

Fix Prompts or Context Instead of Output of AI

There are user prompts and system prompts. When the output doesn't work well, we should add the rule in the user prompts instead of fixing it every time.

As mentioned in a previous essay, Names in the Translation Output of AI, there are many duplicate Chinese names. Even humans can't always translate English names to Chinese names correctly, so it's even more difficult for machines.

It's not hard to define the rules for name translation. In real life, context is often needed. For example, in a class of about 30 people, if two people have the same English name (Pinyin translation), it becomes hard to translate between them. Additional information, such as their appearance, is necessary.

However, in certain contexts—like someone's blog, a lecture platform, or a user's contact list—the list of possible target names becomes much smaller. In these cases, AI can translate perfectly.

Therefore, we should define these rules in our API calls. Don't change the output; fix the root cause instead of adjusting the temporarily generated content. Focus on fixing the reason, not the result.

AI code editors are smarter than AI chatbots because they have a larger context. This allows them to make better inferences about their output.

The same applies to images, audio, and video. These tools should provide a larger context. For example, if AI creation tools are given a bunch of videos, audio clips, and podcasts, they can generate new ones much more effectively.

This is similar to RAG (Retrieval-Augmented Generation). For AI creation tools, they must strike a balance—creating results that are neither too specific nor too generic. However, they should offer tools or functionalities to make the results more specific. ChatGPT, for instance, has project functionality that lets you upload files and then interact with them. Other creation tools need similar features.