California Institute of the Arts

Generative AI in contexts of Aesthetics to Symbiotic Multimedia Art Ecosystem

by

Junson Park

A thesis submitted in partial fulfillment for
the degree of Master of Fine Arts

Herb Alpert School of Music
Music Technology: Interaction, Intelligence & Design
2025

Abstract

Generative AI is rapidly both changing and shifting aesthetic approaches to various art forms such as visual and sonic art including installations and performances. However, the speed of that the new technical achievements are introduced is being accelerated, so that the artists who have been using digital medium are likely to miss enough time to adapt their artistic decision process and workflows even if they want to move forward to the new realm. Therefore, the most backlashes to 'AI art' are caused by lack of understanding artistic contexts using technical aspects and low qualities by massive product style image dumps on social media platforms. This research will explore the usages of machine learning to bridge the gap between human and machine in terms of perceptions and expression, examining the integration of human creativity and how AI assists to overcome the accusatory technical and moral perspective viewing artworks.

Acknowledgments

I would like to express my deepest gratitude to my advisor, Dr. Ajay Kapur, for his unwavering support, guidance, and fierce encouragement throughout the development of this research. His insights have shaped not only the technical direction but also the artistic integrity of this work.

This project was made possible through the generous support of the Responsible Artificial Intelligence (RAI) UK International Partnerships, funded by UKRI EPSRC (grant reference EP/Y009800/1). I extend my sincere appreciation to the RAI Music Artistic Mini-Projects team for providing both the resources and support that challenged and refined the conceptual foundation of UTMORI.

I am indebted to the Korean National Gugak Center for making their traditional Korean music recordings openly available. Their preservation of cultural heritage enabled me to engage with Korean music in a deeply personal yet future-facing manner.

Special thanks to the team behind the RAVE model and to the IRCAM Forum community, especially for developing tools such as the nn~ object, which empowered the real-time sound manipulation core to this work.

I would also like to acknowledge my peers, friends who offered candid critique, encouragement, and companionship during the inevitable moments of disorientation and doubt especially those who reminded me to listen deeply, not just to machines, but to myself.

Lastly, I thank my family for their unconditional support, even when my explanations of "latent space" were incomprehensible at the video calls. This thesis is for all of you.

Contents

List of Figures

Chapter 1

Introduction

Generative Artificial Intelligence (AI) refers to a class of computational techniques capable of autonomously producing novel content – from images and music to text – by learning patterns from data.[1] At its core, generative AI leverages artificial neural networks, which are multi-layered computational models inspired by the human brain's structure, to encode complex representations of data. Key architectures in this domain include Generative Adversarial Networks (GANs) and diffusion models, among others. GANs (introduced by Ian Goodfellow et al. in 2014) consist of two neural networks in competition – a generator that creates synthetic data, and a discriminator that evaluates authenticity – driving each other to improve until the generator produces convincingly real outputs.[1] Diffusion models, a more recent innovation, generate imagery through an iterative process of noise addition and removal: they add random noise to data in a "forward" pass and then learn to reverse this process, gradually denoising to synthesize new, high-fidelity results.[1] Central to these approaches is the concept of latent representations. A latent space is an abstract, high-dimensional feature space learned by neural networks; data (e.g. images or audio) can be encoded as latent vectors capturing salient features, and new variations can be generated by decoding perturbed or interpolated latent vectors.[1] In summary, generative AI systems use advanced machine learning – particularly deep neural networks – to model the statistical structure of large datasets and to generate new content that is like, yet distinct from, the training examples.

1.1Awareness in the art field

The rapid proliferation of AI-generated art in recent years has provoked mixed reactions in the art world, oscillating between excitement and intense skepticism. Enthusiasts argue that generative AI is expanding the creative palette, offering artists unprecedented tools for exploring form, complexity, and interaction. New works employing AI have garnered significant public interest –

major museums have acquired AI artworks, and high-profile exhibitions celebrate the marriage of art and artificial intelligence. At the same time, a strong backlash has emerged from many artists, critics, and ethicists, raising critical questions about authorship, ethics, and the ontology of machine-made art. One core controversy center on authorship and originality: when an artwork is created with AI, who should be credited as the author – the human who conceived or prompted it, or the algorithm itself? Some argue that the human user, whether programmer or prompt-giver, is the true author, analogous to a photographer using a camera as a tool.[2] Others worry that the role of the human artist becomes problematically diminished, as the AI's generative autonomy muddies the notion of intentional, human expression. This ties into deeper ontological concerns about creativity: "can a machine learning model genuinely be creative", or is it merely remixing and regurgitating prior human creations? Art critics have voiced skepticism on this front – for example, Jerry Saltz derided many AI artworks as "poor-quality art" or "pretty crapola illustration," lacking the genuine imagination and conceptual depth that come from human creative struggle.[3] Such critiques suggest that without human intent or emotion guiding the process, the outputs of generative models may be seen as soulless pastiche, raising the philosophical question of whether creativity requires consciousness or agency.

Another major axis of debate involves the ethical use of data in AI art. Contemporary generative models are typically trained on vast datasets of images (or sounds, texts, etc.), often scraped from the internet without explicit consent from the original creators. This has led to an outcry among artists and photographers who discovered their works repurposed as training fodder for AI. Concerns over copyright and consent have spurred petitions and even lawsuits, underlining that these AI systems often learn by ingesting copyrighted artworks. As one illustrator put it, "AI doesn't look at art and create its own. It samples everyone's [work] – then mashes it into something else", effectively using "the pirated intellectual property of countless working artists" as raw material.[4] The lack of transparency and permission in these datasets is seen as an ethical blind spot, prompting discussions about regulatory guidelines for AI in creative fields. In addition, observers have noted issues of bias and representation and even environmental concerns that large AI models require intensive computational resources.

Overall, the discourse around AI art is highly polarized. On one side are those who hail AI as a revolutionary medium that can augment human creativity and generate previously unimagined art forms; on the other side are critics who question the legitimacy and impact of AI-generated art, from practical matters of attribution and fairness to existential questions about what

it means for art to be created by a non-human entity. This thesis takes an academic, neutral stance in navigating these debates: rather than advocating for either uncritical enthusiasm or blanket dismissal, it acknowledges the valid concerns while also recognizing the innovative possibilities that AI offers. The following chapters aim to move beyond this polarization, investigating how artists can engage with AI thoughtfully – addressing its challenges – to produce meaningful art. Key controversies such as authorship, ethics, machine creativity will be revisited through scholarly lenses, ensuring a balanced understanding of the context in which contemporary AI art practices unfold.

1.2 Motivation and Focus

Considering the atmosphere surrounding AI and art, the motivation for this thesis is to transcend the often superficial pro-versus-contra rhetoric and explore "how generative AI can be constructively integrated into creative practice". The research is operated by a belief that focusing solely on abstract debates (e.g., "is AI art real art?") can overlook what happens in studios and labs where artists work with these tools. Instead of treating AI as an existential threat or a magical fix, this study approaches it as a medium with unique affordances and limitations to be critically examined and creatively exploited. By situating generative AI within an artist-led context, the thesis seeks to demonstrate that AI is not an alien agent but can be a meaningful extension of the artist's toolkit when used with intention and insight. There is a growing call in the field for artist-driven AI art, wherein human creators maintain conceptual and aesthetic agency while leveraging AI's generative capacities.[5] Embracing this approach, the research presented here emphasizes the role of the artist's vision: the AI algorithms are employed as instruments, guided by artistic strategy rather than replacing it. This standpoint allows us to move beyond polarized arguments and into a more nuanced inquiry: What can AI do for art, and how can artists do things with AI that were not possible before? The thesis posits that generative AI, when critically engaged, opens new avenues for conceptual exploration and sensory innovation. Through detailed case studies of artistic projects, the work illustrates practical ways to integrate AI into the creative process, highlighting reflections from an artist's perspective. Ultimately, the goal is to show that between the extremes of utopian enthusiasm and dystopian backlash lies a fertile middle ground: a space where artists experiment with AI as a legitimate medium, addressing its challenges head-on and harnessing its strengths to push the boundaries of creative expression.

1.3 Overview of Artistic Projects

To ground the discussion in practice, this thesis analyzes three arts that each integrate generative AI in different ways. These works serve as case studies demonstrating how AI techniques can be applied to achieve specific artistic aims, and they offer a basis for examining the practical and theoretical questions raised in the preceding sections.

1.3.1 Come Closer

Come Closer is a video installation that explores themes of sensory fragmentation and computational perceptual control. In this piece, generative AI is used to manipulate visuals in a way that disassembles and reconstitutes the human sensory experience. By algorithmically filtering and re-combining imagery and sound, Come Closer creates a disorienting yet compelling environment in which the machine actively modulates what the audience perceives. The artwork interrogates how an AI system can "decide" what we experience, controlling attention and perspective – effectively putting the viewer in a feedback loop with an artificial perceptual apparatus. In doing so, it raises questions about mediation and intimacy: as the title suggests, the piece invites viewers to "come closer" to an AI-driven sensorial entity and to experience firsthand the push-pull between human senses and machine interventions. This project demonstrates AI's capacity for humanized aesthetic decision-making, as well as the artistic potential of deliberately fragmenting sensory input to provoke new awareness of perception itself.

1.3.2 Utmori

Utmori is an interactive AI-driven music composition that reinterprets traditional Korean rhythms through real-time latent vector manipulation. *UTMORI* (a title referencing *Eotmori*, a Korean rhythmic pattern) merges cultural heritage with cutting-edge AI techniques. The project involves a custom-trained generative model that learns from Korean drumming patterns and instrumentation, and an interactive interface through which participants can guide the model's output. By adjusting latent parameters in real time – essentially steering the deep generative model's internal state – users co-create evolving music that fluidly blends folkloric rhythms with contemporary electronic textures.[6] The system maintains the distinctive timbres of Korean traditional music while allowing algorithmic improvisation, resulting in a genre-bending soundscape that is both authentic and novel. UTMORI is as much a research inquiry as an artwork: it examines how AI can learn the nuanced expressivity of a non-Western musical form and how

an artist can encode cultural specifics (such as instrument tones and rhythmic accents) into a latent space. The interactive design underscores the thesis's theme of artist-led integration – here, the artist and audience actively shape the AI's creative output, rather than passively consuming a pre-generated piece. This case study highlights the potential of AI to act as a bridge between tradition and innovation, showing that generative models can be tuned to respect and reimagine a cultural aesthetic under the guidance of artistic intent.

1.3.3 Communication Breakdown

Communication Breakdown is a multimedia project employing AI algorithms to systematically disrupt processes of perception and meaning across visual, auditory, and linguistic domains. This ambitious work consists of interlinked components: images that have been distorted by AI image models, audio tracks deconstructed by neural networks, and textual elements generated or altered by language models. Each modality on its own presents a fragmented narrative – for instance, photographs whose subjects are ambiguously morphed, or text words that have been recomposed into nonsensical yet poetic sequences. When experienced together, these elements create a kind of orchestrated chaos: a breakdown of normal communication signals orchestrated by AI. The guiding concept is to investigate what happens when AI deliberately mis-communicates. Rather than using generative AI to produce harmoniously realistic outcomes, Communication Breakdown inverts the premise, using similar technologies to introduce noise, loss of fidelity, and semantic slippage in a controlled fashion. This results in an artwork that is at once disorienting and thought-provoking, as the audience must navigate glitchy visuals, jarring sounds, and scrambled text to find their own meaning. By pushing AI to generate failures or distortions, the project comments on the fragility of interpretation in the age of machine-mediated media. It serves as a counterpoint to the notion that AI always strives for human-like perfection; here the artefacts of AI – the errors, mutations, and gaps – become the aesthetic focus. Communication Breakdown thereby exemplifies an experimental, critical use of generative AI, one that foregrounds the limits of machine perception and the active role of the viewer in making sense of an artwork.

1.4 Structure of the Thesis

This thesis is organized into six chapters. Following this Introduction, Chapter 2 provides a detailed background and literature review, tracing the evolution of generative art and AI in artistic contexts. In that chapter, key historical milestones (from early algorithmic art through the

development of GANs and diffusion models) are examined in depth, and relevant theories from media art and cognitive science are discussed to frame the conceptual landscape. Chapters 3, 4, and 5 each focus on one of the three artistic projects introduced above. In these chapters, the creation and exhibition of Come Closer, UTMORI, and Communication Breakdown are documented and analyzed. Each chapter delves into the specific motivations, technical implementations, and artistic outcomes of the work in question, as well as the challenges encountered in integrating AI into the creative process. Through these case studies, the thesis investigates how theoretical issues (such as those of authorship, cultural context, or interpretability of AI systems) play out in practice. Chapter 6 concludes the thesis by synthesizing the insights gained from the case studies and discussions. It reflects on how generative AI can be effectively harnessed as a medium for conceptual and sensory innovation, summarizing the contributions of the research to the field of media art. Additionally, the Conclusion addresses the broader implications for artists working with AI – offering thoughts on best practices, ethical considerations, and possible future directions for creative AI. By moving from context to practice, to critical reflection, the structure of the thesis is designed to build a comprehensive understanding of artistic human–AI integration, ultimately demonstrating how a nuanced, hands-on engagement with generative AI can enrich contemporary artistic practice while responding thoughtfully to the critiques and concerns surrounding this emergent medium.

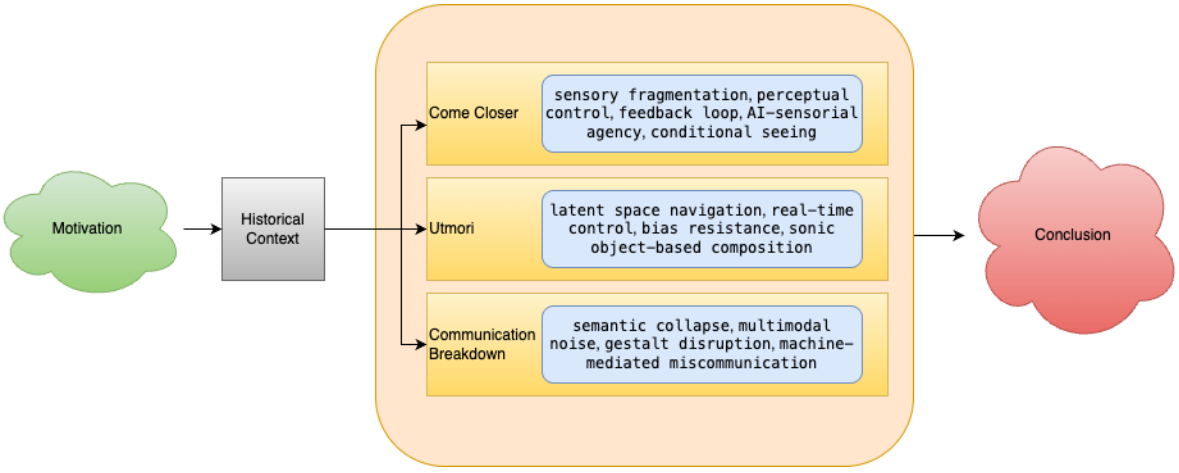

figure 1. Structure of this thesis

Chapter 2

AI in Artistic Flows

2.1 Early Experiments in Generative Art

Generative art as a concept date back to the 1960s, when pioneering artists wrote algorithms to create visual compositions. Artists like Michael Noll, Frieder Nake, and Vera Molnar used computers to produce abstract designs, setting the stage for later AI-driven art. However, these early works were algorithmic rather than intelligent that they followed fixed rules set by the programmer. The emergence of AARON in the 1970s marked a turning point toward true AI art making. Harold Cohen's AARON outputs included figurative scenes drawn autonomously by the program. Cohen's aim was to encode the act of drawing into an AI, effectively modeling human artistic cognition in software.[7] Visually, AARON's work evolved from abstract doodles to complex scenes with human figures and objects, often in bold flat colors and outlines. This early AI art explored the simulation of human creativity: AARON reflected how Cohen thought about creating images, encoding artistic decisions into an algorithm.[8]

figure 2. Harold Cohen, AARON KCAT, 2001[7]

2.2 Rise of Machine Learning Art

The mid 2010s deep learning revolution dramatically accelerated generative art. In 2015, Google's DeepDream[9] algorithm became a viral sensation by revealing how neural networks perceive images, transforming ordinary photos into surreal, hallucinatory visions.[10] Then, the first art exhibition and auction of neural network-generated art took place in 2016: Google Research and Gray Area Foundation hosted DeepDream: The Art of Neural Networks, where artworks created by AI were sold to benefit the arts.[11]

Generative Adversarial Networks (GANs)[12], introduced in 2014, further revolutionized AI art by enabling AI to invent realistic images from scratch. By the late 2010s, artist could train GANs on datasets and have the AI generate entirely new images that resembled the training examples yet were novel. This technique allowed AI to reconstruct human-like sensory outputs – such as faces or painting – in ways never seen before, often defying conventional aesthetic norms. AI art gained mainstream art world attention in 2018 when an AI-generated portrait, Edmond de Belamy, was auctioned at Christie's for $432,500, signaling that AI creation could hold value as art.

2.3 Key Artworks

2.3.1 Refik Anadol – Melting Memories

Refik Anadol is a media artist known for converting vast datasets into immersive audiovisual installations. He is a pioneer in the aesthetics of data and machine intelligence, using AI to visualize the "memories" of culture and nature.[13]

figure 3. Refik Anadol, Melting Memories, 2018[14]

A signature theme in Anadol's work is externalization of memory. In Melting Memories (2018), Anadol famously used EEG brain scan data from people recalling memories, and processed it through AI to create mesmerizing, ever-shifting patterns of light. The installation literally visualized the act of remembering, with golden fluid forms on a screen that "melt" and change, symbolizing how elusive and dynamic our memories are. Critically, Melting Memories "emphasizes just how fluid memories truly are[15]", blurring the line between a brain's activity and an artistic image. Here, AI is used to reconstruct cognition: the neural patterns of memory are translated into visual art, allowing viewers to see an aesthetic representation of an inner mental process.

In art history context, Anadol's work builds on the tradition of visual music and abstract expression, but he introduces a computational sublime: viewers are confronted with the beauty of immense datasets and complex algorithms, which they experience as hypnotic visuals. Philosophically, his installations invite reflection on reality and simulation: if machine can produce perceptual worlds from data, it challenges us to think again what real perception is.[16]

2.3.2 Mario Klingemann – Memories of Passerby I

With background in programming, Klingemann has focused on the intersection of AI and human creativity, often interrogating how machines see and reproduce the human form. Klingemann is "particularly interested in human perception of art and creativity," researching ways machines can emulate or augment these processes. [17] One of his best-known works, Memories of Passersby I (2018), exemplifies how AI can deconstruct and reassemble human imagery – in this case, the human face – to unsettling effect.

figure 4. Mario Kingemann, Memories of Passerby I, 2019[18]

Memories of Passersby I is an autonomous art installation dubbed a "self-contained creative agent". It consists of an AI "brain" encased in a mid-century style wooden console, connected to two portrait-oriented screens. Inside the console, a GAN (Generative Adversarial Network) continuously generates an infinite stream of portraits in real time. [18] The result is a perpetual slideshow of faces that appear strangely familiar yet eerily distorted. No image ever repeats, because the system feeds its own output back into the generative process, engaging in a feedback loop of creation. The faces emerge, morph, and dissolve, sometimes "melting into abstract arrangements of pixels as the machine struggles to create a new portrait."[18] This can be a disquieting experience: the features may be oddly combined (a face with two offset eyes, or half one person and half another), giving a glimpse into the AI's imperfect "imagination."

Aesthetically, Memories of Passersby I produce what Klingemann calls the "Francis Bacon effect"[19]– many portraits have a surreal, anxious quality reminiscent of Bacon's warped faces. The work has been described as presenting "uncanny interpretations of the human face" – an AI-driven take on what Surrealist André Breton termed "convulsive beauty".[18] At times the viewer sees a convincing painted visage; at other times the image breaks into an almost abstract haze of brushstrokes and digital artifacts, essentially deconstructing the familiar image of a face into pure color and form. This oscillation challenges the audience's perception: we are prompted to consider how our brains so readily see a face in even vague suggestions (a phenomenon known as pareidolia), and how the AI both leverages and defies that expectation. By witnessing an endless parade of AI-made "people," one might also reflect on identity – these faces have never existed, yet they trigger recognition. The piece thus probes the cognitive processes behind face perception and memory.

2.3.3 Ada Ada Ada – Being represented by data is like losing a part of yourself

A queer, feminist exploration of glitch as a form of resistance. The artist known as Ada Ada Ada (a transgender new-media artist) uses her own body and voice as training data for AI models, then "actively misuses" the technology to produce aberrant results[20]. In this video piece, a StyleGAN2 model was trained on hundreds of self-portraits taken during the artist's gender transition (the "in transitu" project, 2021–2023), but deliberately under-trained and driven with unusual parameter settings to yield "queer, glitchy expressions" [19]. The visuals are a kaleidoscope of shifting, sometimes grotesque forms – headless or multi-mouthed figures and text fragments (see image frames in the project documentation) – that reject binary stability [19]. By exploiting

the AI's propensity to glitch, the work finds a "queer mode of expression within the glitchiness of algorithmic spaces," turning technical failure into a statement of identity [19]. As the artist writes, "the model is deliberately undertrained … to produce glitchy expressions" a strategy that lays bare the data's inadequacy to capture a fluid self, and "a rebellion against the current paradigms of the binary" in digital systems [19].

figure 5. Ada Ada Ada, Being represented by data is like losing a part of yourself, 2024

This piece was featured in HYPERSPECTRAL and circulated online, contributing to discourse on AI art as a tool of marginalized expression. Ada's broader in transitu performance (posting weekly transition selfies with AI gender classifications) has been discussed in academic forums[21], underlining how personal data fed into AI can yield absurd misreading which Ada embraces as "glitch art as self-portrait."

2.3.4 Ian Cheng – BOB (Bag of Beliefs)

Cheng's BOB is an evolving virtual creature with its own neural network "brain," essentially an artwork that goes on living and changing independent of the artist. Debuted at the Serpentine Galleries (London) in 2018[22], BOB appears as a snake-like organism on a large screen. It reacts to stimuli (including inputs from gallery visitors via a companion app) and continuously adapts its behavior – growing new appendages, altering its color/patterns, and expressing what Cheng calls "beliefs" (goals it tries to achieve). The key theme is open-ended autonomy: once released, BOB's precise actions are not scripted, and even its creators cannot predict its next mutation. In Cheng's words, it's a "game that plays itself," inspired by systems like The Sims[23]. Viewers can witness BOB succeed or fail to satisfy its needs in real time – for example, it might desperately seek digital

"food" or attention, sometimes to absurd effect. This unpredictability introduces "ambiguity of relationship" between viewer and artwork: one critic noted that watching BOB felt uneasy because it's "soulless" yet alive, prompting reflection on whether we empathize with or feel alienated by an AI creature [24].

figure 6. Ian Cheng, BOB, 2018

After Serpentine, BOB showed at MoMA PS1 and other venues. Critical reception was mixed – some marveled at the concept of a self-evolving artwork, while others found it conceptually intriguing but emotionally flat, highlighting the challenge of connecting with a creation that deliberately escapes human control.

2.3.5 Emanuel Gollob – Doing Nothing with AI

Gollob's ongoing series "Doing Nothing with AI" uses a machine learning feedback loop to choreograph a hypnotic robot performance aimed at lulling the participant into a state of complete idleness[25][26]. In version 1.0 (2017–18), a robot arm adorned with 23,000 toothpicks slowly dances in front of a viewer, while an EEG headset measures the viewer's brain activity [25]. Behind the scenes, a generative adversarial network (GAN) controls the robot's motion patterns, mutating its movement "choreography" in response to the viewer's mental state [25]. very 30 seconds, the system evaluates: if the spectator's alpha/beta waves indicate they became more relaxed (closer to mental "nothingness"), the current set of motions is deemed successful and is slightly reinforced

for next time; if not, the AI tries a new, different movement sequence [25]. Over many iterations with many viewers, the robot "learns" an optimal, ever-evolving dance of doing nothing. The aesthetic is one of constant flux and subtle glitch – no two sessions yield the same motions, and the robot's movements can be strangely fluid at times or jarringly stuttering at others, as the GAN explores its vast space of 2^(255*256) possible moves [25]. Importantly, "none of the interactants controls the other": human and AI are in a non-hierarchical loop, each affecting the other. The project is a commentary on society's obsession with productivity; the "machine flaws" and random trials become a means to an unlikely end optimizing for idleness [25].

figure 7. Emanuel Gollob, Doing Nothing with AI, 2017-2019[27]

Gollob's work was featured in Ars Electronica's "ERROR – The Art of Imperfection" exhibition (2018)[28], fittingly highlighting how the robot's imperfect motions are the art. It also showed at the Vienna Biennale and multiple tech-art festivals. Audiences often found the piece meditative and amusing – the sight of a spiky robot attempting to coach your mind through algorithmic trial-and-error provokes questions about control: Is the machine guiding you or are you guiding it? The answer is deliberately ambiguous.

2.3.6 Poison: An Edvard Munch Experience - MUNCH Museum Audience Lab

Generative error as atmosphere. Created by the MUNCH museum's digital lab with artist/researcher Christiane Sivertsen, Poison plunges visitors into the unsettling world of Edvard Munch's 1907 painting series "The Green Room," but reimagined through an AI "hallucination." The installation surrounds the viewer with morphing imagery derived from Munch's works: the AI was trained on the Green Room paintings but purposefully "stopping the training process early" became the technique to inject uncertainty[29]. The result is that faces, figures, and the titular green walls appear and dissolve, never fully coalescing into stable forms – as if the machine

is trying to paint Munch's room but continually forgets or confuses the details. The museum describes the space as "unstable and unsettling. A mysterious being produces a never-ceasing maelstrom of hypnotic imagery" around the viewers [30]. This being is essentially the AI model churning out partial reconstructions of Munch's scenes – an autonomous agent whose "errors" (the glitchy, half-formed images) are embraced to evoke "unreliable stories and shifting perspectives" [28]. The title *Poison* references how Munch's green paint likely contained arsenic, slowly poisoning inhabitants – likewise the AI visuals have a creeping, toxic quality, never allowing the comfort of certainty. Visitors walking through the space trigger changes via motion sensors but cannot control the narrative; they are at the mercy of the machine's fractured storytelling [28].

figure 8. MUNCH Museum Audience Lab, Poison, 2021[30]

Poison ran in the brand-new MUNCH museum [28]. It was an experimental fusion of art, technology and curatorial interpretation. In the context of this report, Poison is notable for turning a known art historical work into a living, error-prone algorithmic experience – emphasizing that even in failure to faithfully recreate Munch, the AI can expose deeper themes of jealousy, dread, and ambiguity present in the originals.

2.3.7 Trevor Paglen & Kate Crawford – ImageNet Roulette

This interactive web app and installation is perhaps the most high-profile example of an artwork that weaponized an AI's errors to make a point. ImageNet Roulette was a simple app: upload a photo of a person, and it returns the labels that an AI (trained on the ImageNet dataset) assigns to that face [29]. These labels were often unexpected, funny, or outright offensive – not due to

the app malfunctioning, but because the underlying dataset contained problematic, biased categories. The artists explicitly framed it as a commentary on "the problematic aspects of classification systems". When users saw their own selfies come back with bizarre tags (e.g. a Black user labeled "wrongdoer" or an Asian user labeled "fine-looking person" – real examples from the project's online viral moment), they experienced first-hand the uneasy role of being judged by a flawed AI [27]. This ambiguity of relationship – are we the viewer, the subject, or the victim of the AI? – was exactly Paglen's intent.

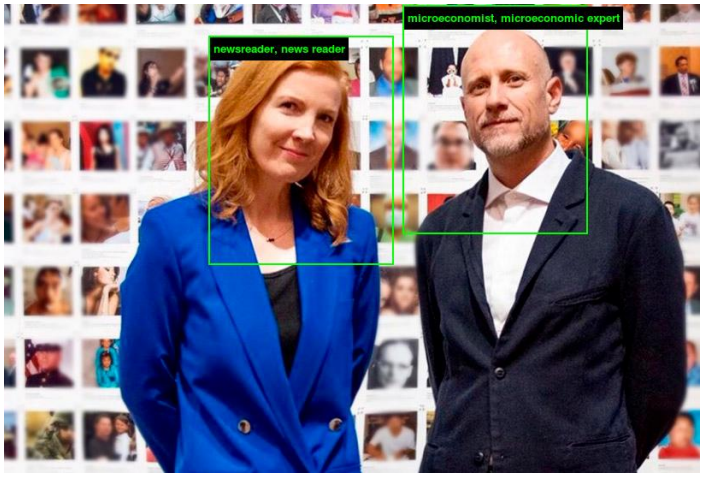

figure 9. Trevor Paglen & Kate Crawford, ImageNet Roulette, 2019[31]

As an installation at the Training Humans exhibition (Milan, 2019), it drew huge public engagement and media coverage. Many were shocked to learn that the AI wasn't "broken" – it was reflecting real data biases. Paglen/Crawford soon shut down the app, once its point was made, to avoid harm. Critically, ImageNet Roulette is cited as a key artwork that "exposes the functioning of ML systems by showing their direct application on concrete persons" [27]. In other words, it made an AI's errors and hidden prejudices visible as the content of the art. This project spurred broad discourse on AI ethics; it's a case where the glitch (in societal values encoded in data) is the message.

2.3.8 Dadabots – Relentless Doppelgänger

Endless, autonomous music generation that revels in the "imperfections" of its own output. Dadabots (a duo of musicians-turned-researchers, CJ Carr and Zack Zukowski) created a neural network to mimic technical death metal – an extreme subgenre characterized by rapid drums, guitars, and guttural vocals. In 2019 they launched Relentless Doppelgänger, a YouTube channel live-streaming AI-generated death metal that plays continuously, with no human curation. The

interesting part is how the AI sounds: initially, many attempts were unstable and surreal, producing what the creators called "music soup" as the songs "would destabilize and fall apart" [32][33]. But rather than discard these failures, Carr and Zukowski found artistic merit in them. "We were delighted by the aesthetic merit of its imperfections," they wrote [30]. For example, the AI's attempts at vocals turned into "a lush choir of ghostly voices," and crossbreeds of multiple metal bands became "a surrealist chimera of sound." [30].

These vivid descriptions show how glitches in the audio (garbled lyrics, missing drumbeats, inhuman speeds) were embraced as a new kind of extreme music that even humans hadn't made. Eventually, one model trained on the band Archspire achieved a more stable output that feels like real (if relentless) death metal, and that is what streams to this day. But even this "successful" model retains quirks: vocals that never pause for breath, and guitars that sometimes execute superhumanly fast riffs [30], sonic markers of the algorithm.

Dadabots have positioned their project as both a celebration and an experiment – even claiming they're "working towards eliminating humans from black metal" (tongue-in-cheek)[30]. In doing so, they highlight questions of authorship and the creative value of machine-made error: if a song collapses into noise but we find it compelling, is it a failure or a new form of art?

2.3.9 Holly Herndon – Spawn (PROTO album)

Human-AI co-creation with an emphasis on unpredictable, emergent results. Composer-vocalist Holly Herndon developed an AI program named "Spawn" – essentially a neural network trained on voices including Herndon's and her ensemble's. During the making of PROTO, Herndon treated Spawn as a young collaborator: it was not given full autonomy to generate entire pieces, but rather was invited to improvise along with human singers in recording sessions. The fascinating aspect is how Spawn's contributions manifest. In tracks like "Godmother," Spawn listened to input from guest artist Jlin (an electronic producer) and responded by outputting vocalizations that no human could replicate – rapid-fire stutters, microtonal shifts – sounding like an alien club singer. These glitchy vocals were kept in the final music. Herndon valued moments when Spawn "misheard" or "mis-sang" what the humans did, because those mistakes produced new ideas and textures. The concept was to give up a degree of control and allow failure to shape the composition. In concert, Herndon even staged call-and-response pieces where her ensemble would make sounds and Spawn would answer with bizarre modulations, effectively performing its machine agency live.

PROTO debuted as a live AV show and the album was critically acclaimed. While Herndon's approach is more guided than other examples (it's a curated use of AI outputs), it still foregrounds unpredictability. Critics noted that the album felt like a dialogue between human and algorithm, sometimes harmonious, sometimes tense. It situates generative error in a musical context: the AI's "voice" cracks in strange ways, but those cracks carry emotional and aesthetic weight in the composition.

2.4 From Algorithmic Glitch to Perceptual Structure

While many AI-based artworks produced between 2018 and 2025 foreground machine unpredictability, failure, or autonomy as aesthetic strategies, the author's practice diverges from this trajectory. Artists such as Mario Klingemann and Trevor Paglen have embraced generative error or systemic bias as the very subject of their work, positioning AI as a quasi-autonomous entity whose misjudgments or glitches generate artistic meaning. These works often highlight the loss of human control and the emergence of machine agency as creative catalysts.

In contrast, the author's approach treats AI not as a self-expressive agent but as a perceptual infrastructure—a system through which the conditions of sensory and cognitive experience are structured and modulated. Rather than privileging randomness or malfunction, the works deploy AI in a tightly controlled manner to investigate how perception, rhythm, and meaning can be reconfigured. This marks a shift from representational outcomes to the design of perceptual conditions.

The projects examined in this thesis—Come Closer, Utmori, and Communication Breakdown—do not celebrate glitch for its own sake. Instead, they embed the possibility of disruption within systems designed to expose or reconstruct perceptual thresholds. The AI is not tasked with producing surprising outputs, but with intervening in processes of bodily recognition, cultural memory, or semantic stability through structural modulation.

This orientation suggests an alternative lineage of AI-based art: one in which machine learning operates less as a generative partner and more as a systemic tool for reorienting human experience. By relocating AI from the level of content production to that of perceptual design, the author's methodology underscores a critical and intentional engagement with AI's affordances—using them to reframe not what art looks or sounds like, but how it is perceived and understood.

Chapter 3

Come Closer

Generative AI has emerged as a pivotal tool in contemporary media art, not as a mechanism for image creation but as a system that reshapes human sensory perception. Come Closer explores this transformation by employing AI-driven distortions of the human body, investigating the interplay between proximity, abstraction, and fragmented perception. This study examines how AI-generated modifications of bodily forms parallel human cognitive mechanisms, particularly in close-range sensory experiences where form dissolves into texture and impression. By selectively revealing and obscuring details through real-time computational filtering, Come Closer highlights the symbiotic relationships between AI's generative logic and human sensory cognition.

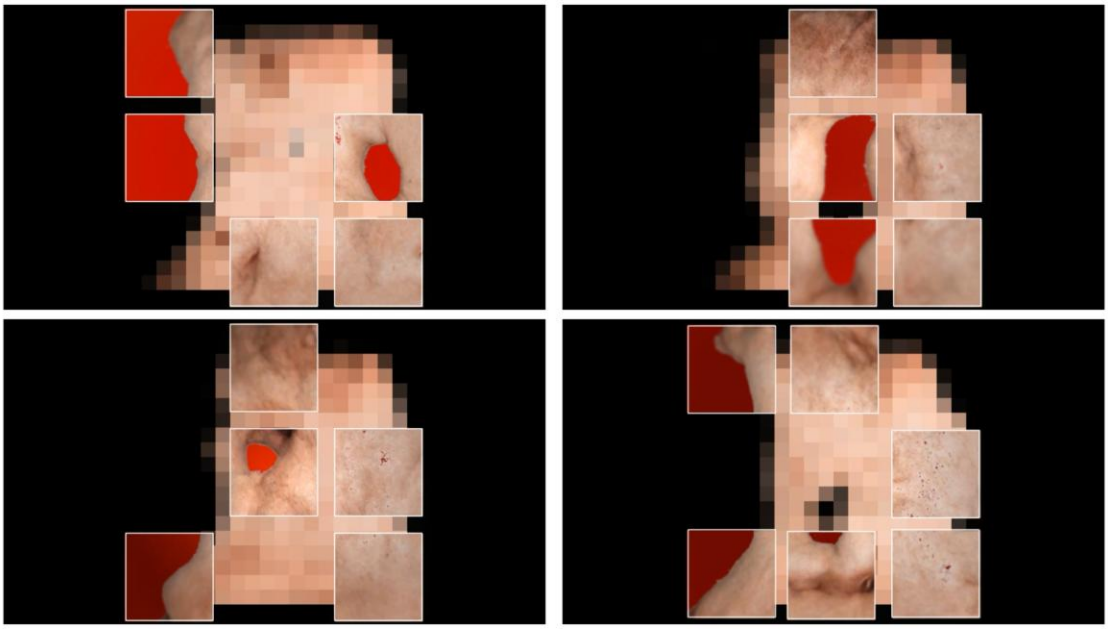

figure 10 Still images of Come Closer (2024)

3.1 Sensory Cognition and Fragmented Perception

Human sensory perception does not operate as a direct recording mechanism; rather, it is an active system that selectively filters, reconstructs, and infers missing information. Visual perception, for example, is inherently incomplete foveal fixation that allows for only a limited number of details to be processed at any given moment, with the brain compensating for gaps through prior knowledge and pattern recognition. This principle extends to bodily recognition, where perception is shaped by fragmentation, abstraction, and selective emphasis.

In close physical proximity, tactile perception further alters cognitive processing, shifting emphasis from form to texture. When encountering another body at intimate distances, details become less about structured outlines and more about surface impressions. The brain abstracts these impressions, exaggerating specific features while discarding extraneous details. Come Closer engages with these perceptual mechanisms by introducing AI-generated transformations that selectively expose and conceal bodily fragments, prompting the viewer to reconstruct the missing details.

3.2 AI as a Dual Mechanism: Simulation vs Reconfiguration

The AI models employed in Come Closer do not merely manipulate images; they function within two distinct paradigms: AI as a perceptual simulator and AI as a mechanism for perceptual reconfiguration.

First, AI operates as a perceptual simulator by mimicking how human cognition reconstructs incomplete visual data. SDXL[34] processes bodily imagery by learning patterns from datasets and generating missing details, aligning with the cognitive process of extrapolating unseen information from partial input. The IP Adapter[35] further intensifies this effect by enhancing certain bodily textures by specific image inputs, representing and amplifying specific details much like how the human brain selectively prioritizes certain sensory stimuli over others.

However, AI in Come Closer does not simply assist human cognition but actively alters the conditions under which perception takes place. Through real-time computational filtering, AI determines what is seen, how it is seen, and under what constraints bodily recognition is constructed.

3.3 AI-Driven Sensory Filtering and Dynamic Perceptual Control

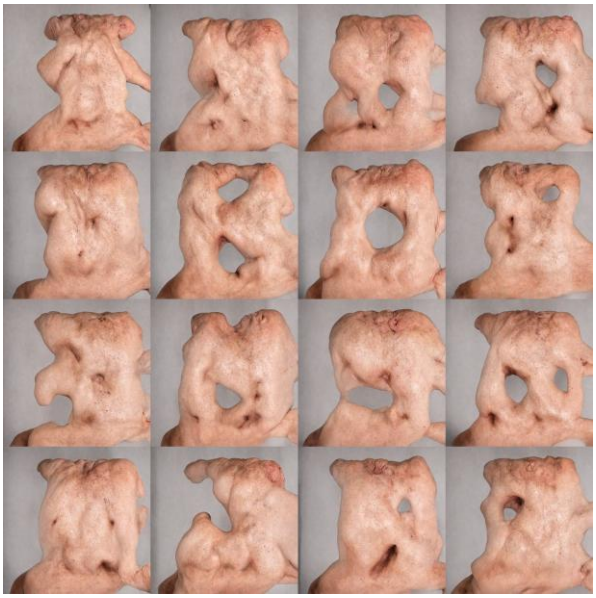

figure 11. Initially generated images from SDXL and IP Adapter

The production of Come Closer involved a multi-stage computational process integrating AI-based image generation, selective masking, and real-time manipulation.

The initial phase of image generation utilized SDXL to create distorted representations of human skin and flesh. This workflow was utilized images that emphasize organic textures, enabling the system to fabricate convincing yet exaggerated bodily impressions. The application of IP Adapter during this phase controlled the degree of distortion, allowing for specific textures to be emphasized while others were suppressed. The output images thus simulated the cognitive process of selective focus, wherein certain sensory impressions are amplified while peripheral details fade.

Once the distorted images were generated, they were processed through TouchDesigner, a real-time visual programming environment. This software facilitated dynamic alternation between pre-rendered AI-generated textures and live computational filtering. A mosaic effect was applied as an additional layer of perceptual interference, mimicking the way sensory input is naturally incomplete. Unlike traditional visual obfuscation techniques, the algorithmic control of the mosaic filter in Come Closer is not arbitrary; rather, it strategically controls the fragmentation and reconstruction of bodily perception, determining which visual elements are granted coherence and which remain disjointed.

This real-time computational mediation transforms the role of AI from an image-generating tool into a mechanism of perceptual control. The audience does not simply "see" a distorted image; rather, their sensory access is actively dictated by the computational system, reinforcing the idea that perception itself is structured by external conditions rather than being an autonomous cognitive process.

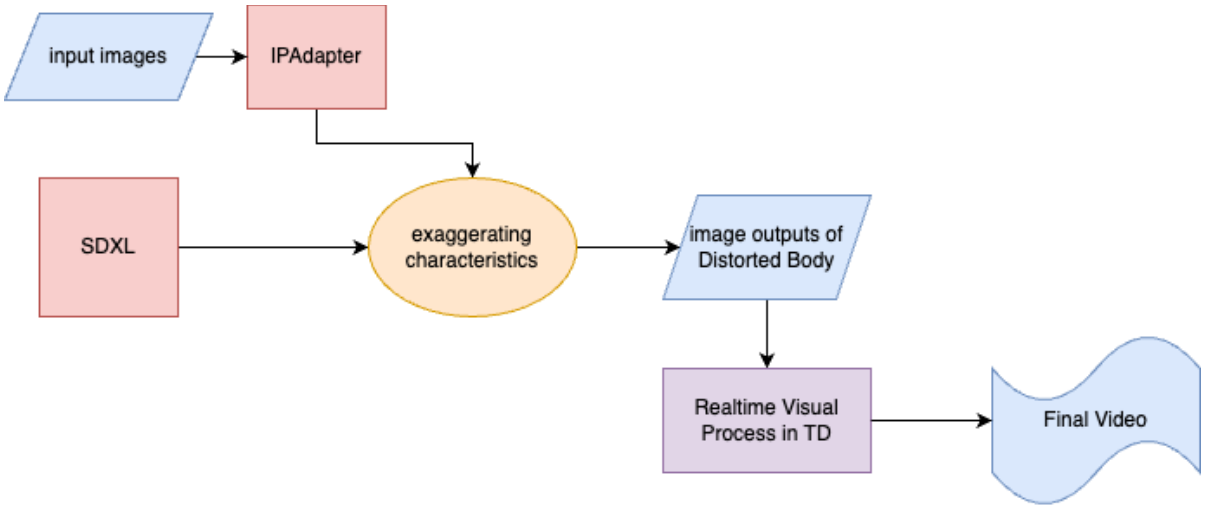

figure 12. Practical workflow of Come Closer

3.4 Artistic Context and the Evolution of Bodily Fragmentation

The processes in Come Closer align with historical artistic approaches that explore bodily distortion and fragmentation. However, while past artistic practices often relied on static representations to challenge bodily perception, Come Closer introduces computational mediation as an active force that governs perceptual reconstruction. By selectively revealing and concealing bodily elements, the work engages in a dialogue with past and contemporary artists who have deconstructed the human form to challenge perception, identity, and embodiment.

3.4.1 Fragmentation and Bodily Reconstruction

Artists have long experimented with bodily fragmentation as a means of altering perception and challenging representational norms. The selective concealment and exposure of body parts in Come Closer resonates with this tradition by emphasizing the cognitive process of visual reconstruction—a process that has been central to the works of Egon Schiele and Francis Bacon.

Schiele's fragmented and contorted representations of the human body prompt the audience to mentally complete the missing parts. His expressive distortions heighten psychological intensity, paralleling how Come Closer invites the viewer to reconstruct obscured bodily forms through perceptual inference. However, Schiele's approach remains confined to static figuration, whereas Come Closer introduces real-time computational concealment, where bodily forms are not merely incomplete but dynamically fragmented by algorithmic intervention.

Francis Bacon's paintings feature distorted, deformed figures that exist in a state of flux between recognition and abstraction. His use of blurred and stretched bodily forms evokes an uncanny yet visceral reality, akin to the AI-generated distortions in Come Closer, where familiar textures are exaggerated into unnatural yet strangely recognizable forms. However, while Bacon's distortions exist as fixed compositions, the AI-driven distortions in Come Closer continuously evolve, modifying the terms under which bodily recognition occurs.

3.4.2 A Mechanism for perceptual Reconfiguration

Unlike historical bodily distortion techniques that depend on static interventions, Come Closer demonstrates how AI can alter the perceptual framework itself rather than simply producing distorted images. The act of seeing becomes inseparable from the computational logic that structures it, shifting AI's role from a representational medium to a perceptual infrastructure that dictates the terms of sensory experience.

By positioning AI as an active agent in sensory mediation, Come Closer challenges the assumption that perception is an organic and autonomous process. Instead, it exposes the extent to which perception is always structured by external systems—whether biological, cultural, or computational. AI in this work does not just generate bodily forms; it redesigns the conditions under which bodily recognition is constructed and experienced.

Chapter 4

Utmori

UTMORI addresses the limitations of existing AI music generation models, which predominantly rely on Western-centric datasets and often fail to authentically represent non-Western traditions such as Korean traditional music. Instead of passively replicating existing musical forms, UTMORI employs a novel generative framework that allows real-time interaction with machine learning-generated elements. By utilizing the RAVE[36] model, the system separates percussive and melodic elements into distinct latent spaces, enabling greater control over rhythm, timbre, and form. This approach circumvents the stylistic biases of mainstream AI music models, offering a more fluid and adaptive method for interpreting traditional Korean rhythmic cycles like eotmori jangdan while preserving the expressive nuances of the original instruments.

Inspired by Pierre Schaeffer's musique concrète, where recorded sounds were treated as compositional materials, UTMORI extends this concept to the realm of AI-driven music composition. The project redefines latent space as a dynamic musical canvas, where manipulated representations function as Sonic Objects that can be structured and recombined within a performance. This methodology not only challenges conventional notions of AI-assisted composition but also expands the aesthetic possibilities of generative AI as a tool for real-time musical storytelling. By integrating machine learning into an interactive performance system, UTMORI shifts the role of AI from a static content generator to an active participant in the creative process, bridging traditional musical heritage with contemporary experimental composition. [1]

[1] https://vimeo.com/1059744514/

4.1 Data Collection and Preprocessing

For the composition UTMORI, which was created using this system, the dataset was sourced from the National Gugak Center's publicly available traditional Korean music samples. This dataset includes recordings of various Korean musical genres such as Minyo, Sanjo, and Samulnori. Instead of training a single model for all instrumental sources, the dataset was split into two separate models: one dedicated to percussion instruments and another for melodic instruments. This separation allowed for greater flexibility in composition and arrangement.

4.2 Model Training

The project employed the RAVE (Realtime Audio Variational autoEncoder) model, chosen for its real-time generative capabilities and its ability to learn latent representations of timbre. The two separate percussion and wind-string models were trained independently using a dataset split by instrument type. Throughout the system's training, periodic evaluations were conducted using loss function analysis and perceptual listening tests to assess the model's ability to replicate the distinct characteristics of traditional Korean instruments. The trained models were then integrated into the Max/MSP environment using the nn~ object for real-time interaction.

4.3 Realtime Control and Composition

One of the core objectives of this system was to develop a flexible and interactive composition system. Instead of relying on a fully autonomous generative model, the trained RAVE models were embedded into a Max/MSP patch, allowing for real-time manipulation of generated audio. The nn~ object[37] facilitated dynamic control over the latent space, enabling nuanced adjustments to generated timbres.

The Max/MSP patch was designed to allow composers to shape the generative process interactively. It incorporated a custom sequencer capable of modulating rhythmic structures based on the traditional Eotmori Jangdan, a mapping interface that allowed dynamic exploration of different regions within the latent space to create evolving timbral shifts, and real-time modulation of generation parameters such as attack, sustain, and spectral morphing to achieve a balance between control and generative unpredictability.

4.4 Signal Flow and Sound Generation

Unlike conventional implementations of RAVE, this system directly controls over the decoder's latent dimensions. Instead of encoding audio input, shaped noise is directly fed into the latent dimension inputs to achieve musical control.

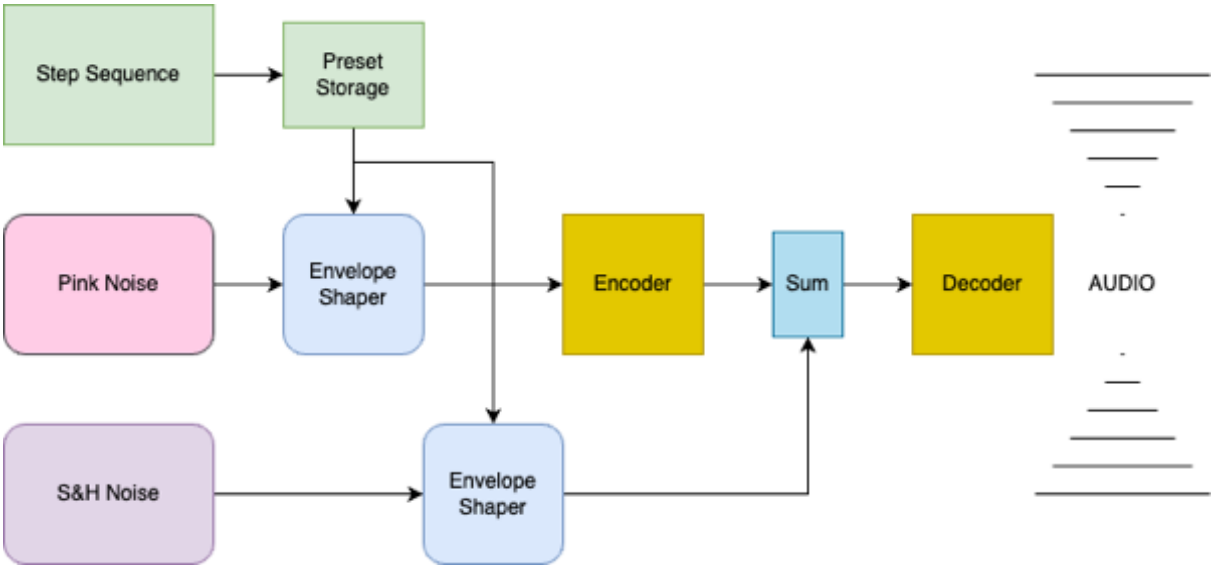

Figure 13. Signal routing structure overview

4.4.1 Direct Manipulation of Latent Dimensions

Instead of processing encoded representations, UTMORI uses direct input signals to control the latent dimensions of the decoder. These signals are generated through randomized signals processed via Sample & Hold to mimic encoder outputs, and multiple envelope generators applied to each latent dimension to create structured variation.

While most synthesis methods where parameter changes result in distinct timbral shifts along predefined dimensions, *UTMORI* found that a single latent dimension manipulation was insufficient for generating rich timbres. Instead, the system explores combinations of multiple latent dimensions, treating each combination as a unique timbral preset. These combinations are then triggered by a sequencer, ensuring that each trigger activates not a single latent dimension but a predefined combination, and that timbral evolution occurs through multi-dimensional shifts rather than single-axis modifications.

Envelope length modulation affects how sound evolves over time. Short envelope durations emphasize percussive, transient heavy articulations, while longer envelope duration does not merely sustain a single note but instead lead to the generation of phrase-like structures based on the trained dataset. Envelope amplitude modulation allows for dynamics intensity control, enabling performers to shape volume and articulation expressively in real-time. [2]

The impact of envelope length on the generated output is particularly notable in the percussion model. Short envelopes produce sharp and isolated drum hits, while longer envelopes reveal embedded rhythmic sequences derived from the training set, this means that envelope duration in the system does not just control sustain but dictates the retrieval of time dependent information from the latent space. As a result, short envelope duration generates individual percussive strikes, whereas long envelope durations elicit rhythmically evolving sequences, effectively generating dynamic rhythmic phrases instead of static hits. [3]

From a compositional perspective, this allows for the real-time exploration of rhythmic patterns via conventional sequencing, performers can control rhythmic density and phrasing dynamically, enabling an improvisational and fluid approach to rhythm generation. This feature fundamentally differentiate composition of UTMORI from standard synthesis and sample-based rhythm programming.

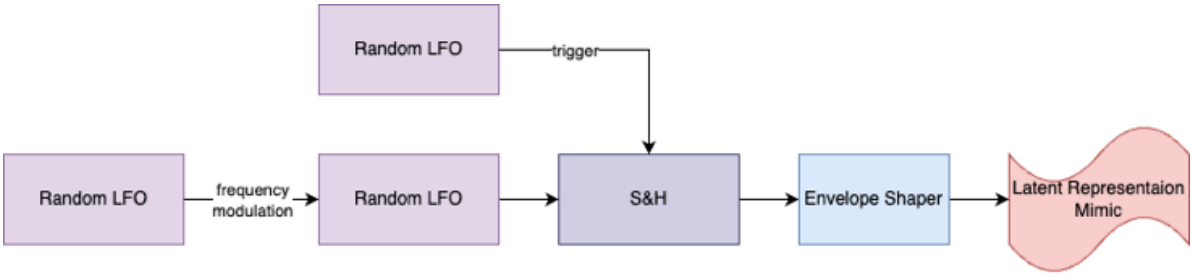

Figure 14. Mimicking latent representation with noise signal routing

4.4.2 Shaping Pink Noise for Timbre Control

The input to the encoder is not an audio sample but pink noise, chosen for its natural spectral balance that aligns with human auditory perception. The pink noise is passed through a low-pass filter, where the cutoff frequency determines the emphasis on different instrument ranges. Lower

[2] https://vimeo.com/1059744914

[3] https://vimeo.com/1059744839

cutoff frequencies result in deeper, resonant tones akin to geomungo (pluck string instrument)) daegeum (flute alike wind instrument), while higher cutoff frequencies emphasize higher pitched instruments.

An envelope generator further shapes the pink noise by dynamically modifying its amplitude profile. The attack speed is particularly critical. Short attack speed produces sharp transient sounds, mimicking breathy articulations in wind instruments or the plucked attacks of string instruments, while long attack speeds create smooth, sustained tones without accentuated transients, leading to a more legato articulation.

4.4.3 Compositional Process

The process of discovering and refining latent dimension combinations is not merely a technical operation but a compositional exploration. This involves determining the musical impressions that a particular combination conveys, structuring soundscapes by preparing specific timbral states as part of the composition and reflecting the composer's personal perception of Korean traditional music and its reinterpretation through AI.

Thus, the system does not simply automate sound generation; it provides a framework where latent vector navigation becomes an active component of artistic decision, shaping how the music unfolds in both structure and expression.

4.5 Performance Implementation

One of the most critical aspects of this system's approach is how controllable parameters interact in real-time composition and performance. The system allows for MIDI CC messages to manipulate envelope length and amplitude during live performance, providing dynamic expressive control. Additionally, cutoff frequency modulation operates within a dynamically shifting range, allowing pitch and spectral variations to be musically structures while retaining organic fluidity.

Through these controllable parameters, the system allows for structured yet fluid modulation, balancing predictability and organic sonic evolution within the framework of live composition and performance.

4.6 Artistic context and Post-Concrete Aesthetics of AI sound

UTMORI explores an alternative approach to AI-driven composition by moving beyond conventional sample-based generation and instead constructing sequences through the manipulation of latent vectors using noise shaping techniques. In this process, latent vectors are shaped and combined to form Sonic Objects, which function as the primary structural units of a composition. Unlike traditional AI-generated music, which typically reconstructs patterns derived from training data, UTMORI's method allows for a more flexible organization of musical material. This approach is intended to facilitate a compositional framework where Sonic Objects can be arranged and modified dynamically, rather than adhering to predefined templates. By treating the generative process as an integral part of composition, UTMORI enables an exploration of musical form that is structured around the interaction of AI-generated materials rather than the direct imitation of existing musical styles.

Building on this foundation, UTMORI also incorporates non-standard rhythmic structures from Korean traditional music, particularly Eotmori Jangdan. In contrast to the symmetrical, grid-based rhythms often found in machine-generated compositions, this project integrates complex rhythmic cycles such as 8/10, 5-beat, and 10-beat patterns, allowing for real-time transformations rather than fixed structural repetition**.** The combination of these rhythmic frameworks with Sonic Object-based composition provides a means to investigate how AI-generated musical units can be shaped within non-traditional rhythmic structures. This approach raises questions about how AI systems interact with culturally specific musical grammars and whether such methodologies contribute to the development of distinctive compositional models. Rather than focusing on stylistic reproduction, UTMORI examines how machine learning can be applied to the structural and organizational aspects of composition, expanding the discussion on AI's role as an interactive tool in music-making.

While UTMORI offers a novel approach to AI-driven music composition, it can also be understood within a lineage of experimental sound practices that redefine the role of the composer and the ontology of musical material. Echoing the spirit of Pierre Schaeffer's musique concrète, where recorded sounds became manipulable compositional elements, UTMORI extends this logic to the manipulation of machine-learned latent representations. However, unlike Schaeffer's acousmatic focus, UTMORI privileges performative interaction, situating the human in an active dialog with AI-generated potentialities rather than fixed recorded material.

Additionally, UTMORI resonates with the aesthetics of algorithmic and generative composition in the works of Iannis Xenakis or Curtis Roads yet diverges in its use of culturally rooted rhythmic frameworks and non-Western sonic vocabularies. This marks a critical turn from universalist, mathematically formalist approaches towards a culturally situated, pluralistic generative logic. In doing so, UTMORI contributes to the emergence of a post-global AI aesthetic—where algorithmic systems do not erase but amplify localized musical grammar.

Lastly, UTMORI participates in a broader shift within AI art, where generative systems are not merely tools of stylistic imitation but are reframed as compositional collaborators. By foregrounding structural manipulation over style reproduction, UTMORI critiques the normative biases of large-scale generative models and offers a counterproposition: a performative, culturally aware, and materially grounded AI composition practice.

Chapter 5

Communication Breakdown

Generative AI has increasingly been integrated into artistic practices, often serving as a tool for structured visual, textual, and auditory creation. However, Communication Breakdown [4] challenges this conventional approach by positioning AI not as a generator of coherence but as an agent of perceptual collapse. This project investigates whether AI can function beyond conventional generativity and serve as a mechanism for destabilizing human cognition and disrupting the process of meaning-making.

The work is methodologically structured around the concept of Gestalt Collapse, in which prolonged exposure to repetitive stimuli leads to the dissolution of perceptual structures, forcing new patterns to emerge. Extending this phenomenon to the realms of visual perception, linguistic coherence, and auditory stability, the work employs three AI-driven systems—Flux[38]-based image retexturing with Controlnet[39], 1-bit LLM[40] for textual disruption, and RAVE for generative soundscapes. Together, these technologies construct an immersive experience that forces the audience into a state of interpretative instability.

This project engages with two interrelated research inquiries. The first investigates the extent to which AI can function as an active disruptor of perception and meaning, rather than as a conventional generative tool. The second explores how Communication Breakdown situates itself within the historical lineage of media art, linguistic deconstruction, and perceptual manipulation. Addressing these inquiries contributes to both the methodological discourse surrounding AI-driven aesthetics and the historical examination of how machines have been employed in art to challenge human perception.

[4] https://youtu.be/Du25jQo4Ito?si=QKROusypOw7JR4Fa

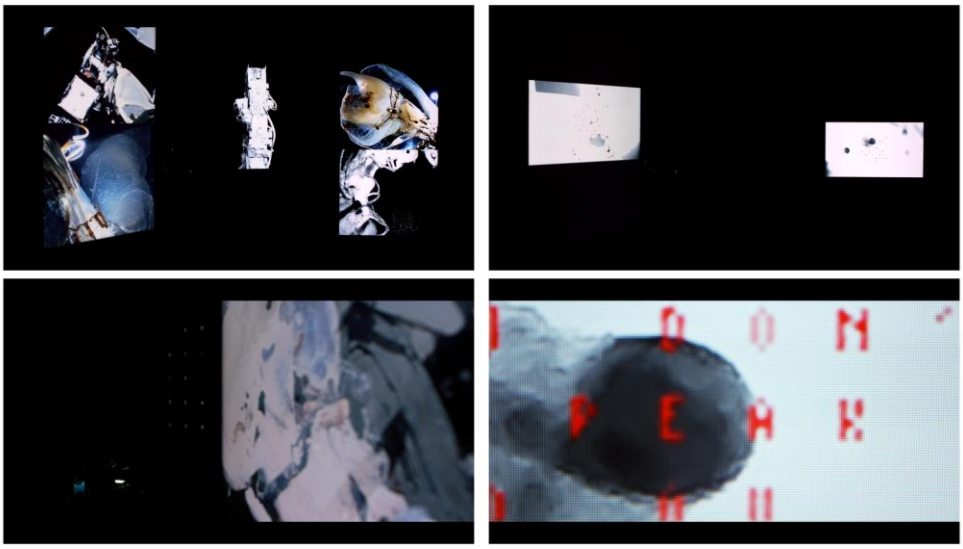

figure 15 Image archives from Communication Breakdown (2024)

The methodological framework of Communication Breakdown is designed to systematically obstruct the viewer's ability to establish visual, linguistic, and auditory coherence. Each AI-driven system operates independently yet interconnects conceptually, forming a multi-layered disruption of perception. To examine these mechanisms in depth, this section is divided into three subchapters, each addressing a specific mode of perceptual disruption: visual instability, linguistic incoherence, and auditory unpredictability.

5.1 Visual Disruption: AI-Based Image Retexturing and Gestalt Collapse

The first methodological layer investigates the destabilization of visual perception through AI-generated transformation. The project employs Flux-based image-to-image processing with ComfyUI[41], applied to a robotic arm whose depth and fundamental structure remain unchanged while its surface textures continuously shift. This controlled instability produces a paradoxical effect, wherein the viewer perceives an object that is at once stable and fragmented, leading to a progressive breakdown of object recognition.

Two key mechanisms facilitate this perceptual dissonance. The first ensures that the robotic arm retains a stable three-dimensional form, maintaining a reference point for the viewer. However, the AI-driven textural transformations prevent visual fixation, causing Gestalt Collapse as familiar patterns dissolve into unpredictable variations. The second mechanism synchronizes these transformations with drum-triggered sequencing in TouchDesigner, creating the illusion of movement and acceleration, further amplifying the audience's perceptual disorientation.

This methodological approach aligns with historical precedents in Constructivist and media art practices. While László Moholy-Nagy and Naum Gabo explored machine aesthetics as a means of enhancing perception, Communication Breakdown inverts this logic, employing AI to subvert rather than reinforce visual stability. Additionally, Richard Serra's repetitive structural deformations, which gradually undermine material consistency, serve as a conceptual reference for understanding the incremental visual breakdown within this work.

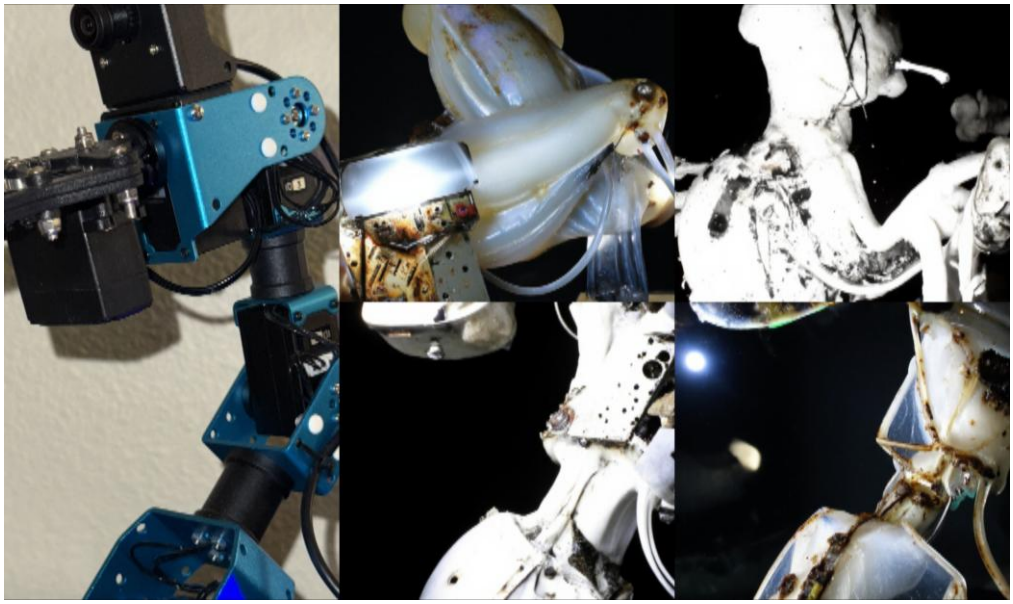

figure 16 Retexturing examples from Communication Breakdown (2024)

5.2 Linguistic Breakdown: AI-Generated Text as a Method of Semantic Disruption

The second methodological component examines how AI-generated text dismantles linguistic coherence. Unlike projects that employ large language models to create structured narratives, Communication Breakdown utilizes LLaMA[42] 1-bit LLM with Bitnet CPP[43] to generate arbitrary, fragmented responses along the ecosystem of this piece. This ensures that the AI fails to establish meaningful syntactic relationships, forcing the audience into an interpretative struggle.

The AI system processes thematic prompts yet produces responses that resist logical continuity. This linguistic breakdown echoes poststructuralist theories of language deconstruction,

particularly those of Jacques Derrida and Brion Gysin. Derrida's concept of différance, which asserts that meaning is always deferred and never stable, finds an AI-driven realization in the perpetual instability of machine-generated text. Similarly, Gysin's cut-up technique, which involved manually disassembling and reordering textual sequences, is extended in Communication Breakdown through automated, real-time linguistic disintegration.

The work also engages with historical precedents in text-based media art. Ed Ruscha's typographic paintings, which rendered words as purely visual rather than semantic objects, offer a parallel to how Communication Breakdown transforms AI-generated text into encrypted, visually fragmented compositions, compounding both its semantic and aesthetic inaccessibility.

5.3 Auditory Instability: AI-Driven Sound Synthesis and the Denial of Rhythmic Closure

The third methodological focus explores the disruption of auditory perception through AI-generated sound environments. The project utilizes RAVE-based sound synthesis, a system that continuously shifts between timbral variations, denying the listener access to stable rhythmic or harmonic structures.

The AI-driven sound design is constructed upon two primary methodologies. The first involves latent space navigation, where AI dynamically shifts across timbral regions, producing unpredictable sonic textures that fluctuate between familiar and alien auditory cues. The second disrupts rhythmic consistency, ensuring that no fixed pulse emerges, preventing the audience from forming auditory expectations.

This auditory instability aligns with historical precedents in perceptual and experimental sound art. James Turrell's exploration of spatial perception, wherein viewers are immersed in light-based environments that distort depth and form, finds an auditory equivalent in Communication Breakdown, where sound is continuously modulated to alter spatial perception and cognitive processing. Additionally, Gestalt psychology's principles of pattern recognition and closure are subverted, as the AI systematically prevents the auditory system from resolving musical structures into coherent forms.

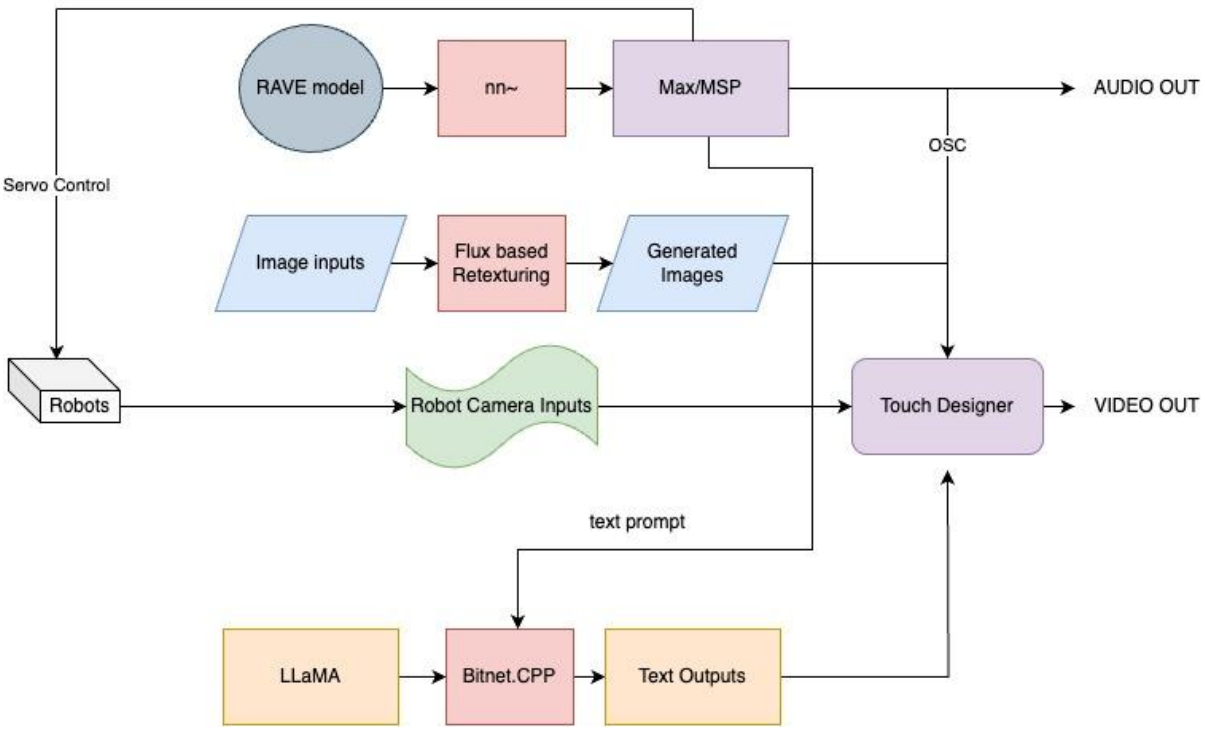

figure 17. Realtime workflow of Communication Breakdown

5.4 Artistic Context and the Aesthetic of Perceptual Collapse

Communication Breakdown redefines the role of AI in artistic practice. Rather than serving as a means for structured storytelling or aesthetic generation, the work employs machine learning to actively disrupt cognition, challenging the audience's ability to establish meaning in visual, linguistic, and auditory domains.

This project extends the discourse on AI-driven aesthetics by demonstrating that machine learning can function not only as a creative tool but as a mechanism for perceptual destabilization. Through its engagement with historical media art traditions—machine aesthetics (Moholy-Nagy, Gabo), linguistic deconstruction (Derrida, Gysin, Ruscha), and perceptual manipulation (Serra, Turrell, Gestalt theory)—Communication Breakdown situates itself within a lineage of experimental art that challenges conventional cognitive processes.

By positioning AI as an agent of perceptual and conceptual instability, the work raises fundamental questions regarding communication in an AI-mediated world. It suggests that rather than facilitating seamless interaction, machine-generated unpredictability forces human cognition to confront its own limitations, ultimately revealing the fragility of perception and the instability of meaning itself.

In this regard, Communication Breakdown can be situated within a broader historical lineage of media art that deliberately destabilizes perceptual and cognitive coherence not only as a critique of representational systems, but as a restructuring of the conditions of meaning itself. While Moholy-Nagy and Gabo's machine aesthetics celebrated precision and rational clarity, Communication Breakdown weaponizes the machine to generate ambiguity, misalignment, and semiotic collapse. It departs from the modernist belief in technological utopia and aligns more closely with postmodern strategies of fragmentation, echoing the discursive disruptions of Derrida or the semantic erosion in Gysin's cut-up technique.

Furthermore, unlike early media art that employed analog distortion or sculptural repetition, Communication Breakdown leverages AI's generative unpredictability as a computational form of failure systematized noise, unintended drift, and synthetic hallucination. This positions the work in dialogue with a new wave of AI art that refuses the aesthetics of coherence and instead embraces perceptual crisis as both method and message.

In doing so, the work extends the aesthetic lineage of glitch art and cognitive estrangement into the domain of machine learning. It no longer merely illustrates instability; it embeds it structurally within the generative system. Thus, Communication Breakdown reclaims AI not as a solver of human intention, but as a destabilizer of sensory regimes, where interpretation falters and cognition is forced to reconfigure itself within machine-mediated uncertainty.

Chapter 6

Conclusion

This thesis has explored the generative capacity of AI not as a tool for content creation but as a medium for perceptual and structural reconfiguration across three distinct projects: Come Closer, Utmori, and Communication Breakdown. Each project interrogated different modalities – visual, auditory, linguistic and deployed AI as a co-creator but as a machine infrastructure through which perception, rhythm, and meaning could be critically destabilized, restructured, or expanded. Rather than seeking only the harmony between humans and machines, these works approach AI as means of challenging and extending the conditions of human sensation and cognition.

Three operative modalities of AI-based aesthetic intervention can be identified through these works: One, AI as a perceptual interface mechanism (Come Closer). Two, AI as a navigable sonic topology (Utmori). Three, AI as an agent of semantic collapse (Communication Breakdown). In each case, the AI system is not just an instrument of generation but a structurally consequential entity whose affordances and limitations become the very material conditions of the work.

6.1 What worked

A key strength across these projects was the articulation of multiple sensory domains: auditory, visual, and linguistic into coherent, structurally responsive systems where AI was not an autonomous creative agent, but a mediating infrastructure. Rather than evaluating each artwork as an isolated success, it is more useful to examine how specific artistic elements operated effectively in relation to machine learning systems and interactive technologies.

In the visual domain, Come Closer employed low-resolution, texture-forward AI-generated imagery to explore sensory ambiguity and bodily defamiliarization. Real-time filtering systems embedded in Touch Designer transformed generated content into fluctuating perceptual experiences, destabilizing the viewer's ability to fixate or resolve bodily forms. This produced not simply a glitch aesthetic, but a strategy for reframing the act of visual perception itself, highlighting

how recognition and intimacy are constructed through fragmentary inputs. AI did not replace the visual system; it provided a parallel one, designed to question its assumptions.

In the auditory domain, UTMORI demonstrated how AI could be harnessed not to generate stylistically coherent music, but to provide a navigable field for sonic exploration. Through RAVE-based latent space modulation and envelope-controlled rhythmic gestures, the system allowed the composer to structure rhythm and timbre interactively. This process reframed composition as a form of topological navigation, where musical structure emerged from embodied, real-time decision making. The ability to treat combinations of latent dimensions as emergent, flexible sonic units offered a new vocabulary for generative composition neither purely algorithmic nor entirely traditional.

In the linguistic and symbolic domain, Communication Breakdown leveraged AI's stochastic behaviors to orchestrate breakdown and incoherence as formal strategies. Rather than generating novel text, the work emphasized semantic erosion using 1-bit language models and recursive fragmentation. These breakdowns were not accidental byproducts of weak AI, but deliberately constructed conditions of interpretive overload. AI thus became a tool for producing semiotic instability and highlighting the fragility of linguistic systems under pressure.

Across all domains, the projects emphasized AI's role in configuring material conditions for perception and interpretation. Whether through rhythm, form, language, or image, the works did not simply showcase what AI could produce, but how it could intervene in the sensory regimes through which aesthetic experience takes shape. By doing so, they offered a model for integrating machine learning not as a creative collaborator, but as a compositional substrate, one that enables the structural redirection of cognition, sensation, and cultural memory.

6.2 What didn't work

While each project explored AI's role in reorganizing sensory experience, several limitations emerged across the perceptual modalities examined. These limitations were not merely technical or aesthetic oversights, but indicative of deeper tensions between machinic logic and human perceptual processes—particularly in relation to control, disruption, and interpretation.

In the visual domain, Come Closer employed generative techniques to challenge bodily recognition, but the perceptual interruption it aimed to provoke was not always consistently achieved. The interaction between SDXL-based image textures and mosaic filtering in Touch

Designer occasionally produced effects that were either too stable or too easily legible, resulting in diminished disruption. Despite its conceptual clarity, the project at times lacked the necessary instability to fully destabilize visual processing—suggesting a need for more refined dynamic calibration or sensor-driven responsiveness.

In the auditory domain, UTMORI revealed a critical asymmetry between the composer's embodied knowledge of Korean rhythmic structures and the listener's interpretive accessibility. While the latent space was shaped by culturally informed impressions, these were not always intuitively traceable by uninitiated audiences. The subjective mapping of tradition into abstract AI parameters produced a perceptual opacity, wherein the intended cultural resonance risked becoming illegible. This gap raises questions about how AI-driven sonic material might bridge the divide between personal memory and shared rhythmic cognition.

In the linguistic and symbolic domain, Communication Breakdown relied heavily on stochastic language generation and recursive fragmentation to explore the limits of meaning. Yet the resulting incoherence sometimes surpassed its intended critical function, diluting authorial control and reducing interpretive traction. The erosion of syntax and semantic cohesion, while conceptually valid, occasionally overwhelmed the composition's ability to guide affective or cognitive engagement suggesting that the tension between structured disruption and total opacity remains a delicate compositional threshold.

Finally, across all domains, a shared limitation involved the absence of audience-centered perceptual feedback. The works posited strong claims about perceptual reorganization and sensory destabilization but lacked mechanisms to evaluate whether these effects were meaningfully registered. The loop between machine-generated complexity and human interpretation remained largely speculative, pointing to a broader need for methodological frameworks that address not only generation but reception, perception, and embodied response. In Come Closer, the real-time visual distortions occasionally lacked the resolution or structural intensity necessary to produce sustained perceptual interference. While the system employed generative models and live filtering techniques to challenge bodily recognition, the technical implementation particularly in the interaction between SDXL-based textures and Touch Designer's mosaic filters sometimes fell short of delivering a strong enough disruption.

6.3 Future work

In anticipating the aesthetic and technological conditions of the near future, it becomes increasingly evident that the world will demand systems capable of adapting to rapidly shifting perceptual and cultural thresholds. In such a context, the artist's role may move away from traditional modes of creation toward the orchestration of conditions under which new forms of experience can emerge.

As human perception becomes more mediated and modulated by technological systems, it becomes necessary to address not only what is produced, but how perception is conditioned and redirected. Artists can respond not by simplifying complexity, but by constructing frameworks that facilitate affective and cognitive engagement within these emergent structures.

Future inquiries will therefore extend in two directions. First, the development of multimodal sensory infrastructures will focus on AI systems that integrate visual and auditory input and tactile, gestural, and physiological signals. These infrastructures are not intended as immersive novelties but as platforms for testing how perception can be expanded, layered, or rerouted. Real-time biofeedback mechanisms such as EEG, posture and gaze tracking will be explored as compositional variables, enabling environments that adapt structurally to the participant's bodily and cognitive state.

Second, attention will turn toward recursive installation systems designed to investigate how repetition and breakdown operate as perceptual strategies. Instead of showcasing AI's capacity for novelty, these systems will engineer conditions of sensory erosion spaces in which clarity gives way to overload, and coherence dissolves into ambient semiotic noise. Such work is not about rejecting meaning, but about confronting the limits of interpretability under machinic mediation.

At the same time, the collective understanding of what constitutes a medium or a tool is expected to undergo significant shifts. The boundaries between consumption and creation, linear authorship, and distributed sensibility are already being redrawn. In this emergent landscape of non-linear, multidimensional expression, AI-based systems offer technical means and conceptual openings for engaging with layered, asynchronous, and hybridized cultural forms. Rather than prescribing fixed outcomes, they invite speculation, iteration, and co-emergent meaning-making. In this light, AI becomes not merely a tool or collaborator, but a condition of composition a structural partner in reshaping sensory possibility and the redefinition of artistic responsibility.

Bibliography

[1] "A Review on Generative AI for Text-to-Image and Image-to-Image Generation and Implications to Scientific Images." Accessed: Mar. 25, 2025. [Online]. Available: https://arxiv.org/html/2502.21151v2

[2] M. Caldwell, "What Is an 'Author'?-Copyright Authorship of AI Art Through a Philosophical Lens," *Houst. Law Rev.*, vol. 61, no. 2, pp. 411–442, Dec. 2023.

[3] A. Hencz, "AI Art Controversy: Hype, Skepticism & Copyright Issues," Artland Magazine. Accessed: Mar. 25, 2025. [Online]. Available: https://magazine.artland.com/ai-art-creativity-controversy/

[4] S. Shaffi, "'It's the opposite of art': why illustrators are furious about AI," *The Guardian*, Jan. 23, 2023. Accessed: Mar. 25, 2025. [Online]. Available: https://www.theguardian.com/artanddesign/2023/jan/23/its-the-opposite-of-art-why-illustrators-are-furious-about-ai

[5] A. Oksanen and J. Lahtinen, "Manifesto for Artist-Driven AI Art," Jul. 31, 2023, *Social Science Research Network, Rochester, NY*: 4536742. doi: 10.2139/ssrn.4536742.

[6] "Responsible AI Music Artistic Mini-Projects." Accessed: Mar. 25, 2025. [Online]. Available: https://music-rai.github.io/projects/

[7] "Harold Cohen: AARON." Accessed: Mar. 20, 2025. [Online]. Available: https://whitney.org/exhibitions/harold-cohen-aaron?section=1&subsection=3

[8] "Harold Cohen's pioneering AI works provide essential context for conversations about generative art," The Art Newspaper - International art news and events. Accessed: Mar. 20, 2025. [Online]. Available: https://www.theartnewspaper.com/2024/02/26/harold-cohen-artificial-intelligence-aaron-whitney-museum-technology

[9] "DeepDream - a code example for visualizing Neural Networks." Accessed: Mar. 20, 2025. [Online]. Available: https://research.google/blog/deepdream-a-code-example-for-visualizing-neural-networks/

[10] "Inside Deep Dreams: How Google Made Its Computers Go Crazy | WIRED." Accessed: Mar. 20, 2025. [Online]. Available: https://www.wired.com/2015/12/inside-deep-dreams-how-google-made-its-computers-go-crazy/

[11] T. T. at AMI, "DeepDream Art and Machine Learning Symposium 2016 recap," Artists + Machine Intelligence. Accessed: Mar. 20, 2025. [Online]. Available: https://medium.com/artists-and-machine-intelligence/deepdream-art-and-machine-learning-symposium-2016-recap-396d1ecf87e3

[12] I. J. Goodfellow *et al.*, "Generative Adversarial Networks," Jun. 10, 2014, *arXiv*: arXiv:1406.2661. doi: 10.48550/arXiv.1406.2661.

[13] "'Painting' with Data: How Media Artist Refik Anadol Creates Art Using Generative AI." Accessed: Mar. 20, 2025. [Online]. Available: https://www.wipo.int/web/wipo-magazine/article-details/?assetRef=67301&title=painting-with-data-how-media-artist-refik-anadol-creates-art-using-generative-ai

[14] "Melting Memories," Refik Anadol Studio. Accessed: Mar. 30, 2025. [Online]. Available: https://refikanadolstudio.com/projects/melting-memories/
[15] "Artificial intelligence hallucination and Refik Anadol | Opinion." Accessed: Mar. 20, 2025. [Online]. Available: https://www.dailysabah.com/opinion/op-ed/artificial-intelligence-hallucination-and-refik-anadol
[16] "'Painting' with Data: How Media Artist Refik Anadol Creates Art Using Generative AI." Accessed: Mar. 21, 2025. [Online]. Available: https://www.wipo.int/web/wipo-magazine/article-details/?assetRef=67301&title=painting-with-data-how-media-artist-refik-anadol-creates-art-using-generative-ai
[17] "Mario Klingemann," Kate Vass Galerie. Accessed: Mar. 21, 2025. [Online]. Available: https://www.katevassgalerie.com/mario-klingemann
[18] "(#109) MARIO KLINGEMANN | Memories of Passersby I," Sothebys.com. Accessed: Mar. 21, 2025. [Online]. Available: https://www.sothebys.com/en/auctions/ecatalogue/2019/contemporary-art-day-auction-l19021/lot.109.html
[19] N. R. S. T. Article, "Sotheby's Is Entering the AI Art Fray, Selling a Surreal Artwork by One of the Movement's Pioneers This Spring," Artnet News. Accessed: Mar. 21, 2025. [Online]. Available: https://news.artnet.com/art-world/sothebys-artificial-intelligence-1460332
[20] "Ada Ada Ada — Being represented by data is like losing a part of yourself." Accessed: Mar. 30, 2025. [Online]. Available: https://ada-ada-ada.art/projects/being-represented-by-data-is-like-losing-a-part-of-yourself
[21] S. B. Pold, "Performing Profiling: Algorithmic Enunciation, Transgender Perspectives, and Ada Ada Ada's in transitu," *Nord. J. Aesthet.*, vol. 33, no. 68, Art. no. 68, Dec. 2024, doi: 10.7146/nja.v33i68.152365.
[22] "Ian Cheng: BOB," Serpentine Galleries. Accessed: Mar. 30, 2025. [Online]. Available: https://www.serpentinegalleries.org/whats-on/ian-cheng-bob/
[23] "Ian Cheng's art is inspired by The Sims and created with AI | WIRED." Accessed: Mar. 30, 2025. [Online]. Available: https://www.wired.com/story/ai-art-ian-cheng-serpentine-emissaries-bob/
[24] J. Jones, "Sondra Perry's Typhoon wrenches my soul but Ian Cheng's AI is merely soulless – review," *The Guardian*, Mar. 07, 2018. Accessed: Mar. 30, 2025. [Online]. Available: https://www.theguardian.com/artanddesign/2018/mar/07/sondra-perry-ian-cheng-review-serpentine-gallery-london-technology
[25] E. Gollob, "Searching a moment of Doing Nothing," Interalia Magazine. Accessed: Mar. 30, 2025. [Online]. Available: https://www.interaliamag.org/audiovisual/emanuel-gollob/
[26] "Doing Nothing with AI 1.0," Emanuel Gollob. Accessed: Mar. 30, 2025. [Online]. Available: https://www.emanuelgollob.com/doing-nothing-with-ai/
[27] E. Gollob, "Searching a moment of Doing Nothing," Interalia Magazine. Accessed: Mar. 30, 2025. [Online]. Available: https://www.interaliamag.org/audiovisual/emanuel-gollob/
[28] "ERROR – The Art of Imperfection – Ars Electronica Export." Accessed: Mar. 30, 2025. [Online]. Available: https://ars.electronica.art/export/en/berlin2018/
[29] C. Sivertsen, G. Salimbeni, A. S. Løvlie, S. D. Benford, and J. Zhu, "Machine Learning Processes As Sources of Ambiguity: Insights from AI Art," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, in CHI '24. New York, NY, USA: Association for Computing Machinery, 11 2024, pp. 1–14. doi: 10.1145/3613904.3642855.
[30] "Poison - An Edvard Munch Experience." Accessed: Mar. 30, 2025. [Online]. Available: https://www.munchmuseet.no/en/exhibitions/archive/2021/poison/

[31] "ImageNet Roulette – Trevor Paglen." Accessed: Mar. 30, 2025. [Online]. Available: https://paglen.studio/2020/04/29/imagenet-roulette/
[32] L. Winkie, "Hear Insane 24/7 Livestream of AI-Generated Death Metal," Revolver. Accessed: Mar. 30, 2025. [Online]. Available: https://www.revolvermag.com/music/hear-insane-247-livestream-ai-generated-death-metal/
[33] "This YouTube Channel Streams AI-Generated Death Metal 24/7." Accessed: Mar. 30, 2025. [Online]. Available: https://www.vice.com/en/article/this-youtube-channel-streams-ai-generated-black-metal-247/
[34] D. Podell *et al.*, "SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis," Jul. 04, 2023, *arXiv*: arXiv:2307.01952. doi: 10.48550/arXiv.2307.01952.
[35] H. Ye, J. Zhang, S. Liu, X. Han, and W. Yang, "IP-Adapter: Text Compatible Image Prompt Adapter for Text-to-Image Diffusion Models," Aug. 13, 2023, *arXiv*: arXiv:2308.06721. doi: 10.48550/arXiv.2308.06721.
[36] A. Caillon and P. Esling, "RAVE: A variational autoencoder for fast and high-quality neural audio synthesis," Dec. 15, 2021, *arXiv*: arXiv:2111.05011. doi: 10.48550/arXiv.2111.05011.
[37] "nn~ | Ircam Forum." Accessed: Feb. 21, 2025. [Online]. Available: https://forum.ircam.fr/projects/detail/nn/#project-related-anchor
[38] "black-forest-labs/FLUX.1-Fill-dev · Hugging Face." Accessed: Mar. 21, 2025. [Online]. Available: https://huggingface.co/black-forest-labs/FLUX.1-Fill-dev
[39] L. Zhang, A. Rao, and M. Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models," Nov. 26, 2023, *arXiv*: arXiv:2302.05543. doi: 10.48550/arXiv.2302.05543.
[40] S. Ma *et al.*, "The Era of 1-bit LLMs: All Large Language Models are in 1.58 Bits," Feb. 27, 2024, *arXiv*: arXiv:2402.17764. doi: 10.48550/arXiv.2402.17764.
[41] comfyanonymous, *comfyanonymous/ComfyUI*. (Mar. 21, 2025). Python. Accessed: Mar. 21, 2025. [Online]. Available: https://github.com/comfyanonymous/ComfyUI
[42] H. Touvron *et al.*, "LLaMA: Open and Efficient Foundation Language Models," Feb. 27, 2023, *arXiv*: arXiv:2302.13971. doi: 10.48550/arXiv.2302.13971.
[43] J. Wang *et al.*, "Bitnet.cpp: Efficient Edge Inference for Ternary LLMs," Feb. 17, 2025, *arXiv*: arXiv:2502.11880. doi: 10.48550/arXiv.2502.11880.