The methodology involved statistical modeling, semiotic analysis, and Systemic Functional Linguistics to tackle the challenges of poetic interpretation and translation. By employing data science practices, including data collection, processing, and model tuning, alongside website development, "Verse Into Vision" emerged as a tool that not only generates visual representations of poems but also fosters an interactive platform for users to engage with the poetry visually. The project underscores the synergy between computational linguistics and the arts, proposing a novel approach to appreciating and understanding Chinese poetry through AI.

Introduction

Images stand as essential mediums through which human engage with and interpret their surroundings, largely perceiving the world through visual stimulation. They represent visual perception, ranging from tangible captures by optical devices—cameras, mirrors, telescopes, and microscopes—to the areas of imagination realized through paintings and digital art. Historically preserved on paper, canvas, and light-sensitive materials like film, the evolution of imaging technology has ushered in a digital era where images are increasingly stored, shared, and manipulated in digital formats. This progression is underpinned by significant advances in digital acquisition technology and signal processing theories, reshaping our interaction with images, and expanding the horizons of artistic and scientific expression. Understanding this background is important because it helps us see how AI and images are changing the way we make, share, and understand pictures today.

The advent of AI-generated art, particularly in the form of drawings, marks a revolutionary chapter in the history of artistic expression and media theory. This innovation traces its basis to the exploration of computational creativity, a field where technology and artistry intersect, challenging traditional notions of authorship and artistic genius. AI drawings originate from the desire to extend the capabilities of human creativity with the precision and versatility of machine intelligence. Early experiments in this domain sought to understand whether machines could replicate or even augment the creative process, leading to the development of algorithms capable of producing visual art. Today, AI's capability in art generation has transcended mere replication, embodying the ability to synthesize novel visual expressions from vast datasets of existing artworks. These systems, powered by advanced machine learning techniques such as Generative Adversarial Networks (GANs) and deep learning models, can now generate images that resonate with human aesthetics, yet are entirely new creations. This breakthrough not only democratizes artistic production, making it accessible to those without formal training, but also stimulates a profound discourse on the essence of creativity, the role of the artist, and the evolving relationship between humans and machines in the creation of art. Through this lens, AI-generated drawings emerge not just as technological feats, but as a medium that challenges and expands the boundaries of traditional media theory, inviting a reevaluation of art's place in the digital age.

‘No matter how long I may look at an image, I shall never find anything in it but what I put there. It is in this fact that we find the distinction between an image and a perception.' Sartre argues that while some believe imagining to be like an internal perception, imagination is nothing like perception. Perception is our study over time of a particular object with our senses. It is necessarily incomplete; one can only see one side of a chair at a time, for example.

Jean-Paul Sartre's theory highlights a fundamental distinction between perception and imagination. As Sartre argued in The imaginary: A phenomenological psychology of the imagination, perception is an observational, incomplete engagement with objects through our senses, inherently limited to one perspective at a time. Conversely, imagination is total and subjective, presenting all aspects of an imagined object simultaneously, based on our synthesis of knowledge and intentions. This conceptualization of imagination as "quasi-observation" suggests that our imaginative constructs are deeply personal and shaped by our desires and prior experiences. In the context of AI art, Sartre's theory offers intriguing parallels. AI-generated art challenges traditional media by introducing a form of 'imagination' that synthesizes vast datasets into creations that might reflect the machine's 'intention' based on its programming. Thus, AI art can be seen as a manifestation of technological quasi-observation, where the output reflects a blend of past data and current algorithmic processes, inviting viewers to engage not just with the art itself but with the underlying intentions and knowledge embedded by its creators.

Under the strict requirements of Yun Lv (rhythm) and Dui Zhang antithesis, Gu Shi (Chinese ancient poetry) uses minimal words to unfold various scenes from magnificent natural landscapes to deeply empathetic experiences of joy and sorrow, separation and union. This is achieved through Yi Xiang (imagery)—a symbol in ancient poetry, and people's imagination. In order to expand upon the nature of classical Chinese poetry within Sartre's theoretical and semiotic framework, it is necessary to delve into how these ancient texts became conduits for profound imaginative engagement. The rigorously structured verses, cleverly restrained by antithesis and rhythm, embody the use of linguistic symbols and evoke the vast world in the reader's mind. Utilizing a concise yet powerful vocabulary, this poetic form reflects Sartre's concept of the imagination: the totality of an object - or in this case, a vivid scene or emotion - emerges through the synthesis of the reader's knowledge and intentions. Present. The poems' language is sparse but evocative, with each word imbued with deep historical and cultural meaning, serving as a semiotic beacon that guides readers beyond the surface of the text to explore multi-layered meanings. Classical Chinese poetry thus demonstrates that ancient poets were adept at creating immersive imagery experiences, elucidating the dynamic interplay between signifier and signified, and emphasizing the imaginative leaps required to fully realize poetic imagery.

Literature Review

1. Chinese poetry is the combination of 意象 (symbols).
2. There is no correct way to “understand” Chinese poetry
3. The interpretation of Chinese poetry reflect people’s difference on their background.
4. What is required to interpret Chinese poetry
5. How can Machine/AI interpret the Chinese poetry

While implementing the tool, we actually concretized the theories related to machine learning semiotics and information theory. Among various topics this implementation of Verse Into Vision may cover, this paper will focus on one topic only, which is whether AI can or cannot “understand” Chinese poetry. However, the definition for the word “understand” is various in different theories and areas. In daily life, the definition of “understand” is to interpret or view something in a particular way or perceive the intended meaning of something. This concept is also reflect in semantic that “understand” means to receive and the message conveyed by words, sentences, and symbols in a context, in another word, to get the meaning of something. In information theory is to receive all the information of something. In semiology, is to decode the symbols to receive the message. During the discussion, all of these definitions will be covered.

As mentioned before, Chinese poetry is profound with the Profound cultural background of China. Thus, it is even hard for people to understand Chinese poetry. In researchers’ approaches to create methodologies to interpret Chinese poetry, there are three directions that are related to our topic. As an emerging means to analyze Chinese poetry, the statistical methods were used to model then Chinese poetry. With Weighted Personalized PageRank (WPPR) to measure the similarity between a given word and all positive and negative reference words simultaneously in a lexical network built from a poetry corpus, the emotions in words can be calculated to provide an analysis of Chinese poetry (Hou and Frank). As Hou and Frank introduced, this graph-based method considers the lexical network as a whole, allowing for a globally optimal solution in identifying sentiment orientations of words within classical Chinese poetry. This research presents a significant advancement in the computational analysis of sentiment in classical Chinese poetry, which may provide a strong and positive prove that AI/machine can understand Chinese poetry. However, like other statistical approaches that based on massive data, WPPR cannot accurately predict the metaphors in a specific poem nor provide accurate interpretation without the background of the times. Though this disadvantages still exist, by using more delicate data and pre-process, the interpretation accuracy for specific metaphors is increasing (Zhang et al.). By applying more delicate BERT-BiLSTM-CRFs Model, the machine seems to have a greater ability to cover details of the training data and apply them in the test.

However, there is a misunderstanding when people use word “understand” while interpreting Chinese poetry, even myself have this misunderstanding till I realized when drafting this paper. The misunderstanding is because people are too eager to figure out the meaning of the poem, which leads them into an obsession to find an answer. However, Chinese poetry does not have a “optimal”, “correct”, or “best” understand. In another word, no one can fully understand the exact message that poets wanted to express at the moment they wrote the poems. Keep discussing mutual understanding will lead the discussion to a philosophical level, which is not the purpose of this paper. Therefore, it is necessary for us to come back to the discussion around information theory and semiotics. Before we spread the topic and try to answer the question, there's a common misconception with the term "understand" in the context of interpreting Chinese poetry. This confusion arises from an intense desire to discern a poem's meaning, leading to an erroneous belief in a single "correct" interpretation. However, Chinese poetry, by nature, does not lend itself to a singular understanding; the poet's original intent can never be fully grasped. Crafted by extraordinary Chinese poets, Chinese poetry is a unique media that is going to express the possibility to the readers. Instead of a giving directed information, Chinese poetry always provide the chance for readers to image and interpret the poems in their own way. This uniqueness of Chinese poetry can be explained with information theory, imaginary and semiology.

A probabilistic theory of communication in which a message consists of a sequence of symbols selected by the sender and decoded by the receiver, without taking into account (which does not mean denying, but certainly ignoring) any meaningful relationship between the messages. notation; the theory that the information conveyed is equal to "the logarithm of the number of available choices" (Shannon & Weaver 1964, 9); finally, the generation of any message depends on "a special case of stochastic processes whose probabilities depend on previous events", i.e. A model called a "Markov chain". In each communicative sequence, the occurrence of a symbol is determined by the symbol chosen previously; therefore, the information conveyed by each symbol is measured against the probability of what is chosen to continue the same sequence. This means that the message conveys more information if more options are available, as will those messages where multiple options yield the same probability when decoded.

Based on Shannon’s information theory, the focus should moved from what Chinese poetry does say to what Chinese poetry could say. The theory developed by Shannon positions communication as a sequence of symbols transmitted from sender to receiver without necessitating an understanding of the symbols’ meanings (623-656). It quantifies information as the logarithmic value of potential symbol choices and utilizes the Markov chain model, suggesting that each symbol’s appearance and the conveyed information are contingent upon preceding selections. This theory implies that messages offering a greater number of decoding options inherently carry more information, especially in scenarios where multiple interpretations are equally probable. In other words, the information entropy of ancient Chinese poetry is very high because it contains images with multiple meanings. Instead of decreasing the information entropy, which is to make the message clear, the Chinese poetry is tried to increase the information entropy. That helps to explain why Chinese poetry is a unique media as it is to express the space for imagination instead of clear information.

在《浅析中国古诗中的隐喻意象翻译》中，赵广发深入探讨了中国古诗中意象的翻译，其中不乏针对一些比较隐晦的意象的讨论。其中，他总结了三种意象隐喻的翻译方法，也可以让我们从中学习到古诗意象的不同种类以及联想，想象，文化背景对于读懂中国古诗是不可或缺的。（1.Object；2. Characteristic；3. Cultural background）不同的意象有不同的理解方法。在另一方面，潘建平对中国古诗词意境审美理论的论述，强调了诗意意象与意境唤起之间的内在联系（Recall the NLP process combines different word vectors to infer the meaning）。但是这个过程需要结合个人的常识（common sense）。

解读中国古诗没有一个标准的答案。解读这些意象的行为是以一种深刻的个人化方式来参与这首诗，并反映一个人自己的经历和世界观的过程。不同人的不同解读其实正反映了每个人的独特性。

Chinese poetry is rich in imagery and easy to interpret, making it an ideal medium for Sartre's theory. These poems provide an "analogy," an imaginative stimulus that transcends literal meaning and embodies the reader's subjective experience, knowledge, and emotions. When readers read a poem, they not only perceive the imagery in the poem; They are imaginatively recreated and imbued with meaning and personal emotion. This process explains why Chinese poetry has been so widely interpreted. Each reader engages in a creative act of "quasi-observation," drawing on their own unique memories, emotions, and knowledge to project onto the poem an interpretation that resonates with their personal perspective. Furthermore, Sartre argued that imagination emphasizes our ontological freedom and the liberating power of poetry. When readers engage with Chinese poetry, they are not limited by what is "real" or concrete; instead, they venture into realms of possibility, evoking meanings and worlds beyond the tangible. This imaginative engagement with poetry thus becomes a liberating exercise, allowing readers to transcend the immediacy of their surroundings and explore the vast landscape of human experience and emotion. In essence, the act of interpreting Chinese poetry from the perspective of Sartre's philosophy demonstrates the infinity of human creativity and the subjective construction of artistic meaning.

Conclude and circle back to discuss how machine can interpret Chinese poem. Three papers:

1. Coon: AI is a participant of meaning making process
2. Monti: Difference of “Information” and “Meaning”. Torpedo’s example.
3. Nandine: “If computation, regardless of its nature (algorithmic or interactive), is not reducible to electric, or quantum, or DNA processes but involves semiotic entities, the question is: What are they? A short answer would be: The same entities that make cognitive processes possible. Somewhere along the line, we end up at the one and only culprit of semiotics – the sign. Thus we close the infamous circle:

The sign as an underlying element of thinking = The sign as a product of thinking, which Boole alluded to while describing language.

Computation has it easier. Bits and bytes (which are only strung-together bits) are processed but not necessarily defined through computation; rather, they are defined beforehand, as a condition of computation.”

“Towards a Semiotic Pyramid: A Linguistic Study of Artificial Intelligence and the Knowledge-Sharing Economy” introduces a groundbreaking model that reconceptualizes the role of artificial intelligence in meaning-making processes (Poschinger and Coon). This three-dimensional symbolic pyramid not only positions AI as an active participant in meaning-making, but also reveals the complex feedback loops between AI-generated expressions and external realities shaped using digital language and messages. The model has profound implications for the interpretation of Chinese poetry, demonstrating that artificial intelligence can play an important role in understanding and even interpreting the rich, detailed, and layered language of poetry. This approach opens up new avenues to explore how AI systems interact with the depth and subtlety of poetic expression, recognizing the complex interplay between form, content and cultural context.

Building on this fundamental understanding, Monti delves into the interrelationships between cybernetics, semiotics, and artificial intelligence, providing profound insights into the potential of artificial intelligence for creative interpretation. Monty's study of the historical trajectory of cybernetics and its impact on the conceptualization of meaning in artificial intelligence systems highlights key challenges inherent in encoding creativity and interpretive depth. Addressing Umberto Eco’s distinction between information and meaning, Monti highlights the complex challenges faced by AI in the symbolic and often ambiguous realm of poetry. This analysis is particularly relevant to the interpretation of Chinese poetry, where the intertwining of metaphors, historical allusions, and philosophical insights requires nuanced understanding rather than mere information processing.

Finally, Nadine delves into how artificial intelligence systems (considered semiotic machines) process and generate meaning through a semiotic lens. Nadine's emphasis on the dynamic and interactive nature of semiotic processes provides a compelling framework for considering artificial intelligence's interpretation of poetry. This perspective is crucial to understanding AI’s ability to decode the complexity and subtlety of Chinese poetry, which is deeply rooted in a rich semiotic tradition. Nadine's work shows that integrating semiotic principles more deeply into AI design can improve the system's ability to address the cultural and linguistic complexities of poetic texts.

Methodology

Introduction to the Creative Process:

1. In this project, I wish to create a tool that visualize the Chinese poetry. This is because that in my early education, the Chinese poetry is always recited instead of understanded, which makes me hate to learn it. However, I realized that the beauty of Chinese ancient poetry soon as I lost the utilitarian external motivation. To let people truly understand Chinese poetry and play with them instead of reciting them mechanically, I want to develop the tool that can turn Chinese poetry into Images with their own modifications.

2. This tool is designed to have basic input of Chinese ancient poem and then it will generate a corresponding English prompt for the AI drawing model. Then it will call the drawig model to draw the image and display it. People can also download the image easily. In order to show capability of the tool, the project also include part that aims to visualize a set of Chinese ancient poetry to have a collection of outcomes. With the outcomes, there must be a way to display them, that’s why another website is designed for people to rate the outcomes and view the images.

3. During this process the knowledge of my major – data science – is fully used: coding in different programming languages, data collection and process, construction of database and website, the model tuning, Human-computer interaction design, and data visualization and outcome presentation.

4. Through the implementation of this tool, I also thought a lot about the way that algorithms understand icons and symbols in Chinese and reflect its meaning to semiotics and information theories.

Second paragraph:

The programming process to translate Chinese poetry into images was a sophisticated routing of exploration and refinement. Beginning with various AI drawing models, though the most suitable drawing model will be the model trained by myself with specific training set customized for visualize Chinese poetry, it is almost impossible for a individual project to accomplish this goal with limited time, funding and equipment. Thus, I navigated through different attempts in popular drawing model: DALL-E, Midjourney, and Stable Diffusion to find the optimal approach for visual translation. While letting different models paint ancient Chinese poems and checking their documentations, I array these three models depend on difficulty of acquisition, price, the compatibility of Chinese poetry, and ability for large-scale amount image conversion. The Midjourney is first abandoned because it does not have an easy access. Therefore, the comparation is mainly between Stable Diffusion and DALL·E-2. The Stable Diffusion is a latent diffusion model that can be deployed locally. It has a great advantage in customized the style of image by applying different style packages. Moreover, Stable Diffusion model uses prompt words in both Chinese and English to generate new images describing elements to include or omit the input prompt, which makes it easy to prepare the input. By using word segmentation models and the Chinese to English translator the required prompt words can be easily generated based on the Chinese poetry. However, during the attempt I did in Stats 302, the words segment models cannot separate sentences in Chinese ancient poem into words accurately enough for this use (see appendix A). Also, the Stable Diffusion hardly understand the nouns and adjectives in ancient Chinese not to mention generate images based on it. Nor the exist translating tools can translate ancient Chinese into English properly. On the other hand, the DALL·E-2 has an official API offered by OpenAI company which is easy to access. Also, by translating Chinese poetry as a whole “paragraph” has higher accuracy since the meaning of words can be inferred by context. Also, the style of the DALL·E-2 can be adjusted by adding specific direction according to The DALL·E 2 Prompt Book (OpenAI 16). Though DALL·E-2 is easy to approach and have many experience summaries for adjusting parameters and input techniques, it charges about $0.020 per image. Consider the pros and cons of two model, I decided to implement the tool by DALL·E-2 because of the work efficiency and leave Stable Diffusion model aside till I finished the prototype.

Central to the methodology was a translating model that parsed the essence of the poetry, focusing on capturing the objects, emotions, and scenes described. I crafted a dynamic script to generate English prompts from Chinese verses, tailored for an AI drawing model's comprehension. Nevertheless, the process to find a perfect language model to do this job is hard. Since DALL-E 2 is determined to be the drawing model, the language model need to produce the translation of the basic content in the poem as well as the mood, the composition and the color tone. According to the exploration during Stats 302 (Appendix A), the existed NLP model like Latent Dirichlet Allocation (LDA) is able to find the topics and emotions from according to the combination of the word vectors. However, the accuracy of the models is not high enough for the application purposes. In order to improve the accuracy, manual labelling is unavoidable, which make these small language models incapable of the job to translate the Chinese poetry with the emotions and objects accurately. Also, it is hard for the current model to construct a composition based on the poem. Therefore, the Large Language Model (LLM) like GPT, LaMDA, LLaMA became the most possible choice to do the translation of the Chinese ancient poetry and produce the text prompts for DALL-E 2. Out of the consideration of convenience and the consistency of the APIs, I choose the GPT model as the language model. Along the attempts, GPT3.5 and GPT4.0 both can finish the same task with similar result. With the lower price, GPT 3.5 is chosen to implement in a large amount of use and to save cost.

<https://medium.com/sciforce/what-is-gpt-3-how-does-it-work-and-what-does-it-actually-do-9f721d69e5c1>

To support the analysis of information theory and semiotics in relation to the tool, it's vital to go through the methodology behind the DALL-E 2 model. This exploration not only contextualizes the tool's capabilities within the broader theoretical landscape but also offers insights into the discussion between technological innovation and symbolic communication. DALL-E 2, developed by OpenAI, uses a sophisticated diffusion model conditioned on CLIP image embeddings, a technique that significantly advances the fidelity and contextual relevance of generated images from textual descriptions (Radford et al.). This model, utilizing 3.5 billion parameters, demonstrates an exceptional capability to interpret and visualize complex textual inputs into detailed and semantically accurate images (Ramesh et al.). Its operational framework is rooted in a transformative approach where text descriptions are translated into visual representations through a process that progressively refines random noise into coherent imagery, echoing the intricacies of human creativity. The training process of DALL-E 2, integrating vast datasets of text-image pairs, equips the model with a nuanced understanding of visual and textual correlations. This endows DALL-E 2 with the unique ability to produce images across a wide array of styles and compositions, from photorealistic renditions to stylized interpretations, effectively capturing the essence of the described scenes or objects (Johnson). The application of DALL-E 2 to translate Chinese poetry into visual art presented an opportunity to explore the convergence of language, culture, and AI-driven creativity. By feeding the model with carefully crafted prompts derived from the poetic texts, the project achieved results that resonate with the symbolic depth and aesthetic appeal of the original poems. Moreover, the integration of CLIP into DALL-E 2's workflow enhances the model's ability to evaluate and select the most appropriate visual outputs based on textual prompts, ensuring that the generated images not only align with the descriptive content but also uphold artistic and cultural fidelity. This nuanced translation from text to image underscores the project's innovative use of AI to bridge traditional artistic domains with cutting-edge technology, offering a fresh lens through which to appreciate the timeless beauty of Chinese poetry..

The GPT model works on the transformer architecture, which uses attention mechanisms to weigh the importance of different words in a sentence, enabling it to predict the next word in a sequence. This model is trained on vast datasets of text, learning patterns, and structures of language. During training, it adjusts its internal parameters (weights) to minimize the difference between its predictions and the actual outcomes. This process allows the GPT model to generate coherent and contextually relevant text based on the input it receives. For more details, please refer to the original sources. To elucidate the process by which GPT's Large Language Model (LLM) transforms Chinese poetry into English prompts for DALL-E 2, an understanding of its complex algorithms and extensive dataset training is imperative. The LLM employs a sophisticated understanding of linguistic structures, cultural nuances, and semantic depths inherent in the poetry, facilitated by its training on a diverse corpus of texts. This enables it to accurately capture and translate the essence, tone, and imagery of Chinese poems into coherent English prompts. Such prompts are then utilized by DALL-E 2 to generate visual representations that are not only contextually relevant but also deeply resonant with the original poetic vision. For example, the translation process for "远芳侵古道，晴翠接荒城," GPT's Large Language Model first analyzes the text, identifying key elements such as "distant fragrance," "ancient path," "bright greenery," and "desolate city." It then synthesizes these elements into a coherent English prompt that reflects the poetic imagery and mood. An illustrative result might be, "A distant fragrance invades the ancient path, where bright greenery meets the remnants of a desolate city," showcasing the model's capacity to maintain the poetic essence while facilitating creative visual generation through DALL-E 2. This example underscores the AI's innovative approach in marrying text with imagery, bridging traditional literary art with modern technological creativity.

After the translation model and drawing model is selected, the code to implement the function is constructed with Python and it's two libraries: 'openai' for calling the models and 'wget' for downloading the images. The program begins by setting up the OpenAI API key, initiating the process of converting poetry into visual art. It prompts the GPT-3.5 model to analyze a given Chinese poem, extracting and interpreting objects, emotions, and the overall setting described within the text. This prompt is carefully crafted to ensure the output is succinct yet descriptive, adhering to a limit of 1000 characters. The program then leverages this refined text to instruct the DALL-E model, resulting in a visual representation that captures the essence of the poem, including specific colors, positions, and ancient Chinese stylistic elements, without including text or watermarks. The outcome is a URL linking to the generated image, demonstrating a seamless integration of linguistic analysis and visual creativity

Through meticulous tuning and experimentation, including adjustments to prompts for clarity and thematic accuracy, I improved the script's ability to produce prompts that resonate with the poem's imagery. The prompt sending to GPT 3.5 model is rough and vage: ‘Take this Chinese poem and turn it into a detailed text prompt within 1000 letters (including space and marks) for DALLE to draw a painting without watermark or text on it:’. In order to specify the content GPT should generate including the theme color, emotion, etc I take hundreds of times revised the prompt and test the result. Here is the outcome of the second version: ‘ake a Chinese poem to analyze the objects in it and the emotions or feelings that the poem describe. Turn it into a text prompt less than 1000 letters for DALLE. The prompt must be shorter than 1000 letters (including space). In the text prompt, describe objects in the picture with their color and their position in the picture. Then, give a brief description about what the background should look like (color and scene). Also, don't forget this is a Chinese poem, so the view and buildings and other things should look like ancient Chinese style. There should not be any text and watermark in this generated image. Here is the Chinese poem:’ Though the second version looks good and test good, it went wrong in the small batch production test. The prompt is too long and it lead the GPT model to loose the ‘concentration’ to the key tasks for example the length limitation and translation. Moreover, GPT model charges based on tokens, which is parts of or a whole word (For example, “GPT” is one token), longer prompt means the exponential cost when running this tool. Thus, the third version of prompt is polished: ‘Take a Chinese poem to analyze the objects in it and the emotions or feelings that the poem describe. Turn it into a text prompt less than 1000 letters for DALLE. The prompt must shorter than 1000 letters(including space). In the text prompt, descirbe objects in the picture with their color and their postion in the picture. Then, gave brief desicription about what background should looks like (color and scene). Also, don't forget this is a Chinese poem so the view and buildings and other things should looks like ancient Chinese style. There should not be any text and watermark in this generated image. Here is the Chinese poem:’. By deleting only a dozen words, the output is completely different.

With the program, a web user interface is developed using Flask, a lightweight WSGI web application framework in Python, to offer a user-friendly tool. Flask enables the creation of web applications with minimal setup, integrating Python functions with HTML through templating. This setup allows users to interact with the application via a webpage, where they can input text to generate images. The Flask application handles requests, processes input using the generated\_image function, and dynamically updates the webpage with the generated image and prompt, enhancing accessibility and engagement for users.

<https://flask.palletsprojects.com/en/3.0.x/>

To showcase the capabilities and quality of Verse Into Vision, the tool was applied to create visual interpretations of Wang Wei's poetry from Quan Tang Shi. This endeavor resulted in a collection of 1053 images for 351 poems, with each poem inspiring three distinct images. This extensive compilation demonstrates the tool's potential for wider public utilization and affirms the high quality of its outputs. The collection's title, "In Poetry, Paintings; In Paintings, Poetry" is inspired by the accolades for Wang Wei, symbolizing the interplay between his poetry and the Verse Into Vision tool.

这个作品集选用王维的诗有以下的几点原因。第一，宋代的大诗人苏轼曾经在《东坡题跋·书摩诘〈蓝田烟雨图〉》中说道：“味摩诘之诗，诗中有画；观摩诘之画，画中有诗。”意思是细细评味王维（字摩诘）的诗，他的诗中仿佛藏着画卷，观赏王维的画作，其中又能感受到如同诗歌般高雅的意境。王维既是诗人，又是画家，其所成就，不仅仅能诗善画，而是把艺术中的诗与画，通过他的他作，给以融化，并且有机的结合。他的诗词代表了中国古诗的一大流派——山水田园诗派。他们在发掘自然美方面，既能概括地描写雄奇壮阔的景物，又能细致入微地刻画自然事物的动态；在自然景物的观察上别有会心，能够巧妙地 捕捉适于表现其生活情趣的种种形象，构成独到的意境。第二，王维的作品被大量收录在全唐诗中，这使得他们能够被完好的保存和读取。《全唐诗》是清康熙四十四年（1705年）在明代胡震亨《唐音统签》和清代季振宜《唐诗》的基础上，旁采残碑、断碣、稗史、杂书，拾遗补缺，巨细靡遗。全书共900余卷，收录2200多人的诗歌作品48900余首。所以如果能找到全唐诗的电子版本，对于数据收集来说就变得轻松了。在前期探索和准备的过程中我发现，全唐诗被十分完整的电子化并且收录在了一个GitHub的库（repository）中，这降低了我搜寻和处理数据的难度。同时，王维和他大量的作品也为我们证明Verse Into Vision这个工具具备应对大批量工作量的处理能力。

In order to create a database of "paintings in poems and poems in paintings", the electronic version of "Complete Tang Poems" in the "Chinese Poetry" GitHub repository was used. The repository contains poems divided into 900 JSON files, corresponding to the volumes of the anthology. Each JSON file separates data into categories such as author, title, and content. Use Python and Pandas to merge these files, filtering basic data such as author, title, and poem content. Use the zhconv library to convert Traditional characters to Simplified characters and store the results in a CSV file. A new CSV file was generated specifically for Wang Wei's poetry, which served as the basis for creating the collection. This methodical approach emphasizes the project's technical and artistic efforts to bridge ancient Chinese poetry with contemporary digital art forms, ensuring accuracy and cultural integrity in the process.

During the attempts, I found that the tool didn’t works well when I input whole poem with multiple sentences. Since there are too many objects included in single poem, not to say that some poems contain dozens of sentences, it is better to only visualize the poem sentence by sentence. To save the time, I write a loop to use GPT 3.5 to select the most poetic sentences in poems. Then using another loop, the sentences selected are put into visualization function of Verse into Vision to generate three pictures. The reason I choose three pictures is because that I found the variation of DALL-E 2 is high, thus to generate multiple pictures can offer more choices of pictures and insure the quality of the collection. While generating the images, I recorded the input prompt for the drawing model and name each picture with the format of: title of the poem, sentence selected, and the index of the image. In this way the images can be well stored and easy to be searched and used.

With more than 1000 images, I faced another challenges which is how to presenting them. There are several plans that I came up with. First, I manually selected some images and publish them as a picture album. Second, I create a website and do a digital gallery which can allow user to search specific picture. However, these two mediums are lacking of interactivity with the audience or users. I want a presenting way that people will enjoy on viewing pictures with corresponding poems while make the interaction easy to use and attractive. Moreover, I also want to let users to grade the pictures in order to study the accuracy of the visualization and the different between machine’s understanding and people’s understanding. Inspired by the dating apps’ user interface, I came up with the design to present the collection "In Poetry, Paintings; In Paintings, Poetry". The logic is quite simple, instead of the whole collection, single images will be present throughout the website with its related information (authors, titles, the sentences correspond to the poem) and two buttons. Like the dating apps like Tinder and Bumble, viewer can choose to ‘like’ and ‘dislike’ the image in judeging from their own standard. With large amount of the viewers and enough data, we can array the images via the ‘like’ rate. This presenting way can present the images with a interesting interactive way while generating the feedback of the images.

Since both website is constructed based on Flask, it is easy to deploy them. In order to allow my peers to use the website, I deployed my website on a server connected to the Duke Kunshan University’s (DKU’s) intranet. This not only provides convenient and fast links, but also increases the security of the operation of the two websites. During the deployment process, I found that DKU's network proxy prevented Verse Into Vision from successfully using various API interfaces of OpenAI, so at the suggestion of my friends zezhen and Loe, I deployed the server to Duke's virtual machine, which can not only provide services for connecting to the Duke intranet (including DKU), but also connect to the API interface. After the successful deployment, I test run the website from November 11th to 17th for my peers to use and comment the online tool Verse Into Vision and the digital work collection "In Poetry, Paintings; In Paintings, Poetry".

Throughout the creative journey, I experimented with the impact of different types of poems on image outcomes. Additionally, encouraged by suggestions from peers, I explored using modern Chinese poetry for image generation to observe potential differences. This process led to numerous intriguing experiments by both my peers and me, fostering exactly the kind of diverse interactions between ancient poetry and artificial intelligence that I had envisioned. Unfortunately, due to time constraints and other factors, many of these remarkable experiments could not be included in this paper.

Outcome:

Two websites

Image collections