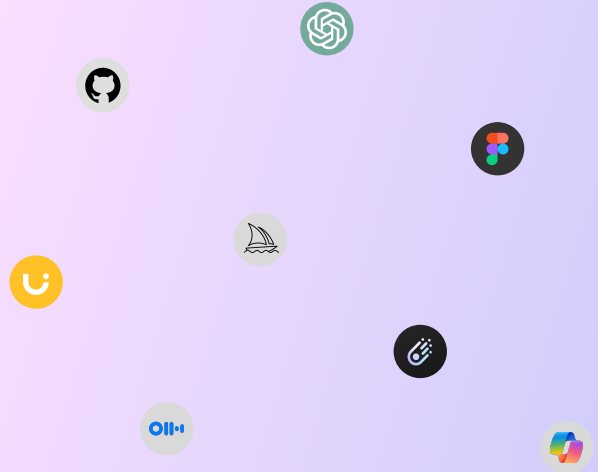


What is the next step for AI interfaces?

September 2024

By Erich Gherbaz, Pranali Shevatekar



- Introduction
- A New Paradigm
- Conveying User Intent
- Multimodal or Hybrid AI Interfaces
- Closing thoughts
- References

With the rapid growth of AI technologies, the question that remains is: where would they go next? How would they continue to improve our users' experience? But perhaps a better question to start with is what has AI done for our users until now? Well, so far these products have been used in a myriad of ways to provide users with the ability to be experts or develop skills in many areas. Consider tools such as ChatGPT, Midjourney, Siri, Alexa, Dall-E, Grammarly, among others, and how they enhance the user's skill to do a specific task.

Authors



Erich Gherbaz
UX Res. & Design



Pranali Shevatekar
Behavioral Analyst

But what is the AI really doing? How is the user interacting with the computer now in comparison to before?

The different ways through which users interacted with a computer throughout history are known as User Interface Paradigms of Computing. In a brief explanation by Jakob Nielsen, they are:

Paradigm 1

Batch Processing

Essentially, users gave a complete set of instructions for a computer to do (as a deck of punched cards, most commonly) through a data centre and was processed at some unspecified time.

Paradigm 2

Command-Based Interaction Design

In short, all we knew before AI came along. Here the user and the computer would alternate, each giving one command at a time.

In essence, the big change with AI is the interaction between the user and the tool, which is different from previous models or paradigms in that the computer must interpret what the user wants and not what they say. This is such a difference from previous models that Jakob Nielsen has called it the first new UI paradigm in 60 years.

A New Paradigm

This new way of interaction between user and computer is no longer defined by needing to tell the computer what to do, but rather recognize what outcome the user wants. In short, it's '**intent-based outcome specification**' as described by Nielsen.

But what does this mean? And why is it so important? Simple, with current generative AIs, the user tells the computer the result they seek but do not specify how this should be accomplished. The AI in response finds patterns of previous and most common answers and gives back a response to the user. However, **this response is determined by what intent the machine could grasp from the user's command.**

Let's take ChatGPT as an example, given that it's likely the most well-known AI. A user can ask any information from the computer, but the computer's response may vary based on the input given or the intent behind the question.

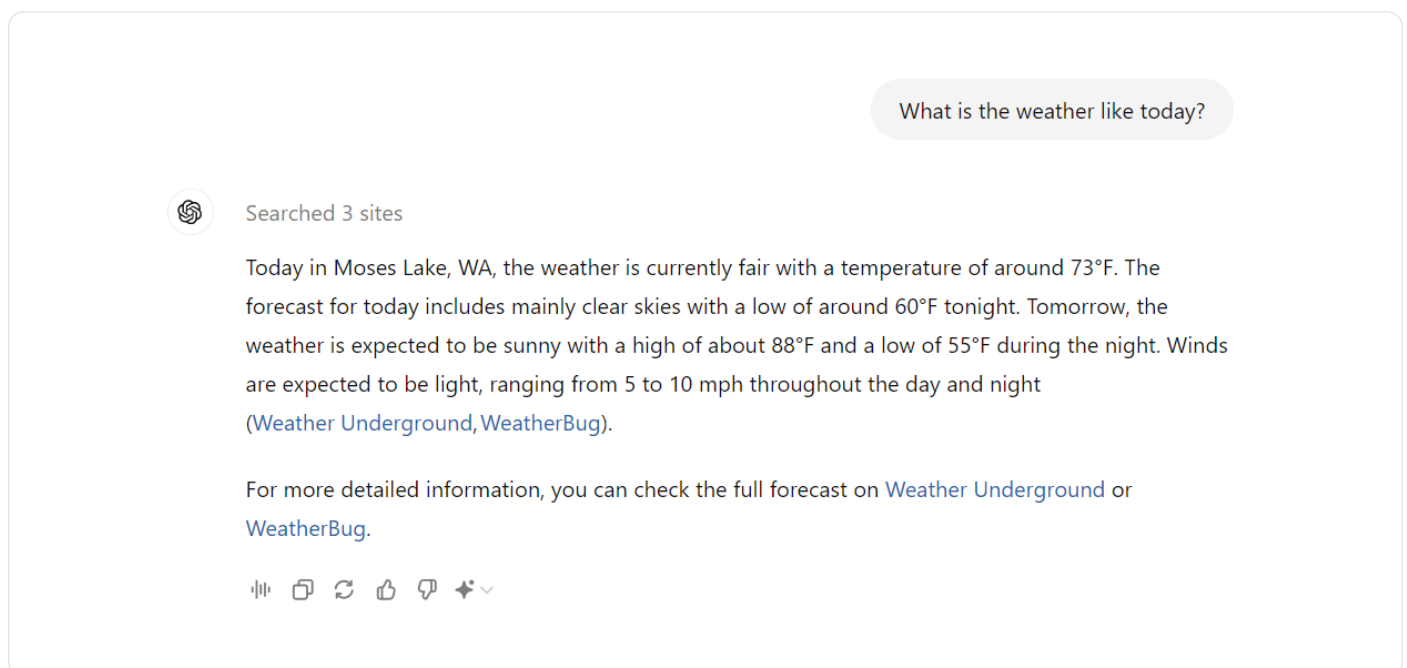


Figure 1 - The user asks, "What is the weather like today?" The AI responds with the weather conditions for a specific location, Moses Lake, WA (Washington, USA). Giving details such as temperature, forecast, expected weather for the following day, and wind information.

Observe Figure 1, here I have simply asked the AI 'what is the weather like today?' as any user might do. However, ChatGPT is not aware of my current location, and has decided, instead of asking for more details, to provide me with a general answer for a different location. It knows part of my intent: know the weather; but it is not aware of other characteristics of my question.

As a result, I am being forced to specify more details in my prompt so that ChatGPT can be better aware of my desired outcome. Therefore, you will see in Figure 2 that I needed to provide the AI with my current location for it to know better my intent.

What is the weather like today in Zurich, Switzerland?



Searched 4 sites

Today in Zurich, Switzerland, the weather is mild with a current temperature of around 21°C (70°F). It's mostly fair with light winds coming from the southeast at 2 km/h. The humidity level is quite high at 84%, and the air pressure is steady at about 1014.9 mbar. The visibility is clear up to 8 km, and there is no precipitation currently. The UV index is low, which indicates minimal risk of harm from unprotected sun exposure.

Later in the day, temperatures are expected to rise to around 26°C (79°F) with a chance of scattered thunderstorms in the afternoon. The chance of rain later today is around 40%, with light and variable winds throughout the day([Time and Date](#), [Weather Underground](#), [Weather Atlas](#)).

For more detailed and updated information, you can check the [Weather Underground](#) or [Weather Atlas](#).

Figure 2 - The AI is responding to the user's more specific weather prompt about the weather in Zurich, Switzerland. Now giving the same level of weather forecast details but based on the location provided by the user's input.

Now, the question that arises is: how can we help users to easily convey their intent to the AI they are interacting with?

Conveying User Intent

We keep mentioning this big word: **intent**. Let's define it a little bit. By intent we understand the goals defined explicitly or implicitly by the users when they give a task to the computer.

In some ways, finding out these goals sounds impossible. Fully being able to catch the true intent of a user might be quite a difficult thing to do, but perhaps we can circumvent this issue for now. Instead, we can **provide ways that would make it easier for them to express their goals**.

But how? Well, as it always happens when new technology emerges, no one really has the answer. We have theories, conjectures, ideas of what might be. Only time will tell. However, one thing is for certain, prompt applications currently have a disadvantage, people are not expert prompt writers. So, what are some potential answers that companies are researching right now?

Multimodal or Hybrid AI Interfaces

According to IBM's AI trends for 2024, the next wave of AI advancements will focus in more than one specific domain, but on models that are multimodal which would allow them to take **multiple types of data as input**. Such is already the case for tools like ChatGPT, Gemini, Claude, among others. This, of course, includes text-to-image models and speech-to-text models that would be able to move naturally between natural processing language (NLP) and computer vision tasks, while bringing video as an addition. **This would in turn make AIs more intuitive, with users having the ability to ask about a picture or give instructions out loud and receive a proper response in turn.**

In this aspect we can mention AI voice assistants on a category of their own. Currently, there are big advancements in more sophisticated voice AI. There's a bet that this will become the primary interface for AI interactions, which is a more natural and intuitive way to interact with an assistant. Part of the promise is that these would become AI companions, attempting to blend seamlessly into our lives.

But how will this help to understand the user's intent more? Simple, these models can describe in words what they see in various formats. These formats can be used to train the AI, therefore expanding the number of sources with which to train it. **The more access to different sources these models have, and the more formats to gather data they can use, the more they learn.** Being trained on multi-modal data gives them a wider range of options when deciding from what to generate, how to edit, and to what to serve.

Other types of hybrid AIs are those that involve graphic user interfaces (GUI). These models are also a consideration when discussing the future, because they combine the powerful capabilities of AI with the user-friendly nature of graphical interfaces. One of the biggest advocates for this type of interface is Jakob Nielsen, who being a 'fan' of GUI, believes that AI interfaces will maintain elements of both intent-based and command-based interfaces while still retaining many GUI elements. Nevertheless, how would they work? The idea is that GUI could lessen the users' memory load by favouring recognition over recall. It could allow users to select options without having to recall 'fancy' terminology.

According to Tarun Mugunthan, Research Assistant for the Nielsen Norman Group, **hybrid interfaces would be able to address important usability issues**, such as recall and discoverability in AI image-generation tools. But given that these problems of discoverability and recall are shared by many other generative-AI tools with prompt-based interfaces, including text generators like ChatGPT and code generators like Github Copilot, they would also benefit from this. This way, gaps in knowledge and awareness can be resolved by surfacing possible commands and actions such that users may discover new ways of using the tool.

Closing Thoughts

Then, what is the next step for AI interfaces? AIs are increasingly becoming more complex. What seems to be the answer to our question is the fact that **AIs will become interfaces with multiple interactions available to them.** As we have seen, these interfaces are called hybrid or multimodal interfaces.

Hybrid interfaces are in essence a combination of different ways to interact with the computer. Imagine that you can work with multiple sets of data inputs, such as images, sounds, text, or video. All in the same place, with the same AI. You could communicate with your AI assistant by voice command and ask questions about an image, and it would be able to understand and respond accordingly. You wouldn't need to type commands, although this still could be a possibility, but rather you could simply talk, and it would respond naturally like a human assistant. This would make every interaction easier and more natural. In addition, the AI could provide tools like prompt generators or images to help or provide users with new information to describe better their commands. **This would make a more user-friendly interaction between human and AI, empowering users with different interaction methods into one system.**

References

- [Andreessen Horowitz - How Generative AI Is Remaking UI/UX Design – by Jennifer Li and Yoko Li - 2024](#)
- [IBM – AI Voice Assistants Evolve, Promising Deeper Interactions - 2024](#)
- [IBM's AI trends for 2024](#)
- [IBM - Top 6 predictions for AI advancements and trends - 2024](#)
- [Jakob Nielsen – AI: First New UI Paradigm in 60 Years – 2023](#)
- [Pragmatic Coders - Designing AI interfaces: Challenges, trends & future prospects - 2024](#)
- [Tarun Mugunthan – Overcoming the Articulation Barrier in Generative AI Using Hybrid Interfaces – 2023](#)