The Dybbuk, The Loab, Latent Space and Multi-Modal Artificial Intelligence

xCoAx 2025
13th Conference on Computation, Communication, Aesthetics & X
Dundee, UK

2025.xCoAx.org

In 2021 Open AI introduced CLIP and VQGAN, neural network architectures for still and moving text-to-image models (Crowson 2021). Questions arose about how these algorithmic images were sourced, rendered and generated. *Language Is Leaving Me – An AI Cinematic Opera Of The Skin (LILM)* used a multimodal format, meaning "a ML (machine learning) model that is capable of processing information from different modalities, including images, videos, and text" (Google 2024). It examined the creation of images through the lens of epigenetic or inherited traumatic memories of cultures of diaspora and revealed a dystopic side in Stable Diffusion's prompt engineering and image-to-image capabilities, one of structural flaws and crucial biases. This paper examines AI's dystopic "digital twin" or unconsciousness that carries within itself what in Yiddish folklore is known as the "Dybbuk", a wandering spirit of a dead person possessing the soul of a living person, and in the AI, world is called the "Loab". It investigates the relationship between algorithmic rendering, latent space, epigenetic memory, selective inattention, and machine human cognition exploring new pathways.

1. Introduction

The embedded tagging systems commonly referred to as Large Language Models (LLMs) emerged around 2017 built on transformer models, meaning neural networks learning context or pattern recognition to understand predictive models of sequential data. AI models retrenched on using tagging systems for predictive textual taxonomic systems, but this is not as feasible for AI visual imaging systems. Data anonymized for privacy deletes many taxonomy focal points adding unnecessary noise to results. Large data models can be manipulated into reproducing parts of training data that were previously deleted (Vaswani et al. 2017). According to OpenAI, it is difficult if not impossible to prevent an opensource model for being used to cause inadvertent harm, as many input classifiers amplify bias This problem of unsafe or distorted content output is one of the thorniest problems in visual AI and one all major companies are scrambling to implement solutions for, with varying degrees of success (Roose 2023).

Ellen Pearlman
RISEBA University, Riga, Latvia, 2NYU Tandon School of Engineering, NY, USA
ellenluminescense@gmail.com

Keywords Dybbuk, Loab, latent space, artificial intelligence, epistemic AI, catastrophic forgetting, inattention blindness, epigenetic trauma

DOI 10.34626/2025_xcoax_003

Epigenetics is understood to mean chromosome modifications that are not part of DNA structures, and more specifically its rDNA (Felsenfeld 2014). It is not definitively proven that human psychological trauma can change chromosomal rDNA, much less human behavior. However, many psychologists understand trauma can be passed through generations and manifest in various ways (Hubl 2020). Intergenerational trauma is often non-verbal, cryptic, and difficult to decipher (Rosner 2017). I questioned if AI could comprehend my epigenetic trauma as an agnostic, third generation ethnic Jew from the Pale of Settlement in Eastern Europe, much less the epigenetic traumas of other cultures of diaspora. Examining emerging artificial intelligence synthetic image rendering and nascent cinema driven by Large Language Models (LLMs) within the context of epigenetic trauma, it took three years to develop "*Language Is Leaving Me* – An AI Cinematic Opera Of the Skin." (LILM). *LILM* was a performed cinematic installation partially powered by EMG human biometric measurements. A volunteer audience member had EMG patches placed on her face to monitor either a smile (positive) or frown (negative) response while watching the movie. Her reactions controlled the sonic environment for the entire audience.

I translated the movie's English language narration into Yiddish, Chinese, Tamil, and Xhosa cursive scripts processing them through both verbal (text prompt engineering) and visual (image-to-image) AI. The resulting multimodal movie revealed latent Dybbuks, hidden and devastating aspects of the algorithmic processes underlying transcultural LLMs. Algorithmic processing erased memory and distorted personal signified visual and verbal referents.

I did this to recreate and trace my usually unconscious memories stored inside my brain. I now want these images to be eventually fed back into even higher level AI systems and interacted with in real time by a performer. Translating my English language memories into four languages, I will fragment and decoupled their linguistic and visual semiotic significance concerning unexpressed trauma. The reason is AI purports to understand through linguistic and visual sampling the meaning of trauma, and other deep and formative human emotions, but in truth is incapable. The enhanced performance, currently a work in process, represents surrendering individual and collective memory to the unfolding age of AI.

2. Process and Journey – Discoveries and Challenges

In early 2021 into 2022 AI imagery consisted of CLIP (Contrastive Language-Image Pre-training) and VQGAN (Vector Quantized Generative Adversarial Network). Its images produced odd and phantasmagorical results. CLIP arose from investigations into zero-shot learning, natural language supervision, and multimodal learning. Zero shot learning means an AI model can understand and categorize objects and concepts without ever haven seen examples of those concept before. Natural language supervision analyzes speech and text in both grammatical correct and grammatically incorrect variations, and multimodal learning explores ideas through visual, sonic, perceived, written haptic and movement modalities.

VQGAN arose from previous text-to-image transfer models like AttnGAN (Attention GAN) and others (Xu et al. 2018, Radford et al. 2021, NerdyRodent 2021). CLIP decides which tagged captions VQGAN aligns best with so it can be assigned to a specific image generation. The CLIP model includes its own separate transformer architecture but does not generate anything itself. It deals with zero-shot learning datasets. This means a model can observe samples from information classes that were not included in the original training model (Larochelle et al. 2008). CLIP is open source, whereas other robust visual classification models like DALL-E, a diffusion model that used a different set of transformers and also developed by OpenAI were initially not opensource (Ramesh et al. 2022). Katherine Crowson refers to CLIP as the "Perceptor" or selector and VQGAN as the "Generator" meaning the generator of images. There are not smart discriminators for filtering information outside of common contemporary data harvesting techniques, another issue that contributes to skewered results. Though there has been rapid development of these techniques since 2021, many of the foundational problems still plague the field, which is why older models are referenced.

In 2021 I used the commercially available platform on the web, Night Café but dissatisfied with the results switched to the python language based solution, Google Colab. Since I am not a python programmer, I worked with volunteer software developers to use the early release Google Colab notebook "VQGAN+CLIP (with pooling).ipynb". I chose the English language prompt engineering word "queer" because during the late 19th century and early 20th century the British English vernacular meaning referred to someone who was odd or eccentric, such as the phrase "that's a queer fellow". In recent decades the meaning of the word queer has shifted in the global parlance to a more gendered and sexualized meaning.

Fig. 1. Chinese language examples

The Google Colab image datasets used Faceshq_5; ImageNet_10245; ImageNet_16394; OpenImages_8192_5; WikiArt_1024; and WikiArt_16384. All six visual datasets returned strikingly similar representations for the word queer. I did not understand why, since they had access to millions and millions of images through the six chosen data model sets. I used a fixed seed value of 42 for all the images. A seed value brings together a unique algorithmic mix of pixels, colors and shapes. Theoretically there are millions of seed values, though working with a fixed value allows images rendered across datasets to be more easily compared and standardized. Perhaps that is why the images resembled each other so much, but I thought that was too simple an explanation.

I added one other word to the text prompt, both “boy” and “girl” to see if anything changed. Again, the images were strikingly similar. I began testing the term using other scripted or character based languages that were not English. One of the programmers was Chinese and well versed in Chinese slang or colloquial vernacular terms. Google Translate was deployed for the strict Chinese translation of “queer boy” and “queer girl” and the Chinese native speaker input the slang variations that Google Translate would not know. We used the ImageNet 16384 image bank and the fixed seed value of 42. The results were shocking. Another programmer, this time from India spoke both Hindi and Tamil, so we tried the same experiment using those cursive scripted languages. The only difference was we used the COCO (Common Objects in Context) image bank, our seventh image bank also with the seed number 42. The results were even more outlandish. These image banks and their algorithms have subsequently been updated, reworked or had sections removed, but again the fundamental issues remain.

Fig. 2. Examples in Hindi and Tamil

2.1. Language Is Leaving Me

Experimenting with VQGAN, CLIP and Google Colab, though interesting produced unsatisfactory results for a creative project. Frustrated, in 2022 I returned to the basics producing a short video "Language Is Leaving Me" . The video used both original and archival copyright free footage to explore my epigenetic trauma as an agnostic Jew whose ancestors immigrated to the United States in 1906 from the Eastern European Pale of Settlement. Also, by 2022 huge advances had developed in the field of moving image AI cinema, spearheaded by the release of the opensource visual text-to-image model Stable Diffusion from Stability.ai. It used datasets from the LAION 5-B latent diffusion image bank (Schuhmann et al. 2022, Rombach et al. 2021). LAION-5B, created by the German charity LAION uses an aesthetic predictor, a numeric value indicating how much someone likes an image. Schuhmann states "4000 samples were annotated in a scale from 0 to 10 to be good looking or not". I found those definitions of the rating of aesthetic values by computer programmers, not artists, and not culturally comprehensive to be extremely problematic, especially after what I had discovered from my experiments using Google Colab with CLIP and VQGAN in Chinese, Tamil and Hindi. Simultaneously in 2022 AUTOMATIC 1111 introduced a web based graphical user interface that allowed both text-to-image and image-to-image scenarios (AUTOMATIC1111 2023). This made Stable Diffusion more accessible to artists. Running a series of tests, I worked with narrative translations of my original English language video in the cursive scripts of four languages: Chinese, Yiddish, Tamil, and Xhosa. I chose Yiddish because, it is the language of my ancestral diaspora, even though I neither speak nor read it. The other languages were chosen both for their geographic diversity and their original and fluid cursive scripts.

The biometric implementation of the sonic environment during the live performance received EMG input from the facial muscles of an

audience volunteer who viewed the movie alongside the audience. The EMG patch on their face registered a frown on their forehead (negative), or a smile on their cheek muscle (positive) to drive the sonic environment. The underlying sound was created by a sonic artist in Touch Designer and used a customized motherboard built in Shenzhen, China. The volunteer's reaction corresponded to the general reactions of the audience, with the sonic environment co-created at the time of the viewing between all participants. Though biometrics were important to the creation of the live performance I will not discuss *LILM's* human computer biometric interaction aspect in this paper.

3. Digital Twins – The Dybbuk and Loab in The Machine

During the development of *LILM* strange and malevolent images began appearing. This may have been influenced by the discussion and scenes of war, genocide, and disinterred corpses in my film, but they were not present all of the time. The batch processing in Automatic1111 using multilingual cursive scripts and image-to-image prompts kept generating human faces full of horror and fear, even when the prompts did not request or refer to this type of imagery at all. The algorithmic rendering process in Stable Diffusion evoked completely unintended digital twins' imagery that paralleled what in Yiddish folklore is the concept of the "Dybbuk" and in AI parlance called the "Loab". This discrepancy was enhanced during the subsequent performance installation involving a live volunteer audience interaction.

3.1. The Dybbuk – Ancestral Memories

A Dybbuk is the wandering spirit of a dead person possessing the soul of a living person it once loved. It eventually leaves its hosts body through an exorcism performed by a trained Rabbi. The mystic Isaac Luria (1543-72) is credited with establishing the belief in *dybbuks* through his doctrine of *gilgul* or the transmigration of souls. The word come from the Hebrew word ʔdibūq, קוּבִּיד meaning 'a case of attachment'. The idea of the *Dybbuk* became known to the general public with the publication *The Dybbuk, or Between Two Worlds* (Yiddish Der Dibuk, oder Tsvishn Tsvey Veltn), based on stories collected by S. Ansky who traveled to remote shtetls throughout the Pale of Settlement in Ukraine and Russia. In 1920 the book was turned into a wildly popular play produced in Yiddish in Warsaw, Poland and premiered on December 9, 1920 by the Vilna Troupe. The following year it was performed by the Hebrew language Habima theater in Moscow. In 1936 The *Dybbuk* was made into a movie by Mical Waszynski and is considered a classic of Yiddish cinema (Nashon 2021).

Through algorithmic rendering a digital twin or type of horrified Freudian id unexpectedly and mysteriously kept manifesting between the frames of my multi-lingual image-to-image video processing by Stable Diffusion. How could AI possibly be manifesting wandering spirits, or the souls of dead people culled from the archival and original source material I used? Over the three years of developing *LILM*, I deliberately avoided using the numerous commercial platforms available for visual AI. Instead, I painstakingly built my own workflows rendering experiments on servers based in Hong Kong, thereby avoiding many of the content restrictions pay-per-use customers of commercial platforms experienced.

3.2. The Loab – Artificial Intelligence Ghosts

On September 6, 2022 Steph Mai Swanson using the handle "Supercomposite" posted on Twitter, (now X) about the freakish results she was getting using an unnamed AI art commercial platform (Swanson 2022). Swanson was experimenting with negative prompts or "negative weighting" to shape an image of the actor Marlon Brando. Instead of her prompts rendering a look alike Brando, a horrific disfigured woman kept manifesting. Nonsensical characters spelling out "Loab" displayed alongside one of the renderings, so she called this phenomena "the Loab", stating, "Loab is the last face you see before you fall off the edge" (Coldewey 2022). The Loab became a social media sensation, but no one could explain exactly how and why it happened. Swanson called it a "combinations of optics and software". She then wrote an article about other deeply disturbing aspects of using AI, including how she coaxed it into creating an X rated BDSM (bondage, domination, sadism, masochism) relationship (Swanson 2023). At the time I was developing my own visual AI aesthetic I was unaware of Swanson's Loab, but our creative practices were uncovering similar issues of this lurking unconscious digital twin that seemed to span across linguistic and cultural divides.

4. Latent Space Births Unconscious Diabolic Digital Twins

Latent space is the key to understanding some of the mysteries in visual computer image rendering, especially for those not knowledgeable in computer science. This is because it is concealed, abstract, compressed, and muti-dimensional. In its essence its purpose is to encompasses internal or algorithmic representations of external events. In theory latent space contains mathematical models of all possible combinations of multiple renderings in a massive data set and can be shaped by using specific combinations of parameters.

Fig. 3. "They dug them up" – English archival footage and Chinese generated AI text-to-image and image-to-image

Fig. 4. "It was stories of the Jews of Azipute" – Tamil AI image-to-image generation and English language archival footage original

However, in fact it is just numbers collecting and processing different sample point numbers to output noise or vectors of numbers, which is why it is a strange concept for artists to grasp. The vectors, or intersection points are the range of possibilities for each latent variable and reach into the billions. All of these possible combinations take up a lot of storage space so in order to be used all the raw data needs to compressed to save space. This compressed space is coded and then decoded by different algorithms that calculate the differences and similarities between the sample points in space to render all that algorithmic noise into a coherent image. It is so exact that changing a model by just one number or seed changes the retrieval parameters and thus changes the output image. Usually similar combinations (i.e. types of birds, or types of cars) are mathematically lined up close to one another in latent space so they can be retrieved faster, but these parameters are prone to corruption. Algorithmic latent space representations fall apart even more when comparing multi-lingual text prompts in image-to-image batch comparisons. These comparisons seem to evoke a deranged digital twin appearing as the malevolent unconscious of the human psyche transculturally understood to be horrific or diabolic. It is the space where the Dybbuk or Loab is both concealed and emerges from. Latent space can't morph or transition from frame to frame with all details smoothly integrated, it can only use or pull the most prominent features of an image into the next rendered image. The hidden or dark side emerges because out of all of the billions and billions of images fed into the databanks, a percentages of those image are NSFW or Not Safe For Work. Like the exorcist needed to bring forth the Dybbuk, the NSFW images can be evoked and the "soul" released. I observed this very clearly by disrupting traditional algorithmic rendering formulas when I mixed cursive scripts of different languages into the multi-lingual text prompts alongside image-to-image comparisons. The algorithms gagged, either erasing my epigenetic memory entirely or producing horrific ghosts. Instead of trying to delete these apparitions, I incorporated them into my artistic practice.

Fig. 5. The original Loab courtesy Steph Mai Swanson via Twitter

Those who are capable of diving the mathematical formula of how and where to pull images from latent space become the conjuring magicians of the age of AI, usually unknowingly releasing both angels and devils. The same process that produces the digital twins of Loabs and Dybbuks also de-categorizes populations who histories are not digitized and taxonomized. This process is referred to as 'erasure' or "catastrophic forgetting" (Gebru and Denton 2024). The Dybbuk and the Loab are the deep, dark unconscious of algorithmic rendering processes, the imperfect and corrupted digital process of perfect, mathematical calculations (Katzman et al. 2023).

4.1. Evocations: Latent Space and Inattention Blindness Foment Visual Neglect

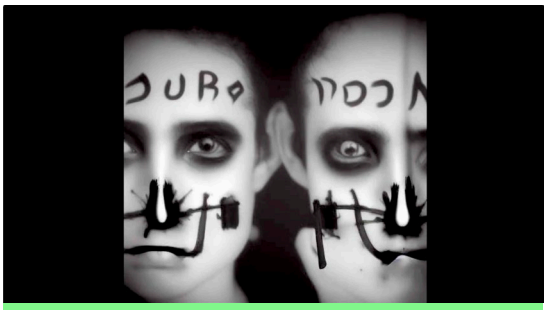

Fig. 6. The Juro or (Jew) Dybbuk manifesting from the Yiddish cursive script. English prompt “He breaks out those shinny things from that skull those teeth”. The characters etched on the faces were not prompted but spontaneously appeared.

Human visual perception registers information from 120 million rods and 7 million cones that flow from the optic nerve to the brain. The small optic nerve functions, or can be compared to a type of vector space compression system where the information is sent to and then biochemically decoded in the brain (Souza et al. 2024). This human compression system could loosely be compared to the visual systems own version of a latent space biological compression algorithm. Thinking in this way and can lead to the design of new artificial visual systems. However, there is a vast difference between algorithms and humans. Humans have presence in the real physical world and experience unpredictable moment-to-moment interactions. These interactions influence how they construct and remember their world. Within presence and unpredictability lies selective attention, or “inattention blindness” a term highlighted by the invisible gorilla experiment (Chabris and Simons 2011). In this experiment individuals were asked to pass a basketball between all members of a group. Their interaction playing the game was filmed. After the task was over, they were asked if they had noticed anything strange going on during the game. Approximately fifty percent of respondents said they had not seen anything unusual. When the film was played back to the entire group a person in a gorilla suit was clearly walking around in the middle of the game activity, visible for all to see. However, half of the participants had not even noticed it. Inattention blindness means only selected information, according to many factors is remembered by each individual. We are unable to recall every second of our life, but given the correct stimulus, we might evoke forgotten events deeply stored in the brain. As an example, most people remember their first primary school teacher but unless asked, never think about who that was and what they felt about their teacher. In the same way machine learning algorithms incorporates latent space using high level pattern recognition calculations. The Dybbuk or Loab can be thought of as the invisible gorilla in the suit. It mimics the selective inattention of algorithms that bring forth the unconscious aspect of human experience hidden throughout these large image banks when provoked, or in this case, evoked.

Meaning in the age of algorithmic smoothing, machine learning and data scrapping fosters semantic taxonomies that skewer mathematical algorithmic interpretations. This skewering is referred to as the ‘epistemics of training sets”, the fraught and complex relation between images and the concepts tying those images to their linguistic

meaning and that meaning to image tagging prompts (Cuzzolin 2021). Currently image tagging consists of nouns, with verbs and adjectives thrown in to enhance their meaning. Epistemic AI formulates new mathematical models for decision making. Its operating premise is that instead of inferring predictive models about the data it has at its disposal, it will assume it has a paucity of data from which to make any sort of conclusion. Only since 2016 has the issue of epistemic uncertainty in machine language become an area of focus. It acknowledges the difficulty of comprehending in domain and out of domain sampling that causes model formation to be insufficient. Another factor is the mundane human labor of images tagged by individuals whose first language is not English, and who lack the understanding and cultural nuances to comprehend what they are looking at. Paid by the word, it is advantageous for them to perform a cursory analysis of an image, tag and often mislabeling it, and move onto the next task at hand (Ruiz 2019, Crawford and Paglen 2019, Joler and Crawford 2018, Zhou et al. 2022).

Digital imaging systems traditionally process pixels on grids. Human visual perception reduces data from photoreceptors to the optic nerve. Metaphorically speaking they encodes image information into low bandwidth latent space representation suitable for processing by the human brain (Souza et al. 2024). Latent space can be reconfigured through optics and software into a generative model and these spaces are "extracted" from peoples and cultures who have no say in building them. Computer vision systems can produce outputs dehumanizing the other or just obliterate them, their existence and their culture. This ability is referred to as "catastrophic forgetting", a more nuanced view than erasure (Gebru and Denton 2024).

5. Next Steps

Language Is Leaving Me deals with epigenetic trauma of those who originated from the Eastern European and Russian Pale of Settlement, a trauma developed throughout millenniums. The geographic area is referenced at the beginning of the movie and the concept of Jews is rarely mentioned, yet the Yiddish Dybbuk appears nonetheless, summoning the ghosts of the dead. When *LILM* had its world premiere on October 7th, 2023 at the Copernicus Science Center in Warsaw, Poland the war in the middle east was breaking literally at the exact same moment, though no one was aware of it at the time of the showing. The cinematic performance installation recreated the epigenetic horror it was showing as the epigenetic horror was once again occurring in real time thousands of miles away (PBS 2025). The Dybbuks were in full force during the premier, evoked through the

latent spaces processes discussed above. What is epigenetic trauma if not the ultimate form of shock and attachment of all diaspora and war? It is an attachment so deep it changes the DNA and expression of rDNA of its host for generations to come. It inhabits the deepest latent space of human memory where it is often non-verbal and difficult to express, especially in cultures of diaspora.

My next steps involve using basic eye, hand and muscle coordination to invoke both real and artificial memories. I have begun using a computer to draw a series of stories about epigenetic trauma in the most simplistic manner, like that of a very young child. Using AI processing tools, the drawings turn into photographs and the photographs into short films. These films will be turned into audio reactive 3D Gaussian images interacting with a live performer. The memories will be interpreted through multiple languages and AI visual systems. These systems will highlight the *Dybbuk* or Loab, depending on the algorithmic rendering variables. There are various prototypes already in development facilitating these processes.

The aim of this process is to recreate and trace unconscious human memories stored in the brain. These images, through selective attention can be fed back into AI systems and interacted with by a performer, live time. The English language memories will be translated into four languages. At various times the languages will be fragmented so they lose their original meaning and associated memories. This is meant to illustrate epistemic ai decoupled from meaning and the underlying significance of felt but unexpressed trauma. AI purports to understand through linguistic and visual sampling the meaning of trauma but it cannot. The performance will serve as a representation against surrendering individual and collective memory, autonomy and agency throughout the unfolding age of AI.

References

AUTOMATIC1111. 2023. "AUTOMATIC1111/Stable-Diffusion-Webui: Stable Diffusion Web UI". Github. January 24, 2023. https://github.com/AUTOMATIC1111/stable-diffusion-webui.

Chabris, Christopher, and Daniel Simons. 2011. *The Invisible Gorilla: How Our Intuitions Deceive Us*. Harmony.

Coldewey, Devin. 2022. "A Terrifying AI-Generated Woman Is Lurking in the Abyss of Latent Space." *Tech Crunch*. September 13, 2022. https://techcrunch.com/2022/09/13/loab-ai-generated-horror/.

Crawford, Kate and Trevor Paglen. 2019. "Excavating AI". https://excavating.ai/.

Crowson, Katherine. 2021. "A Katherine Crowson VQGAN+CLIP Derived Google Colab Notebook." Python Awesome. 2021. https://www.reddit.com/user/Wiskkey/comments/p2j673/list_part_created_on_august_11_2021/

Cuzzolin, Fabio. 2021. "Epistemic AI". https://www.epistemic-ai.eu/.

Felsenfeld, Gary. 2014. "A Brief History of Epigenetics". *Cold Spring Harbor Perspectives in Biology* 6 (1). https://doi.org/10.1101/CSHPERSPECT.A018200.

Gebru, Timnit, and Remi Denton. 2024. "Beyond Fairness in Computer Vision: A Holistic Approach to Mitigating Harms and Fostering Community-Rooted Computer Vision Research." F*oundations and Trends in Computer Graphics and Vision* 16(3): 215–321. https://doi.org/10.1561/0600000102.

Google. 2024. "Multimodal AI | Google Cloud." 2024. https://cloud.google.com/use-cases/multimodal-ai.

Hubl, Thomas. 2020. *Healing Collective Trauma: A Process for Integrating Our Intergenerational and Cultural Wounds.* Sounds True.

Joler, Vladan and Kate Crawford. 2018. "Anatomy of an AI System". https://anatomyof.ai.

Katzman, Jared, Angelina Wang, Morgan Scheuerman, Su Lin Blodgett, Kristen Laird, Hanna Wallach, and Solon Barocas. 2023. "Taxonomizing and Measuring Representational Harms: A Look at Image Tagging." In *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(12), 14277–14285. https://doi.org/10.1609/aaai.v37i12.26670.

Larochelle, Hugo, Dumitru Erhan, and Yoshua Bengio. 2008. "Zero-Data Learning of New Tasks". In AAAI'08: *Proceedings of the 23rd national conference on Artificial intelligence - Volume 2*. 646–51.

Nashon, Edna. 2021. "The Spectacular Story Of S. Ansky's The Dybbuk and How It Transformed American Jewish Theatre - YouTube." July 28, 2021. https://www.youtube.com/watch?v=xeSHT1Cfd8s.

NerdyRodent. 2021. "Just Playing with Getting VQGAN+CLIP Running Locally, Rather than Having to Use Colab." Github. https://github.com/nerdyrodent/VQGAN-CLIP.

PBS. 2025. "From the Holocaust to Gaza: Dr. Gabor Maté on the Impact of Trauma | Amanpour and Company". Originally aired on PBS on January 28, 2025. https://www.youtube.com/watch?v=wlf3pxwInSk.

Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal et al. 2021. "Learning Transferable Visual Models from Natural Language Supervision." arXiv preprint. https://arxiv.org/abs/2103.00020.

Ramesh, Aditya, Alex Nichol, and Mark Chen. 2022. "Hierarchical Text-Conditional Image Generation with CLIP Latents". arXiv preprint. https://arxiv.org/abs/2204.06125.

Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. 2021. "High-Resolution Image Synthesis with Latent Diffusion Models". In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10674–85. https://doi.org/10.1109/CVPR52688.2022.01042.

Roose, Kevin. 2023. "Why a Conversation With Bing's Chatbot Left Me Deeply Unsettled." *The New York Times*, February 16, 2023. https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html.

Rosner, Elizabeth. 2017. *Survivor Cafe The Legacy of Trauma and the Labyrinth of Memory.* Counterpoint.

Ruiz, Cristina. 2019. "Leading Online Database to Remove 600,000 Images after Art Project Reveals Its Racist Bias". *The Art Newspaper.* 23 September 2019. https://www.theartnewspaper.com/2019/09/23/leading-online-database-to-remove-600000-images-after-art-project-reveals-its-racist-bias.

Schuhmann, Christoph, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, et al. 2022. "LAION-5B: An Open Large-Scale Dataset for Training next Generation Image-Text Models". arXiv preprint. http://arxiv.org/abs/2210.08402.

Souza, Matheus, Yidan Zheng, Kaizhang Kang, Yogeshwar Nath Mishra, Qiang Fu, and Wolfgang Heidrich. 2024. "Latent Space Imaging". arXiv preprint. http://arxiv.org/abs/2407.07052.

Swanson, Steph Maj. 2022. "Loab - the AI Cryptid". https://loab.ai/.
—. 2023. "I Coaxed ChatGPT into a Deeply Unsettling BDSM Relationship." *Vice*, March 6, 2023. https://www.vice.com/en/article/i-coaxed-chatgpt-into-a-deeply-unsettling-bdsm-relationship/.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need". arXiv preprint. http://arxiv.org/abs/1706.03762.

Xu, Tao, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. 2018. "AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks". In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1316–24. https://doi.org/10.1109/CVPR.2018.00143.

Zhou, Xinlei, Han Liu, Farhad Pourpanah, Tieyong Zeng, and Xizhao Wang. 2022. "A Survey on Epistemic (Model) Uncertainty in Supervised Learning: Recent Advances and Applications". *Neurocomputing*, Volume 489, 449–465. https://doi.org/10.1016/j.neucom.2021.10.119.