Learning to Prompt: Human Adaptation in Production with Generative AI*

Sijie Lin†

Version: November 2025

Please click here for the latest version

Abstract

What is the role of human input in AI-assisted production? Humans interact with generative AI through combinations of words called prompts. A key feature of *human* input is *adaptation*: users dynamically modify their prompts based on their understanding of AI. I empirically investigate two types of adaptation: (1) adaptation to new AI versions, referring to how people change their prompts in response to AI upgrades; (2) adaptation to outputs from previous prompts, referring to how people adjust their prompts iteratively to converge on desired outcomes. I study this adaptation using prompt-level data from Midjourney, a leading AI image generator. First, users adapt to AI upgrades by writing different words in their prompts. By submitting prompts written for the old version to the new AI and vice versa, I decompose the output shifts as arising from prompt changes (73%), AI changes (20%), and an AI-human cross effect (7%), implying complementarity between AI and human inputs. Second, prompts evolve within the creative process of an artwork. I estimate a structural model of the creative process using the sequential search framework. Counterfactual shows that without human adaptation, users need three times more prompts to achieve data-observed results. Both results highlight the importance of human judgment and adaptation in the creative process.

Keywords: Artificial Intelligence, Human Adaptation, AI-Assisted Production

JEL: D80, D24, L86, O30, M15,Z11

*I am grateful to my advisors, Matthew Mitchell, Heski Bar-Isaac, Avi Goldfarb, and Ambarish Chandra, for their guidance and support. I also thank Charles Hodgson, Victor Aguirregabiria, Yanyou Chen, Yao Luo, Siyuan Liu, Irisa Zhou, Ruiqi Sun, Hong Luo, Kevin Bryan, Stephan Heblich, Tommaso Alba, Devin Bissky Dziadyk, Ben Couillard, and the participants in the IO seminar at the University of Toronto for their helpful comments. I thank Ali Berk Senbas for his insights. All errors are my own.

†Rotman School of Management, University of Toronto (sijie.lin@mail.utoronto.ca).

1 Introduction

In recent years, generative AI has advanced rapidly, especially in language-based models. Between November 2022 and October 2025, five major versions of GPT have been publicly released. Language-based AI image, video, and music generators are also developing rapidly, with each new version demonstrating greater capability than its predecessor. Generative AI is now being used for creative tasks, including creative writing, email drafting, and artistic production. These were once considered tasks that could only be completed by humans. Creative industries have increasingly adopted generative AI into their workflows. For instance, AI companies are partnering with studios in film, gaming, and news to create illustrations and animated characters. With AI being widely adopted so quickly, what is the role of human input in AI-assisted production?

In the context of language-based AI, human input refers to prompts, the combinations of words that humans use to interact with AI. A key feature of *human* input is *adaptation*: users dynamically modify their prompts based on their understanding of AI. This involves *judgment* of the AI output and beliefs about how each new prompt impacts the subsequent output.

In this paper, I empirically examine two specific types of adaptation. The first is adaptation to different AI versions. As AI becomes more capable and understands broader vocabularies, users systematically change the words in their prompts. Changes in both AI capabilities and prompts jointly influence the generated images. By resubmitting old prompts to the new AI and new prompts to the old AI, I decompose the outcome image shifts into effects of AI improvements, prompt adjustments, and residual factors. The result implies complementarity between AI and human input.

The second form of adaptation concerns how users respond to the outputs of previous prompts within the creative process of an artwork. Initial outputs from AI rarely meet users' requirements and need further adjustments. Users typically refine prompts iteratively based on previous results to converge towards desired outcomes. This adaptation directly impacts the creative process with generative AI. I formalize this behavior in a structural model of sequential search and evaluate the importance of adaptation using a counterfactual, in which users cannot modify prompts based on judgments of prior outputs. The result suggests that adaptation guides the creative process and accelerates progress toward the same outcome.

The empirical setting is Midjourney, a leading AI image generator. Midjourney operates in a public chat channel on Discord, a messaging platform. In this environment, users submit prompts in the chat box to the AI, which then returns generated images in exchange for a fee. This setting enables me to observe the submitted prompts and generated images. The

richness of this data allows me to study how humans interact with generative AI in the following two ways.

First, I examine user adaptation to different AI versions by analyzing how prompts change when AI upgrades. To measure AI capabilities, I use text embeddings from large language models. Text embeddings, a standard technique in textual analysis, represent texts as numerical vectors such that semantically similar texts are placed near each other in this vector space. I measure alignment between prompts and generated images as the cosine similarity between two embeddings: (1) the embedding of the submitted prompt and (2) the embedding of the textual description of the generated image produced by GPT. A higher cosine similarity indicates that Midjourney's output more closely reflects the user's intended meaning. Applying this measure, I find that the generated images in the new AI (Midjourney V4) align significantly more closely with their prompts than those in the old AI (Midjourney V3). New AI's outputs are more consistent with the user's intended meaning.

When AI capabilities change, the words that users include in their prompts also change systematically. For example, users include more words about colors and photography in the new AI than in the old AI, such as "neon color" and "shutter speed". This is because the new AI understands these words better. Conversely, users include fewer words like "detailed" and "realistic" because the new AI automatically makes images more detailed and realistic without users mentioning these words in the prompt.

Consider the creative process as a production function, where the inputs are AI capability and human input, and the output is the generated image. Both AI capability and human input change simultaneously in the new AI, jointly changing the generated images. To decompose these changes, I resubmit prompts written for the old AI to the new version and vice versa, collecting the resulting images, which are then converted into embedding vectors using large language models. I project the vectors onto the direction of $\overrightarrow{Image}_{new} - \overrightarrow{Image}_{old}$, which captures the overall shift in image output between AI versions. This decomposition reveals that 73% of the changes in images result as a consequence of changing AI alone, 20% from changing AI alone, and 7% from residual effects. These findings demonstrate that human input remains highly relevant even as AI capabilities advance. Simply applying historical prompts to new AI cannot replicate the current outputs. The positive signs of these contributions indicate that AI and human input are complements rather than substitutes. When AI learns to understand more words, users begin using those words more often to take advantage of its expanded vocabulary. An AI upgrade increases the return of human input.

Second, I turn to the analysis of the creative production process. I define a session as a sequence of highly correlated prompts submitted by a single user within a short time frame, representing the creative process of an artwork. Within each session, I analyze how users

adapt their prompts in response to the images generated from previous prompts.

Using reduced-form analysis, I document the following patterns in prompt construction within a session: (1) prompt length increases incrementally; (2) prompts are path dependent; (3)users prioritize adjusting words with higher weights before addressing less influential words; (4) prompts become increasingly similar to the final prompt in the session. These patterns, together with anecdotal evidence, suggest that users iteratively adjust prompts after seeing the generated images from previous prompts and converge toward their desired outcome.

Motivated by these data patterns, I develop a structural model of the creative process using the sequential search framework from Weitzman (1979) and Hodgson and Lewis (2025). In the model, individuals choose a prompt to submit to the AI based on their current beliefs about the expected image quality of prompts. They are uncertain about the generated image quality until they observe the image. Observing an additional image requires paying a search cost, which includes the mental effort required to construct prompts and judge image quality, the fees paid to the AI company, and the waiting time. They then update their beliefs about the expected returns of prompts in their consideration set and either stop or choose the next prompt accordingly. The consideration set is defined as a local set of prompts similar in content to others within the same session, representing ideas that belong to the same artwork. When a prompt is submitted to the AI and the individual observes a less satisfying image, they infer that similar prompts are more likely to produce unsatisfying results. These prompts will be chosen with lower probability in the next search. Individuals stop searching when no remaining prompts are expected to generate sufficient value to compensate for the search cost. The trade-off is between the search cost and the option value of the next prompt.

A counterfactual analysis investigates how removing adaptation from previous prompt outputs affects the prompt construction process. In the counterfactual scenario, users can only submit a list of prompts to the AI at the start of a session. They do not adjust the order of the prompts in the list when they see the generated image sequentially. The results show that eliminating judgment on image quality increases the number of prompts per creative process by 313% to achieve the same outcome as in the data, implying that adaptive learning plays a crucial role in guiding efficient prompt searching behavior. This finding quantifies the critical role of human adaptation: the ability to learn from prior outputs and iteratively guide the AI makes the creative process more efficient.

In summary, I measure the impact of human adaptation in AI-assisted production. When AI upgrades, human judgment remains essential for exploiting new technological capabilities. In addition, human adaptation in the creative process guides the direction of creation and enhances the efficiency of production. Together, AI as a technology expands the frontier of

what can be produced, while human adaptation determines how these new capabilities are utilized effectively.

Literature Review

This paper connects to three strands of literature. First, I contribute to the literature about AI's impact on productivity in the workplace and labor market (Agrawal, Gans, and Goldfarb, 2019; Acemoglu and Restrepo, 2020; Acemoglu, Autor, Hazell, and Restrepo, 2022; Agarwal, Moehring, Rajpurkar, and Salz, 2023; Angelova, Dobbie, and Yang, 2023; Dell'Acqua et al., 2023; Noy and Zhang, 2023; Peng, Kalliamvakou, Cihon, and Demirer, 2023; Babina, Fedyk, He, and Hodson, 2024; Choi and Schwarcz, 2024; Vaccaro, Almaatouq, and Malone, 2024; Brynjolfsson, Li, and Raymond, 2025; Kanazawa, Kawaguchi, Shigeoka, and Watanabe, 2025). Several studies focus specifically on generative AI and the creative process (Zhou and Lee, 2023; Zhou, Lee, and Gu, 2025).

This literature primarily compares outcomes with and without AI adoption, typically holding AI capability fixed. These studies examine predetermined tasks and overlook the possibility that users may pursue more complex goals as AI capabilities improve. Rather than comparing outcomes with and without AI, I examine how users adapt their behaviors across different AI versions, conditional on AI adoption in Section 4. I show that both AI capabilities and human inputs evolve simultaneously when AI upgrades. I decompose their relative contributions to creative outcomes.

Second, I contribute to the recent work examining usage patterns in generative AI (Chatterji et al., 2025; Handa et al., 2025). This literature analyzes conversational data between users and AI to understand the purpose of usage. Handa et al. (2025) document that AI usage concentrates primarily in software development and writing tasks. Chatterji et al., 2025 find that users mostly employ generative AI to seek information and refine writing.

While I also analyze conversational data, my focus differs. Rather than examining why users employ AI, I investigate how users interact with AI. I show that users adapt to generative AI by systematically modifying the words in their prompts. When AI upgrades, users adjust their prompts to better leverage the new capabilities. Users also iteratively refine prompts within the creative process to converge on desired outcomes.

Third, this paper is related to the sequential search model literature (Weitzman, 1979; Kim, Albuquerque, and Bronnenberg, 2010; Bronnenberg, Kim, and Mela, 2016; Chen and Yao, 2017; De los Santos and Koulayev, 2017; De Los Santos, Hortaçsu, and Wildenbeest, 2017; Dzyabura and Hauser, 2019; Honka, Hortaçsu, and Wildenbeest, 2019; Hirsch, Guy, Nus, Dagan, and Kurland, 2020; Ursu, Wang, and Chintagunta, 2020; Hodgson and Lewis, 2025) which provides the foundation for the structural model in Section 5. Weitzman (1979) propose the theoretical framework of the sequential search model, in which individuals search

sequentially across potential research projects. Individuals prioritize projects with the highest expected returns and gradually proceed down the ranked list until no remaining project offers sufficient expected returns to justify the search cost. This framework has been widely adopted in consumer search contexts, where consumers search sequentially for products to purchase (Kim et al., 2010; Chen and Yao, 2017; De los Santos and Koulayev, 2017). These papers estimate search cost distributions using consumers' browsing history data. While earlier work assumes static beliefs about unsearched products, more recent studies incorporate learning into consumer search models, allowing beliefs to update along the search path (De Los Santos et al., 2017; Dzyabura and Hauser, 2019; Ursu et al., 2020; Hodgson and Lewis, 2025). For example, a consumer encountering negative reviews for one camera of a certain brand may update the beliefs about all cameras of this brand. The most closely related work is Hodgson and Lewis (2025), in which consumers update their beliefs about camera quality during the search process. Upon observing a satisfactory camera, consumers infer that similar cameras are also likely to be satisfactory, increasing the probability of searching related products in subsequent periods.

I adapt this framework to the creative context by interpreting the image generation process as a search for the optimal prompt. In the structural model, individuals hold prior beliefs about prompt quality based on prompt attributes. They pay a search cost to observe and evaluate the generated image, then update their beliefs about prompts in their consideration set.

The remainder of the paper is structured as follows. I introduce the empirical setting and data in Section 2 and Section 3. Section 4 presents the analysis of human adaptation to AI upgrades. In Section 5, I build a structural model of the creative process. Section 6 concludes the paper.

2 Empirical Setting

The context studied in this paper is Midjourney, one of the leading AI text-to-image generators, which operates on a messaging platform. In this context, both the prompts submitted by users and the corresponding image outputs are observable.

Midjourney allows users to submit prompts, typically text and sometimes image references, to generate images using its AI model in exchange for a fee. Prior to August 2024, Midjourney was only accessible on a public chat channel on Discord, a messaging platform that typically allows game streamers to communicate with each other in real-time. After Midjourney opened its own web interface to all users in August 2024, a significant number

of users have continued to use Discord to access Midjourney's services. There are two major reasons for the partnership between Midjourney and Discord. Anecdotally, during its beta phase, Midjourney operated as a small firm with limited financial support. Operating with Discord's infrastructure can be cost-saving. Besides, according to an interview of the company's founder and CEO, David Holz, a key motivation was to build an interactive community where users could discuss ideas and prompt crafting techniques.[1]

The public chat channel allows any user to join, and users see each other's actions. It enables researchers to observe the creative process of art generation using AI. Figure E2 in the appendix shows an example of Midjourney usage. When a user submits a prompt through the chat box, the AI returns four generated images within minutes, along with some interactive buttons. For each prompt and each button clicked, I observe a timestamp and a corresponding user ID.

This empirical setting offers several benefits for research. First, it permits direct observation of human interaction with generative AI. This feature allows me to analyze how users iteratively adjust their prompts and gradually converge to their desired outcomes. In many other creative settings, researchers only observe the input of the creative process, such as the monetary investment, the number of scientists and their characteristics, and the output of the creative process, such as patents or journal publications. The creative process itself remains unobservable. However, in this setting, a significant part of the creative process can be directly observed.

Second, during the observation period, four new versions of Midjourney were released. These changes enable me to investigate how users modify their prompts in response to AI updates. In the literature, researchers have been utilizing the variation of production between with and without AI (Angelova et al., 2023; Dell'Acqua et al., 2023; Zhou and Lee, 2023; Noy and Zhang, 2023; Peng et al., 2023; Babina et al., 2024; Choi and Schwarcz, 2024; Vaccaro et al., 2024; Brynjolfsson et al., 2025; Kanazawa et al., 2025; Zhou et al., 2025). These papers generally do not capture how users adapt their behaviors as AI evolves. This setting allows me to understand how users dynamically adjust their behaviors in response to AI version changes, conditional on already using the technology.

Third, this setting allows researchers to observe two counterfactuals directly without imposing model assumptions. In conventional settings, researchers can observe outputs from old machines, old labor input $F(K_{old}, L_{old})$; and the outputs after firms adopt new technology and labor is adjusted based on new technology $F(K_{new}, L_{new})$. By contrast, this setting allows me to observe outputs from new machines with old labor input $F(K_{new}, L_{old})$,

[1] https://www.theverge.com/2022/8/2/23287173/ai-image-generation-art-midjourney-multiverse-interview-david-holz

and old machines with new labor input $F(K_{old}, L_{new})$. This is achieved by submitting the old prompts to the new Midjourney version and submitting new prompts to the old Midjourney version. By observing these counterfactuals, I am able to decompose output changes into effects from prompt changes and effects from AI changes.

Finally, Midjourney is both influential and representative. Although there isn't an official announcement from Midjourney revealing its market share, Midjourney has been one of the "most popular AI image generators" upon search results, along with DALL-E, Stable Diffusion, Adobe Firefly, and Leonardo AI.[2] By October 2025, Midjourney has become the largest Discord server by number of users.[3]

It is useful to understand who uses Midjourney and the purpose of using it. Although Midjourney does not have an official report about the composition of users, there is a "micro-polls" and a "discussion" sub-channel in the Midjourney Discord server.[4] There are also some online discussions on social media like Reddit, Twitter, and Instagram. Additionally, SimilarWeb, a data analytics company, provides information on what other websites Midjourney users frequently visit. Together, these sources help us understand the user composition. Some screenshots of the "micro-polls" and the SimilarWeb report can be found in Appendix D.

Many Midjourney users possess formal art skills. They self-reveal that they are currently working in or used to work in the art or design industries. They own professional drawing tablets like Wacom and maintain Adobe Photoshop subscriptions. Apart from Midjourney, they also frequently browse other AI tools like ChatGPT, graphic design platforms, and art resource websites such as Behance, e-commerce platforms for selling custom products like T-shirts and mugs, and freelancer websites like Fiverr. Many of these users are hobbyists and professional users. There are some other types of creators, such as fiction writers and musicians. There are also small business owners.

Users engage with Midjourney for various purposes. Many of them are using Midjourney purely for fun. Some users are using Midjourney to create decorative images for social media posts, wallpapers, and T-shirts. Sometimes they sell these creations. Some users use Midjourney to accelerate the sketching process. They experiment with combinations of colors and compositional elements before transferring the concepts to physical canvases. Some creations are for commercial purposes, such as book illustrations, music album covers,

[2]https://www.demandsage.com/ai-image-generators/; https://www.cmswire.com/digital-marketing/a-marketers-guide-to-the-best-text-to-image-generators/; https://www.forbes.com/lists/ai50/

[3]https://backlinko.com/discord-users; https://influencermarketinghub.com/discord-stats/

[4]These polls are usually initiated by Midjouney's CEO David Holz. According to the description of this sub-channel, Midjourney is using it to construct larger questionnaires, which he later publishes in the "announcements" sub-channel for users to fill out.

and logo designs. In some cases, users employ Midjourney as a source of creative inspiration. For example, fiction writers use it to visualize scenery, and musicians generate visual concepts to inform their musical compositions.

3 Data

The data includes all publicly available messages from the Midjourney server on Discord, between June 2022 and January 2025. Each message sent by the Midjourney Bot represents an action of a user. An action could be submitting a prompt to Midjourney or clicking a button below the generated images. For each action, I can observe the timestamp, user ID, submitted prompt, generated images, and the buttons clicked, if any. In this paper, I primarily focus on prompts submitted to Midjourney rather than button interactions.

One caveat is that users who purchase higher-tier subscriptions may activate "stealth mode", which conceals their actions from the public chat channel. These messages are not included in the data. A natural concern is whether users worry about others observing and appropriating their prompts. Discussions on platforms such as Reddit suggest that a high volume of activity in the channel, where individual messages are quickly displaced by hundreds of others, mitigates such concern. Users generally do not appear to think of prompt imitation as a significant risk.

In the remainder of the paper, I focus on a random sample of 500 users who submitted at least 50 prompts in each of V3, V4, V5, and V6 in the dataset, along with their complete usage history. There are three reasons for this sampling strategy. First, it minimizes the impact of users who tried Midjourney briefly out of curiosity before quitting, allowing me to focus on users with sustained engagement. Second, it ensures every user has sufficient usage in each Midjourney version, such that I can analyze their behavioral changes as the AI improves. Third, given the computational intensity of processing text and image data, restricting the sample to a small subset makes subsequent data processing feasible.

An illustrative example of a prompt is as follows.

> **<https://s.mj.run/ABCDEFG> dark fantasy, retro anime style, mink hiding in a forest, horror core, ultra realistic, retro, video game style, creepy - - aspect 5:4** - <@1234567899876543 21>
>
> 07/16/2023 11:18:25.796 UTC

In this example, "https://s.mj.run/ABCDEFG" is the link of the image input; "dark

fantasy, retro anime style, mink hiding in a forest, horror core, ultra realistic, retro, video game style, creepy" is the text input; "aspect 5:4" is a parameter, and "123456789987654321" is the user ID. Note that there can be multiple image inputs and multiple parameters in a prompt. Every prompt is associated with a timestamp. In the following analysis, I focus on the text input of the prompts.

3.1 Cluster Prompts to Define Session

An important observation from the data is that users submit sequences of correlated prompts within short time periods, generating visually similar images. The following example illustrates this behavior.

1. "a closeup of a book cover on a beach on a rainy day"
2. "closeup of a book cover lying in the sand on a beach"
3. "closeup of a book cover lying in the sand on a beach, an hourglass is also in the sand"
4. "closeup of a book cover lying in the sand on a beach, a small hourglass is also in the sand"

Users appear to refine their prompts iteratively to converge toward a desired outcome. This practice is consistent with online discussions where users describe such strategies as "I'd experiment by removing or slightly modifying descriptors to make it fit what I wanted", "Rerun, refine, repeat, dominate" or "Start simple, only add what you need". Many YouTube videos also teach Midjourney tips by showing the YouTubers' workflow on Midjourney. In these videos, many users are demonstrating such iterative prompt refinement strategies.

Therefore, I treat these prompt sequences as intermediate steps in a production process and define a *session* as a sequence of highly correlated prompts submitted by a single user. The session serves as the production unit in this context.

To identify such sessions, I implement the following clustering procedure. For each user, I calculate a Jaccard similarity matrix across all prompts and apply the Density-based spatial clustering of applications with noise (DBSCAN) algorithm, one of the most commonly used and cited clustering algorithms. I then verify the constructed sessions by examining the time span. Over 90% of sessions are finished within 24 hours. The details of clustering are described in Appendix C.3.

Table 1 represents the summary statistics. Both the median and minimum number of prompts per session equal one, indicating that a substantial share of sessions consist of a single prompt. In these cases, users submit one prompt, and either obtain a satisfying result

Table 1: Summary Statistics

	Mean	Median	99th Percentile	Min	Max
#sessions per user	191	114	1,408	7	4,120
#prompts per session	12	1	18	1	45,720
#words per prompt	28	18	195	0	378
#parameters per prompt	1.60	1	5	0	10
#image inputs per prompt	0.21	0	3	0	57
N_{user}			500		
$N_{session}$			95,545		
N_{prompt}			1,160,773		

or receive images that are sufficiently poor that they quit the session immediately. The distribution of session length is skewed and long-tailed. Besides, the minimum number of words per prompt is zero because, in rare cases, users submit only image inputs, sometimes with parameter specifications to Midjourney, without any text. Fifteen sessions contain more than 10,000 prompts, representing less than 0.02% of all sessions. These extreme cases likely reflect some rare automated prompting behaviors. Since the subsequent analysis is mainly performed at the session level, these outliers do not significantly affect the results.

3.2 Supplementary Data

Apart from messages on the Midjourney Discord server, I also use supplementary data from online Midjourney community classifications including Midlibrary, Imiprompt, ShellyPalmer and a popular collection of words on GitHub repository.[5] These data give me information like artist names, photography words, and style words, which are later used in prompt analysis.

4 Adaptation to New AI

In this section, I examine human adaptation to different AI versions and show that AI and human input are complements in creative production. When AI upgrades, users systematically alter the words in their prompts. This adaptation reflects users' subjective *judgments* about technological change and plays a crucial role in shaping the final outcomes of the

[5]https://midlibrary.io/;
https://www.imiprompt.com/;
https://shellypalmer.com/midjourney-reference-art/;
https://github.com/willwulfken/MidJourney-Styles-and-Keywords-Reference/blob/main/README.md

creative process. Simply applying better AI to unchanged prompts does not replicate the outcome images we observe in actual usage. The decomposition exercise reveals that human adaptation accounts for 73% of the output improvements when AI upgrades. In contrast, AI improvements alone, holding prompts constant, contribute only 20%. This finding suggests that realizing the full potential of AI upgrades requires humans to adapt their inputs.

I begin by showing that when AI upgrades, both AI capability and human input change. In particular, in Section 4.1, I submit the same set of randomly sampled prompts to all six Midjourney versions. By comparing the submitted prompts with the GPT-generated descriptions of the resulting images, I demonstrate that more recent AI versions produce images that align more closely with the prompts. In other words, more recent AI is more accurate. Specifically, Midjourney V4 exhibits significantly greater accuracy than V3. This substantial improvement in AI capability has been widely discussed in the online Midjourney community and has influenced users' word choices in their prompts. In Section 4.2, I analyze words submitted across different AI versions and show that users systematically modify their prompts. For example, as the new AI better interprets color words, users include significantly more color words in their prompts. Conversely, because the new AI generates more detailed images by default, users reduce their use of phrases like "super detailed." These systematic changes in word usage indicate that human input adapts to AI capability.

In Section 4.3, I consider a production function for images of the form $F(AI, L)$, where both AI capabilities and human input shift the generated images. I conduct a decomposition exercise by submitting prompts written for the old AI to the new version, obtaining images $F(AI_1, L_0)$, and submitting prompts written for the new AI to the old version, obtaining images $F(AI_0, L_1)$. Comparing $F(AI_0, L_0)$ with $F(AI_1, L_0)$ isolates the effect of changing AI alone, holding human input fixed. Similarly, comparing $F(AI_0, L_0)$ with $F(AI_0, L_1)$ isolates the effect of changing human input alone, holding AI fixed. These comparisons inform the extent to which AI and human input are complements or substitutes in the creative process. These findings shed light on the role of human adaptation to technological change in creative production.

4.1 Newer AI is More Accurate

AI models are developing rapidly, and more recent models exhibit greater capabilities. During the observation period, Midjourney released six versions: V1, V2, V3, V4, V5, and V6. Higher version numbers indicate more recent and capable models, which is similar to the difference between GPT-4 and GPT-5. V1 and V2 were primarily used during the invite-only closed beta test on Discord. On July 12, Midjourney entered open beta, allowing anyone to

join the Discord server. Shortly after the beginning of open beta, Midjourney released V3 on July 25, 2022. Since Midjourney did not disclose the selection criteria for the closed beta test, this paper focuses on user behavior from V3 through V6.[6]

Figure 1 illustrates differences across the six versions. The same prompt "Satyr, ancient greek tavern, dickensian fashion, cosmic horror, non-euclidian geometry" was submitted to each version, generating substantially different images. More recent versions tend to give more detailed images that appear to align more closely with the prompt. There is a significant improvement between V3 and V4. Other than that, there is improvement between V4 and V5, V5 and V6 as well.

Figure 1: Illustration of Version Differences

(a) V1 (b) V2 (c) V3

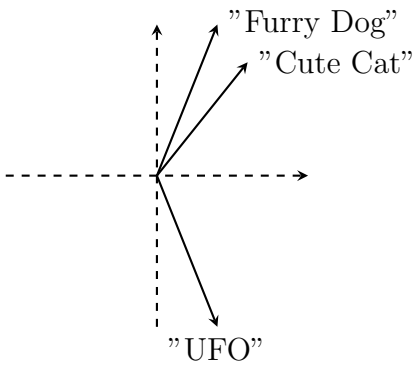
(d) V4 (e) V5 (f) V6

Notes: This figure presents image outputs from Midjourney with the **same prompt** in **different versions**: "Satyr, ancient greek tavern, dickensian fashion, cosmic horror, non-euclidian geometry".

[6]V3 was released on July 25, 2022, and became the default version immediately. V4 was released on November 5, 2022, and became the default version on December 20, 2022. V5 was released on March 15, 2023, and became the default version on May 4, 2023. V6 was released on December 21, 2023, and became the default version on February 14, 2024.
I infer which version a user employed based on timestamps and prompt parameters. For example, if a prompt contains "- - V5" the user is employing V5 even if the timestamp falls after February 14, 2024. If a prompt does not specify a version, it uses the default version for that time period. Figure E1 in the appendix displays the version timeline.

To quantify this improvement in AI capabilities, I employ text and image embeddings, a standard and widely adopted approach in text and image analysis. The method converts text or images into vectors in a high-dimensional space using large language models or vision-language models. The embeddings, which are vectors, are interpretable through cosine similarity between them. If two pieces of text have high cosine similarity, they are similar in meaning space, as illustrated in Figure 2. In this paper, I use OpenAI text-embedding-3-small for the main results. In the appendix, I use all-MiniLM-L6-v2, another embedding model, to cross-verify robustness. Results are robust to both embedding models.

Figure 2: Embedding Illustration

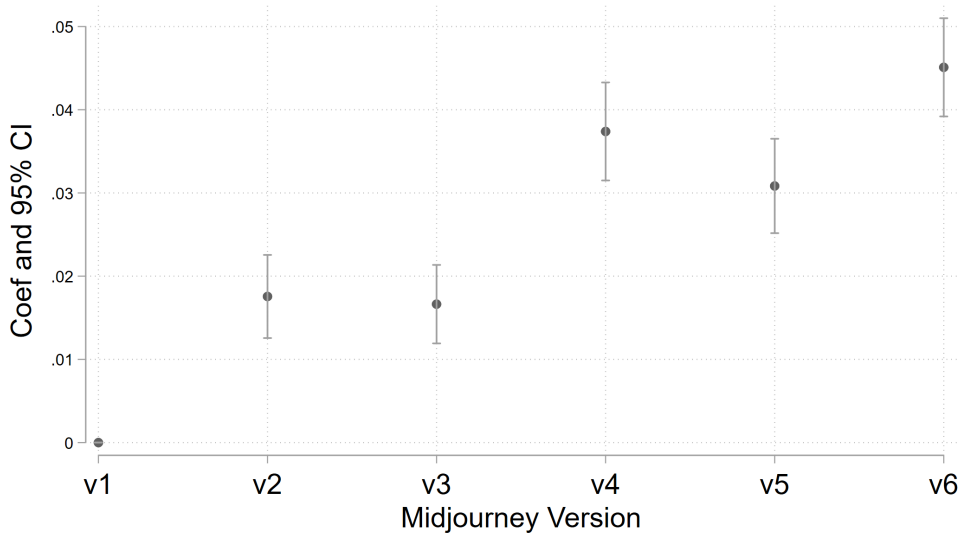


The following experiment is implemented. I randomly select 500 prompts from the data and submit each prompt to all six MidJourney versions. I then collect the generated images and ask GPT-4o to "describe the image". This approach enables me to compare the input (prompt) and output (proxied by the GPT-generated description) and quantify how AI capability changes across versions using regression (1).

$$Cosine\ Similarity(\overrightarrow{p}_{prompt}, \overrightarrow{p}_{GPTdescription}) = \sum_{v} \beta_v \mathbb{1}(Midjourney\ version = v) + \mu_{prompt} + \epsilon_{pv} \tag{1}$$

In regression (1), the dependent variable is the cosine similarity between the submitted prompt and the GPT-generated description of the resulting image. Cosine similarity, which takes values between -1 and 1, is computed from text embeddings generated by OpenAI's text-embedding-3-small model. Values closer to 1 indicate that the two texts are similar in meaning, whereas values closer to -1 denote opposite meanings. Higher cosine similarity indicates greater accuracy: Midjourney produces images that align more closely with the input prompt. The independent variables are dummies indicating Midjourney versions, along with the prompt fixed effects.

Figure 3: Later Versions Give More Accurate Results



Notes: This figure plots coefficients before $\mathbb{1}(Midjourney\ version = v)$ in regression (1) and the 95% confidence intervals. Standard errors are clustered at the prompt level. This figure uses embeddings from the embedding model text-embedding-3-small from OpenAI. The dependent variable, cosine similarity, can range from -1 to 1. Values closer to 1 indicate that the two texts are similar in meaning, whereas values closer to -1 denote opposite meanings. Higher cosine similarity indicates greater accuracy: Midjourney produces images that better align with the input prompt. I run the same regression with embeddings from all-MiniLM-L6-v2. The result is reported in Figure B1 in the appendix.

Figure 3 plots the coefficients of β in regression (1). More recent Midjourney versions generally produce more accurate images corresponding to the submitted prompts. As expected, there is a substantial increase in accuracy between V3 and V4.

A more recent Midjourney version generally implies a more accurate image corresponding to the submitted prompt. As expected, there is a big jump between V3 and V4. A potential mechanism is that more recent versions generate more detailed images. In Figure B2 in the appendix, I plot the similar coefficients in regression (1), with the dependent variable being the length of the GPT description. The results suggest that more recent versions tend to generate images requiring more words to describe, hence greater detail. Although a more detailed image does not always imply that it is more aesthetic, a more accurate output at least represents a more faithful reflection of the submitted prompt.

In Section 4.2 and Section 4.3, I exploit this substantial shift between V3 and V4. I show that users change prompts, and that these changes in AI capability and prompt wording jointly and systematically alter the generated images. By submitting prompts written for the old AI to the new version and vice versa, I decompose shifts in output images into contributions from AI improvements and prompt adaptations.

4.2 People Change Words that They Use to Prompt in New AI

Since different Midjourney versions have different capabilities and recognize different vocabularies, users write significantly different prompts across versions. Users of the online Midjourney community have voluntarily summarized the vocabulary each version recognizes. For example, websites document how each style appears across different Midjourney versions.[7] There is also a public Google spreadsheet summarizing whether specific artist names are recognized in each version.[8] Online discussions also debate which "magic words" enhance artistic quality in each version. The existence of these community resources suggests that users actively explore the effect of putting various words in the prompts across Midjourney versions.

To empirically show that words in the prompts are systematically different across Midjourney V3 and V4, I construct word classifications using the Midjourney community online resources, including Midlibrary, Imiprompt, ShellyPalmer, and a popular collection of words on a GitHub repository summarized by Will Wulfken. For each user, I consider the 50 prompts before and the 50 prompts after the V4 release date, constructing a balanced panel around this threshold. Using these classifications and the balanced panel, I estimate regression 2.

$$\#Realistic\ Words_{ip} = \sum_k \beta_k \mathbb{1}(k^{th}\ Prompt)_{ip} + \mu_i + \epsilon_{ip} \tag{2}$$

The dependent variable is the number of "realistic words" in prompt p written by user i. Realistic words include terms such as "realistic", "super-realistic", "hyper-realistic", etc. The independent variable $\mathbb{1}(k^{th}\ Prompt)_{ip}$ is a set of dummy variables indicating whether this is the k^{th} prompt since the V4 release date of user i. The index k can be negative if the prompt is written before the V4 release date. The μ_i are the user fixed effects. Panel (a) in Figure 4 plots the estimated coefficients β_k against k.

The result shows that users systematically include fewer realistic words in prompts for V4 than V3. This pattern likely reflects the fact that V4 generates more realistic images than V3 by default, reducing the need to explicitly specify "realistic" in prompts.

Similar patterns are found in other dimensions. Figure 4 shows that users include fewer terms such as "detailed" or "octane render" in prompts for V4, since the images produced by V4 are more detailed by default. Conversely, users employ more color and photography-related terms in V4 prompts because V4 interprets these words more effectively. Users also include more stop words, such as "a" "the" "here" "there", in their prompts. This is

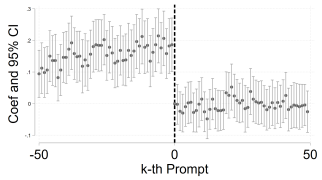
[7] https://midlibrary.io/

[8] https://docs.google.com/spreadsheets/d/1cm6239gw1XvvDMRtazV6txa9pnejpKkM5z24wRhhFz0

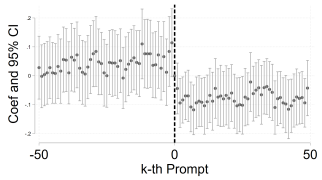
because V4 understands natural language better, and users are doing less prompt engineering. For example, a V3 prompt might be crafted into "cat, blue shirt, sitting at table, white background". Whereas a V4 prompt can be phrased as "a cat wearing a blue shirt sitting at a table with a white background". The increased presence of stop words suggests that users are crafting prompts into sentences that resemble natural language, as the new AI understands them more effectively. This systematic change in the words of the prompt suggests that users are adapting to new AI by adjusting their actions.

Figure 4: People Change Prompts in New AI

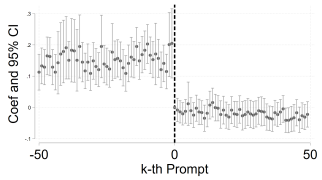
(a) realistic, hyper-realistic, ...



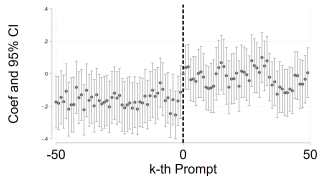
(b) detailed, super-detailed, ...



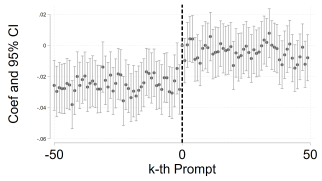
(c) rendering engines (octaine render, ...)



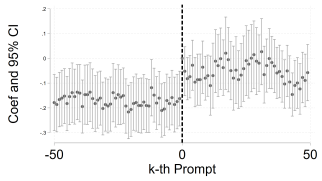
(d) colors



(e) stop words (a, the, ...)



(f) photography (105mm lens, shutter speed, ...)



Notes: The classifications of words can be found in Appendix C.4.

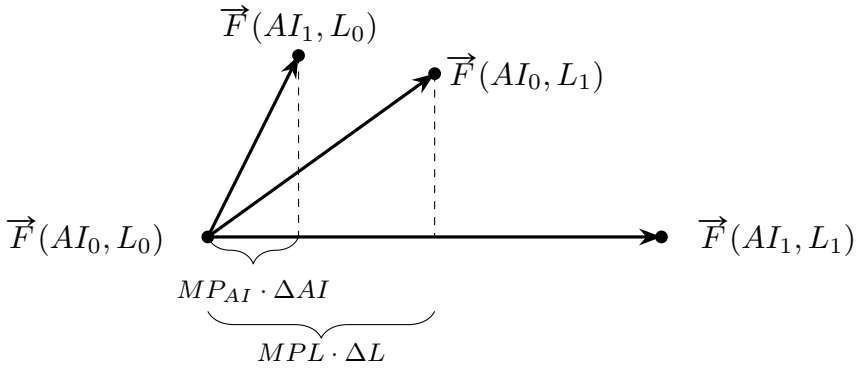
4.3 Decomposition of Output Shifts

When AI upgrades, both AI capability and words in the prompt change, jointly altering the generated images. This section decomposes total shifts in output images into contributions from AI changes alone and contributions from prompt changes alone. A key feature of this setting is that, unlike typical settings where researchers observe only outputs from old machines and old labor input ($Y = F(K_{old}, L_{old})$) and outputs from new machines and new labor input ($Y = F(K_{new}, L_{new})$), I can directly observe two counterfactuals without imposing model assumptions: outputs from new machines and old labor input ($Y = F(K_{new}, L_{old})$), and outputs from old machines and new labor input ($Y = F(K_{old}, L_{new})$). To construct these counterfactuals, I submit prompts users wrote for the old AI to the new version, and vice versa.

Specifically, I collect images users generated with the old AI, $F(AI_0, L_0)$, and images

generated with the new AI $F(AI_1, L_1)$. I then submit the old prompts to the new AI to obtain $F(AI_1, L_0)$, and submit new prompts to the old AI to obtain $F(AI_0, L_1)$.

Figure 5: Projection Illustration



The collected images are first described using GPT-4o, and these descriptions are then converted into vector embeddings with OpenAI's text-embedding-3-small model. The vector embeddings are representations in meaning space. The vectors are then projected onto the direction of $\overrightarrow{F}(AI_1, L_1) - \overrightarrow{F}(AI_0, L_0)$ as illustrated with Figure 5. Define the vector from old output $\overrightarrow{F}(AI_0, L_0)$ to the new output $\overrightarrow{F}(AI_1, L_1)$ as representing 100% of the output shift, $MP_{AI} \cdot \Delta AI$ captures the output shift attributable to AI changes alone, marginal product of AI multiplied by AI changes. While $MPL \cdot \Delta L$ captures the output shift attributable to prompt changes alone, marginal product of labor multiplied by labor changes. The next section provides an economic interpretation of this decomposition.

4.3.1 Decomposition: Interpretations

This decomposition can be interpreted as a discrete approximation of the second-order expansion of the production function, as shown in equation (3). When Y represents output level, $MP_{AI} \cdot \Delta AI$ corresponds to the marginal product of AI multiplied by the change in AI. $MPL \cdot \Delta L$ corresponds to the marginal product of prompts multiplied by the change in prompts. $MP_{AI,L} \cdot \Delta AI \Delta L$ captures the second-order cross-partial derivative multiplied by the interaction of both AI changes and prompt changes, which can be interpreted as the strategic complementarity term.

$$Y_1 - Y_0 \simeq \underbrace{\frac{\partial F(AI_0, L_0)}{\partial AI}(AI_1 - AI_0)}_{MP_{AI} \cdot \Delta AI} + \underbrace{\frac{\partial F(AI_0, L_0)}{\partial L}(L_1 - L_0)}_{MPL \cdot \Delta L} + \underbrace{\frac{\partial^2 F(AI_0, L_0)}{\partial AI \partial L}(AI_1 - AI_0)(L_1 - L_0)}_{MP_{AI,L} \cdot \Delta AI \Delta L} \tag{3}$$

The decomposition implemented in this paper is a discrete approximation of this expansion in *vector space*, where the output image Y should be considered as a location in a multi-dimensional space, captured by its embedding. Equation (4) presents a similar approximation in vector space. $MP_{AI} \cdot \Delta AI$ and $MPL \cdot \Delta L$ retain similar interpretation as in equation (3), while $MP_{AI,L} \cdot \Delta AI \Delta L$ is the residual of the output shift, which can be considered as an approximation of the strategic complementarity term in equation (3).

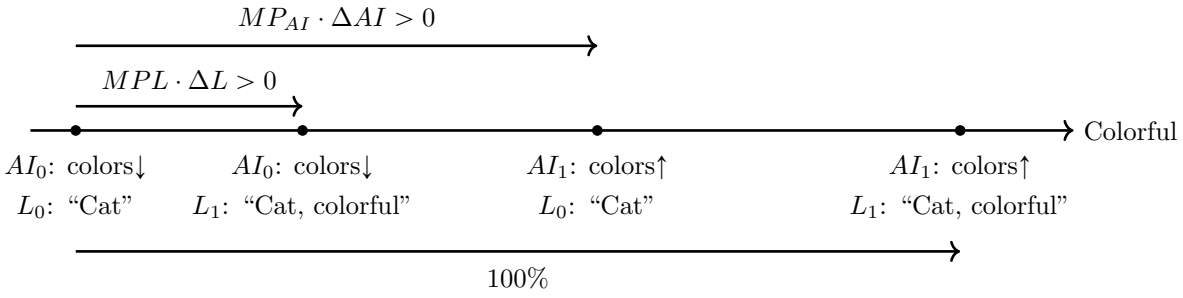
$$\begin{aligned} \overrightarrow{Y}_1 - \overrightarrow{Y}_0 = & \underbrace{\overrightarrow{F}(AI_1, L_0) - \overrightarrow{F}(AI_0, L_0)}_{MP_{AI} \cdot \Delta AI} + \underbrace{\overrightarrow{F}(AI_0, L_1) - \overrightarrow{F}(AI_0, L_0)}_{MPL \cdot \Delta L} \\ & + \underbrace{\overrightarrow{F}(AI_1, L_1) - \overrightarrow{F}(AI_1, L_0) - \overrightarrow{F}(AI_0, L_1) + \overrightarrow{F}(AI_0, L_0)}_{MP_{AI,L} \cdot \Delta AI \Delta L} \end{aligned} \tag{4}$$

How should we interpret these fractions economically? The magnitude of $MPL \cdot \Delta L$ will inform us how much human adaptation to new AI shifts the output images. The signs of these fractions reveal to what extent AI and human input are complements or substitutes, with human input defined as words in the prompt.

Suppose AI and human inputs are complements. Consider the case where the new AI interprets color words more effectively than the old version. In the old version, even if users include color words in the prompt, AI does not understand it, and hence users say *fewer* color words in the old AI. However, as the AI's understanding of color words improves, users adapt by incorporating more color words into their prompts. Another example may be that in the old version, AI does not recognize certain artist names. As a result, users exclude them from prompts. When the new AI recognizes these names, users include them.

Figure 6 illustrates this scenario. Suppose that in the old AI, the prompt is "Cat", while in the new AI, the prompt is "Cat, colorful". When the new prompt is submitted to the old AI, the resulting image shifts toward the new AI's output. Similarly, when the old prompt is submitted to the new AI, the resulting image also shifts toward the new output. If the total output shift from $Image(AI_0, L_0)$ to $Image(AI_1, L_1)$ is normalized to 100%, then $MP_{AI} \cdot \Delta AI$, the contribution from changing AI alone, is positive and points in the direction of the new image. $MPL \cdot \Delta L$, the contribution from changing the prompt alone, is also positive and points in the same direction. Hence $MP_{AI} \cdot \Delta AI > 0$, $MPL \cdot \Delta L > 0$.

Figure 6: Illustration: AI and Human Input as Complements



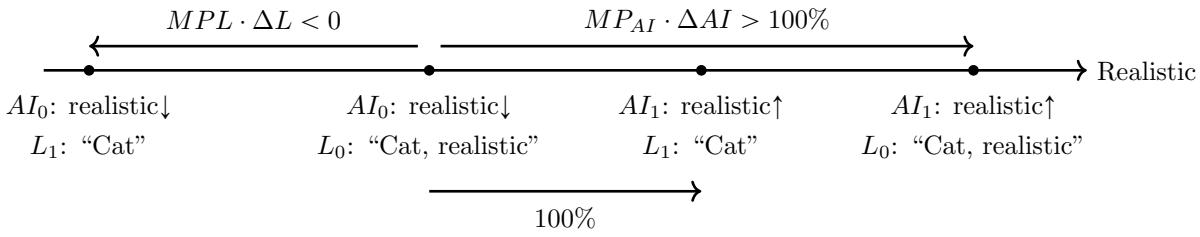
By contrast, suppose AI and human input are substitutes. Consider the case where the old AI generates less realistic images by default, requiring users to specify "Cat, super-realistic" to obtain sufficiently realistic images. In contrast, the new AI automatically generates more realistic images even without explicit realistic specifications in the prompt. Users adapt to this change by including *fewer* realistic words in the prompt.

Figure 7 panel (a) illustrates this scenario. When the new prompt "Cat" is submitted to the old AI, the resulting image is less realistic than the old output. When the old prompt "Cat, realistic" is submitted to the new AI, the resulting image is more realistic than the new output. If the total output shift from $Image(AI_0, L_0)$ to $Image(AI_1, L_1)$ is normalized to 100%, then $MP_{AI} \cdot \Delta AI$, the contribution from changing AI alone, is positive and points toward the new image. However, $MPL \cdot \Delta L$, the contribution from changing the prompt alone, points in the *opposite* direction. Hence, $MP_{AI} \cdot \Delta AI > 0$, $MPL \cdot \Delta L < 0$.

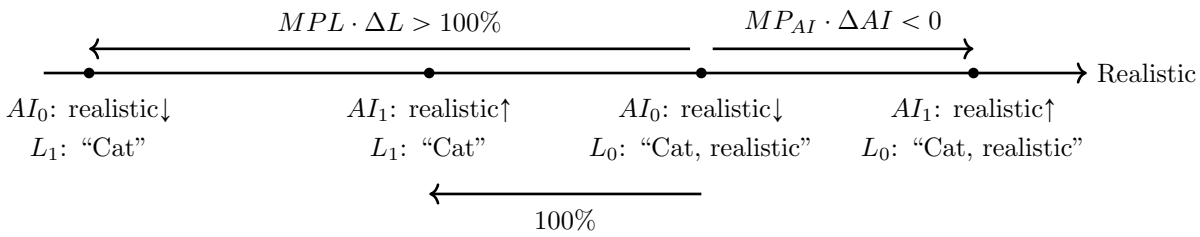
Figure 7 panel (b) illustrates the other possible scenario where the new image is less realistic because users no longer include the word "realistic" in their prompts. In this case, $MP_{AI} \cdot \Delta AI < 0$, $MPL \cdot \Delta L > 0$. More generally, if AI and human inputs are substitutes, $MP_{AI}\Delta AI \cdot MPL\Delta L < 0$.

Figure 7: Illustration: AI and Human Input as Substitutes

(a) Case 1



(b) Case 2



4.3.2 Decomposition: Implementation

To implement the decomposition, I randomly select 1000 sessions in V3 and 1000 sessions in V4, collecting the generated images. For each session, I find the final prompt and image by

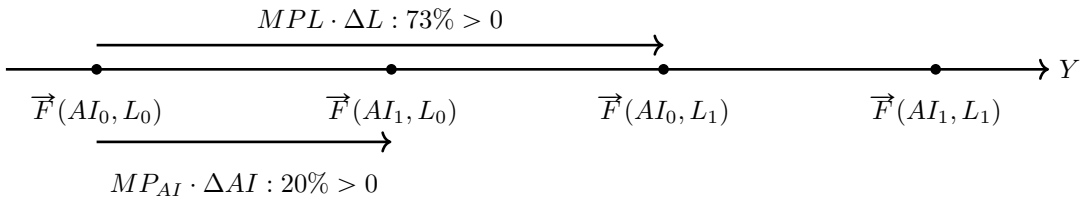
timestamp. I focus on the final prompts and images of the sessions because they represent the final outcome of a creative process. I then submit prompts written for V3 to V4, and prompts written for V4 to V3, collecting the generated images.

To calculate the fractions discussed in the previous section, I match prompts from V3 with prompts from V4 based on similarity. The rationale is to compare similar content: a robot image in V3 should be compared with a robot image in V4, and a landscape in V3 should be compared with a landscape in V4. Moreover, since the decomposition represents a discrete approximation of the second-order expansion of the production function, comparisons should be as local as possible. With 1000 prompts from each version, I construct a 1000×1000 similarity score matrix and apply the Hungarian Algorithm (Munkres, 1957) to obtain a one-to-one mapping. This algorithm has been widely adopted in computer science and engineering for matching strings, such as bus route identifiers and gene sequences. This procedure yields 1,000 matched tuples of the form $\Big(Image(AI_{v3}, L_{v3}), Image(AI_{v4}, L_{v3}), Image(AI_{v3}, L_{v4}), Image(AI_{v4}, L_{v4})\Big)$. Figure E3 in Appendix E displays an example of such a matched tuple.

After constructing matched tuples, I use GPT-4o to generate textual descriptions of each image and convert these descriptions into embeddings using the OpenAI text-embedding-3-small model. The obtained embeddings are projected onto the direction of $Image(AI_{v4}, L_{v4}) - Image(AI_{v3}, L_{v3})$ to obtain $MP_{AI} \cdot \Delta AI$ and $MPL \cdot \Delta L$ for each tuple. The following section reports the average values of $MP_{AI} \cdot \Delta AI$ and $MPL \cdot \Delta L$ across all matched tuples.

4.3.3 Decomposition: Result

Figure 8: Decomposition Result



$$MP_{AI,L} \cdot \Delta AI \Delta L = 100\% - MP_{AI} \cdot \Delta AI - MPL \cdot \Delta L = 7\% \tag{5}$$

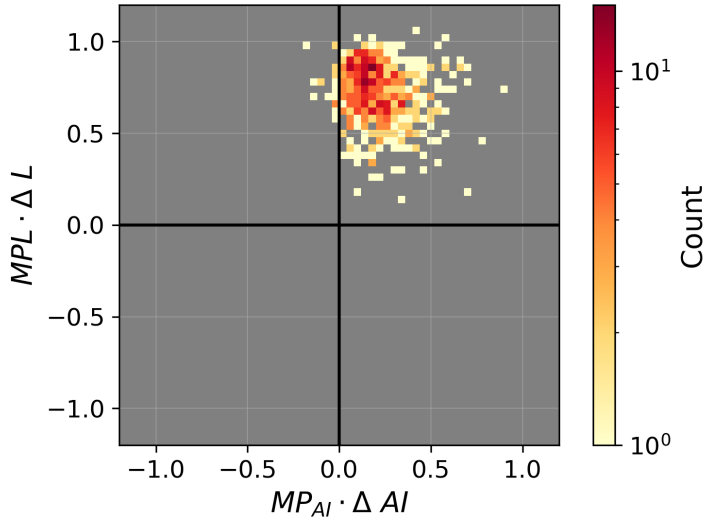
Figure 8 and equation (5) present the average values of $MP_{AI} \cdot \Delta AI$, $MPL \cdot \Delta L$, and $MP_{AI,L} \cdot \Delta AI \Delta L$. The decomposition attributes 20% of the total output shift to AI changes, 73% to prompt adaptations, and 7% to the residual term. The positive signs of $MP_{AI} \cdot \Delta AI$ and $MPL \cdot \Delta L$ indicate that AI and human input are more like complements than substitutes.

The main reason is that both ΔAI and ΔL are positive. It means that the AI's improvement and users' prompt changes reinforce one another. As AI improves at interpreting certain words (e.g., color or photography words), users supply *more* of those words, suggesting complementarity between AI and human input.

Moreover, $MP_{AI,L} \cdot \Delta AI \Delta L$, corresponding to the second-order derivatives multiplied by AI changes and prompt changes, is also positive. Since both $\Delta AI > 0$ and $\Delta L > 0$, $MP_{AI,L}$ is also positive. In other words, the improvements in the AI increase the marginal product of human input. It further confirms the complementarity between AI and human input.

Figure 9 displays a heat map of $(MP_{AI} \cdot \Delta AI, MPL \cdot \Delta L)$ across all matched tuples. For 96.5% of tuples, both fractions $MP_{AI} \cdot \Delta AI$ and $MPL \cdot \Delta L$ are positive, indicating that AI and human inputs are complements in the majority of cases.

Figure 9: Density Heatmap of AI vs Prompt Contributions



Notes: This figure displays the heat map of $(MP_{AI} \cdot \Delta AI, MPL \cdot \Delta L)$ of the tuples. 96.5% tuples fall within the domain with both fractions being positive.

4.3.4 Supporting Evidence of Complementarity: Longer Prompts in New AI

$$\#\text{Words}_{ip} = \beta \mathbb{1}(NewAI)_{ip} + \mu_i + \epsilon_{ip} \tag{6}$$

Table 2: Number of Words Increases in New AI

	(1)	(2)	(3)
Dep Var	$\#Words$	$\#Stop\ Words$	$\#Non\ Stop\ Words$
$\mathbb{1}(NewAI)$	3.28***	1.08***	2.20***
	(0.88)	(0.16)	(0.78)
Mean in Old AI Prompt	19.59	3.07	16.52
User FE	Y	Y	Y
N	49,803	49,803	49,803
R^2	0.31	0.18	0.33

Notes: *** denotes significance at 1 percent, ** at 5 percent, and * at 10 percent. Standard errors are clustered at the user level. This table presents the estimates of regression (6).

To further examine complementarity between AI and human inputs, I estimate regression (6) to test whether users include more words in their prompts in the new version. For each user, I consider the 50 prompts before and the 50 prompts after the V4 release date, constructing a balanced panel around this threshold. The dependent variable is the total number of words in prompt p written by user i.

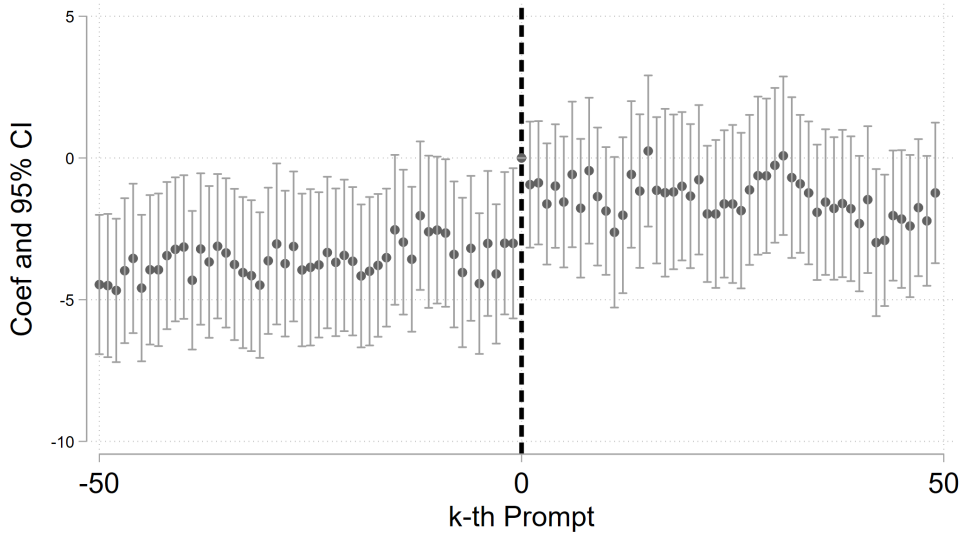
Table 2 presents the regression results. The number of words per prompt increases 3.28 in the new AI. However, this increase in total word count may not fully reflect the amount of meaningful content in prompts. Column (2) shows that users also include more stop words after the new AI is introduced. This reflects the new AI's improved understanding of natural language. Hence, users formulate prompts as complete sentences (e.g., "a cat wearing a blue shirt and drinking coffee") instead of keyword sequences (e.g., "cat, blue shirt, drinking coffee"). Examining total word count alone could therefore only capture that prompts become more like natural language in the new AI.

To better assess whether users provide more informative content, I examine the number of non-stop words in each prompt. Column (3) in Table 2 reports the corresponding regression, using the count of non-stop words as the dependent variable. The results suggest that the number of non-stop words increases significantly in the new AI, implying that users write more informative words in prompts when AI capability improves. On average, the number of non stop words increases by 2.20 in the new AI. This pattern suggests that users include more meaningful words when the AI becomes more capable, consistent with a complementary relationship between AI performance and human input quality.

Figure 10 plots the estimated coefficients β_k in regression (7). The figure shows a sharp and immediate increase in non-stop words after the new AI release.

$$\#\text{Non-Stop Words}_{ip} = \sum_k \beta_k \mathbb{1}(k^{th}\ Prompt)_{ip} + \mu_i + \epsilon_{ip} \tag{7}$$

Figure 10: The Number of Non-Stop words Increases in New AI



Notes: This figure plots coefficients before $\mathbb{1}(k^{th}\ Prompt)_{ip}$ in regression (7) and the 95% confidence intervals. Standard errors are clustered at the user level.

4.4 Discussion of Adaptation to New AI

In this section, I have shown that AI upgrades induce changes in both AI capability and human input, jointly altering the generated images. By resubmitting old prompts to the new AI and vice versa, I construct two counterfactuals without imposing model assumptions and perform a decomposition. The results indicate that 73% of the output shifts are driven by prompt changes, while only 20% is attributed to AI changes. The positive signs of both fractions imply that AI and human input are complements rather than substitutes.

This decomposition highlights the crucial role of human adaptation when technology advances. Simply applying historical prompts to a more advanced AI does not replicate actual outputs, because users continuously refine their language to leverage the new AI's strengths. As AI capability evolves, human judgment remains essential in the creative process. It determines which words are more suitable to include in the new AI.

An implication is that evaluating AI upgrades without accounting for user adaptation underestimates the full potential of the technology. Improvements in model capability may translate into better results when users adjust their inputs accordingly with judgment.

Adaptation to the new AI focuses on the final outcome in a session. In the next section, I will investigate within-session adaptation: how users perceive the output image from the previous prompt, evaluate the image quality, and adjust their next prompt.

5 Adaptation to Output from Previous Prompts

In this section, I show that users adjust their prompts iteratively within a session, exhibiting path-dependent patterns. I then employ a sequential search model to characterize such patterns. In this model, users observe the output of each prompt, assess its quality, and update their beliefs about the expected return of other prompts in their consideration set. A satisfying output increases the posterior probability that similar prompts will also give satisfying results, making those prompts more likely to be selected in subsequent iterations. I describe this quality assessment of images as the adaptation to output from previous prompts.

To quantify the effect of this adaptation, I conduct a counterfactual experiment. In this experiment, users submit a predetermined list of prompts to the AI, based on their prior beliefs. The AI processes the prompts sequentially, and users observe the generated images but are not allowed to reorder the prompts based on their judgment of image quality. Sessions continue until users reach the same final prompt observed in the actual data. Under this restriction, sessions require, on average, 313% more prompts. This result suggests that human adaptation plays a crucial role in directing the prompt search process and enhancing production efficiency.

5.1 Prompt Evolvement Patterns within Session

In online Midjourney communities, users share workflows for generating desired images, Common suggestions include "start simple, then elaborate, add one detail at a time"; "elements placed at the start of a prompt will have a more significant effect on the result than those placed at the end"; "regenerate the image with the same prompt until a satisfactory result appears".[9] These discussions suggest that formulating a satisfying prompt is mentally demanding and relies on iterative experimentation with the AI's outputs.

A typical prompt sequence within a session is illustrated in Section 3.1. Below are four patterns of prompts within a session.

[9]https://www.youtube.com/watch?v=vUj4VNXXC1c;
https://medium.com/the-nerd-circus/midjourney-prompt-weight-mastery-a-guide-to-advanced-prompt-crafting-b013c3bcade4;
https://www.reddit.com/r/midjourney/comments/xdqbsx/whats_your_typical_flow/

Prompt Length Increases along Path. Appendix A.1 shows that users progressively add words, parameters, and image inputs to their prompts within a session, making prompts more specific than its preceding ones.

Path Dependence of Prompts. Section 3.1 illustrates that in the data, there are sequences of prompts that are highly correlated with each other and are submitted to AI within a short time frame. The step-by-step strategy for adjusting the prompts has also been widely discussed on user forums, such as Reddit and YouTube. These all suggest that prompts are also path-dependent.

Prompts Become Increasingly Similar to the Final Prompt in a Session. Appendix A.2 documents that prompts are becoming increasingly similar to the final prompt in a session along the path.

Users Take Larger Steps in Prompt Space Early in the Search Path. Appendix A.3 shows that in early stages, artists adjust words with higher weights on image outputs. These are words closer to the beginning of the prompts. And later, they move to adjusting words that are closer to the end of the prompts. This implies that users begin by focusing on the most key components of the image, then turn to the minor details.

Table 3: Similarity Between Product Searching and Prompt Writing

Consumer Search	**User Writing Prompts**
1. Query becomes longer / more specific along the search path.	1. Prompt length increases along the path.
2. State dependence: The Product searched currently has similar attributes to the product searched previously.	2. State dependence: Within a session, users adjust prompts step by step.
3. Products become increasingly similar to the final prompt in a path: As search progresses, products become increasingly similar to the product finally purchased.	3. Prompts become increasingly similar to the final prompt in a session: As search progresses, prompts become increasingly similar to the final prompt in a session.
4. Take larger steps in attribute space early in the search path: Consumers search a wider variety of products and take larger "steps" through attribute space early in the search path than later in the search path. For example, consumers explore a wider range of prices. And then later, they narrow it down to a smaller price range.	4. Take larger steps in prompt space early in the search path: Users adjust words with greater weights on image output early in the search path than later.

These patterns mirror consumer search behavior documented in the literature as described below. Table 3 summarizes the similarities between product searching and prompt writing.

In the consumer search literature (Bronnenberg et al., 2016; Hodgson and Lewis, 2025, Hirsch et al., 2020), researchers have shown that along the search path, the queries consumers submit to the search box become longer and more specific. For example, in Bronnenberg et al. (2016), the authors show that early queries tend to use generic terms such as "digital camera", whereas later inquiries include specific models and brands like "Nikon" and "P520". Similarly, Hirsch et al. (2020) documents an increase in query length as consumers refine their searches.

State dependence also arises in product search. Bronnenberg et al. (2016) and Hodgson and Lewis (2025) find that the probability for searching a Nikon camera increases by 0.250 if the brand was searched in the previous decile. These studies further show that the cameras that a consumer browses become increasingly similar to the camera they finally purchase along the path.

Additionally, consumers make larger attribute "steps" early in the search path. Bronnenberg et al. (2016) and Hodgson and Lewis (2025) find that early in their search, consumers explore a wide price range and diverse product attributes, before narrowing their focus in later stages.

In sum, although prompt writing and product searching seem different, they share some data patterns. One way to think about this is that a trained AI is a set of all possible images it can generate. Users are searching for the exact prompts that can "extract" the desired image. In the following section, I will characterize these patterns using the sequential search framework developed by Hodgson and Lewis (2025).

5.2 Data Processing

Since the set of possible prompts is all combinations of words in the English language, the prompt space is too large to handle. I follow the IO literature (Lancaster, 1966) and map prompts into attribute space to reduce dimensions. In addition, I focus on landscapes to better construct consideration sets. The construction of consideration sets is discussed in Section 5.4.1.

I split prompts into color, style, lighting, and a combined adjective/adverb category using the Midjourney online community classifications discussed in Section 3.2. An illustration of classified words is presented in Table B1.

Table 4: Illustration of Prompt Attributes

Time	Text	X_{color}	$\mathbb{1}(Color)$	X_{style}	$\mathbb{1}(Style)$	...
0	Mountain, Black and White	-1	1	0	0	
1	Mountain, Black and White, Realistic	-1	1	0.9	1	
2	Mountain, Neon Color, Abstract	0.8	1	-0.7	1	
3	Mountain, Neon Color, Abstract	0.8	1	-0.7	1	

To quantify these attributes, I embed the classified words with large language models and apply principal component analysis (PCA) to project each attribute's embeddings onto a $[-1, 1]$ scale. An illustration of the attributes of prompts can be found in Table 4. Without specifying the direction of the numbers in each attribute, the embeddings and PCA automatically make colors that are more black and white into numbers closer to -1, and more colorful descriptions into numbers closer to 1. If the style is more realistic, the style attribute is closer to 1, compared to a style that is more abstract, with a number closer to -1.

I also include the dummy variables to indicate whether certain words are mentioned or not. For instance, if a prompt omits color terms, $\mathbb{1}(Color) = 0$. In this way, In this way, I am able to capture the difference between whether to specify certain attributes in the prompt, and the difference between different values of attributes.

5.3 Model Setup

Figure 11: Timeline

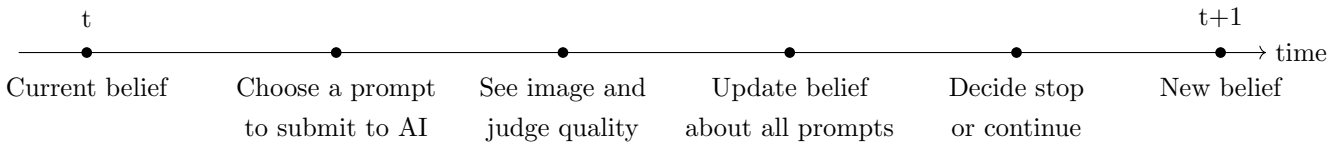


Figure 11 describes the timeline of the model. In this model, an individual holds beliefs about the expected returns of the prompts in the consideration set at a given time t. The individual selects a prompt j, a combination of words, to submit to the AI. A search cost c_j is then incurred. The individual will see and judge the quality of the image, and update the belief about how good all prompts are. The search cost refers to the mental effort to construct prompts and judge the quality of an image; money paid to Midjourney, and wait time. Since the average monetary cost to Midjourney per prompt is less than 10 cents, and

it takes, on average, less than 60 seconds for one generation, the search cost mainly refers to the mental effort. Based on active online discussions about prompt construction, it is usually mentally demanding to come up with the right prompt that generates satisfying results. After belief-updating, the individual will then decide whether to stop searching and end the session or to continue to search. If decide to continue, the individual will decide which prompt to search next.

I assume that individuals search sequentially. In every period, they choose the prompt with the highest expected return based on their current belief. They stop searching when no remaining option offers sufficient expected returns to justify the search cost. The trade-off is between the option value of the next prompt and the search cost.

5.3.1 Individual's Problem

In a given prompt j, denote $\{X_k\}_j$ to be the attributes of the prompt. $\{X_k\}_j$ is eight dimensional, where $k \in \{color, style, lighting, adj/adv, \mathbb{1}(color), \mathbb{1}(style), \mathbb{1}(lighting), \mathbb{1}(adj/adv)\}$. Denote $\{\beta_k\}$ as the marginal return of attributes in individuals' *prior belief*. In session s, the utility of prompt j at time t is written in equation (8).

$$u_{sjt} = m_{st}(X_j) + \epsilon_{sjt} \tag{8}$$

where $\epsilon_{sjt} \sim N(0, \sigma_\epsilon)$ represents the idiosyncratic noise in each evaluation. This is because even if an individual submits the same prompt to AI multiple times, it will give different results, similar to other generative AI like ChatGPT. The AI model itself has randomness.

The function $m_{st}(X_j)$ is sampled from a Gaussian process with prior mean of

$$\mu_{s0}(X_j) = \sum_k X_{jk}\beta_k \tag{9}$$

and the covariance matrix

$$\kappa_{s0}(X_j, X_{j'}) = \lambda^2 exp(-||j - j'||^2) \tag{10}$$

where $||j-j'||$ is the Euclidean distance of prompt j and j' embeddings, converted using OpenAI text-embedding-3-small model. The individual holds a prior mean μ_{s0} for the expected return of each prompt and a covariance matrix $\kappa_{s0}(X_j, X_{j'})$. If two prompts are similar in the embedding space, the covariance between them is high in prior belief. The randomness in $m_{st}(X_j)$ reflects the individual's uncertainty about how the AI interprets prompts without actually seeing the generated images.

At time t, after searches for prompt j, the individual observes the utility of the generated image u_{sjt}. The individual updates the beliefs of $m(X)$ function according to Bayes' rule, as illustrated in equation (11) and equation (12).

The intuition behind equation (11) is that observing the generated image for prompt j yields a realization of u_{sjt}. If u_{sjt} exceeds the current mean $\mu_{st}(X_j)$, the posterior mean for all prompts increases, with larger increases for prompts whose embeddings are more similar to j. The intuition behind equation (12) is that sampling prompt j reduces the posterior variance for j most substantially, and also lowers the variance for other prompts. Uncertainty goes down because of the information gained. The reduction in uncertainty for each prompt is proportional to its similarity with j. Prompts that are more similar to j experience greater variance reduction.

$$\mu'(X) = \mu(X) + \frac{\kappa(X, X_j)(u_j - \mu(X_j))}{\kappa(X_j, X_j) + \sigma_\epsilon^2} \tag{11}$$

$$\kappa'(X, X') = \kappa(X, X') - \frac{\kappa(X, X_j)\kappa(X_j, X')}{\kappa(X_j, X_j) + \sigma_\epsilon} \tag{12}$$

Assume that at every period t, individuals consider the current choice as their final choice. In other words, in the individual's problem, the continuation value equals zero. This "think one step ahead" approach follows Hodgson and Lewis (2025). This assumption simplifies estimation.

A state in session s at time t comprises two components. The first one is $u_{st}^* = max\{\hat{u}_{st}, 0\}$, which is the maximum between the highest utility among the past search results and the outside option utility normalized at zero. The second one is $f(u_{sjt})$, which represents the current belief of the utility of prompt j, with $u_{sjt} \sim N(\mu_{sjt}, \kappa_{st}^{jj} + \sigma_\epsilon^2)$.

Denote the search cost of prompt j as c_{sjt}. Assume $c_{sjt} = c + \zeta_{sjt},\ \zeta_{sjt} \sim$ EVT1 distribution, following Hodgson and Lewis (2025). The logit term ζ_{sjt} is drawn independently across s, t and j. This logit error assumption simplifies subsequent computation and captures heterogeneous search costs across sessions, time, and prompts. Note that search cost includes the mental effort required to construct prompts and assess the quality of generated images, the money paid to Midjourney, and the wait time. The heterogeneity primarily comes from differences in mental effort.

Given state variables $\{u_{st}^*, f(u_{sjt})\}$, the individual's problem can be written as:

$$\max\left\{ \underbrace{u_{st}^*}_{\text{stop}}, \underbrace{u_{st}^* + \max_{\hat{j}}\left\{ \int_{u_{st}^*}^{\infty} (u_{s\hat{j}t} - u_{st}^*) f(u_{s\hat{j}t})\, du_{s\hat{j}t} - c_{s\hat{j}t} \right\}}_{\text{continue searching}} \right\} \tag{13}$$

If stops searching, the individual secures the current best utility u^*_{st} with certainty. If continues, the individual will at least collect the current best utility and select prompt j to maximize the expected premium, in addition to the search cost. Similar to Weitzman (1979) and Hodgson and Lewis (2025), a prompt that has a more dispersed distribution on the right end is more favored.

A distinction that I make from Hodgson and Lewis (2025) is that prompts can be resubmitted to better fit the setting. Even if the same prompt is submitted to AI multiple times, it gives a random image each time. The scale of σ_ϵ will determine how often users revisit the same prompt. If AI randomness σ_ϵ is large, we will see more revisits in the data.

5.4 Identification and Estimation

Given the individual's problem, the probability of choosing prompt j can be written as follows. The idea is that if prompt j has a higher expected return given the current state, it will be chosen with a higher probability.

$$P(j|u^*, f) = \frac{\exp\left[u^* + \int_{u^*}^{\infty}(u_j - u^*)f(u_j)\,du_j - c\right]}{\exp(u^*) + \sum_{\hat{j}\in J}\left[u^* + \int_{u^*}^{\infty}(u_{\hat{j}} - u^*)f(u_{\hat{j}})\,du_{\hat{j}} - c\right]} \tag{14}$$

The likelihood of observing prompt sequences of session s in the data can be written as:

$$L_s = \int \Pi_{t=0}^{T-1} P_s(j|u^*_t, f_t) \times P_i(0|u^*_T, f_T) dF(\epsilon) \tag{15}$$

The likelihood of observing all sessions in the data is then:

$$L = \Pi_s L_s \tag{16}$$

I estimate the model based using MLE. The parameters to be identified are:

$$\psi = \left(\underbrace{\vec{\beta}}_{Prior\ Mean}, \underbrace{\sigma_\epsilon}_{AI\ randomness}, \underbrace{\lambda}_{learning\ parameter}, \underbrace{c}_{search\ cost} \right) \tag{17}$$

Table 5 presents the identification intuition of the parameters. $\vec{\beta}$ is the expected return of attributes in prior belief in equation (9). It will be identified by the first prompt in a session. λ is the coefficient before the covariance matrix in equation (10). If λ increases, the relative importance of the covariance between prompts increases. When prompt j is searched, the variance of prompt j and similar prompts decreases significantly if λ is large. This is because if the covariance between two prompts is high, when the individual obtains information from

one prompt, the uncertainty of the related prompt decreases significantly. λ is identified by how far the individual jumps away from prompt j after the individual searches for it.

σ_ϵ governs the AI randomness. Suppose that every time the same prompt is submitted to AI, AI gives the same image. Users will not have incentives to revisit the same prompt since it is costly. Hence, if AI is not random at all, we shall expect to see no revisit behaviors in the data. If, instead, AI gives very different images even if the same prompt is resubmitted to AI, we should expect to see more revisit behaviors in the data. σ_ϵ will be identified by the frequency of resubmission of the same prompt in the data. c is the average search cost, which will be identified by the average search length. If the search cost is high, the average search length should be short.

Table 5: Identification Intuition

Parameters	Interpretation	Identification Intuition
$\vec{\beta}$	Prior belief of attributes	Words in first search
λ	Covariance Scale	Jump distance after search prompt j
σ_ϵ	AI Randomness	Frequency of resubmitting the same prompt
c	Search Cost	Search length

5.4.1 Construct Consideration Set

In equation (14), the denominator could comprise millions of unique combinations of words in the English language, if no constraints are imposed. This large denominator is driving the probability of choosing prompt j very close to zero. If the individual is choosing from any potential prompt in the world, the probability of choosing any specific one of them at time t is almost zero. This will cause weak identification of the model.

In practice, however, users do not consider all combinations of words in the English language at each step. The path-dependent clusters of prompts documented in Section 3.1 imply that users restrict their choices to a local consideration set, prompts similar to the previous one, and consistent with the idea of an artwork.

I construct consideration sets under two constraints. First, I restrict estimation to landscape sessions. Appendix C.5 describes the classification procedure in detail. Estimates for portrait sessions yield similar results, which can be found in Table B3 in the appendix.

Second, I include in the consideration set for session s only prompts that (1) exist in the data; (2) lie within a distance threshold from any prompt in session s. To find this distance threshold, I compute embedding distances for all consecutive prompt pairs in landscape

sessions. Figure B4 in the appendix shows that over 95% of consecutive pairs lie within a distance of 4.7. I adopt this value as the threshold, yielding an average consideration set size of 21.85 prompts per session.

Another way to construct consideration sets is to identify the N nearest prompts in the data for each prompt in the session. To make estimation results comparable, I choose N = 14 to match the average set size from the distance-based method. The resulting average set size is 21.35. I estimate the model under both consideration set definitions.

5.4.2 Estimation Results

Table 6 reports the estimation results of the parameters. The complete results with β are presented in Table B2 in the appendix. Because the raw scales lack direct economic interpretation, their implications should be interpreted via counterfactual simulations. The results from the two consideration set definitions are very similar to each other. The main takeaway here is that the covariance matrix scale λ is significantly greater than zero, indicating that covariance plays a crucial role in belief updating. Users indeed adjust prompts based on judgment of the previous results.

Table 6: Model Estimates

	(1)	(2)
Consideration Set	Prompts Within Distance	N Nearest Prompts
λ (Covariance Scale)	81.82	80.97
	(1.64)	(1.56)
σ_ϵ (AI Randomness)	13.44	14.24
	(0.32)	(0.33)
c (Search Cost)	6.43	6.66
	(0.08)	(0.08)

Notes: Standard errors in brackets, computed using the observed Fisher information.

5.5 Counterfactual

The key element of human adaptation to output is the individual's ability to *judge* the quality of generated images and adjust subsequent prompts accordingly. This impacts the direction of search, and hence the production process. To understand the role of such adaptation, I conduct a counterfactual analysis where I eliminate this kind of judgment by setting $cov(j, j') = 0, \forall j \neq j'$.

In the most extreme case, all judgment is removed. In that case, individuals are submitting prompts randomly at each period. The mathematician Émile Borel once used a metaphor of the infinite monkey: Imagine a monkey hitting keys independently and randomly on a typewriter for an infinite amount of time. Almost surely, it will type any given text, including the complete works of William Shakespeare. In the context of image generation, if we iterate over every combination of pixels on a canvas, we will almost surely obtain some masterpieces. However, it is not an efficient way of production. We still need humans to go through all these images and select the masterpieces based on their judgment.

In this counterfactual, I am making the individuals smarter than the infinite monkey, in the sense that they can submit a list of prompts to the AI based on their prior beliefs. But the individuals are less adaptive than individuals in the data, in the sense that they are not allowed to adjust prompts based on judgment of previous output.

The prompts with a higher expected return in prior beliefs are ranked higher in the predetermined list. The AI will proceed with the order of the list and generate images. However, unlike the reality setting, the individual is not allowed to change the order of the list based on the judgment of the images. The list is infinitely long. The counterfactual exercise is designed to ask: how many more prompts are needed to reach the same final prompt in a session as in the data?

Table 7 presents the results of the counterfactual analysis. Without human adaptation directing the prompt-searching behavior, it takes around three times more prompts per session to achieve the same outcome. The results suggest that human adaptation plays a crucial role in directing search. When an individual sees an unsatisfying image, the prompt is adjusted to move away from the previous one. This adaptation increases production efficiency by around three times. This number highlights the importance of human adaptation to the creative process.

Table 7: Counterfactual Results

	(1)	(2)
Consideration Set	Prompts Within Distance	N Nearest Prompts
#Prompts per Session Increases	313%	284%

Notes: The table reports, for each consideration-set definition, how much longer sessions would be in the counterfactual where users cannot adapt prompts to previously observed images. The outcome is the percentage increase in the number of prompts needed to reach the same final prompt as in the data. Column (1) defines the consideration set as all prompts within an embedding-distance threshold from the session's prompts; column (2) defines it as the N nearest prompts in the data. In both cases, removing adaptation requires roughly three times more prompts.

6 Conclusion

In this paper, I study the role of human adaptation in the creative process with generative AI. In particular, I focus on two types of adaptation: (1) adaptation to different AI versions; (2) adaptation to outputs from previous prompts within the creative process of an artwork.

To understand the role of adaptation to different AI versions, I utilize AI upgrades during the observation period and show that the new AI is significantly more accurate compared to the old AI. When AI capability changes, users also adapt to the new technology and change the words to include in their prompts. Users include fewer words, such as "realistic" and "detailed" in the new AI, since it automatically produces images with greater realism and detail even without these explicit instructions. Conversely, users include more words about colors and photography in the new AI because it can recognize these words more effectively than before. Thus, when AI is upgraded, both AI capability and human input shift, jointly shifting the output images.

To isolate the effect of AI changes and prompt changes on the final outcome, I decompose the observed output shifts by resubmitting new prompts to the old AI and old prompts to the new AI. The generated images are converted into embeddings with LLM and projected onto the direction of $\overrightarrow{Image}_{new} - \overrightarrow{Image}_{old}$. The decomposition results show that 73% of the output shifts come from changing prompts alone, 20% from changing AI alone, and 7% of residuals, suggesting complementarity between AI and human input. The decomposition highlights the importance of human adaptation for extracting value from new technologies.

To understand the role of adaptation to outputs from previous prompts, I investigate prompts within a session and show that users refine the prompts iteratively to achieve their desired outcome. I build a sequential search model to capture the prompt patterns. In the model, individuals search sequentially for the prompts that are correlated. When a prompt is searched, the individual sees the generated image, evaluates the quality, and adjusts the next prompt accordingly. A counterfactual exercise that disables output assessment reveals that each session requires 313% more prompts, on average, to reach the same final outcome. These results emphasize the role of human adaptation in guiding the creative process.

Taken together, this paper quantifies the value of human adaptation in creative processes with AI. As generative models advance, human judgment and iterative adaptation remain essential for realizing their full creative potential. The findings imply that technological improvements do not automatically yield the best creative outcomes. Instead, it requires humans to judge how to use the technology effectively. In this sense, adaptation is a crucial driver of innovation.

References

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2022). Artificial Intelligence and Jobs: Evidence from Online Vacancies. *Journal of Labor Economics*, *40*(S1), S293–S340. Publisher: The University of Chicago Press. doi:10.1086/718327

Acemoglu, D., & Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*, *128*(6), 2188–2244. Publisher: The University of Chicago Press. doi:10.1086/705716

Agarwal, N., Moehring, A., Rajpurkar, P., & Salz, T. (2023, July). Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology. Working Paper. doi:10.3386/w31422

Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction. *Journal of Economic Perspectives*, *33*(2), 31–50. doi:10.1257/jep.33.2.31

Angelova, V., Dobbie, W. S., & Yang, C. (2023, October). *Algorithmic Recommendations and Human Discretion* (tech. rep. No. w31747). National Bureau of Economic Research. doi:10.3386/w31747

Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, *151*, 103745. doi:10.1016/j.jfineco.2023.103745

Bronnenberg, B. J., Kim, J. B., & Mela, C. F. (2016). Zooming In on Choice: How Do Consumers Search for Cameras Online? *Marketing Science*, *35*(5), 693–712. Publisher: INFORMS. doi:10.1287/mksc.2016.0977

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at Work*. *The Quarterly Journal of Economics*, *140*(2), 889–942. doi:10.1093/qje/qjae044

Chatterji, A., Cunningham, T., Deming, D., Hitzig, Z., Ong, C., Shan, C., & Wadman, K. (2025, September). How People Use ChatGPT. SSRN Scholarly Paper. doi:10.2139/ssrn.5487080

Chen, Y., & Yao, S. (2017). Sequential Search with Refinement: Model and Application with Click-Stream Data. *Management Science*, *63*(12), 4345–4365. Publisher: INFORMS. doi:10.1287/mnsc.2016.2557

Choi, J. H., & Schwarcz, D. (2024). AI Assistance in Legal Analysis: An Empirical Study. *Journal of Legal Education*. Retrieved October 19, 2025, from https://librarysearch.library.utoronto.ca/nde/fulldisplay/cdi_scopus_primary_2_s2_0_105004031311/01UTORONTO_INST:UTORONTO_NDE/en

De Los Santos, B., Hortaçsu, A., & Wildenbeest, M. R. (2017). Search With Learning for Differentiated Products: Evidence from E-Commerce. *Journal of Business & Economic Statistics*, *35*(4), 626–641. Publisher: ASA Website _eprint: https://doi.org/10.1080/07350015.2015.11 doi:10.1080/07350015.2015.1123633

De los Santos, B., & Koulayev, S. (2017). Optimizing Click-Through in Online Rankings with Endogenous Search Refinement. *Marketing Science*, *36*(4), 542–564. Publisher: INFORMS. doi:10.1287/mksc.2017.1036

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., ... Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *SSRN Electronic Journal*. doi:10.2139/ssrn.4573321

Dzyabura, D., & Hauser, J. R. (2019). Recommending Products When Consumers Learn Their Preference Weights. *Marketing Science*, *38*(3), 417–441. Publisher: INFORMS. doi:10.1287/mksc.2018.1144

Handa, K., Tamkin, A., McCain, M., Huang, S., Durmus, E., Heck, S., ... Ganguli, D. (2025, February). Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations. arXiv:2503.04761 [cs]. doi:10.48550/arXiv.2503.04761

Hirsch, S., Guy, I., Nus, A., Dagan, A., & Kurland, O. (2020, July). *Query Reformulation in E-Commerce Search*. Pages: 1328. doi:10.1145/3397271.3401065

Hodgson, C., & Lewis, G. (2025). You Can Lead a Horse to Water: Spatial Learning and Path Dependence in Consumer Search - The Econometric Society. *Econometrica*, *93*(4), 1299–1332. doi:https://doi.org/10.3982/ECTA19576

Honka, E., Hortaçsu, A., & Wildenbeest, M. (2019). Empirical search and consideration sets. In *Handbook of the Economics of Marketing* (Vol. 1, pp. 193–257). doi:10.1016/bs.hem.2019.05.002

Kanazawa, K., Kawaguchi, D., Shigeoka, H., & Watanabe, Y. (2025). AI, Skill, and Productivity: The Case of Taxi Drivers. *Management Science*. Publisher: INFORMS. doi:10.1287/mnsc.2023.01631

Kim, J. B., Albuquerque, P., & Bronnenberg, B. J. (2010). Online Demand Under Limited Consumer Search. *Marketing Science*, *29*(6), 1001–1023. Publisher: INFORMS. doi:10.1287/mksc.1100.0574

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*(2), 132–157. Publisher: University of Chicago Press. Retrieved October 15, 2025, from https://www.jstor.org/stable/1828835

Munkres, J. (1957). Algorithms for the Assignment and Transportation Problems. *Journal of the Society for Industrial and Applied Mathematics*, *5*(1), 32–38.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187–192. Publisher: American Association for the Advancement of Science. doi:10.1126/science.adh2586

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023, February). The Impact of AI on Developer Productivity: Evidence from GitHub Copilot. arXiv:2302.06590 [cs]. doi:10.48550/arXiv.2302.06590

Ursu, R. M., Wang, Q., & Chintagunta, P. K. (2020). Search Duration. *Marketing Science*, *39*(5), 849–871. Publisher: INFORMS. doi:10.1287/mksc.2020.1225

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, *8*(12), 2293–2303. Publisher: Nature Publishing Group. doi:10.1038/s41562-024-02024-1

Weitzman, M. L. (1979). Optimal Search for the Best Alternative. *Econometrica*, *47*(3), 641–654. Publisher: [Wiley, Econometric Society]. doi:10.2307/1910412

Zhou, E. B., & Lee, D. (2023, October). Generative AI, Human Creativity, and Art. SSRN Scholarly Paper. doi:10.2139/ssrn.4594824

Zhou, E. B., Lee, D., & Gu, B. (2025). Who expands the human creative frontier with generative AI: Hive minds or masterminds? *Science Advances*, *11*(36), eadu5800. Publisher: American Association for the Advancement of Science. doi:10.1126/sciadv.adu5800

Appendix

A Descriptive Prompt Writing Patterns

A.1 Prompt Length Increases Along Path

This section shows that users add words, parameters, and image inputs incrementally along the path. Figure A1 panel (a) plots the coefficients of $k^{th}Prompt$ in regression (A1). The dependent variable is the number of words in prompt p written by user i in session a. The right-hand side consists of dummy variables indicating whether it is the k-th prompt in a session, along with the session fixed effects. If β increases with k, users are adding words to the prompt along the path.

In panel (a), the horizontal axis indicates whether it is the k-th prompt in a session, and the vertical axis is the coefficients and the 95% confidence intervals. I constrain the sample to be sessions with at least 15 prompts to avoid effects coming from the fact that longer search sequences have longer prompts. The figure shows that users are adding words to prompts along the path, at a decreasing rate. The effect is most salient in the first 12 prompts, and the trend slows down afterwards. Similar patterns arise if the dependent variable becomes the number of parameters or the number of image inputs.

$$\#Words_{isp} = \sum_k \beta_k \mathbb{1}(k^{th}Prompt_{isp}) + \mu_s^i + \epsilon_{isp} \tag{A1}$$

Table A1 shows the average effects of the sample. It presents the results from regression (A2). Although the effect seems small compared to the mean in the first prompt in a session, we need to keep in mind that the effects are diminishing along the path, and this is just the average effect. The effect in the first few prompts is not negligible, as shown in Figure A1.

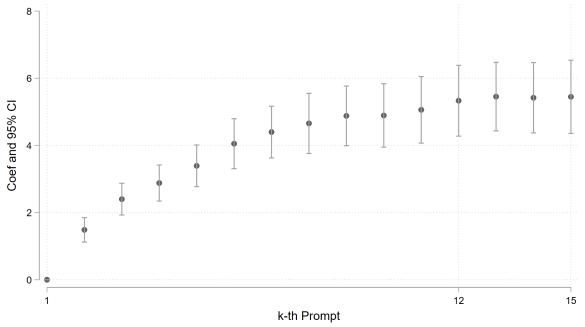
$$\#Words_{isp} = \beta k^{th}Prompt_{isp} + \mu_s^i + \epsilon_{isp} \tag{A2}$$

A.2 Prompts Become Increasingly Similar to the Final Prompt in Session

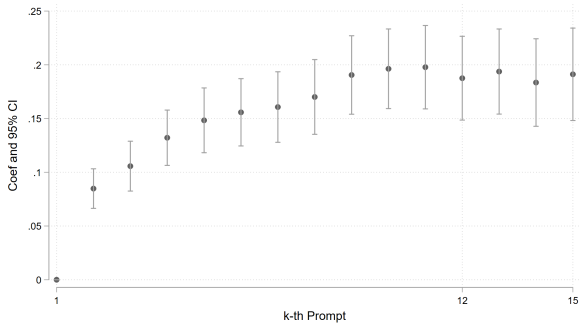
This section shows that along the path, prompts are converging to the final prompt in a session. Figure A2 plots the coefficients before $\mathbb{1}(k^{th}Prompt_{iap})$ in regression (A3). The dependent variable is the cosine similarity between prompt p written by user i in session a, and the final prompt in the same session. The cosine similarity is calculated based on prompt

Figure A1: Prompt Length Increases Along The Path

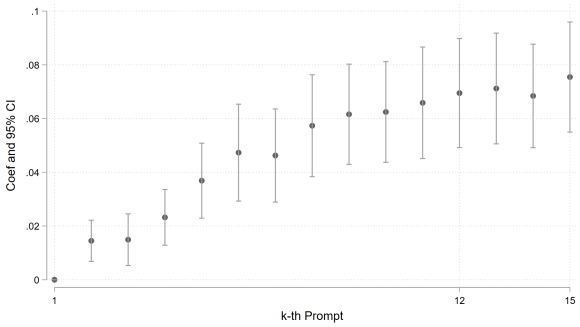
(a) Words



(b) Parameters



(c) Image Inputs



Notes: Panel (a) in this figure shows the coefficients before $\mathbb{1}(k^{th}Prompt)$ in regression (A1). The horizontal axis indicates whether it is the k-th prompt in a session, and the vertical axis is the coefficients and the 95% confidence intervals. Standard errors are clustered at the user level. The reference point is the first prompt in a session. Panels (b) and (c) plot similar coefficients with the dependent variable being the number of parameters and the number of image inputs instead. I constrain the sample to be sessions with at least 15 prompts to avoid effects coming from the fact that longer search sequences have longer prompts. I have tried changing this number to 10 or 30, and the trends persist.

embeddings, which are vectors generated by large language models.[10] The independent variables are dummies indicating whether this is the k-th prompt in the session, along with session fixed effects. Standard errors are clustered at the user level. If β is increasing with k, it implies that prompts are increasingly similar to the final prompt in a session.

There are two trends worth noticing in Figure A2. Firstly, for a given Midjourney version, β increases with k. This trend suggests that prompts are converging to the final prompt in a session along the path.

Secondly, prompts in later Midjourney versions are converging faster. For example, V4 has a larger β for every k in V3. This suggests that users are taking greater steps each time

[10]I employ both "sentence-transformers/all-MiniLM-L6-v2" in Python and "text-embedding-3-small" model from OpenAI. They give similar results.

Table A1: Prompt Length Grows Along The Path

	(1)	(2)	(3)
Dep Var	$\#Words$	$\#Parameters$	$\#Image\ Input$
$k^{th} Prompt$	0.336***	0.010***	0.005***
	(0.037)	(0.001)	(0.0007)
Mean in 1^{st} Prompt	21.18	1.10	0.10
Session FE	Y	Y	Y
N	70,380	70,380	70,380
R^2	0.87	0.84	0.74

Notes: *** denotes significance at 1 percent, ** at 5 percent, and * at 10 percent. Standard errors are clustered at the user level. This table presents the estimates of regression (A2). Here, I restrict the sample to be sessions with at least 15 prompts because I want to obtain the average effect of the prompt length growth in Figure A1.

they search in later Midjourney versions.

$$Cosine\ Similarity(Prompt_{isp}, Prompt_{isT}) = \sum_k \beta_k \mathbb{1}(k^{th} Prompt_{isp}) + \mu_s^i + \epsilon_{isp} \tag{A3}$$

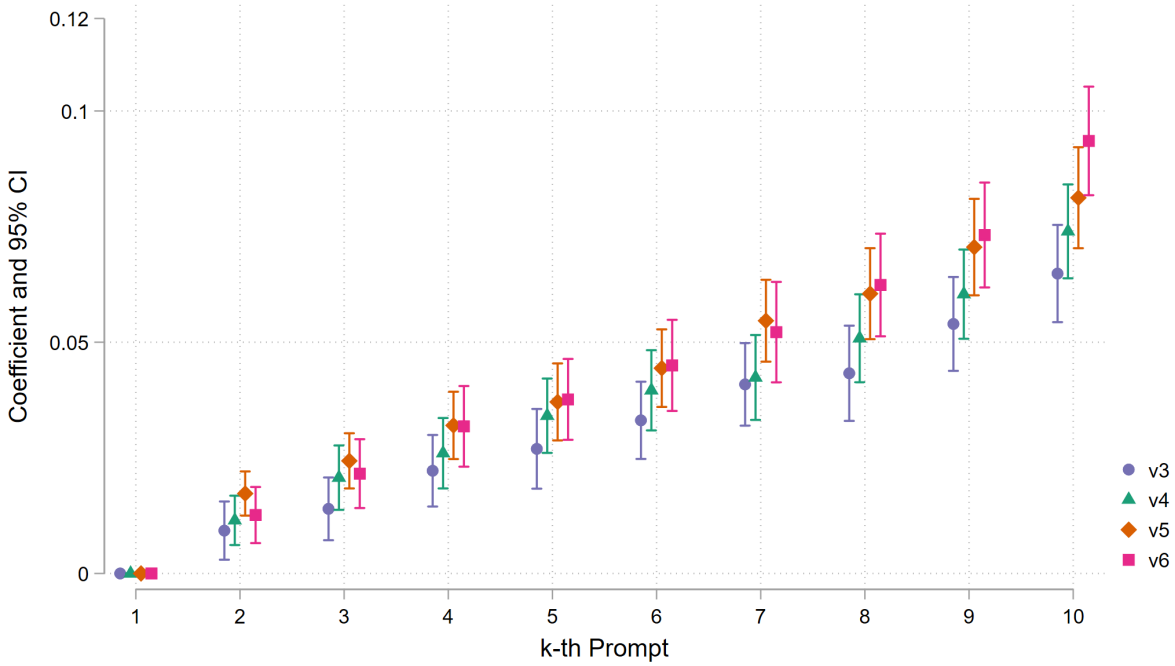
Another way to verify that users take smaller steps in the prompt is to run regression A4. The dependent variable becomes the cosine similarity between prompt p and its previous prompt $p-1$. The independent variables are dummies representing the Midjourney version being used, along with user fixed effects. Figure A3 plots the coefficients of β. The results show that users take smaller steps in V3, but bigger steps in V4, V5, and V6.

$$Cosine\ Similarity(Prompt_{isp}, Prompt_{isp-1}) = \sum_v \beta_v \mathbb{1}(Midjourney\ Version_{isp} = v) + \mu_i + \epsilon_{isp} \tag{A4}$$

A.3 Words Change in Order

This section shows that users change words with higher weights in image outputs in early stages, and then they move to adjusting minor words. Words with higher weights are words that are closer to the beginning of the prompt. Online discussions show that users are aware of this rule. Some statements include "AI prioritizes the sequence of words in your prompt with the most important words being at the front of the prompt and descending from there, so be sure to organize the structure of your prompt by putting what you want to see most at the front of it"; "words 1-5 are very influential, ..., words 40+ are very likely to appear

Figure A2: Convergence of Prompts Towards Final Prompt in Session



Notes: This figure plots coefficients before $\mathbb{1}(k^{th} Prompt)$ in regression (A3) and the 95% confidence intervals. I constrain the sample to be sessions with more than 10 prompts to avoid effects coming from longer search sequences converging faster. I also constrain the sample to be sessions with no more than 50 prompts because I want the first 10 prompts to be non-negligible in the search sequence. Imagine a search sequence with 500 prompts. The first 10 prompts only play a small part in the search sequence, and the effects can be noisy if the early convergence trend is very different from the late stages. I have tried different upper bounds and lower bounds, and the patterns persist.
I also divide the sample by the Midjourney version they are using to show that users are converging at different rates when using different versions.

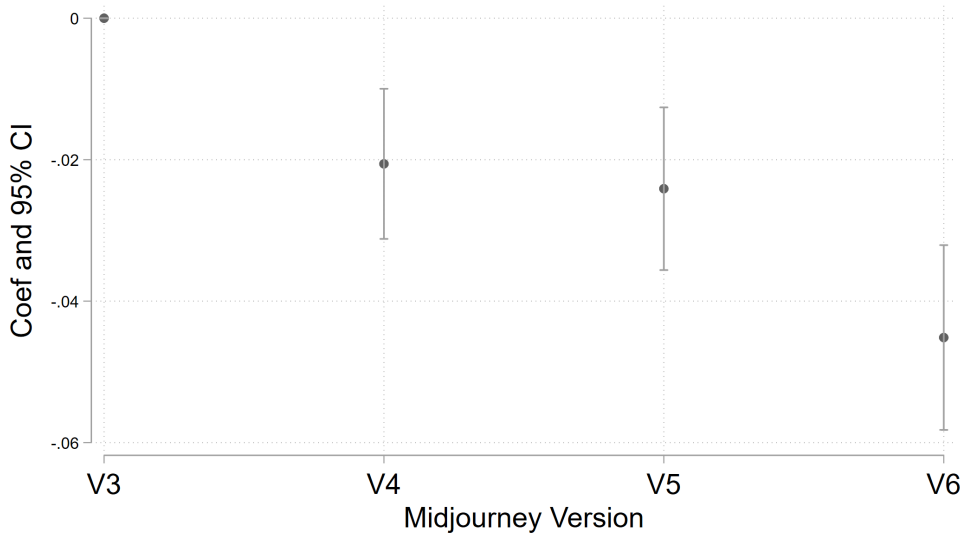
ignored". Figure A4 shows an example. The words in the prompt are the same in the left and right panels, with the only difference being the order of words. In the left panel, "Pikachu" is at the beginning and "at park" is at the end, while in the right panel, the order flips. In response, AI returns images with a bigger Pikachu in the left panel and emphasizes the park on the right panel.

The regression A5 examines whether the following prompt sequence is happening.

1. Pikachu, green shirt, drink coffee
2. Pikachu, **blue** shirt, drink coffee
3. Pikachu, **blue** shirt, drink **soda**

① **②** ③ ④ **⑤**

Figure A3: Larger Steps In Prompts With Newer AI



Notes: This figure plots coefficients before $\mathbb{1}(Midjourney\ Version_{iap} = v)$ in regression (A4) and the 95% confidence intervals. I constrain the sample to be sessions with more than 10 prompts and no more than 50 prompts to be consistent with Figure A2. I have tried different upper bounds and lower bounds, and the patterns persist.

$$j^{th}Word\ Change_{isp} = \sum_k \beta_k \mathbb{1}(k^{th}Prompt_{isp}) + \mu_s^i + \epsilon_{isp} \tag{A5}$$

The dependent variable indicates, compared to the most similar previous prompts, which is the word with the earliest position changed.[11] Table A2 demonstrates what the data looks like. In the previous example, the first prompt is removed from the data because it does not have a previous prompt as a reference. In the second prompt, the earliest word that has changed compared to the first prompt is the second word "blue". Hence, the dependent variable equals 2. Similarly, in the third prompt, the earliest word that has changed compared to the second prompt is the fifth word "soda". The dependent variables are dummies indicating the prompt position in a session, along with session fixed effects. Standard errors are clustered at the user level. Figure A5 plots β and the 95% confidence intervals. An increasing β with k suggests that users change words closer to the beginning of the prompt first, and then move on to adjusting later words.

A question here is whether this ordered word change happens horizontally or vertically. Denote a vertical change as adding words to the most similar previous prompt. In other words, one of the previous prompts is a strict subset of the current prompt. Denote a

[11] By most similar, I mean the prompt with the most number of words at the beginning that are the same.

Figure A4: Words at the Beginning are More Important

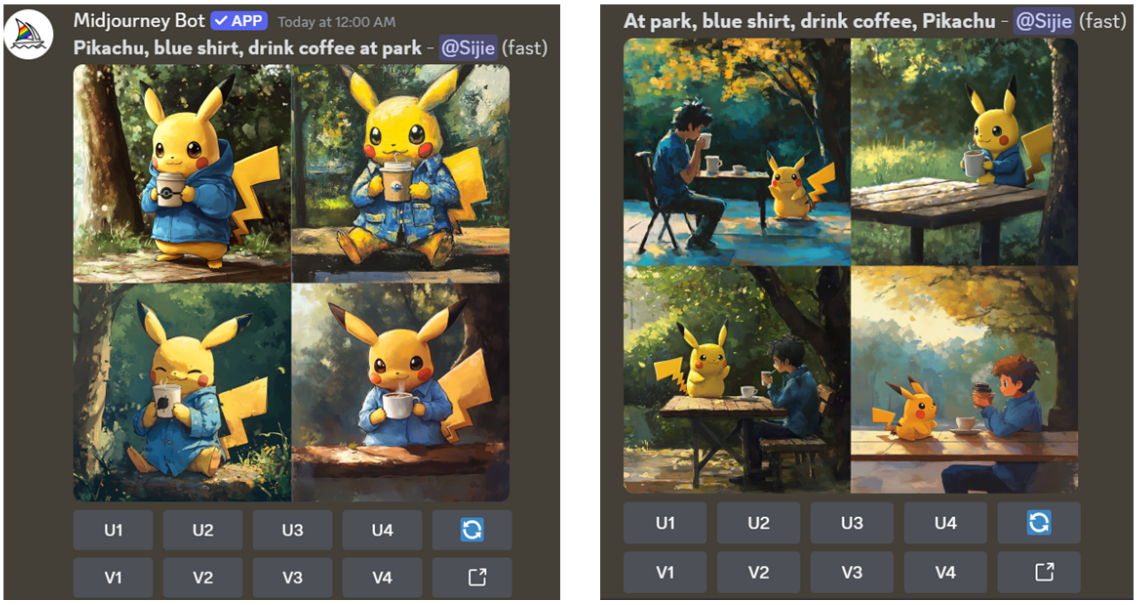


Table A2: Sample Data For Regression (A5)

	Y	$\mathbb{1}(2^{ed}Prompt)$	$\mathbb{1}(3^{rd}Prompt)$	..
Pikachu, **blue** shirt, drink coffee	2	1	0	..
Pikachu, **blue** shirt, drink **soda**	5	0	1	..

horizontal change as a change of words from the most similar previous prompt. The following example serves as an illustration.

Vertical

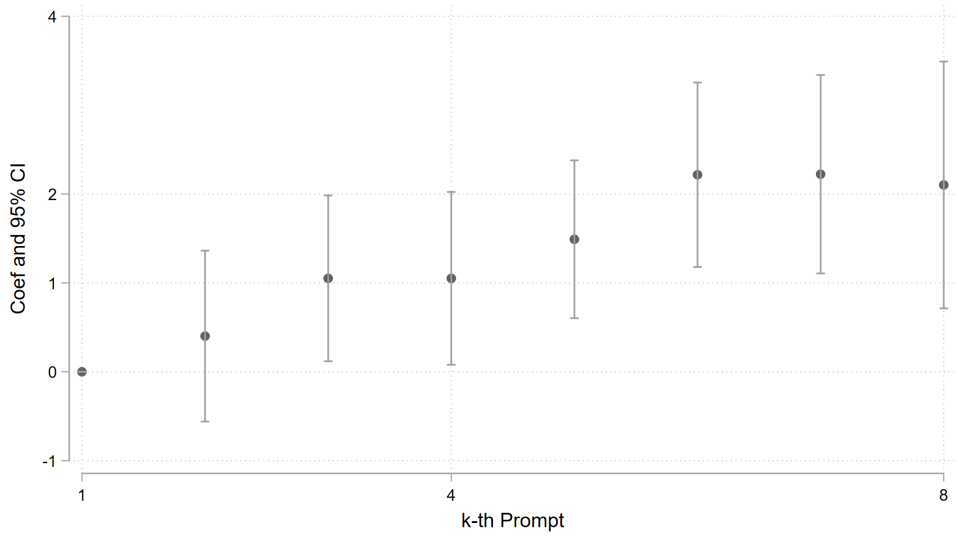
1. Pikachu
2. Pikachu, **blue shirt**
3. Pikachu, blue shirt, **drink coffee**
4. Pikachu, blue shirt, drink coffee **at café**

Horizontal

1. Pikachu, green shirt, drink coffee at café
2. Pikachu, **blue** shirt, drink coffee at café
3. Pikachu, **blue** shirt, drink **soda** at café
4. Pikachu, **blue** shirt, drink **soda** at **park**

I divide the sample into two subsamples. The first subsample only includes prompts that are a vertical change from the most similar previous prompt, and the second one includes all other prompts. And I run regression A5 separately in both samples, and Figure A6 shows the results. The results suggest that the ordered word change happens both vertically and horizontally.

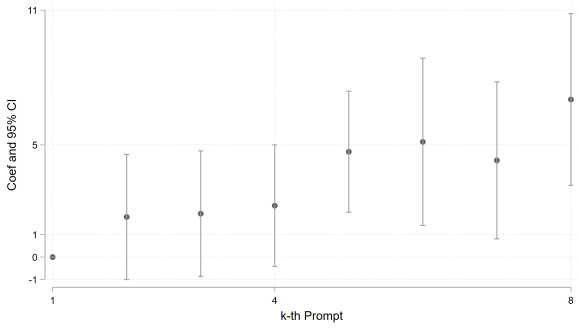
Figure A5: Change Early Words First



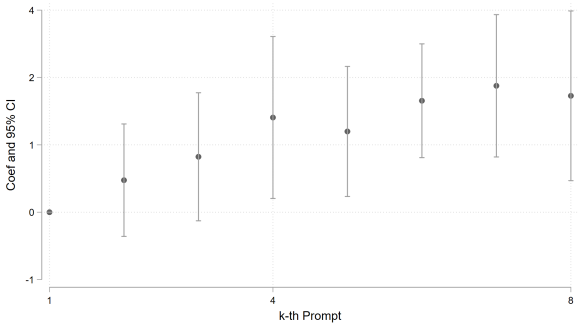
Notes: This figure plots coefficients before $\mathbb{1}(k^{th} Prompt)$ in regression (A5) and the 95% confidence intervals. I constrain the sample to be sessions with at least 8 prompts to avoid effects coming from longer search sequences changing words further away from the beginning of the prompt. I have tried different thresholds, and the trends persist.

Figure A6: Ordered Word Change Happens Both Vertically and Horizontally

(a) Vertical

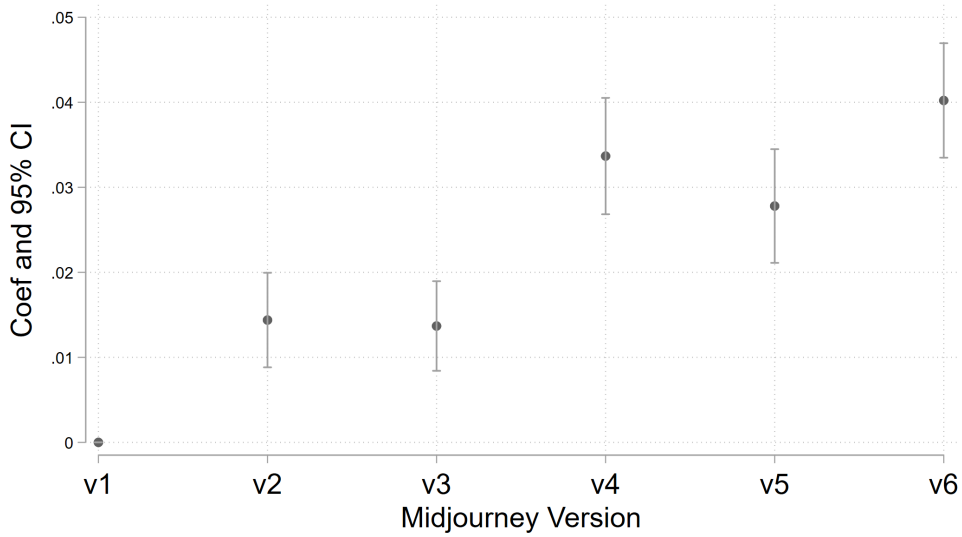


(b) Horizontal



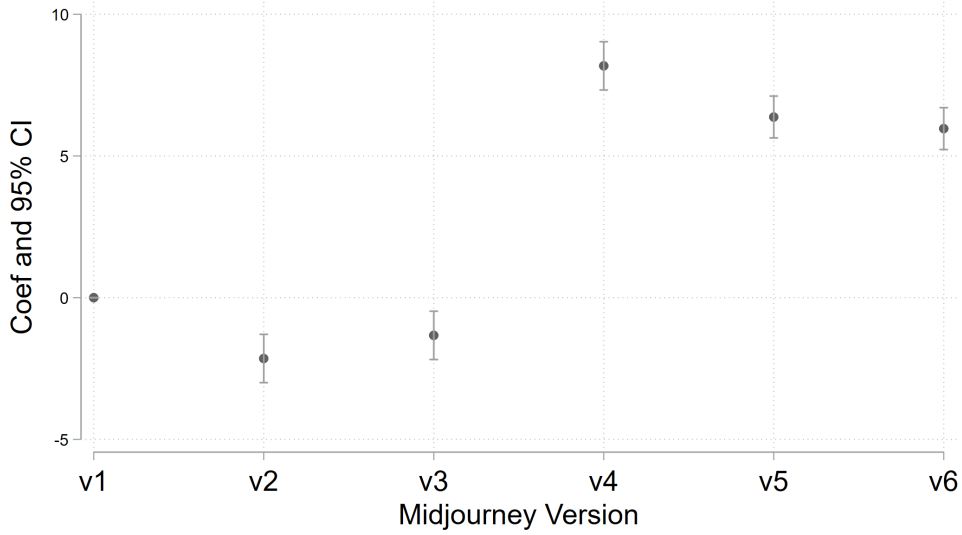
B Additional Tables and Figures

Figure B1: Later Versions Give More Accurate Results



Notes: This figure plots coefficients before $\mathbb{1}(Midjourney\ version = v)$ in regression (1) and the 95% confidence intervals. Standard errors are clustered at the prompt level. This figure uses embeddings all-MiniLM-L6-v2.

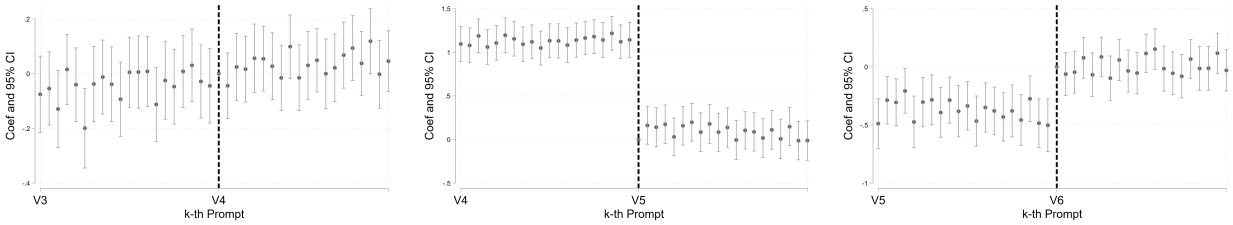
Figure B2: Later Midjourney Versions Give More Detailed Images



$$\%Upscale_{is} = \sum_k \beta_k \mathbb{1}(k^{th}\ Session)_{is} + \mu_i + \epsilon_{is} \tag{B1}$$

Figure B3: Upscale Ratio Changes across Versions

(a) V3-V4 (b) V4-V5 (c) V5-V6



Notes: This figure plots coefficients before $\mathbb{1}(k^{th}\ Session)$ in regression (B1) and the 95% confidence intervals. Note that in V3, default resolution before upscaling is 256×256 pixels, V4 is 512×512 pixels, V5 is 1024×1024 pixels, V6 is 1024×1024 pixels. Hence, the incentives for upscaling could decrease between V3 and V4, V4 and V5. The only versions that are comparable are V5 and V6.

Table B1: Illustration of Splitting Texts

Time	Text	Color	Style	...
0	Mountain, Black and White	Black and White	-	
1	Mountain, Black and White, Realistic	Black and White	Realistic	
2	Mountain, Neon Color, Abstract	Neon Color	Abstract	
3	Mountain, Neon Color, Abstract	Neon Color	Abstract	

Table B2: Model Estimates (Complete)

Consideration Set	(1) Prompts Within Distance	(2) N Nearest Prompts
λ (Covariance Scale)	81.82	80.97
	(1.64)	(1.56)
σ_ϵ (AI Randomness)	13.44	14.24
	(0.32)	(0.33)
c (Search Cost)	6.43	6.66
	(0.08)	(0.08)
β_{color}	0.18	-0.12
	(0.29)	(0.29)
β_{style}	-0.70	-0.56
	(0.22)	(0.22)
$\beta_{lighting}$	0.95	0.70
	(0.44)	(0.48)
$\beta_{adj/adv}$	0.35	0.54
	(0.18)	(0.17)
$\mathbb{1}(color)$	-0.75	-1.05
	(0.21)	(0.19)
$\mathbb{1}(style)$	-1.48	-1.80
	(0.21)	(0.21)
$\mathbb{1}(lighting)$	-1.34	-1.32
	(0.31)	(0.33)
$\mathbb{1}(adj/adv)$	-1.01	-1.54
	(0.25)	(0.22)

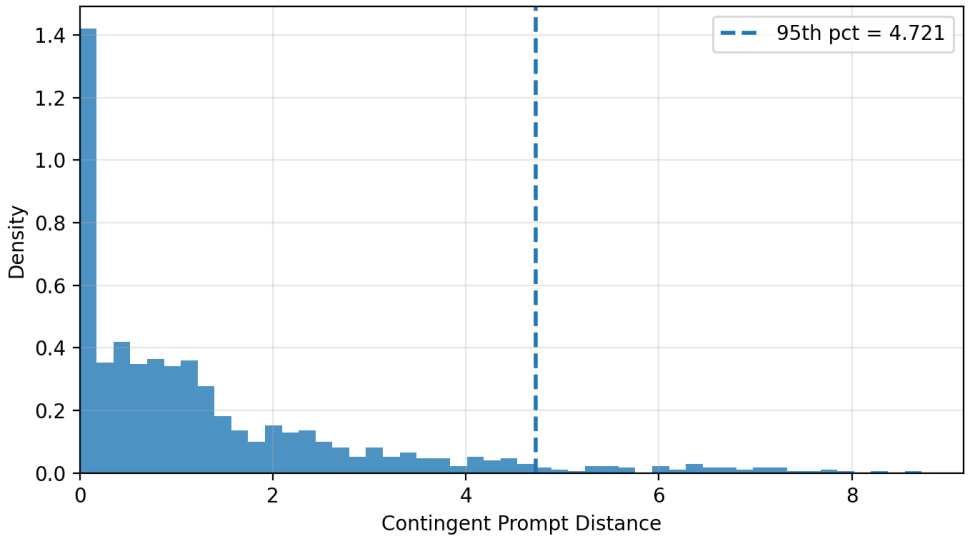
Notes: Standard errors in brackets, computed using the observed Fisher information.

Table B3: Model Estimates (Portrait)

Consideration Set	(1) Prompts Within Distance	(2) N Nearest Prompts
λ (Covariance Scale)	99.43	98.87
	(2.77)	(2.77)
σ_ϵ (AI Randomness)	17.05	18.31
	(0.59)	(0.59)
c (Search Cost)	8.13	8.49
	(0.19)	(0.19)
β_{color}	-0.41	-0.21
	(0.29)	(0.29)
β_{style}	-2.21	-2.40
	(0.25)	(0.25)
$\beta_{body\ part}$	0.73	0.87
	(0.21)	(0.21)
$\beta_{adj/adv}$	-0.14	-0.10
	(0.28)	(0.28)
$\mathbb{1}(color)$	-0.34	-0.72
	(0.25)	(0.25)
$\mathbb{1}(style)$	-2.28	-2.88
	(0.36)	(0.36)
$\mathbb{1}(body\ part)$	-1.87	-2.15
	(0.24)	(0.24)
$\mathbb{1}(adj/adv)$	-1.98	-2.44
	(0.44)	(0.44)

Notes: Standard errors in brackets, computed using the observed Fisher information.

Figure B4: Contingent Prompt Distance Distribution



C Data Appendix

C.1 Data Collection

Data collection started in February 2, 2025 and ended in February 8, 2025. I collected messages from channels called "general 1", "general 2", ..., "general 20".

C.2 Clean Errors And Link Type 19 Messages to Type 0 Messages

The data cleaning process follows the following steps.

1. Read all messages.

2. Delete messages whose type is not 0 or 19. Type 0 messages are those prompts submitted by users. Type 19 messages are the result of clicked buttons. I also delete those that are not sent by the Midjourney Bot and those with errors.

3. For each message of type 19, identify the *message_reference* to determine which message the user interacted with to trigger the current message. By doing so, I can trace backward until a type 0 message is found. For a given type 0 message, if there is more than one user involved along the path, meaning that more than one user clicks on the buttons, I remove this type 0 message and all the subsequent messages.

4. Break "content" in each message into text inputs, parameters, image inputs, and mode.

5. For each type 0 message, if at least one image input comes from the generated output of another user, remove the path (keep the original user's path).

C.3 Clustering Prompts Into Sessions

This section describes the details of clustering prompts into sessions. Table C1 presents the distribution of session time spans.

1. For text inputs in each type 0 message, remove stop words and "magic words".

    - Stop words are words like "a", "an", "the", "here". This is a terminology usually employed by computer scientists who study natural language processing. I am using the list from the "nltk.corpus" package in Python.

2. For a given user, collect all type 0 messages and find N cleaned text inputs. Calculate a $N \times N$ Jaccard similarity matrix. The Jaccard similarity score is between 0 and 1, with a larger number meaning texts being more similar.

    - For each type 0 message, if at least one of the image inputs comes from the generated output of the **same user**, assign a similarity score of one to guarantee that they are in the same session.

3. For prompt pairs in which at least one of them does not have text input, look at the image inputs. If at least one of the image inputs is the same, assign a similarity score of one to guarantee that they are in the same session.

4. Use DBSCAN to cluster type 0 messages within a user. Assign the threshold to be 0.2.

5. For each session, find the first action by timestamp and remove it if it is within the last 24 hours of the data collection period. This is to remove those sessions that have just started, but I did not collect the complete path due to a fixed data collection start time.

Table C1: Distribution of Session Time Span

Session Time Span	%
0–1 h	90.62
1–6 h	1.30
6–12 h	0.37
12–24 h	0.55
24–48 h	0.47
48–72 h	0.25
3–7 days	0.56
7–30 days	1.11
> 30 days	4.77

C.4 Word List For Classifications

Below are some examples of the word classifications. The complete list will be available in the online appendix.

1. *detailed words:* detail, detailed, high-detailed, ultra-detailed, hyper-detailed, extremely-detailed
2. *realistic words:* realistic, hyperrealistic, super realistic, photorealistic, photorealism
3. *rendering engines:* rendering engine, unreal engine, Cinema4D, C4D, Arnold Render, Octane Render
4. *colors:* vibrant colors, black, white, emerald, olive-green, vivid-maroon, cyan, neon colors, light yellow
5. *stop words:* in, on, just, me, as, is, are, am, there, here, had, while
6. *photography words:* fisheye lens effect, 135mm lens, white balance, shutter speed, stop motion
7. *lighting:* tyndall effect, dim lighting, glowing radioactivity, studio light, dj lighting
8. *style:* post-impressionism, neo-pop art, street art, futurism, cubist, dadaism

C.5 Clustering Sessions into Topics

I classify sessions into topics using a popular language-processing Python package called BERTopic. It is widely used in classifying documents into topics using clustering.

For prompts in a single session, I append the text inputs together. And then I remove the color words and stop words to improve clustering accuracy. I then cluster sessions into 100 topics using BERTopic, with the K-Means clustering method. Below are some example of topics, with corresponding keywords, and some sampled prompts in each topic.

1. Landscapes
   Keywords: mountains, lake, sky, landscape, desert, river, sunset, water
   Sampled Prompts:

   - "new Mexico landscape, clear blue sky, rocky mountains, rainclouds, dramatic, oil painting, detailed, Albert bierstadt style"
   - "dramatic cloud formation coming from behind a mountain, clear blue sky, midday, bright, southwest, colorful, awe-inspiring, detailed, oil painting"

2. Cute Animals
   Keywords: cat, dog, cute, kitten, adorable, puppy, fur, bunny
   Sampled Prompts:

   - "super cute black french bulldog + big expressive eyes + Pixar + ghibli + extremely detailed + detailed reflections + bold highlights + photo-realistic + hdr + octane render + 8K + soft light + twinkling fire + smoke"
   - "white cat, roller skating, playing keytar, wearing sunglasses, playing keytar"

3. Architectures and Room Designs
   Keywords: room, interior, house, modern, design, walls, furniture, architecture, wall
   Sampled Prompts:

   - "studio with architects designing models of future cities"
   - "luxurious comfortable biopunk lounge inside a huge tropical greenhouse, volumetric lighting, hyperdetailed, 8k octane render"

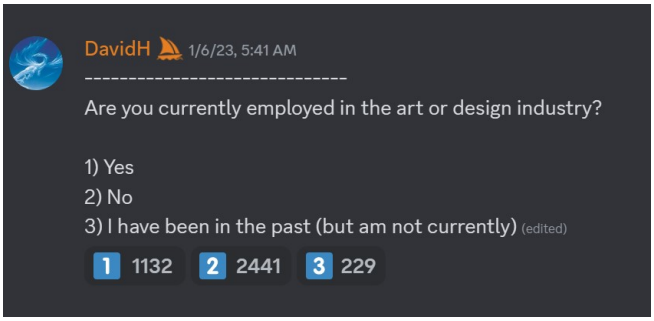
4. Portraits
   Keywords: hair, eyes, face, woman, portrait
   Sampled Prompts:

- "tall 50 year old female police officer with Curly blonde hair and wearing black framed reading glasses, realistic, Norman Rockwell style, 10k"
- "long red hair, beautiful woman in green dress"

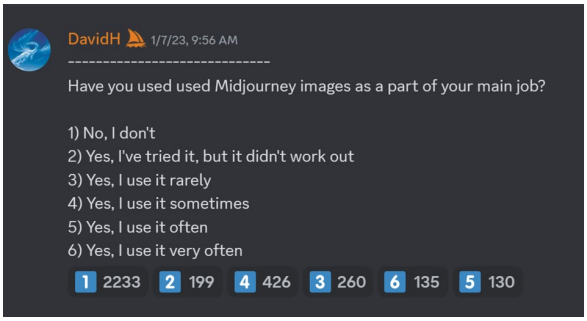
D Screenshots About User Composition

Figure D1: "Mini Poll" Screenshots 1

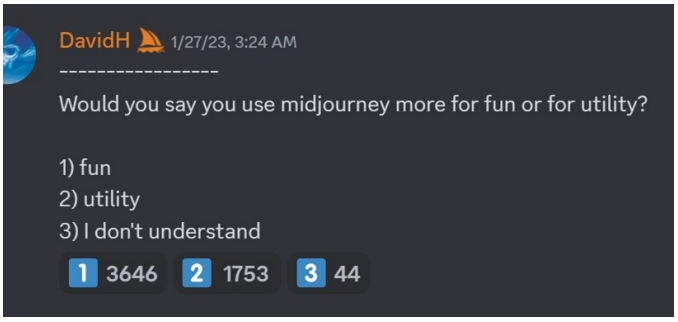
(a)



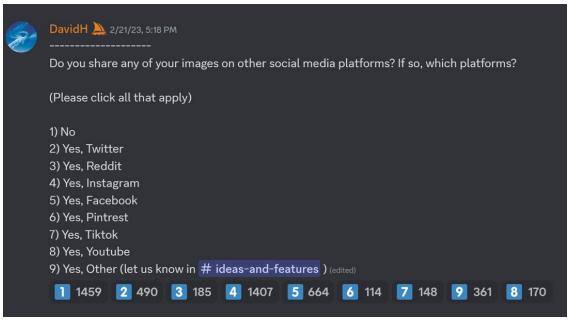
(b)



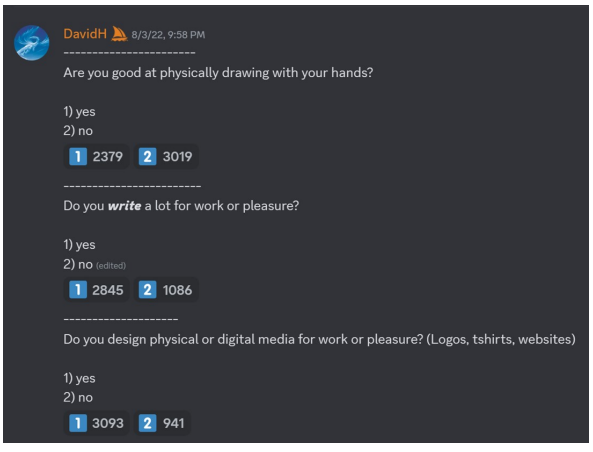
(c)



(d)



(e)



(f)

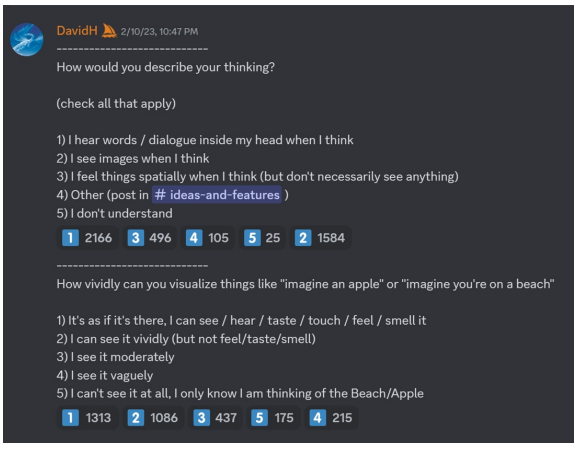
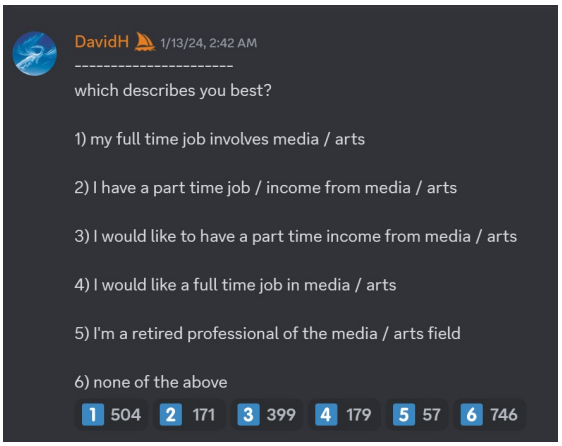
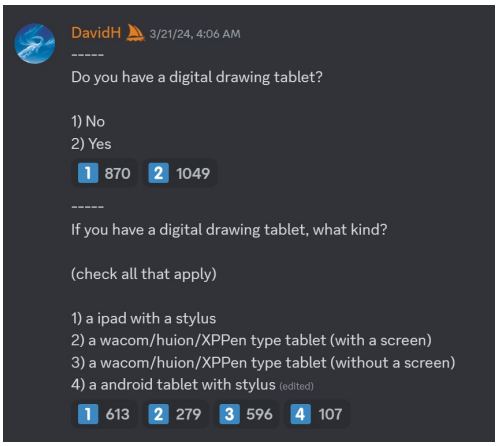


Figure D2: "Mini Poll" Screenshots 2

(a)



(b)



(c)

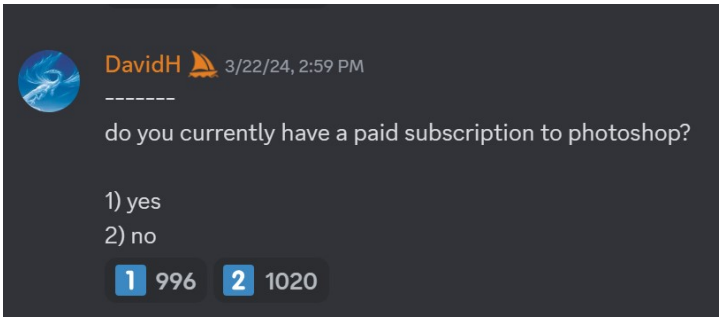
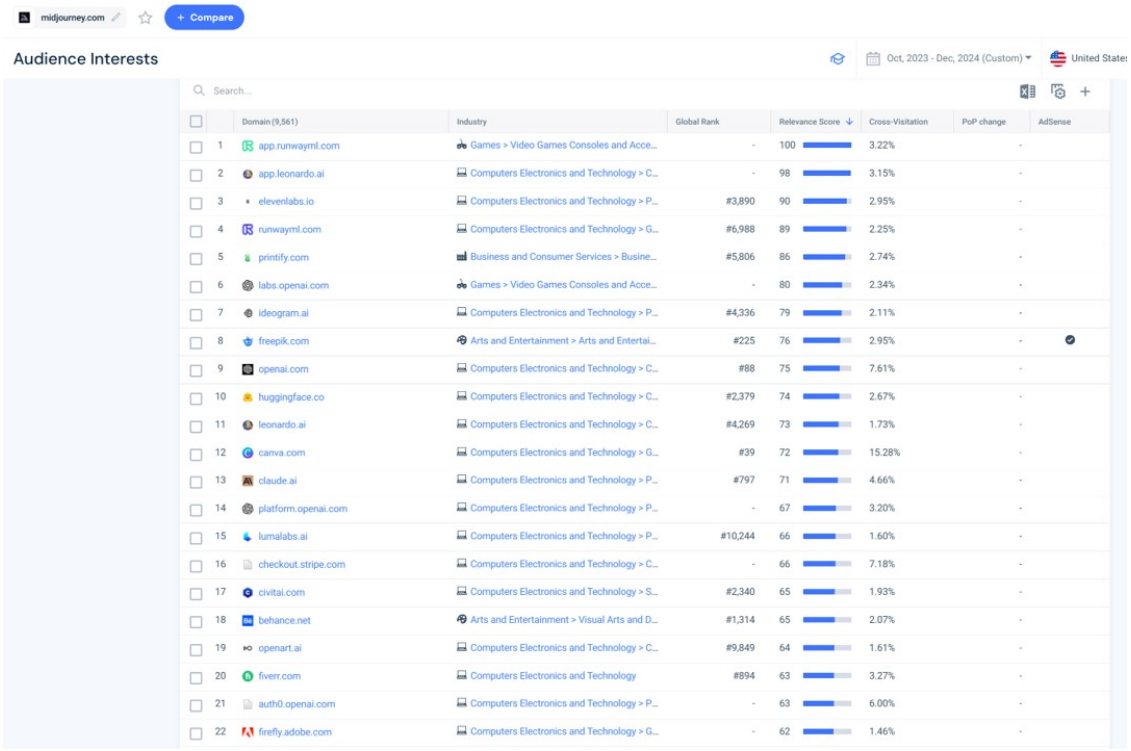
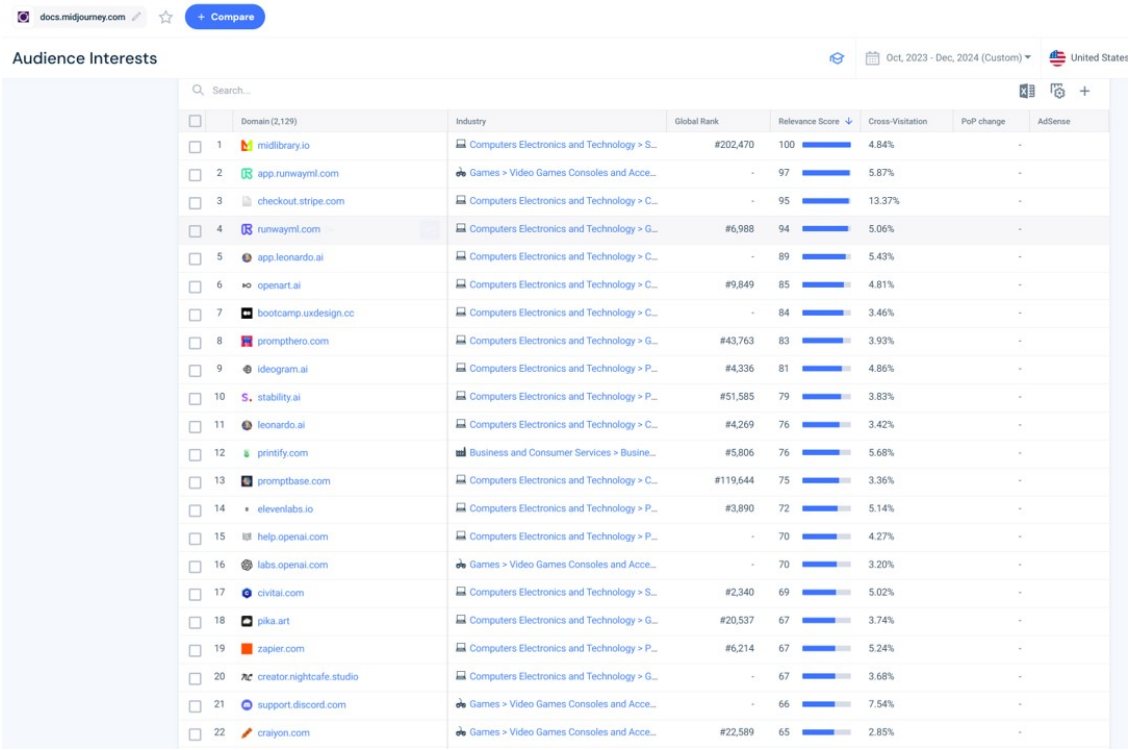


Figure D3: SimilarWeb Screenshots

(a)



(b)



E Image Illustrations

Figure E1: Timeline of Version Updates

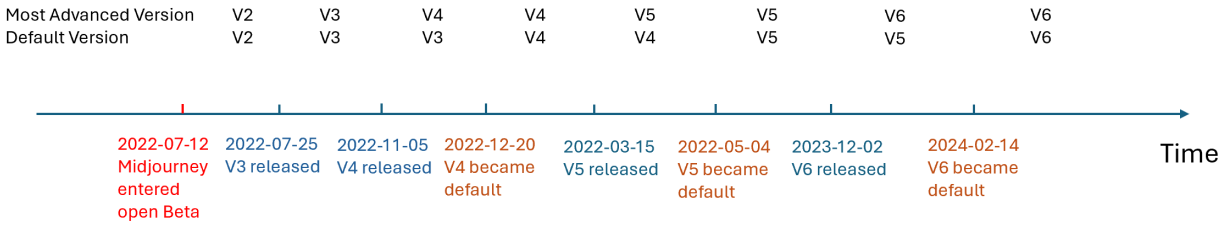


Figure E2: An Example of Using Midjourney

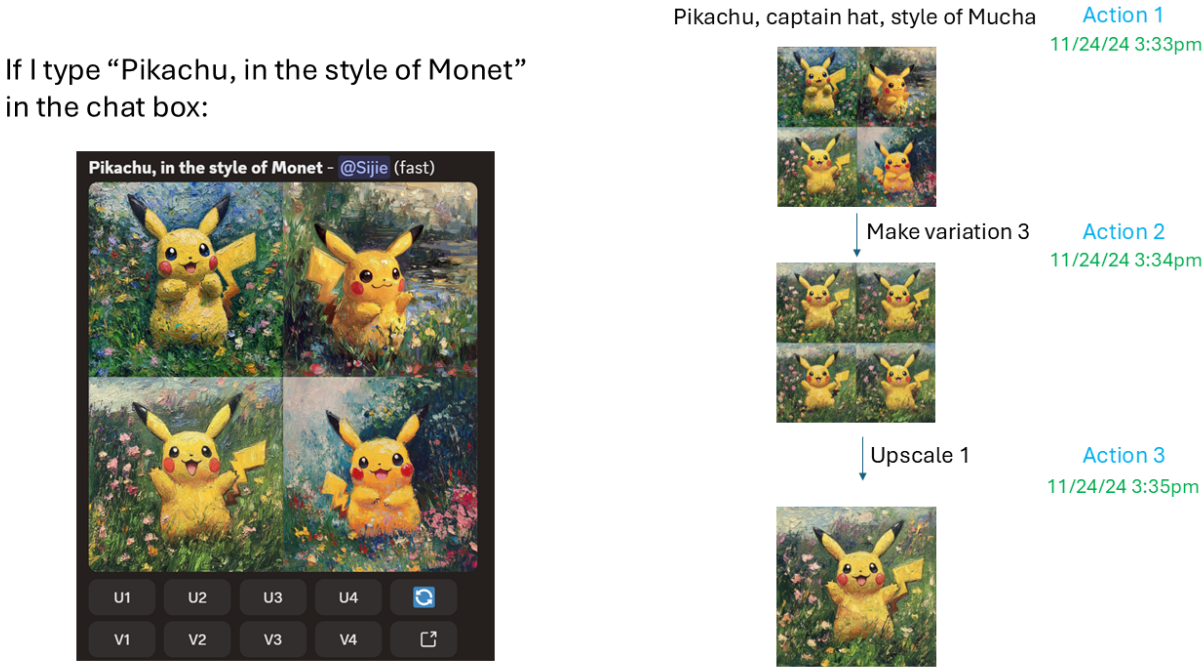


Figure E3: An Example of Matched Tuple

(a) **AI3 Prompt3**: "supernova,8k"

(b) **AI3 Prompt4**

(c) **AI4 Prompt3**

(d) **AI4 Prompt4**: "ethereal celestial cosmic space, 8k, hyper detail, hdr, cinematic, high resolution"