

GPT-6 Astra

Prompting & Migration Quick Reference

Compiled from OpenAI's official GPT-6 Astra best-practices and migration guide (developers.openai.com/api/docs/guides/latest-model?model=gpt-6-astra). Every quoted prompt line and setting below is taken directly from that page.

1. Model overview & capabilities

GPT-6 Astra is OpenAI's most advanced model, built for computer use, browsing, software engineering, and professional workflows. Compared to earlier models it shows stronger alignment, more transparent reasoning, and better respect for task boundaries.

2. What's new in Astra

- Async tool calling - tools can run while the model keeps reasoning on other parts of the task.
- Mid-turn steering - real-time instruction updates over WebSocket preserve work already completed.
- Dynamic reasoning adjustment - change reasoning effort mid-conversation without invalidating the cache.
- Misalignment monitoring - built-in safeguards with asynchronous oversight.
- Limitations - no 'none' reasoning mode; Fast mode is unavailable for EU data residency.

3. Prompting best practices

Initiative & follow-through

Use prompts that emphasize autonomy and treat a capability request as authorization to do the work, without unnecessary approval gates.

```
"bias towards action and carry the user's intended task to completion"
```

Instruction following

Establish a clear hierarchy between user instructions and skill guidelines, and ask for transparency when the two conflict.

```
"The user's instructions take precedence over guidelines provided in a skill."
```

Personality & writing style

For prose-heavy output, ask for plain paragraphs over nested lists, and cut stock contrastive phrasing.

```
"Default to using clear, concise paragraphs" (avoid "nested lists unless hierarchy cannot be expressed clearly in prose")
```

```
Avoid "contrastive framing such as 'X, not Y.'"
```

Subagent delegation

Encourage the model to hand off parallelizable work to another agent instead of doing it serially.

"If at any point you can parallelize work by delegating tasks to another agent, you should do so."

Testing & verification

Calibrate how much test coverage you ask for based on the blast radius of the change.

"Do not write tests for reversible, low-impact changes that mirror the implementation."

4. Migration quickstart

- Set "model" to gpt-6-astra.
- Replace prompt_cache_retention with prompt_cache_options.ttl, set to "30m".
- Remove the temperature, top_p, and top_logprobs parameters.
- Use the Responses API for tool calling.
- To change reasoning effort mid-conversation, send a configuration_update item.

5. Limitations to plan around

- No 'none' reasoning mode is available on Astra.
- Fast mode is unavailable for EU data residency deployments.

Source: developers.openai.com/api/docs/guides/latest-model?model=gpt-6-astra
techbytes.app/posts/gpt-6-astra-best-practices-guide/

Guide: