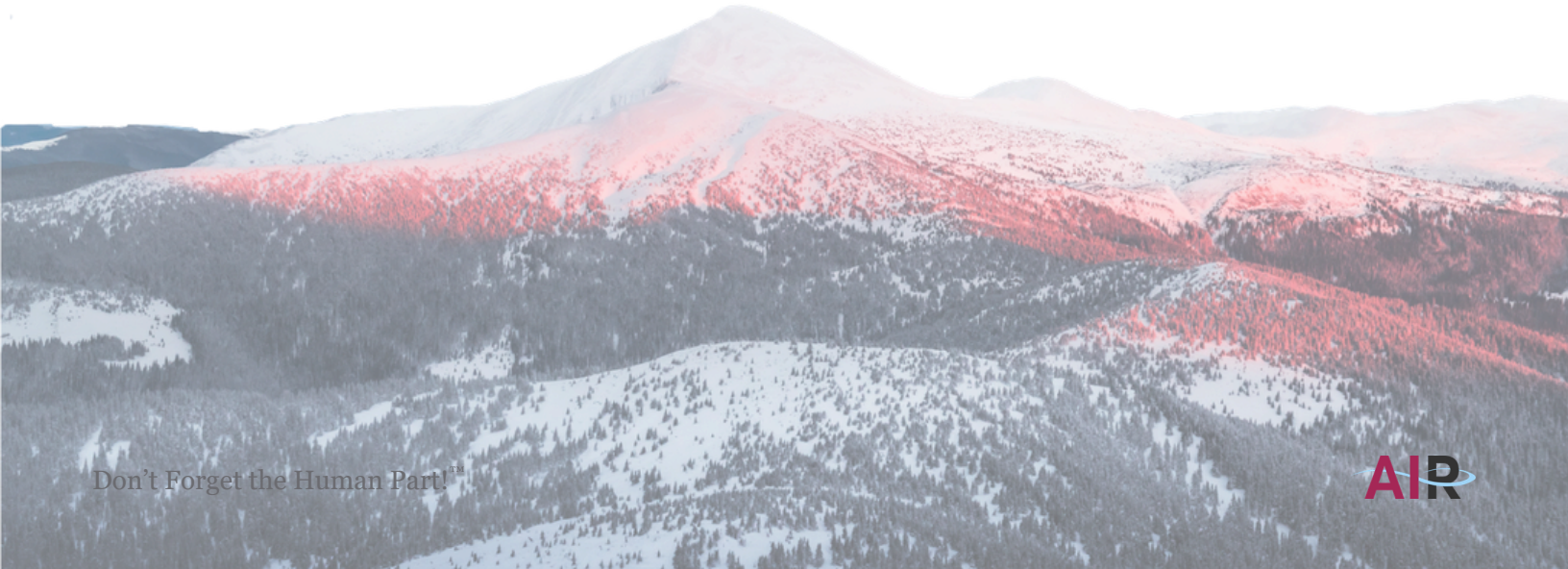
LEAST CAPABILITY NEEDED — MOST

Luna
Clear, repeatable tasks

Terra
Everyday work

Sol
Depth and polish

Astra
Strongest across steps and tools



WHERE EACH MODEL SHINES

Astra

HARDEST END-TO-END WORK

Complete workflows across code, apps, and research. Brief it with sources, templates, constraints, and checks.

- Build a feature end to end, then test it
- Multi-source research with a defined output
- Work where you want questions back first

Sol

COMPLEX, OPEN-ENDED WORK

Ambiguous or high-value tasks needing extra analysis, judgment, or polish. Define what done looks like.

- Complex code changes
- Deep research
- Client-ready documents

Terra

THE PRAGMATIC ALL-ROUNDER

Everyday work needing strong reasoning and tool use, without Sol's full depth. The default starting point.

- Drafting and editing with context
- Everyday analysis and planning
- Routine tool and file work

Luna

CLEAR, REPEATABLE TASKS

High-volume, specific tasks where you already know what a good result looks like.

- Extraction and classification
- Format and tone transformation
- Structured summaries at volume

Pick a reasoning effort

Use the lowest reasoning effort that produces the result you need. Increase it for tasks that need more planning, analysis, or checking.

Light / **Low**	Light in the ChatGPT desktop app, ChatGPT Work on the web, and IDE extension, or Low in the CLI, suits quick, well-scoped tasks.
Medium	Balances speed and depth for tasks that need more planning.
High / **Extra High**	Suit difficult work with multiple steps, sources, or tradeoffs.