Fostering AI Literacy Through Simple Prompting Exercises Using Dall-E

Tatiana Ringenberg
Computer and Information Technology
Purdue University
West Lafayette. USA
tringenb@purdue.edu

Nidhi Bhardwaj
Computer and Information Technology
Purdue University
West Lafayette. USA
bhardwan@purdue.edu

Nathan Kim
Computer and Information Technology
Purdue University
West Lafayette, USA
kim4344@purdue.edu

Paul J. Thomas
Computer and Information Technology
Purdue University
West Lafayette, USA
pjosekut@purdue.edu

***Abstract*—This innovative practice full paper describes an AI prompting exercise to enhance students' AI prompting skills by providing immediate visual feedback.The exercise was executed within an introductory systems development and organizational security course, consisting of approximately 54 students who were enrolled in a cybersecurity program at a large Midwestern university. The exercise was conducted in the classroom setting, followed by a discussion of observed patterns in successful and unsuccessful prompts. This paper delves into the exercise's methodology, offering an in-depth exploration of the prompts generated by the cohort of 50 students, supplemented with a content analysis. Majority of students initiated their prompts with elaborate descriptions of the intended subject, showcasing a clear inclination towards a subject-centric approach. There was also variance around explicitly addressing the task of image generation, with only a minority of prompts articulating the core objective. Spatial details of background features were prevalent, particularly in terms of their positioning relative to the subject or within the overall image composition. Moreover, the comprehensive nature of subject descriptions encompassing elements such as gender, age, physical features, ethnicity, and clothing, was a ubiquitous trend observed across nearly every prompt. We also consider future directions for enhancing and extending the exercise in courses with larger enrollments.**

***Index Terms*—AI literacy, prompting, image generation**

I. Introduction

With the proliferation of artificial intelligence (AI) tools into different facets of everyday life, there is a renewed emphasis placed on developing AI literacy [1] and incorporating it as a part of digital literacy skills [2]. AI literacy has been defined as "...a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI" [3]. AI literacy also emphasizes the ability of individuals to utilize AI tools both at home and the workplace [4]. Applications of AI, as well as ethics and safety elements associated with AI are considered to be subsets of AI literacy [5]. There have been efforts in recent years to promote AI literacy but most of them have taken a computer science-centered approach [6] and have been situated in a K-12 context. It is also worth noting the computer science centered approach towards AI literacy may not be suitable for most audiences since they will likely not need to know how to program AI to interact with it [4]. Efforts to improve AI literacy are needed given how organizations from around the world are recognizing the importance of AI literacy at an individual and organizational level [7]. Image generation systems utilizing generative AI have also risen to mainstream prominence in recent years [8]. The primary allure of this AI-enabled image generation is the ability for anyone, regardless of artistic ability, to create digital images using prompts in a natural language. It must be noted that the quality of the images generated depend greatly on the nature of the prompts provided with users often needing to interact with the AI tool in an iterative fashion to arrive at the desired output [9]. This approach for interacting with AI tools isn't limited to the realm of image generation and as such, there has been a renewed emphasis on the notion of prompt engineering across various disciplines [10] [11]. Prompt engineering encompasses not just the design and implementation of prompts but also the refining of prompts to shape the output of large language models (LLMs) or AI tools [11]. Prompt engineering plays an integral role in the effective utilization of LLMs or AI tools as it can be leveraged to control parameters such as length, complexity, or style of the output [12]. Some studies have also indicated that novices may not be able to construct effective prompts and often end up creating ambiguous prompts or those prompts lacking in context [10] [12]. Given the emerging nature of AI, there is a relative dearth of literature pertaining to AI literacy specifically in the topic of promoting AI literacy or the conjunction of higher education and generative AI have for the most part only provided frameworks, systematic reviews, or a better understanding of what other researchers perceive as ”fostering AI literacy”. There is also a lack of public datasets detailing college student interaction with generative AI or LLMs. This is in turn reflected in the lack of research analyzing such data. Very few studies in the field of human-computer interaction, specifically with educational applications of AI tools, have sampled their own data for analysis of human behavior and as such leaves gaps for this study to have relevance. This study was guided by the research question: What are student approaches to prompt engineering when tasked with image creation using a generative AI tool? This paper presents the results of a qualitative data analysis on image generation prompts from students.

II. Methodology

A. Context and Participants

The exercise conducted in this study was deployed in an undergraduate systems development and organizational security course in a Cybersecurity program at a large Midwestern university during the Spring 2023 semester. Over the semester, students received instruction on topics including but not limited to the risk management, the confidentiality-integrity-availability (CIA) triad, information systems modeling using Unified Modeling Language (UML) as well as database design and development [13]. The course was held in an in-person modality with two fifty-minute meetings for lecture and one hour and fifty-minute lab meeting held every week. Vast majority of the students enrolled in this course were in their first year of undergraduate education. We developed this exercise to understand how students with limited AI experience use multi-modal models to create prompts. By analyzing how students construct these prompts and learn from the hands-on activity, we aim to gather insights to develop educational resources in prompt-engineering that align with our students' AI literacy and skill levels. The university's IRB approved this study under protocol IRB-2024-52. During the exercise, students indicated, at most, a cursory knowledge or experience of working with AI tools. Most students indicated a vague familiarity with ChatGPT with little to no formal experience or training in prompt engineering.

B. Exercise

For the purposes of the exercise, students were presented with an image of their course instructor. The image contained prominent foreground and background elements. The number and diversity of elements in the foreground and background served to add complexity to the task of image generation. The students were then tasked with replicating the image using the Dall-E image generation tool. It must be noted that there are several AI Image Generation platforms that are available for public use with one notable example being MidJourney [8] [14]. Dall-E was released over a year ahead of Midjourney and as such has been more widely utilized in prior studies [15] [16] [17] [18]. Given the general dearth of literature in this nascent field, Dall-E with its relatively sound foundation in previous studies was selected as the most suitable tool for this exercise.

Once the task was explained, the students were given time to try various approaches towards generating the provided image. They were asked to share their prompts and were encouraged to discuss with the instructional team and their peers as to the patterns that they observed in successful and unsuccessful prompts.

C. Data Collection

At the conclusion of the exercise, students were asked to submit the prompt which they felt produced the closest output image in comparison to the provided sample image. Data for the qualitative analysis was gathered from a pool of 30 textual prompts written by students. These prompts were generated in response to the exercise detailed in section IIB. The purpose of these prompts was to elicit responses from an AI system to generate an image resembling the one shown to the students. The prompts were collected and compiled into a single file for analysis. Each prompt contained textual descriptions aimed at guiding the Dall-E model in generating an image based on the visual stimulus presented to the students.

D. Data Analysis

The textual prompts were reviewed to standardize formatting. The prompts were kept as gathered without changing any spelling error or the kind of language used (formal or informal) to maintain the authenticity of the prompts written. The prompts were devoid of any identifiable information. The prompts were subsequently analyzed used inductive content analysis.

We chose content analysis for this study as our goal was to identify and quantify meaningful and recurring patterns in the students' responses. The flexibility and descriptive-nature of the categories generated through content analysis is thus ideal for this task [19]. Content analysis consists of three main phases: preparation, organization, and reporting of results. During preparation, the team makes initial decisions regarding the unit of analysis of the study and familiarizes themselves with the data. During the analysis phase, the team engages in open coding of the data, creates categories based on coding, and revises and links categories based on the structure of the categories [19].

Our preparation and organization phases were conducted as follows:

- Preparation Phase: During the preparation phase, two student coders (one undergraduate, one graduate) were responsible for the coding of the data. For the unit of analysis, the coders and the two instructors chose words or phrases representing a recurring descriptor within the data.

  Following selection of the unit of analysis, the two student coders read the prompts in detail to familiarize themselves with the content of the prompts. Following the initial reading of prompts for familiarization, the two coders independently engaged in open coding of a subset of 10 of the prompts to identify categories that represented descriptors within the prompts.

  The team, including coders and the two instructors, met following the initial open coding round to align on coding practices. This step was crucial, as the initial definition

TABLE I
MASTER CATEGORIES AND SUB-CATEGORIES FOR DALL-E PROMPTS

Master Category	Sub-Category	Examples	# of Prompts	# Instances
Description of Subject	Facial Feature	*face is shining, slight facial hair, looking to the right*	16	26
	Gender	*he, man, male*	28	39
	Hair Description	*black haired, has a short haircut, black beard*	20	22
	Height	*average tall, average height, 5'10*	3	3
	Descriptors of Ethnicity	*indian, south asian, medium brown*	25	31
	Clothes	*turtleneck sweater, fur collar, windbreaker*	27	80
	Age	*youngish, middle aged, 30 year old*	9	10
	Physique	*narrow shoulders, skinny, average build*	6	8
Background Details	Features in Background	*part of the house, shadow, luscious trees*	28	83
	Location	*wooded area, forest campsite, outdoors*	14	14
	Time	*during spring, noon*	2	2
	Weather	*sun, sunny and warm, sun is shining*	3	3
Positional Descriptions and Relationships	Position with respect to Subject	*behind him, surrounded by, to their right*	21	30
	Arm position of Subject	*his hands folded, crossing his arms, left arm is crossed over his right*	25	27
	Subject Watch Position	*apple watch, black apple watch, black watch*	8	14
	Subject Body Position	*standing, 45 degree angle, facing the camera*	17	19
Task Instructions		*create a photo, photo realistic, generate an image*	10	10

of the unit of analysis was broad. Discussion at this stage helped the annotators to align on the unit of analysis and the granularity of codes. Initial discussions of coding practices resulted in our draft codebook, which we would use in the analysis phase to code the remainder of the data. The draft codebook was used to align the coders. Any additional codes emerging during the open coding of the remaining prompts would be noted by the coders and discussed during coding meetings.

- Analysis Phase: Following the preparation phase, the two student coders engaged in the final open coding round, where they coded the remainder of the 30 prompts. The research team met frequently throughout this process to discuss emerging codes and align coding practices.

Following the coding of the data, the team met to discuss the final codes. Codes which contained significant overlap with other codes were merged together. For instance, *subject clothing* and *subject clothing features* were merged together as both codes were ultimately related to the subject's clothing. A *subject activity* code was also removed from the analysis as the content within the *subject activity* code was present in other codes. For instance, smiling was coded as both a *subject activity* and a *facial feature*. Thus, we felt the *subject activity* code was too vague and was removed.

Once the codes were finalized, the research team used inter-rater reliability to assess the trustworthiness of the coding process. The level of agreement between coders was examined using the simple agreement metric. Simple agreement measures the number of agreed upon instances out of all coded instances. We calculated inter-rater reliability based on 20% (6) of the prompts. Over the 6 prompts, we found agreement between the student coders to be 78.9%.

Following assessment of inter-rater reliability and refinement of the final codes based on disagreement, the instructors reviewed the categories. The categories were grouped by related topics into master categories. Aggregation of these sub-categories into master categories was based on relatedness of the subject matter, not quantity of a sub-category. Master categories and sub-categories are summarized in Table 1.

RESULTS AND DISCUSSION

The master categories we identified in the student responses were *descriptions of the subject*, *details about the background of the image*, *positional descriptions and relationships*, and *task instruction*. *Descriptions of the subject* refers to language which was used to describe the instructor in the image in terms of physical features. *Details about the background* refers to all language which described the physical environment around the subject including objects, lighting, location, and so on. *Positional descriptions and relationships* refers to language which was used to describe the positioning of items in the background and the instructor. This includes both absolute location within the image (e.g., right side of the picture) and relative location within the image (e.g., in front of a campsite). Finally, *task instruction* refers to words or phrases used to describe the goal of the prompt or exercise. Sub-categories of each of the master categories are provided in Table 1 above.

As seen in Table 1, several of the categories occurred in almost all of the prompts. The categories which were identified in the greatest number of prompts included *gender (28), features in the background (28), clothes (27), descriptors of ethnicity (25), and arm position of the subject (25).* These categories were specific to descriptions of both the instructor who was in the image and the details which were in the background. As the students were asked to recreate the image completely, descriptors which provide detail around the instructor, position of the instructor, and the background details all seemed salient. Several other categories occurred in the majority of the prompts. These categories included *hair description (20), position with respect to the subject (21), the subject's body position (17), and facial features (16).* The following example represents the typical prompt where both background elements and the subject are described:

> Man smiling, arms crossed, wearing blue jacket with beige collar, South asian decent, standing in front of a house with trees. short hair, wearing a black apple watch, looking to the right, slight facial hair, standing near a charcoal grill.

While many of the prompts had a similar structure in which both some characteristics of the subject and the background were described, others only focused on the task of describing the subject, as in the following example:

> A young Indian man in a quarter zip with a windbreaker looking to the right side with his hands folded.

In the above prompt, the student described several elements of the subject including the clothing worn, the age of the subject, and the gender of the subject. Furthermore, the prompt had descriptions of position, with the description of the subject's arm position and the direction in which the subject's eyes are looking. However, no information was provided regarding the subject's background. This led to both blank backgrounds and backgrounds which were entirely made up and not representative of the original image.

In some cases, categories occurred in less than half of the prompts with some only occurring in a few of the prompts. These included *task instructions (10), age (9), position of the subject's watch (8), physique (6), height (3),weather (3), and time of day (2).* These categories were important for improving this activity and AI literacy for our students, as many of these categories align with current recommendations for producing images using multi-modal models. For instance, it is generally recommended to describe what type or style of image that is requested. It is also generally recommended to describe the environment in terms of time of day, weather, and the general tone or mood of the image. We saw very little use of time of day or weather. We also saw no references to tone of the image or mood. Instead, elements of mood were incorporated into the description of the subject, as demonstrated by the following prompt:

> Draw a man standing and smiling and having his arms crossed. there is a little smoke from bbq behind him and a part of the house. add a bunch of trees behind all that.

The above example demonstrates how the mood of the image was generally handled by describing the subject's face. Many of the prompts described the subject as having a smile or a slight smile. However, none of them appeared to refer to the subject's mood. Additionally, this prompt also demonstrates some of the issues which arise based on task descriptions. In this instance, the student used ”draw” to describe the task. The resulting image was thus not a photo but rather a drawing. Finally, the above prompt demonstrates the general lack of description of a time of day or weather. Specificity in task was any issue for many of the prompts.

While many of the categories of descriptors occurred in the majority of prompts, the number of instances within those prompts varied. Information regarding the number of times a sub-category appeared in the data can be found in Table 1. Students placed a distinct emphasis on background features and the clothes worn by the subject as 83 instances of the former across 28 prompts and 80 instances of the latter across 27 prompts. This stood in sharp contrast to categories relating to subject features such as height, age, and physique or background details such as time of day and weather.

Debrief with Students

Following the activity, we conducted a group discussion with the students. Students indicated their difficulties primarily arose from finding the right amount of detail and Dall-E filling in assumptions. The difficulties with the amount of detail arose from students providing too much detailed information in a single prompt. Those who provided very detailed prompts found the resulting images that were generated left out other details that the students felt were also important. Furthermore, too much detail led to the model preserving details which were not crucial to the image (e.g., watch color). Some students posited such details should be explicitly provided in the order of importance.

On the other hand, students who provided few specific details found the Dall-E model did not handle ambiguity as intended, leading to unexpected features of both the subject and the background. The largest concerns were related to the categories which we found to occur in the least frequent numbers of prompts. For instance, students found that when they did not give a task description, they often ended up with images which were either in a painted or animated style, instead of a photo. Others found that not indicating a time of day or weather resulted in random weather elements being inserted into the image or the wrong lighting being used on the background and subject.

Conclusion, Limitations, and Future Work

In summary, the exercise was successful in enabling students to engage with the Dall-E image generation tool to the end of recreating the provided image. Students were able to gain insights into how to craft prompts that resulted in images with the elements that they felt were important and they were able to engineer prompts at the required level of detail. The analysis of the student prompts revealed a distinct focus on certain aspects of the subject itself such as their gender, hair, ethnicity, and clothes worn. This was in tandem with an emphasis on some of the background details and the subject's positioning relative to these elements.

The study is subject to the limitation that it did not formally quantify or qualify students' preexisting AI literacy. As such, the effects of the exercise in terms of bolstering AI literacy was not evaluated either. Future inquiry could incorporate the development of survey instruments to formally measure and operationalize AI literacy. Additionally, the study includes the analysis of only 30 prompts that were generated students, most of whom working in teams. As such, each prompt does not represent the independent work of an individual, but rather the collective work for the group. In future iterations, we will conduct the exercise on an individual level to understand the AI literacy and prompting needs of individuals. Additionally, our data only consisted of 30 prompts. This, in addition to conducting the activity in a single class, reduces the generalizability of results. Future research could evaluate the efficacy of the exercise with larger number of students, perhaps across multiple classes.

Future iterations of this class exercise could also incorporate training elements based on the feedback of the students during the debrief. Based on the students' issues with right-sizing their prompts, they could be provided training on prompting strategies such as iteratively chaining multiple prompts to reach a final image. Given that students' prompts frequently did not align with modern guidance on salient descriptors for image prompts (e.g., image style, mood, time of day, and weather), students could also be provided with guiding materials which will help them to identify the weaknesses of their prompts and as well as areas of improvement. The exercise could be extended to have two parts: (1) an exploratory portion where students try to recreate prompts based on their own experiences and prior knowledge, followed by (2) a guided exercise in which students improve their prompts using suggestions from modern prompt engineering literature. This will allow students to visually see improvements in their prompts and help them to build their own prompting style that works for them.

In addition to these changes, we would also like to expand our activity to include a segment in which the students use automated prompting tools to help them improve their prompts and generate better images [20]. These tools can allow the student to optimize their prompts and pinpoint problematic elements of the resulting images. Such tools are becoming more common and we posit, given the visual component of these tools, that they could be utilized to provide automatic feedback to the student.

Beyond developing AI competencies, the exercise of generating and refining prompts for image generation tools like DALL-E fosters several higher-order cognitive skills that are valuable across various domains. These skills include critical thinking, problem-solving, and effective communication. By engaging in this activity, students learn to articulate their ideas clearly and precisely, which is essential for any STEM professional. Additionally, the ability to tailor communication to different audiences and contexts is a crucial skill in both academic and professional settings. Understanding how to "right size" prompts not only enhances students' ability to work with AI but also prepares them to tackle complex real-world problems where clear and concise communication is paramount. This focus on audience and emphasis is a transferable skill that individuals in STEM can apply in diverse scenarios, making it a compelling reason for educators to integrate such exercises into their curriculum.

References

[1] Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, 100041.

[2] Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities. Computers and Education: Artificial Intelligence, 4, 100124.

[3] Druga, S., & Ko, A. J. (2021, June). How do children's perceptions of machine intelligence change when training and coding smart programs?. In Proceedings of the 20th Annual ACM Interaction Design and Children Conference (pp. 49-61).

[4] Long, D., & Magerko, B. (2020, April). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-16).

[5] Wong, G. K., Ma, X., Dillenbourg, P., & Huan, J. (2020). Broadening artificial intelligence education in K-12: Where to start?. ACM Inroads, 11(1), 20-29.

[6] Kong, S. C., Cheung, W. M. Y., & Zhang, G. (2021). Evaluation of an artificial intelligence literacy course for university students with diverse study backgrounds. Computers and Education: Artificial Intelligence, 2, 100026.

[7] Cetindamar, D., Kitto, K., Wu, M., Zhang, Y., Abedin, B., & Knight, S. (2022). Explicating AI literacy of employees at digital workplaces. IEEE transactions on engineering management.

[8] Oppenlaender, J. (2022, November). The creativity of text-to-image generation. In Proceedings of the 25th International Academic Mindtrek Conference (pp. 192-202).

[9] Turchi, T., Carta, S., Ambrosini, L., & Malizia, A. (2023, May). Human-AI co-creation: evaluating the impact of large-scale text-to-image generative models on the creative process. In International Symposium on End User Development (pp. 35-51). Cham: Springer Nature Switzerland.

[10] Giray, L. (2023). Prompt engineering with ChatGPT: a guide for academic writers. Annals of biomedical engineering, 51(12), 2629-2633.

[11] Meskó, B. (2023). Prompt engineering as an important emerging skill for medical professionals: tutorial. Journal of Medical Internet Research, 25, e50638.

[12] Heston, T. F., & Khun, C. (2023). Prompt engineering in medical education. International Medical Education, 2(3), 198-205.

[13] Jaiswal, A., Thomas, P., & Karabiyik, T. (2023, October). Characterizing Intercultural Competence among Cybersecurity Majors. In 2023 IEEE Frontiers in Education Conference (FIE) (pp. 1-5). IEEE.

[14] Tan, L., & Luhrs, M. (2024). Using Generative AI Midjourney to Enhance Divergent and Convergent Thinking in an Architect's Creative Design Process. The Design Journal, 1-23.

[15] Cheong, M., Abedin, E., Ferreira, M., Reimann, R., Chalson, S., Robinson, P., ... & Klein, C. (2023). Investigating gender and racial biases in DALL-E Mini Images. ACM Journal on Responsible Computing.

[16] Amri, M. M., & Hisan, U. K. (2023). Incorporating AI tools into medical education: harnessing the benefits of ChatGPT and Dall-E. Journal of Novel Engineering Science and Technology, 2(02), 34-39.

[17] Leivada, E., Murphy, E., & Marcus, G. (2023). DALL· E 2 fails to reliably capture common syntactic processes. Social Sciences & Humanities Open, 8(1), 100648.

[18] Kapelyukh, I., Vosylius, V., & Johns, E. (2023). Dall-e-bot: Introducing web-scale diffusion models to robotics. IEEE Robotics and Automation Letters.

[19] Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. Journal of advanced nursing, 62(1), 107-115.

[20] Wang, Z., Huang, Y., Song, D., Ma, L., & Zhang, T. (2024, May). PromptCharm: Text-to-Image Generation through Multi-modal Prompting and Refinement. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1-21).