How are resources used by agents?

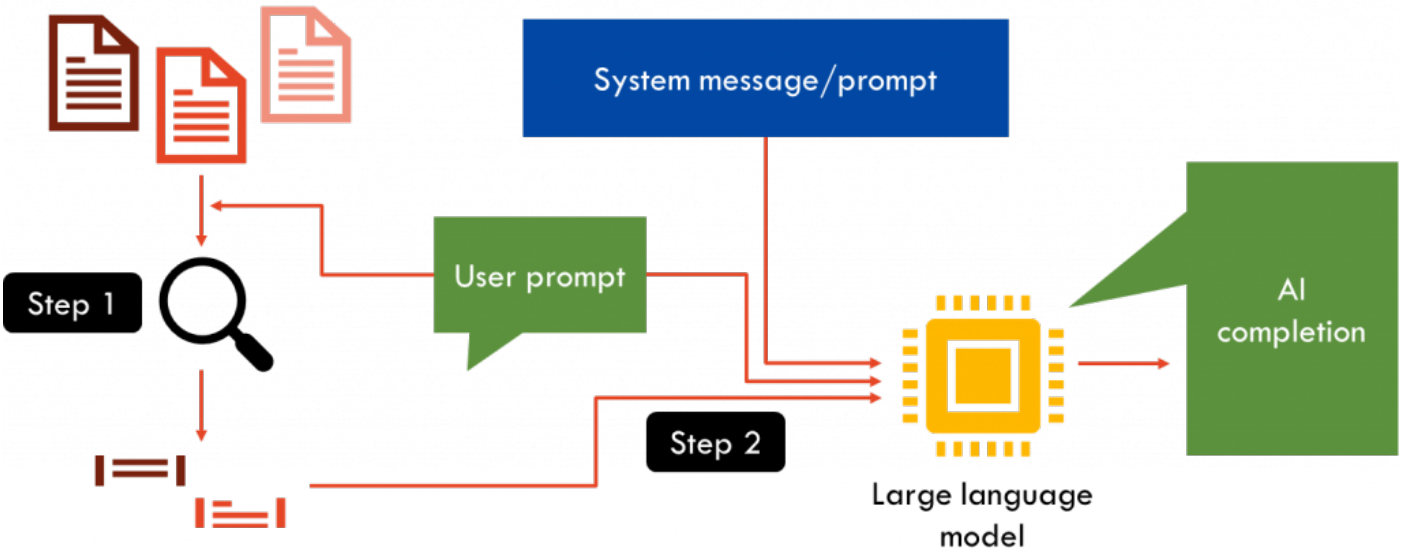
Agent fundamentals · https://cogniti.ai/docs/how-are-resources-used-by-agents/

Resources are documents that can be provided to assist Cogniti agents in creating more contextualised and accurate responses. If you use resources in agents, it's important to understand how Cogniti agents use them.

Importantly, telling your agent to "refer to the resources" or "use the resource notes.pdf" or "use the docs provided" will **not** work. This is because resources are **not** read in their **entirety** when the agent generates a response.

How are resources brought into an agent?

When a user types in a prompt, Cogniti uses this approach:



Step 1 - The **user's prompt** is used to perform a text search over the agent's resources. This search looks for textual and semantic similarity. For example, if the user's prompt is "What do chickens eat?", Cogniti will search through that agent's resources to find a few snippets (basically, paragraphs of text) that match that prompt. It may find nothing, or it may find irrelevant snippets - this very much depends on the user's prompt. For example, if the user's prompt is "Teach me everything", the snippets that Cogniti finds will likely be rather irrelevant.

Step 2 - The top few snippets that are found (ranked according to relevance) are then fed into the large language model (e.g. GPT-4), along with the system message, and the user's prompt. The large language model then generates the response ('completion') for the user.

Implications

What this means for the behaviour of your agent

There are a few implications for using resources in Cogniti agents:

- Adding resources does slow down completions slightly
- Only snippets of the resources are sent to the large language model for consideration, **not** the full resource

What this means for the design of your agent

If you have information that is critical to the function of your agent, ensure you put that information into the system message.

In your system message, do not refer to your resources - this will not make sense to the underlying AI, because all it sees are little snippets that are derived from the resources you provide. The underlying AI will **not** know what you mean if you say in the system message, "Refer to the document provided" or similar.

Importantly, the system message is **not** used to search over the agent's resources. Even if you tell the agent in the system message to use the resources, this does not mean it will. The system message has **no** bearing over how the agent engages with the resources provided to it.