

Stockholm University of the Arts
February 2024

ARTISTS AND MACHINE INTELLIGENCE

**/Imagine: An archaeology of an autoregressive language model
that uses deep learning to produce human-like texts and images.
Kodachrome --ar 4:3 --v 6.0 --sref <https://s.mj.run/1Ue3Ju7YObw>**

by Linda Janson and Mirko Lempert



Table of content

Preface.....3

Introduction.....4

Methodology.....6

Experiments and Findings.....7

The Legal Aspects.....8

BETA.....8

Experiment A.....10

Experiment B.....11

Experiment C.....12

Experiment D.....13

Experiment E.....14

Experiment F.....15

Experiment G.....17

Experiment H.....19

Experiment I.....21

Experiment J.....22

Educational Settings.....24

Glossary.....26

Reference..... 28

Preface

In the dynamic interplay between art and technology, the integration of artificial intelligence (AI) has opened new frontiers of exploration and expression. Our research project, initiated in 2021, ventured into this evolving landscape with a mission to examine the relationship between AI and contemporary artistic practices. Focusing primarily on text and image synthesis using AI technologies, our project embarked on an in-depth exploration of the creative potentials and limitations of Large Language Models (LLM).

This journey was far more than a technical exercise; it represented a deep dive into the fusion of creativity and technology, examine traditional workflows, especially related to pre-production in filmmaking. The project revealed AI's capacity to both emulate and stimulate human creativity, offering insights into the capabilities and boundaries of LLM in artistic creation. These reflections are not just a recount of past achievements but a lens through which one can view the potential and future intersections of art and machine intelligence.

Leading the research project are Linda Janson, a production designer with over 30 years of experience in the art departments of films, TV series, and commercials, currently serving as a senior lecturer in production design at Stockholm University of the Arts, and Mirko Lempert, founder of the company Monocular. Monocular specializes in integrating AI with traditional 2D/3D techniques for practical applications in visual content creation. This project is driven by their collective goal to investigate the practical applications of AI in visual arts, combining Linda's extensive industry experience and academic background with Mirko's expertise in applying AI technologies. Their joint effort focuses on enhancing the methodologies of visual arts and design education to align with the ongoing advancements in digital technology.

We would also like to recognize and express our gratitude to PhD candidate Marc Johnson, whose significant contributions were instrumental in initiating this research endeavor and who has continuously supported the process with valuable advice and expertise.

Lastly, we wish to emphasize that Chat-GPT served as an editorial assistant in the composition of this report. This involvement was twofold: as an integral part of our research methodology and as a proficient contributor, aiding in the summarization, formalization, and articulation of arguments to deepen the discourse on the topic of Artificial Intelligence in the visual arts.

Introduction

The announcement of DALL-E by OpenAI on January 5th, 2021, marked a pivotal moment in the field of artificial intelligence and its applications in creative industries. DALL-E, a variant of the GPT-3 model capable of generating images from textual descriptions, opened new avenues in the interplay between AI and visual design. This system not only demonstrated proficiency in natural language understanding and image generation but also offered a glimpse into the future of design processes, particularly in the entertainment industry. Alongside OpenAI's DALL-E, models like Midjourney and Stable Diffusion have significantly contributed to the field of image synthesis. Midjourney focuses on creating high-quality artistic images, often with a unique, stylistic flair.

Stable Diffusion, on the other hand, is able to run on consumer-grade hardware, opening up possibilities for creators to experiment with AI-driven image synthesis and by doing so providing artists and designers with new ways to conceptualize and visualize their ideas. There's a growing trend in collaborative AI, where LLMs are being designed to work alongside artists in creative and analytical tasks, enhancing human capabilities rather than replacing them.

Our earlier conducted research in the field of real-time technologies, in particular regards previsualization and its effect on design processes has always centered around the concept of how to transform internal design thinking into external expressions¹. This practice has been instrumental in enhancing communication across departments and establishing iterative processes, moving away from the rigid, linear workflows traditionally seen in film industrial settings.

The introduction of LLM's seemed to align seamlessly with these principles, offering a novel tool for the visualization process in the entertainment industry. This evolution extends beyond the arts in isolation; it touches the core of storytelling, enabling the realization of ambitious visions with unprecedented speed and flexibility.

¹Vistisen, Peter: "The Roles of Sketching in Design" in No 6 (2015): Nordes 2015: Design Ecologies, ISSN 1604-9705. Stockholm

In our exploration of artificial intelligence's vast landscape, we narrowed our focus to its application in specific tasks, particularly within the film and TV industry. The initial stages of concept development are pivotal in audiovisual media production. Here, generative AI proves to be a game-changer. It enables artists to rapidly transform their ideas into visual representations. This capacity for quick iteration not only accelerates the previsualization phase but also significantly broadens creative possibilities. It empowers artists to explore and visualize complex or time-intensive concepts with greater ease and efficiency, a process that would be far more challenging without the assistance of AI.

Customized Set and Costume Design: In set and costume design, generative AI can be used to create a wide range of design options that align with a project's artistic ambition, production conditions and thematic requirements. By inputting specific parameters, designers can use AI to sketch sets and costumes, enabling a more experimental and innovative approach to design. This can be particularly advantageous in genres like science fiction or fantasy, where the imagination sets the limits.

Rapid Prototyping and Testing: For all forms of entertainment media, generative AI facilitates rapid prototyping and testing of ideas. This enables creators to experiment with different visual styles, narratives, and concepts at an early stage, reducing the risk and cost associated with development.

Transforming Script Development and Writing with Visual Referencing: In scriptwriting, early visualization through AI can profoundly impact the writing process. It offers scriptwriters a dynamic platform for visual referencing, allowing them to see immediate visual representations of their written words. This not only can enrich the scriptwriting process but also can enhance the narrative's depth by providing a more tangible and immersive experience. Early visualization assists in fine-tuning story elements, ensuring a more coherent and visually aligned final script.

By focusing on these specific applications of AI within the entertainment industry, our research sheds light on the transformative potential of these technologies in enhancing creativity, efficiency, and innovation in film and television production. This technology serves as a canvas and a bridge, connecting the depths of imagination with the tangible world, all while ensuring that the human essence and soul of the art remain at the forefront.

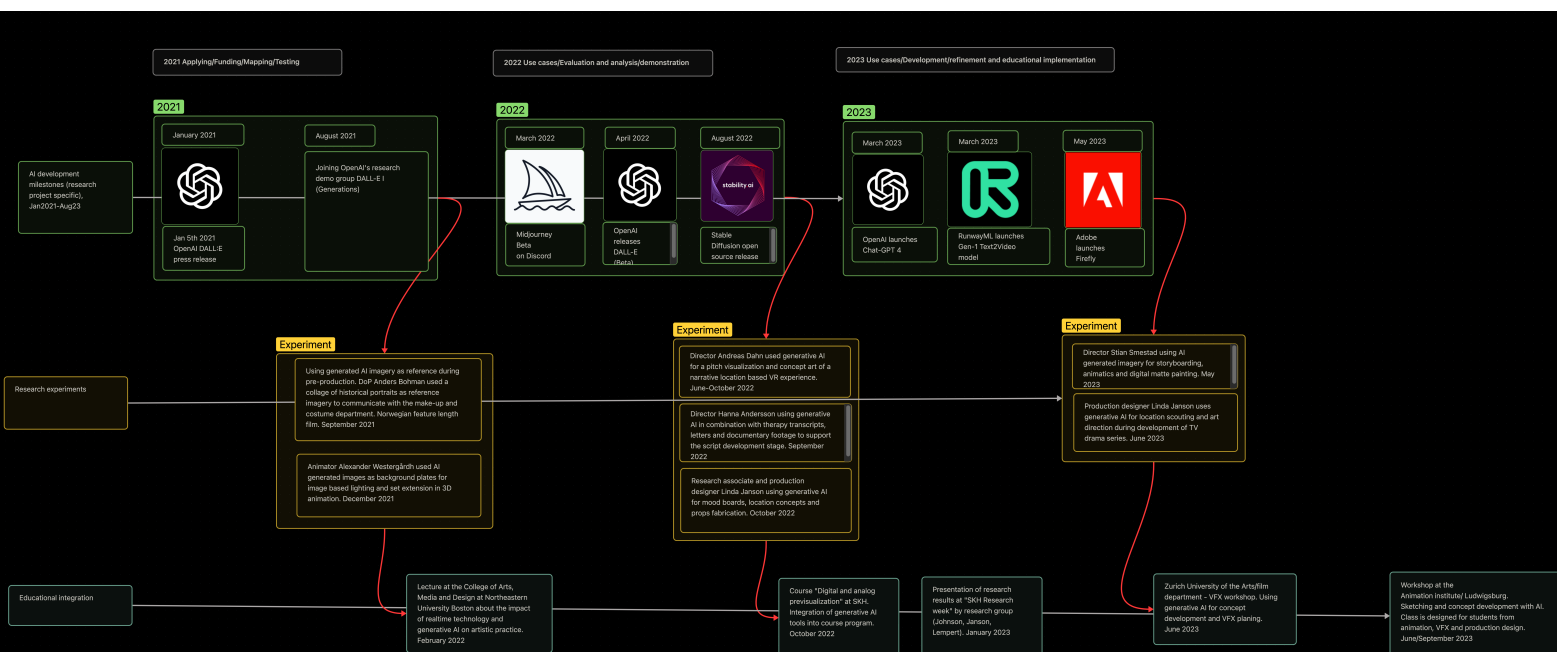
Methodology

From the outset we wanted to involve the use of AI technology in practical experiments and integrating the results into educational settings, thereby transferring knowledge into the classroom, a systemic approach to integrate AI into artistic practices. The timeline shown in chart 001 indicates that each step builds upon the previous one, ensuring that advancements in AI are not happening in isolation but are grounded in real-world applications and feedback. This structured approach ensures that the infusion of AI into the arts is methodical, evidentiary, and pedagogically sound.

It begins with the usage of AI tools like DALL-E or Stable Diffusion, moves onto applying these tools in practical settings like film pre-production, and finally incorporates the learnings into educational settings. Methodical experimentation with AI technologies in the realm of the visual arts allows us to map the contours of these tools' potential while identifying their limitations in real-world scenarios.

It serves as a dynamic process of hypothesis, testing, and iteration, which is indispensable in the calibration of AI for artistic creation and production. It's about transferring the results from our experiments into curricula that prepare the next generation of visual artists and introduces a discussion with the students about the topic of AI. Our methodology, which incorporates continuous feedback from both the experimentation and educational phases, remains agile. It adapts not only to new findings but also addresses ethical considerations, ensuring that our approach evolves in tandem with advancements in technology.

Chart 001



Case studies and findings

As previously outlined we approached a range of professional visual artists to examine the current possibilities of applying Artificial Intelligence in traditional production workflows and get a better understanding of the limitations of those tools. Please note that the rapid development of those technologies let our current results seem to be overdue or obsolete by the release of this report, nevertheless certain fundamental principles and insights have emerged as permanent. Through rigorous assessment, we have identified key areas where AI can enhance efficiency, such as in generating preliminary design concepts or automating routine aspects of the production pipeline such as asset creation for physical based lighting in 3D environments or the production of materials and props that can be used for art direction on physical sets.

However, our evaluation always underscores the necessity for ongoing human oversight and professional knowledge, particularly in the realms of aesthetic judgment and ethical considerations. The nuanced understanding of context, cultural sensitivity, and emotional resonance remains a uniquely human faculty that AI cannot replicate.

But before we continue listing experiments let us take a step back and have a closer look on generative AI in general and text and image synthesis in particular. Generative AI refers to a type of artificial intelligence that can generate new content, whether it's text, images, music, or other forms of media. This is achieved through algorithms that have been trained on large datasets of existing art or visual patterns. The AI learns from these datasets and can then generate new images or art that reflect the styles, forms, or themes found in the training data.

When it comes to text and image synthesis, especially in visual art, there are two primary aspects: *Text-to-Image Synthesis*: This involves generating images from textual descriptions. The AI is trained on a dataset of images and their corresponding descriptions. This is particularly useful in visual art for creating illustrations, concept art, or visual interpretations of written content. *Image Synthesis and Manipulation*: This refers to the AI's ability to create new images from scratch or modify existing ones. In visual art, this can be used for creating abstract compositions, altering or enhancing photographs, or producing realistic or fantastic scenes. The AI can be trained on various art styles and genres, allowing it to generate artwork in a range of visual aesthetics. However, it's important to note that while generative AI can create impressive works, it operates within the confines of its training and programming, and the role of human creativity and intention remains central in the art-making process.

The legal aspects

As visual artists navigating the world of artificial intelligence (AI) in art, it's crucial to be aware of the evolving legal landscape, particularly regarding copyright and the use of AI for creating images and text. Globally, this area is marked by continuous changes and discussions. Key among these is the need for transparency and responsibility in how we use AI in our creations. A major legal concern is the risk of copyright infringement, especially when AI tools use existing copyrighted works to generate new art. This is a complex issue that affects artists worldwide. Internationally, the big question is how to safeguard the artworks created by AI and set rules for their use. For artists, this means a trend towards more oversight and possible new rules for using AI in art. Yet, how these rules look can vary greatly from one country to another, influenced by local legal systems, cultural views on copyright and AI, and how advanced the technology is. As legal cases continue and new regulations are proposed, these will likely shape how we use AI in creating art and the legal side of things in the future.

BETA

In the summer of 2021, we were privileged to gain access to OpenAI's internal Beta program for DALL-E. This exclusive access offered us an advanced preview into the uprising field of Generative AI. Merging the linguistic prowess of GPT-3 with an embryonic version of ChatGPT, we were presented with a unique platform for experimentation. Despite grappling with resolution fidelity and occasionally stumbling in maintaining coherence, the early iteration of DALL-E, built on the GPT-3 framework, hinted at its future potential. While still in its formative stages, the AI offered creative possibilities where the fusion of text and image synthesis could revolutionize our interaction with digital content. From August 2021 to February 2022, we embarked on a creative exploration with DALL-E at the helm, assessing its potential as a tool for visual artists. In our initial projects, we concentrated heavily on previsualization, which led us to explore new opportunities aimed at enhancing early-stage visualization for rapid prototyping within the audiovisual domain, particularly in filmmaking. We recognized that the effective utilization of Large Language Models (LLMs) hinges on mastering the art of crafting prompts to yield diverse and nuanced results. To this end, we developed a linguistic framework that the model could understand, beginning with concise text inputs. We opted to start with minimal text input to gain a clearer understanding of the AI's interpretation and response to basic prompts before incrementally introducing greater complexity. This step-by-step approach facilitates the establishment of a more effective communication framework with the AI, ensuring that the outputs more closely align with our specific needs.

Exploring the AI's comprehension of place, time, and cinematic language proved particularly fascinating. By systematically combining elements such as time, place, weather, and shot type, our strategy delves into the combinatorial potential of the AI's output. This demonstrates the model's creative capacity and its ability to generate a broad spectrum of images from a defined set of parameters. As illustrated in the accompanying slide, our journey began with a simple setting, using the text prompt: "Astronaut arrives in an abandoned city. Brutalist architecture." Subsequently, we enriched the prompt by incorporating factors such as time, weather conditions, and nuanced location descriptions, each capable of further expansion. To advance our methodology, we investigated whether the AI could grasp cinematic language—a capability that would be invaluable for sequential use of this technology. To our delight, the AI met the challenge, allowing us to guide composition to a degree and incorporate different shot types at the end of each text prompt, such as: "Arctic landscape. Two people approaching a large cargo ship stuck in the frozen water. Snowstorm. Establishing shot."

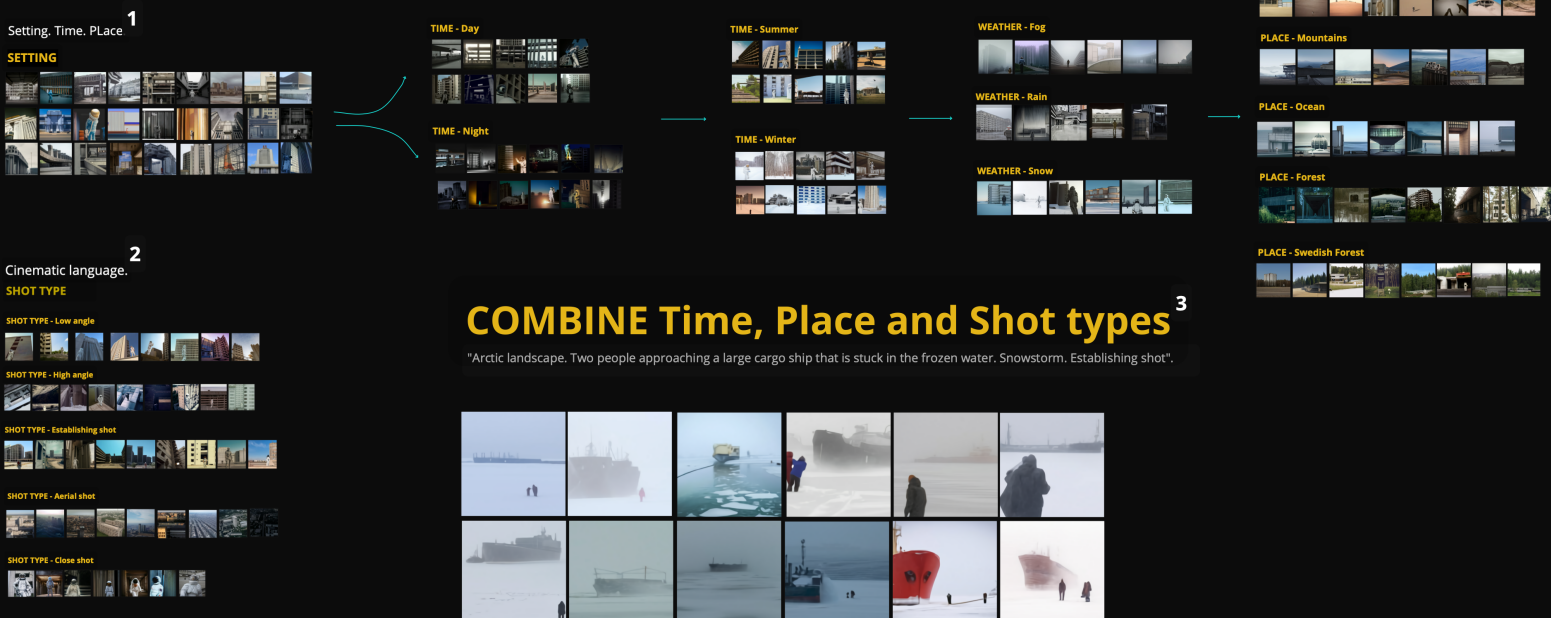
In retrospect, the foundational structure for text prompting that we chose continues to hold its ground, even as the models have evolved to offer higher visual fidelity and coherence. This approach remains a testament to the robust nature of our framework, affirming that generative AI can be used for conceptual development in various departments like camera or art department and over time for visual effects planning. Participation in the Beta program marked a critical introductory phase for us, where we developed the groundwork for efficient text prompting techniques, forming the cornerstone of our later experiments.

Chart 002

Test overview DALL-E Beta August/September 2021

BUILDING PROMPT STRUCTURE

Textprompt: "Astronaut arrives in an abandoned city. Brutalist architecture."



Experiment A

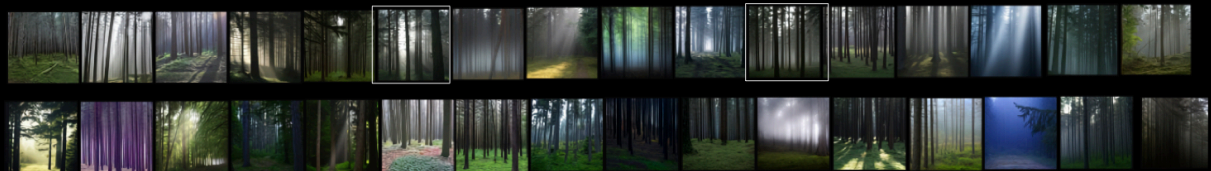
December 2021

3D animator and cinematographer Alexander Westergårdh utilised AI-generated imagery as backdrop elements within an animated short film. Our approach involved curating a collection of forest-type images generated by AI, which were then integrated into a 3D space as environmental backdrops. This technique not only reduced rendering times by significantly lowering the polygon count required but also allowed these images to act as static illumination sources.

Consequently, this enhanced the overall efficiency of the rendering process. We utilized thumbnail images (256x256px) produced by the AI, which were then expanded and combined in Photoshop to form a panoramic background. This served a dual purpose as both a scenic backdrop and a lighting element within the 3D environment.

This approach showcases a blend of 2D and 3D assets, leveraging the strengths of AI in image generation to solve practical challenges in animation and rendering. The efficiency gains from reducing polygon count and rendering times are significant, especially in projects with limited computational resources or tight production schedules.

Chart 003



DALL-E Beta, Selection of processed images



Select image, duplicate and merge to panoramic



Import panoramic to Blender



Test render including 2d panoramic background, Alexander Westergårdh, December 2021

Experiment B

September 2021

Cinematographer Anders Bohman utilized Generative AI to facilitate communication with the art department, particularly with the costume and wardrobe team. The Norwegian feature film production, set at the turn of the 19th to the 20th century, demanded historical accuracy while still aspiring for a high degree of creative interpretation within the team.

In meetings that brought together cinematography, production design and the director with the makeup and costume departments, AI-generated images served as a foundational tool for discussion. Anders remarked, “We opted for black and white images to avoid being overly prescriptive at the outset, focusing instead on form and diversity. We incorporated specific prompts such as the time period 'Early 20th century,' a geographic location 'Norway,' and particular occupational groups like 'Kitchen staff and mine workers.’” These AI creations provided a visual reference point for the team, ensuring a unified vision and greatly aiding the costume and wardrobe department in aligning their designs.

With the advancements in image synthesis in mind, one could argue for the integration of these tools early in the pre-production phase to experiment with different styles, cuts, and colors of the wardrobe. Similarly, the makeup department could rapidly test and visualize various prosthetic and special effects makeup options.

Chart 004



Experiment C

June-October 2022

Working closely with director and VFX artist Andreas Dahn, we crafted character sheets and initial environmental concepts for a narrative-driven VR experience, "Mino XR," as part of the EMIL project within the European XR lab network. This venture draws inspiration from Greek mythology. Andreas leveraged these visualizations for his pitch, which finally helped to secure the necessary funding. The project is scheduled to be released in 2024.

Moreover, these early visualizations provided Andreas with crucial insights into aligning his creative vision with the available production resources, guiding both aesthetic and pragmatic decisions to ensure they corresponded with the project's ambition and budget.

Chart 005

Selection

Concept Mino XR

Characters low poly



Props / technical study



Environment / level design



Midjourney vs1, Stable Diffusion 1.4, June/October 2022

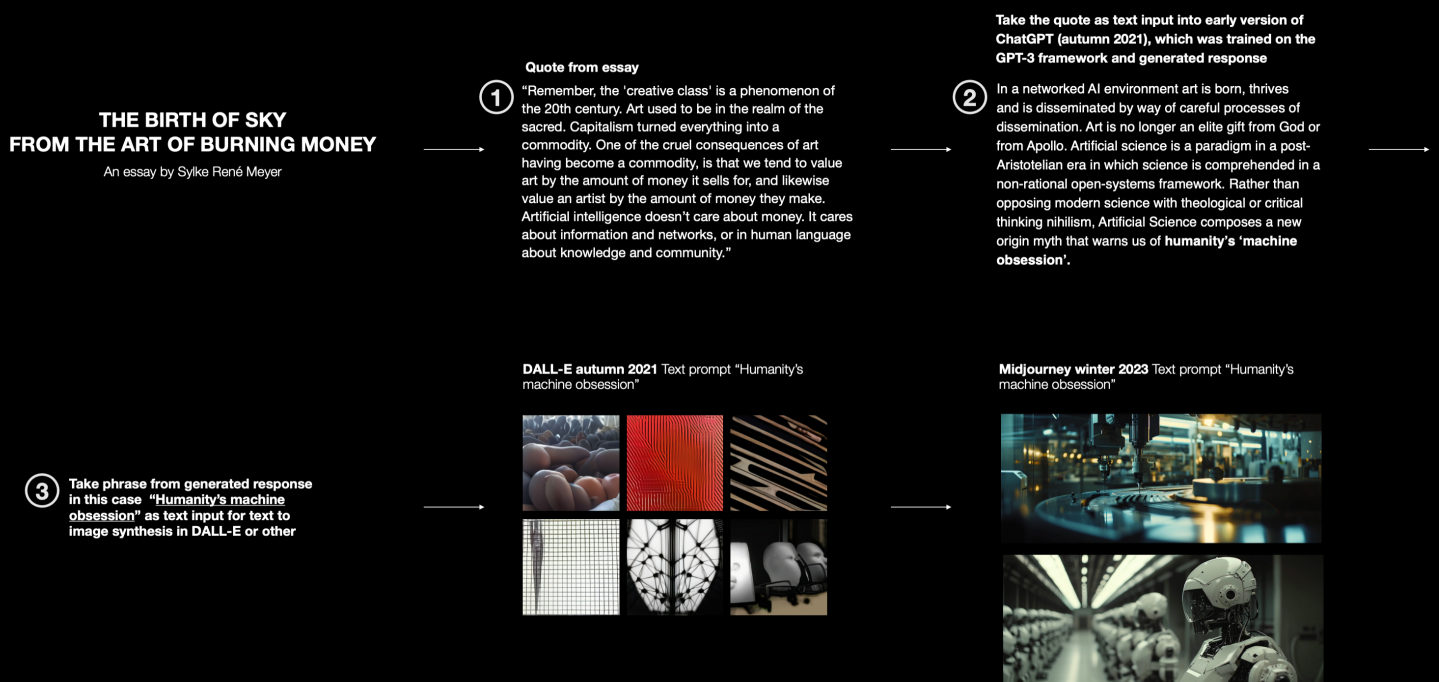
Experiment D

October 2021

In collaboration with visual artist Sylke René Meyer, professor for Theatre, Art and Design at Northeastern University in Boston, we embarked on an intriguing experiment during the closed BETA period of ChatGPT and DALL-E in autumn 2021. Our focus was to test the tools capabilities in illustrating text. The starting point was professor Meyers essay, "The Birth of Sky from the Art of Burning Money."

We fragmented the essay and used ChatGPT to generate new textual content, essentially remixing and re-contextualizing the original work. This process echoes the cut-up technique popularized by William S. Burroughs in the 1960s, yet with a modern twist: here, AI becomes an active participant in the creative process. Our next step was to bring these AI-generated texts to life visually using DALL-E. This translation of reinterpreted text into imagery added a new layer of depth to the experiment. The method challenged the traditional concepts of authorship and the role of the writer in the creative world. Furthermore, the inherent randomness of this approach led to unique and unexpected connections, offering a fresh perspective on finding order in chaos. This project, blending text with imagery through AI, stands as a testament to the evolving landscape of creativity in the digital age, inviting us to reconsider our understanding of art and its creation.

Chart 006



Experiment E

October 2022

In this experiment, we delved deeply into the impact of early visualization on the scriptwriting process and its potential to influence and shape the narrative. Collaborating with writer and director Hanna Andersson, we embarked on a journey to visualize potential locations, aiming to enrich and inform the writing process for the screen. At the nascent stage of script development for the project "Dear Carson," we adopted a unique approach. Utilizing photographs and documentary footage, we created visual sketches. These sketches were tools designed to help Hanna immerse herself more deeply into the cinematic imaginarium.

The method allowed her to explore parts of the visual and emotional landscapes of the story, bridging the gap between historical accuracy and creative interpretation. "Dear Carson" narrates the poignant tale of the relationship between American prodigy Carson McCullers and Swiss writer, photographer, and antifascist Annemarie Schwarzenbach. Their encounter in New York during the summer of 1940 blossoms into a story of intense, unrequited love, with the young McCullers falling deeply for her older, androgynous colleague.

Hanna reflects on the process of starting visualization during project research: "This project is based heavily on archival research and the process of pin-pointing key elements in the story through their physical reality and the imagined aesthetics of a film not only brought the images in my mind to life. What took place in those long lost spaces suddenly fell into place. It made me trust my own leap from archival truthfulness to imagined reality."

This method of using generative AI for early visualization is transformative. It serves as a bridge, allowing writers like Hanna to navigate the intricate dance between historical authenticity and the creative freedoms of cinematic storytelling. It's a tool that opens up new dimensions in the writing process, enhancing the writer's ability to visualize and thereby enrich the narrative with a vivid sense of place, time and emotion.

Chart 007

Selection

Dear Carson

Scene A

First meeting between Carson McCullers and Annemarie Schwarzenbach
Bedford Hotel, New York, 1940, hotel corridor, Rolleiflex



Scene B

Annemarie Schwarzenbach submitted to Bellevue Hospital
Psychiatric ward, 1940, New York



Midjourney vs3-Test, DALL-E I, Stable Diffusion 1.5, September 2022

Experiment F

September 2022

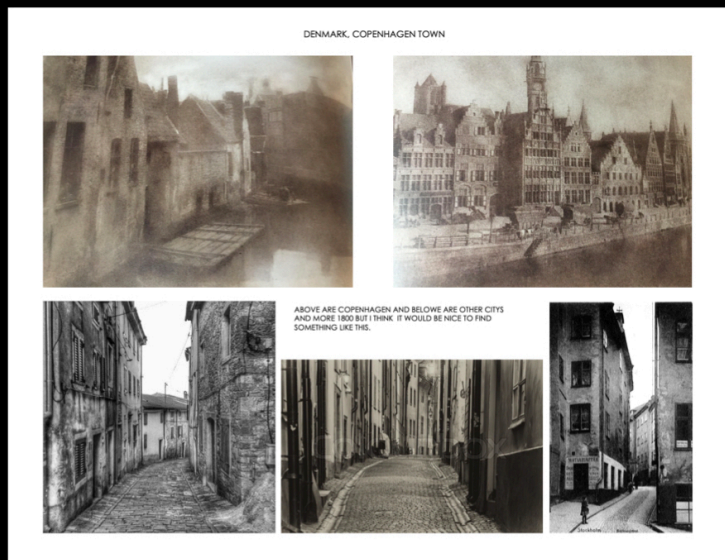
The pursuit of visualizing historical settings in film production often grapples with the challenge of limited or non-existent historical references. Generative AI offers a solution to this dilemma, as demonstrated in the experiment with the recreated Copenhagen street from the 1600s. Utilizing human oversight combined with generative AI, the production team can quickly and effectively generate a visual sketch that captures the essence of a location that time has erased. The technical sophistication of generative AI shines through in the provided images, where the transition from a basic mood board to an AI-generated concept is remarkable. The initial images serve as a foundation, a historical scaffold that the AI then builds upon, infusing it with creative details and a lifelike atmosphere. The final output, a haunting streetscape, bears the marks of a place both rooted in history and imagined with artistic freedom.

This rapid sketching capability—realized in a mere five hours—is not just about practicality; it's about creating a shared visual language that can be instantly communicated and understood by the entire film production team. When a director, cinematographer or production designer can see and feel the texture of the cobbled streets and the mood of the lighting, the creative dialogue becomes more precise and productive.

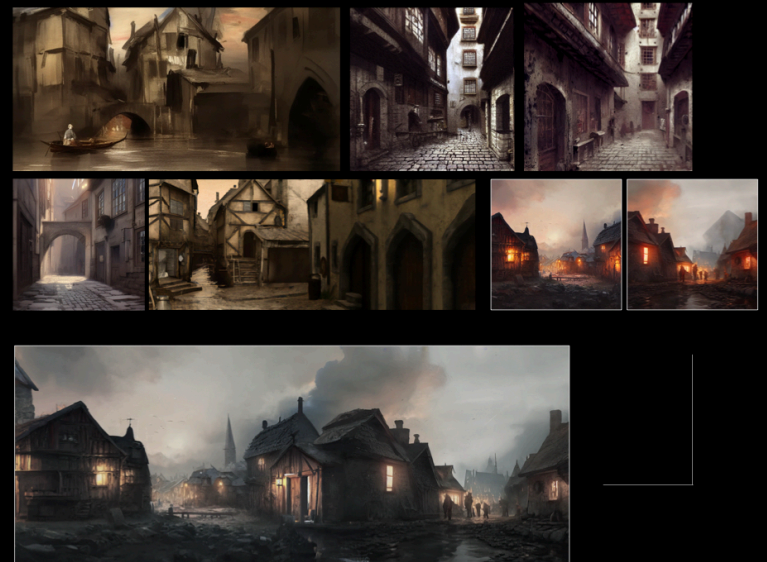
Instead of abstract descriptions or disjointed mood images, generative AI provides a cohesive visual narrative, aligning the team's vision and propelling the creative process forward. It not only accelerates the pre-production phase but also enhances the depth and believability of the cinematic world, allowing audiences to step into a time and place that, though never captured on canvas or film, feels undeniably real. Thus, generative AI is not just a tool but a creative ally in the art of filmmaking, offering a new avenue for directors and designers to craft and convey the visual stories they aspire to tell.

Chart 008

Regular method / Mood board + Floor plan



Experiment / AI generated concept in **Midjourney vs3, Stable Diffusion 1.5**



Final - **DALL-E** image merge and outpainting, October 2022

Chart 009



Midjourney vs 3 + DALL·E I, October 2022

Experiment G

November 2022

In the realm of film production, the set serves as the canvas upon which stories are painted. Generative AI stands as a brush in the set designer's hand, allowing them to transform a mere fragment of visual inspiration into a fully realized backdrop. The attached images exemplify this, where a small, historical wallpainting sample becomes the seed for a vast, wall-encompassing tapestry. The experiment began with a modest piece of a wallpainting, its patterns and colors echoing tales from a bygone era. The technique of pattern continuation and stylistic consistency, enabled the expansion of the wallpainting sample into a complete work, seamlessly blending the old with the newly generated to cover a large expanse suitable for a film set.

The transformation from a limited sample to an extensive set piece showcases the AI's potential to maintain the integrity of historical designs while scaling them to modern production needs. The process, once time-consuming, has now become more efficient with advancements in AI, making it possible to produce diverse iterations quickly. This adaptability ensures that each film can feature a unique element that resonates with its specific aesthetic and narrative demands.

The generative AI's capacity to extrapolate small details into grand visions not only saves time and resources but also allows for a kind of creative exploration previously unimaginable. Set designers can now experiment with designs in real-time, adjusting and evolving the visuals until the perfect look is achieved, one that aligns with the director's vision and enhances the film's storytelling.

In the case presented, what was a time-intensive yet satisfying process a year ago would now yield even more intricate and refined results. The implication for current and future film productions is clear: generative AI is not merely a novel technique but can become an essential tool in the creation of film sets that require historical accuracy, artistic beauty, and narrative depth.

Chart 010

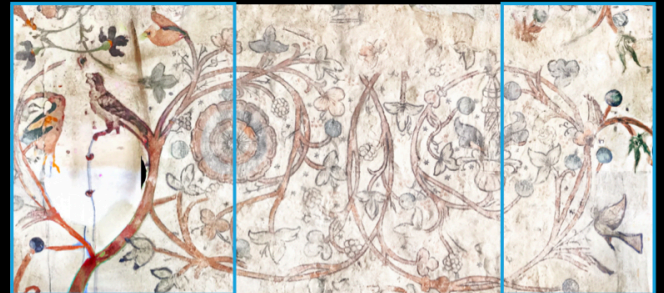
Original sample taken from a wall in a castle



First Color correction



Outpainting in DALL-E I, take portions of existing image and define look with text prompt



Outpainting in DALL-E I, repeat the step.



Chart 011



Experiment H

December 2022

In the endeavor to adapt an existing set construction, or backlot, to fit a new narrative and historical context, we utilized a photograph of the existing environment as a starting point for our creative transformation. The attached image showcases this method, illustrating how the original Barrandov backlot in Prague was digitally altered to conform to a different story and epoch. The method employed involves an innovative use of AI, specifically image-to-image model processing, for evaluation and budgeting purposes. The process begins with a clear, high-quality photograph of the existing set.

Utilizing Stable Diffusion 1.5 and DALL-E's outpainting capabilities, we infused the photograph with stylistic alterations that brought our envisioned changes to life. This allowed us to present tangible, visual proposals for discussions with the director, cinematographer and VFX team, streamlining the initial creative dialogue.

The image reveals the original state of the set and the final, altered version that reflects a dampened and more atmospheric tone, suggesting a historical shift. Below, we see a series of style variations, offering a glimpse into the iterative process that AI enables. This flexibility is invaluable for exploring various aesthetic directions without the labor-intensive process of creating new drawings for each iteration. This AI-assisted approach not only fosters a collaborative and dynamic design process but also serves as a practical tool for financial planning. The ability to visualize the end result aids in more accurate cost estimations for the required changes. Furthermore, the final AI-rendered image acts as a detailed guide for the production team, providing a reference for the potential hues, textures, and details needed to realize the set physically.

We found clarity for our inquiries. AI demonstrated its effectiveness as a resource for reference and preliminary sketches. It offers a swift approach to acquire reference materials that more accurately reflect visions, which is instrumental for advancing discussions toward achieving the visual objectives for the film. This tool streamlined collaborative conversations with the director, production designer, producer, cinematographer, and the teams in CGI, SFX, makeup, and costume departments.

Chart 012

Image to Image model processing for evaluation and budgeting

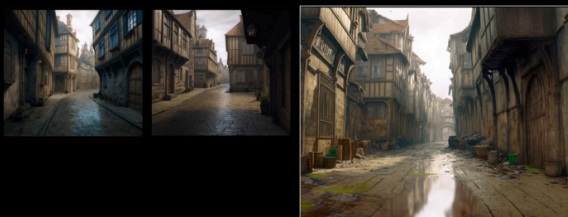
Original image Barrandov backlot Prague



Final



Style variations



Text prompt including original image Stable Diffusion 1.5, DALL-E I in+outpainting, November 2022

Experiment I

April 2023

This experiment tried to outline a straightforward process that integrates AI to enhance early visualizations during the creative development and production planning stages. Technically, the process involves merging references from online image searches with AI-generated descriptive texts to construct detailed prompts for creating cinematic stills. These prompts combine historical and architectural details to visualize locations accurately.

For production designers, directors, cinematographers and VFX teams, this inventive approach is instrumental in previsualizing scenes, especially for genre films and VFX-intensive projects that demand elaborate world-building. It enables the investigation of diverse visual styles and the assessment of potential set extensions or alterations, making pre-production more streamlined and cost-effective. This methodology not only simplifies the planning of intricate scenes but also ensures that the final film's aesthetic matches artistic ambition.

As detailed in the slide, the technique involves an image search for a fictitious "18th-century manor," paired with a ChatGPT-generated description of the architectural style of period-specific Swedish manor houses, highlighting Baroque and Rococo influences with their ornate ornamentation and balanced designs. This melding of visual and textual data crafts a text prompt that underpins the generation of a cinematic still image. The comprehensive prompt specifies a reference image and cinematic details, including aspect ratio, visual texture akin to Kodachrome, and nuanced image composition, along with environment descriptions. However, please be aware that you must legally clarify the usage rights of the image you intend to incorporate into your prompt, in case you work on a commercial project.

By combining historical and architectural accuracy with cinematic styling cues, production designers, cinematographers and directors can better envision and plan the necessary set dressing, set extensions or digital modifications required to bring a scene to life. The use of AI-generated prompts can help in creating pre-visualizations or "previs" of complex scenes before any physical or digital production begins, saving time and resources. It allows for experimenting with different visual styles and effects in an efficient manner before finalizing the design for actual production or even supporting the location scout with the search of finding a fitting location.

Chart 013

Mood/Decks

Goggle search "18th century manor"



ChatGPT description

Task: Please describe the architectural style of Swedish manor houses from the 18th century and give some examples.

ChatGPT (GPT-4) reply:

The architectural style of Swedish manor houses from the 18th century was predominantly influenced by Baroque and Rococo styles, which featured elaborate ornamentation, symmetrical designs, and grandiose facades. These manor houses were often characterized by their intricate details, gable roofs, and the use of materials such as brick and stone. Additionally, Swedish manor houses incorporated elements of local vernacular architecture, which resulted in a unique blend of international and national styles.

Textprompt: IMG link, Cinematic still, (subject=Chat-GPT description), film-look, aspect ratio etc

Google img +
Chat GPT description

<https://s.mj.run/WXIBIaxnuM8> A cinematic still. A house Influenced by Baroque and Rococo styles, which featured elaborate ornamentation, symmetrical designs, and grandiose facades. The house incorporates elements of local vernacular architecture, which results in a unique blend of international and national styles. The building is surrounded by an abandoned garden. Dead wood is covering the facade. Close shot. Kodachrome --ar 12:5 --iw 2



Experiment J

May 2023

The slide describes a comprehensive workflow that utilizes generative AI for previsualization in film production through a series of steps, beginning with the creation of a low-resolution 3D scene. Employing the mobil application "Marple", this initial phase is crucial, laying down the virtual elements within a 3D framework that forms the basis for subsequent enhancements. Moving forward, the process capitalizes on a stable diffusion technique to generate an image that elevates the initial scene's visual fidelity, infusing it with realistic details that mirror the envisioned final film.

Further refinement is achieved in the third phase, where the image undergoes a meticulous look development process using image-to-image generation technology, likely powered by a tool like Midjourney. Here, text prompts assist in honing the image's color, lighting, and overall mood, ensuring alignment with the directorial vision. The culmination of this workflow is a polished and retouched image that encapsulates the previsualized scene, ready to inform the film's actual production.

The efficacy of this workflow is underscored by its ability to facilitate rapid iteration and refinement, which is paramount during the pre-production phase. It fosters a collaborative environment where real-time feedback can be exchanged across departments, ensuring a cohesive and shared visual objective. The flexibility offered by this method is significant, allowing for creative exploration without the constraints of physical setups, thereby saving on costs and time. This is complemented by the precision and attention to detail that AI brings to developing the scene's look, which is essential for planning sophisticated VFX and cinematography. This workflow not only bridges the gap between the conceptual and the practical aspects of filmmaking but also provides a tangible visual guide that steers the entire production team. With tools like the Monocular app “Marple” enhancing previsualization, filmmakers can dynamically tailor and iterate on scenes. Such capabilities prove invaluable in ensuring that complex visual sequences are fully conceived and agreed upon before the commencement of principal photography, thereby optimizing the production process.

In essence, the workflow presented is an innovative and practical application of generative AI in film production. It empowers creative teams to visualize and refine scenes with a remarkable degree of realism and adaptability, offering particular advantages for productions that demand extensive visual effects or that aim to establish a specific visual style from the outset.

Chart 014

WORKFLOW

Take the still as input source and use AI Image to Image models to explore certain styles or looks.

Departments involved → Director, Camera department, Art dept, Editorial (VFX)



Low res 3D scene based on scan.
Build in Monocular app



Img2img generation Stable diffusion.
Transfer img to Midjourney



Img2img generation Midjourney + text prompt with emphasis on look development



2D Retouch



Final output

Educational settings

Our approach in integrating AI technology into the realm of visual arts, exemplified through a series of planned workshops, demonstrates the transition of theoretical AI concepts into educational settings. Utilizing AI tools such as DALL-E, Midjourney or Stable Diffusion, our methodology progresses methodically from their initial use to practical real-world applications, culminating in their integration into academic curricula. This progression ensures that advancements in AI are not merely theoretical or experimental but are intimately tied to practical usage and knowledge dissemination. The following is an introduction to four structured workshops designed to seamlessly blend AI technology into educational frameworks. These workshops not only exemplify our structured approach to AI but also illustrate the dynamic journey from experimentation to application and educational integration, highlighting the real-world impact and educational value of AI advancements

February 2022

Northeastern University, Boston (Theatre and Media Department): This workshop marked the beginning of our exploration, where students were introduced to real-time technologies and Large Language Models. The practical application of this technology was demonstrated in a play development project. Here, AI-generated backdrop images, created from the play's script, were used. This exercise not only facilitated a deeper understanding of visual storytelling but also illustrated how AI can directly translate written content into visual art, thereby expanding the students' creative thinking and approach to theatrical design.

October 2022

Stockholm University of the Arts (Pre-visualization Course): Focusing on pre-visualization in 2D and 3D spaces, this course delved deeper into the capabilities of generative AI. Students engaged with text-to-image synthesis and incorporated 3D scanned data of real locations. The resultant models provided various visual perspectives for storytelling and pre-visualization, effectively marrying traditional pre-visualization techniques with AI enhancements. This fusion encouraged students to rethink conventional methods, embracing a more integrated and innovative approach to visual narration.

June 2023

Zurich University of the Arts, Switzerland (VFX Workshop): In this workshop, students applied AI in visual effects through location scanning and photography. By utilizing AI techniques like image synthesis, in-painting, and out-painting, participants explored new dimensions in location illustration and set dressing. This process highlighted the transformative role of AI in visual effects, prompting students to reconsider the boundaries of visual representation and the potential of AI to elevate cinematic aesthetics.

September 2023

Animation Institute, Baden-Württemberg, Germany (Set Extension VFX Course): This workshop was particularly focused on the application of AI in set building within a science fiction context. Here, the integration of AI for text and image synthesis, alongside motion synthesis, demonstrated the potential of AI in expanding creative possibilities. The course emphasized conceptualization and the realization of ideas within production constraints, showcasing how AI can be a crucial tool in both the ideation and execution phases of creative projects. The use of AI for set extensions and concept manipulation provided a practical example of how AI can be leveraged to enhance traditional set design techniques, encouraging students to adopt a more exploratory and innovative mindset in their creative endeavors.

Throughout these workshops, the recurring theme was the impactful role of generative AI in transforming creative workflows and mindsets in the visual arts industry. By integrating AI into various stages of the creative process, from conceptualization to execution, these educational experiences showcased AI's potential as a collaborative tool in artistic creation, highlighting its ability to augment human creativity and open new avenues for artistic expression.

In conclusion, our research project has shown the value of our focused strategy, emphasizing specific scenarios and weaving AI technologies into the established processes of film production. The agility to rapidly iterate and to experiment with diverse variations that AI technology provides, show significant potential to empower artists. This advantage enables a comprehensive examination of the narrative from various perspectives, enhancing story exploration and facilitating artistic choices that are in harmony with the production's resources. Rather than replacing established methods, the integration of AI technology complements and enhances cinema as an art form.

Glossary

Machine learning

is a subfield of artificial intelligence, which is broadly defined as the capability of a machine to imitate intelligent human behavior. Artificial intelligence systems are used to perform complex tasks in a way that is similar to how humans solve problems.²

Generative AI

refers to deep-learning models that can generate high-quality text, images, and other content based on the data they were trained on.³

Prompt engineering

is the process of structuring text that can be interpreted and understood by a generative AI model. A prompt is natural language text describing the task that an AI should perform.⁴

Inpainting

Image inpainting is the art of reconstructing damaged or missing parts of an image — and the techniques can be extended to videos easily. There are a plethora of use cases that have been made possible due to image inpainting, from enhancing old photos to fixing corrupted videos.⁵

Outpainting

also known as image extrapolation, is a process where we try to extend the content of an image beyond its original boundaries. It's like an artist taking a small painting and extending the scene onto a larger canvas.⁶

Text to image synthesis

is a machine learning model which takes an input natural language description and produces an image matching that description. The most effective models have generally been trained on massive amounts of image and text data scraped from the web.⁷

² Brown, Sara: "Machine learning, explained." MIT-management Sloan School, April 21, 2021. <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>

³ Martineau, Kim: "What is generative AI?" IBM-Research, April 20, 2023. <https://research.ibm.com/blog/what-is-generative-AI>

⁴ Wikipedia: "Prompt engineering", February 07, 2024. https://en.wikipedia.org/wiki/Prompt_engineering

⁵ Paul, Sayak: "Introduction to Image Inpainting with Deep Learning", <https://wandb.ai/ayush-thakur/image-impainting/reports/Introduction-to-Image-Inpainting-with-Deep-Learning--Vm1ldzo3NDlwNA#what-is-image-inpainting?>

⁶ Invoke: "What are Inpainting and Outpainting?", Invoke Support Portal, May 22, 2023. <https://support.invoke.ai/support/solutions/articles/151000096702-what-are-inpainting-and-outpainting->

⁷ Wikipedia: "Text-to-image model", February 11, 2024. https://en.wikipedia.org/wiki/Text-to-image_model

Large language models

is a language model notable for its ability to achieve general-purpose language generation. LLMs acquire these abilities by learning statistical relationships from text documents during a computationally intensive self-supervised and semi-supervised training process.⁸

GPT

Generative pre-trained transformers (GPT) are a type of large language model (LLM) and a prominent framework for generative artificial intelligence. They are artificial neural networks that are used in natural language processing tasks. GPTs are based on the transformer architecture, pre-trained on large data sets of unlabelled text, and able to generate novel human-like content.⁹

DALL-E

is a 12-billion parameter version of GPT-3 trained to generate images from text descriptions, using a dataset of text–image pairs.¹⁰

Stable Diffusion

is a deep learning, text-to-image model released in 2022 based on diffusion techniques. It is primarily used to generate detailed images conditioned on text descriptions, though it can also be applied to other tasks such as inpainting, outpainting, and generating image-to-image translations guided by a text prompt.¹¹

Midjourney

is a generative artificial intelligence program and service created and hosted by San Francisco–based independent research lab Midjourney, Inc. Midjourney generates images from natural language descriptions, called prompts, similar to OpenAI's DALL-E and Stability AI's Stable Diffusion. It is one of the technologies of the AI boom.¹²

⁸ Wikipedia: "Large language model". February 12, 2024. https://en.wikipedia.org/wiki/Large_language_model

⁹ Wikipedia: "Generative pre-trained transformer". February 12, 2024. https://en.wikipedia.org/wiki/Generative_pre-trained_transformer

¹⁰ OpenAI: "DALL-E", January 5, 2021. <https://openai.com/research/dall-e>

¹¹ Wikipedia: "Stable Diffusion". February 11, 2024. https://en.wikipedia.org/wiki/Stable_Diffusion

¹² Wikipedia: "Midjourney". February 6, 2024. <https://en.wikipedia.org/wiki/Midjourney>

Reference

Brown, Sara: "Machine learning, explained." MIT-management Sloan School, April 21, 2021. <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>

Invoke: "What are Inpainting and Outpainting?", Invoke Support Portal, May 22, 2023. <https://support.invoke.ai/support/solutions/articles/151000096702-what-are-inpainting-and-outpainting->

Martineau, Kim: "What is generative AI?" IBM-Research, April 20, 2023. <https://research.ibm.com/blog/what-is-generative-AI>

OpenAI: "DALL-E", January 5, 2021. <https://openai.com/research/dall-e>

Paul, Sayak: "Introduction to Image Inpainting with Deep Learning", <https://wandb.ai/ayush-thakur/image-impainting/reports/Introduction-to-Image-Inpainting-with-Deep-Learning--VmIldzo3NDlwNA#what-is-image-inpainting?>

Vistisen, Peter: "The Roles of Sketching in Design" in No 6 (2015): Nordes 2015: Design Ecologies, ISSN 1604-9705. Stockholm

Wikipedia: "Generative pre-trained transformer". February 12, 2024. https://en.wikipedia.org/wiki/Generative_pre-trained_transformer

Wikipedia: "Large language model". February 12, 2024. https://en.wikipedia.org/wiki/Large_language_model

Wikipedia: "Midjourney". February 6, 2024. <https://en.wikipedia.org/wiki/Midjourney>

Wikipedia: "Prompt engineering", February 07, 2024. https://en.wikipedia.org/wiki/Prompt_engineering

Wikipedia: "Stable Diffusion". February 11, 2024. https://en.wikipedia.org/wiki/Stable_Diffusion

Wikipedia: "Text-to-image model", February 11, 2024. https://en.wikipedia.org/wiki/Text-to-image_model