

Article

Edge Knowledge in Cognitive Art: Munch Digital Twin

Iana Fominska ^{1,*} , Gerardo Iovane ^{2,*}  and Marta Chinnici ³ 

¹ Department of Education, Cultural Heritage and Tourism Sciences, University of Macerata, 62100 Macerata, Italy

² Department of Computer Science, University of Salerno, 84084 Fisciano, Italy

³ ICT Division-HPC Lab, Department of Energy Technologies and Renewable Energy Sources (TERIN), ENEA C.R. Casaccia, 00123 Rome, Italy; marta.chinnici@enea.it

* Correspondence: i.fominska@unimc.it (I.F.); giovane@unisa.it (G.I.)

Abstract

In an era where artificial intelligence is rapidly expanding into creative domains, the challenge of modeling human-like cognition and emotion in generative processes becomes increasingly central. The present study was made in connection with the exhibition of Munch's works held in Rome from February to June 2025. Indeed, the paper introduces the concept of a Cognitive Digital Twin grounded in the Super Time-Cognitive Neural Network (STCNN) framework and applies it to the case of Edvard Munch, the iconic Norwegian expressionist. The proposed system—Munch Digital Twin—goes beyond static generative models by integrating temporal, emotional, and cognitive dimensions through a complex-valued time representation $t = a + i \cdot b$, where a denotes chronological time and b encodes imagination, memory, and creativity. We define Edge Knowledge as an output-stage re-ranking criterion that admits a generated response only where corpus evidence, knowledge-graph constraints and the LLM surface jointly agree (the boundary, or 'edge', between documented identity and machine inference). STCNN allows this twin to process real inputs (text, visual prompts, emotional cues) and generate outputs that reflect both the rational and expressive styles of Munch. The imaginary components of the network enable speculative and affective expansions of known artworks—such as reinterpreting *The Scream* under new emotional or social contexts. This paper presents the theoretical underpinnings of cognitive digital twins, the architecture of the STCNN-based model, and a prototype implementation trained on Munch's paintings, letters, and critical essays. The system—comprising a GPT-4-Turbo cloud profile and a 4-bit LLaMA-2-13B edge profile for language, Stable Diffusion 1.5 + LoRA for image generation, a Neo4j knowledge graph, and FAISS retrieval—is trained on approximately 600 letters, 100 artworks, and Munch's diaries and criticism, and evaluated across 100 interactive sessions with 14 students and expert raters. Headline results against an unconditioned baseline include CLIPScore +13.8%, FID −25.5% (small-sample, indicative), and emotion-cosine similarity +44.9%. Ethical implications surrounding posthumous digital emulation, authorship, and emotional manipulation are also discussed. The Munch Digital Twin represents a new paradigm in AI-driven art, where machines do not merely replicate, but collaborate across time with human legacies, enabling an anticipatory and emotionally intelligent form of computational creativity. This work is primarily a conceptual and architectural contribution, supported by a proof-of-concept prototype and a preliminary, non-controlled user study; the quantitative results are indicative and not yet confirmatory.



Academic Editor: Stefan Fischer

Received: 6 May 2026

Revised: 23 June 2026

Accepted: 23 June 2026

Published: 26 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

Keywords: digital twin modeling; cognitive computational architectures; affective generative AI; multimodal neural encoding; STCNN framework; emotional simulation in art; expressionist style reconstruction; epistemic plausibility modeling; knowledge-infused AI agents; human–machine creative interaction

1. Introduction

This work was created in connection with the exhibition of Munch’s works held in Rome from February to June 2025. The Inner Scream, the title of the exhibition, was conceived for the commemorations of the 80th anniversary of Munch’s death (Edvard Munch, 1863–1944), as part of a touring exhibition shown in Milan (Palazzo Reale, September 2024–January 2025) and Rome (Palazzo Bonaparte, 11 February–2 June 2025). It was a major retrospective promoted by Italian institutions, with the patronage of the Royal Norwegian Embassy in Rome, in collaboration with the MUNCH Museum, Oslo, produced by Arthemisia. The exhibition was curated by Patricia G. Berman, an American art historian and museum curator and international expert on Munch, with scientific collaboration by Costantino D’Orazio. The exhibition narrated the entire universe of Edvard Munch, his human journey, and his artistic production through 100 works, including one of the lithographic versions of *The Scream* (1895) kept in Oslo, *The Death of Marat* (1907), *Starry Night* (1922–1924), *The Girls on the Bridge* (1927), *Melancholy* (1900–1901), and *Dance on the Beach* (1904). Therefore, this work was inspired by the intention to further explore how the artist’s precious work can be immortalized using artificial intelligence in general, and artificial agents and digital twins in particular. The system we will describe consists of a GPT-4-Turbo cloud model and a 4-bit LLaMA-2-13B edge model for natural language processing [1], Stable Diffusion 1.5 + LoRA for image generation [2], a Neo4j knowledge graph, and the FAISS retrieval system.

Artificial intelligence is no longer confined to mathematical optimization or computational logic; it now speaks the language of aesthetics, metaphors, and memory. Contemporary AI models are increasingly deployed in fields such as literature, visual arts, and philosophy [3]. While remarkable, many of these systems operate through correlation-based generation rather than any deep sense of understanding or emotional grounding [4]. This limitation becomes especially evident when attempting to emulate an artist like Edvard Munch, whose work is steeped in personal suffering and existential themes [5]. Reproducing his artistic style requires more than brushstroke imitation, it demands the recreation of affective depth. The ambition of this paper is to introduce a digital twin of Munch, not only stylistically accurate but cognitively and emotionally resonant.

Digital twins originated in engineering contexts, where the replication of physical systems in virtual environments allowed for simulation and predictive maintenance [6]. Their migration into cognitive science marks a paradigm shift: a digital twin is no longer just a structural replica, but a semantic and interpretive agent [7]. When this agent is infused with STCNN’s temporal-cognitive capabilities, it becomes capable of imagination and internal reasoning [8].

For clarity, we define the four key concepts used throughout: Cognitive Digital Twin—a generative agent conditioned on a documented corpus that reconstructs plausible (not conscious) states of a historical subject; STCNN (Super Time-Cognitive Neural Network)—a structured conditioning framework adding complex-time gating and a resonance check to a standard transformer pipeline; complex time $t = a + i \cdot b$ —where a is the chronological anchor and b is a learnable salience/recall weight; Edge Knowledge—an output-stage

re-ranking criterion admitting a response only where corpus evidence, knowledge-graph constraints and the LLM surface jointly agree.

The STCNN architecture is based on a novel representation of time, formulated as a complex variable where real components correspond to chronological sequences and imaginary components encode memory and creativity [9]. This structure allows the AI to “remember” like a human—not as exact storage, but as selective, weighted recollection [10].

For Munch, such a system means modeling not only his visible works, but also his reactions to trauma, solitude, and inspiration. The STCNN framework is uniquely positioned for this challenge due to its hybridization of recurrent, convolutional, and attentional modules [11]. These modules collaborate to simulate how a memory might influence a future output, a key feature of human-like artistic reasoning [12]. This aligns with philosophical interpretations of cognition that emphasize the entanglement of time, memory, and meaning [13]. Creative production, especially in emotionally charged contexts, cannot be decoupled from lived experience. Thus, a system that hopes to emulate Munch must possess not only aesthetic capability but a temporal-cognitive anchor [14].

From a technical perspective, the Munch Digital Twin ingests multimodal data—paintings, letters, voice logs, diary entries—and integrates them into a latent space structured by emotional intensity and thematic recurrence [15]. This latent structure enables the model to generate new works “in the style of” Munch while anchoring them in plausible emotional and historical trajectories [16].

Recent advancements in generative art—especially through large language models and image diffusion networks—have shown potential in style transfer and visual synthesis, but they lack the ability to integrate real-time feedback from an internal identity model [17]. STCNN addresses this by incorporating memory traces and cognitive feedback loops that mimic introspective processes [18]. An important advantage of this model lies in its dialogic affordances. Users can engage the Munch Twin in conversation, asking about the meaning of a specific painting or posing hypothetical scenarios, to which the twin responds with emotionally grounded and stylistically coherent reflections [19]. Such responses are not hard-coded but emerge from the interaction between learned history and internal reasoning dynamics [20].

Ethical considerations naturally arise when a digital entity appears to think or feel. This is especially sensitive when simulating historical figures, where the line between tribute and exploitation becomes thin [21]. To mitigate this, the Munch Digital Twin explicitly marks generative outputs as hypothetical and offers transparency about the data and reasoning processes behind its responses [22]. Scholars have debated whether computational creativity can be genuinely novel or merely derivative [23]. Our position is that systems like the Munch Twin offer a third path: generative fidelity. That is, the system does not innovate for innovation’s sake but extends the artistic worldview of Munch into contexts he never lived to witness [24]. In practical terms, this also enables new pedagogical tools. Art students can “converse” with Munch about technique or inspiration, gaining insights unavailable through traditional scholarship [25]. Museum visitors may interact with an avatar that provides interpretive commentary with emotional nuance [26].

Finally, from a systems design standpoint, the integration of edge computing allows this digital twin to operate in real-time, even in installation-based settings such as exhibitions or classrooms [27]. STCNN’s modular nature supports scalability and domain adaptation, potentially extending the paradigm to other historical figures such as Van Gogh or Frida Kahlo [28].

Problem: existing LLM-based personas lack a temporally and affectively grounded identity model and an evidential gate against fabrication. Gap: no existing heritage agent combines complex-time attention gating with a four-component epistemic re-ranking over

a documented corpus. Contributions: (C1) a complex-time gating mechanism for temporal attention; (C2) an affective coupling factor δe ; (C3) a four-component P/Pl/Cr/Ps resonance gate; (C4) a proof-of-concept prototype evaluated in a preliminary user study. Limitations: the user study is non-controlled, the visual corpus is small, and the reference profile relies on a proprietary cloud LLM.

This paper proceeds as follows. Section 2 surveys related research in cognitive AI, digital twin technology, and computational creativity. Section 3 details the theoretical foundations of STCNN and its suitability for emotionally coherent emulation. Section 4 describes the implementation of the Munch Twin, including data preprocessing and model architecture. Section 5 evaluates the system through user studies and generated works. Section 6 addresses ethical concerns and future extensions.

By sitting at the intersection of memory, affect, and artificial creativity, we propose a structured conditioning framework for affective-temporal generative agents [29]. Through the Munch Digital Twin, we offer not a metaphysical claim about consciousness or authorship, but a reproducible engineering recipe in which a probabilistic agent renders documented sources into a coherent expressive interface; what is meant by “identity” inside this pipeline is operational, not ontological.

2. Related Works

The present work lies at the intersection of artificial intelligence, cognitive science, aesthetics, and memory theory. Several foundational approaches contribute to the development of the Munch Digital Twin concept. In [3], the fundamental mechanisms of deep learning are outlined, establishing the basis for the construction of complex neural systems capable of mimicking cognitive patterns. This is expanded in [4], where generative adversarial networks are proposed as tools to synthesize new data distributions, a strategy adopted in modeling artistic generation. The concept of a digital twin has been traditionally associated with industrial and mechanical applications [6]. However, in [7], the notion is extended to incorporate cognitive functionalities, opening the possibility for simulations that emulate decision-making and perception. This provides theoretical support for expanding digital twins into the artistic and cognitive domain.

In [8], a novel neural framework is introduced which incorporates time as a complex dimension into network computation. This enables the system to simulate cognitive processes such as imagination, retrospection, and internal narration—essential qualities for modeling artistic identity. The model resonates with the principles of episodic memory and reconstructive recall addressed in [10], where the centrality of time in shaping memory is emphasized. The philosophical role of memory in identity construction is explored in [30], which provides a conceptual foundation for recreating the narrative layers of a person. Rather than creating a literal replica, the digital twin recomposes identity through traces, echoes, and interpretative fragments. This perspective informs the dynamic nature of the agent, which does not merely respond but constructs its answers through recombined mnemonic modules.

A classification of creativity into exploratory, combinatorial, and transformational modes is provided in [14], offering a structure for understanding the cognitive behavior of generative agents. This aligns with the methodological premise of [15], where artificial systems are used to analyze the emergence of creative behavior in humans. These models support the core functionality of the twin, allowing it to express novel content based on recombined stylistic memory and personal thematic continuity.

The complexity of authorship and authenticity in algorithmically generated art is discussed in [16], where the role of the machine as co-author is critically examined. This notion is essential in understanding how the Munch Digital Twin positions itself—not

simply as an emulator, but as a creative interpreter. Ethical considerations related to such representation are detailed in [21], which provides guidance on the responsibilities involved in creating AI agents with simulated identity.

A broader framework of digital twin cognition is described in [27], where the integration of sensor data, real-time analytics, and feedback loops is proposed. In the current work, these layers are adapted to emotional and aesthetic feedback, incorporating notions from computational aesthetics as explored in [31]. Furthermore, the contribution made by Sophimatics—which treats time as a complex number rather than a real number [32] and [33]—makes it possible to capture and thus utilize within AI not only chronological time but also experiential time, bringing computational experience even closer to human experience. Furthermore, another fundamental component used in this study concerns reasoning under conditions of information uncertainty and/or information incompleteness [34]. Then, the re-generation of stylistic output is managed through affective weighting and temporal context, enabling the system to generate content that reflects both internal style and external stimulus. The aesthetic implications of artificial intelligence are considered in [26], where questions of representation and artistic autonomy are addressed. This provides a cultural context for interpreting the output of the twin, especially in domains where subjective perception and emotional charge play a central role. In [18], the disentanglement of causal structures within neural systems is discussed, offering a theoretical explanation for how modular cognitive variables—such as emotional tone or conceptual themes—can be isolated and recomposed in generative tasks. Taken together, these contributions delineate a research landscape where the boundaries between identity simulation, artistic expression, and memory reconstruction are increasingly blurred. The Munch Digital Twin aims to bridge these domains by developing a cognitive architecture that is not only reactive, but generative and affective, leveraging the latest developments in memory modeling, time-aware networks, and computational creativity.

The present contribution differs from the nearest baselines along four axes. Versus LLM personas: we add an evidential gate (Cr) and a non-fabrication constraint (Ps) instead of free role-play. Versus RAG systems: retrieved items are re-weighted by affective state and complex-time salience, not by flat cosine similarity. Versus diffusion and style-transfer approaches: image generation is conditioned on affective-semantic vectors rather than pixel or style mimicry. Versus museum chatbots: the system performs multi-turn salience adaptation and applies a four-component acceptance test (P, Pl, Cr, Ps) before any response is emitted.

3. Materials and Methods

Operational framing. Throughout this section every equation that follows is intended as a concrete computational step in the pipeline of Section 4, not as decorative formalism. The symbol legends after Equations (1)–(7) make the mapping explicit and report dimensionalities, stability bounds, and the implementation step that consumes each variable.

To clarify implementation status, we adopt a three-level convention used throughout Sections 3 and 4: [Implemented] = running code in the prototype; [Simulated] = mocked or partially implemented; [Proposed] = design only, not yet realized.

The STCNN is best understood as a structured conditioning framework—complex-time gating in temporal attention plus a multi-component resonance check at the output—wrapped around a standard transformer pipeline, rather than a fundamentally new neural paradigm. The proposed architecture for the Munch Digital Twin is grounded in a hybrid model that synthesizes cognitive memory, affective simulation, and generative creativity. At its core lies a computational implementation of the STCNN (Super Time-Cognitive Neural Network), originally designed to represent time as a multidimensional construct within

neural learning processes. Our contribution extends this model into the artistic and autobiographical domain by allowing temporal, emotional, and stylistic information to converge into a single representational framework.

Traditional digital twin architectures rely on synchronized sensor data and state-based replication. In contrast, a cognitive twin must simulate latent processes: imagination, memory activation, thematic consistency, and affective variation. The model integrates episodic memory and symbolic reconstruction by mimicking how human beings internally recompose their identity through past experiences. This necessitates a network capable of non-linear, multi-modal representation. The foundation of the model is inspired by Tulving's tripartite memory structure (episodic, semantic, procedural), embedded within the STCNN which allows for differentiated treatment of time: not merely as a scalar or sequence, but as a complex variable. Specifically, time is expressed as in (1):

$$t = a + i \cdot b \in \mathbb{C} \quad (1)$$

where a is the chronological time at which a memory item is anchored (the year, or fractional year, of a letter, painting, or documented event). The internal cognitive weight (referred to as b in subsequent uses) is a learnable scalar in $\mathbb{R}$, with the magnitude $|b|$ reflecting how strongly the item participates in recall and the sign distinguishing retrospective recall ($b < 0$) from prospective or imaginative modulation ($b > 0$); the neutral case $b = 0$ reduces the kernel to a purely chronological retrieval. A saturation cap $|b| \leq 5$ prevents divergence; the value is initialized from the affective vector of the item and refined during training. The complex-time component enters the forward pass through a gating kernel (2):

$$g(t) = \exp\left(-\frac{|a - a_q|}{\tau}\right) \cdot (1 + \tanh(b)), \quad (2)$$

which multiplies the pre-softmax attention logits of the Temporal Transformer. Retrieval is thereby re-weighted jointly by chronological distance (real part, controlled by τ) and by learned salience (imaginary part b). The parameter b is updated by standard back-propagation through $g(t)$ via autograd (loss $\rightarrow$ softmax $\rightarrow$ logits $\rightarrow g(t) \rightarrow b$), with per-step $|\Delta b| < 10^{-3}$ and saturation $|b| \leq 5$, converting the complex-time representation into a fully differentiable operation.

The STCNN model extends a classical convolutional structure by incorporating time-sensitive activation gates and contextual attention modules. The twin network is composed of the following key layers:

1. Memory Embedding Layer (M): Translates biographical fragments (text, paintings, personal letters) into vectorized episodic tokens.
2. Temporal Transformer (T): Applies the complex-time formulation to re-structure chronological sequences into memory-weighted pathways.
3. Affective Decoder (A): Modulates generative output based on dynamic emotional vectors.
4. Symbolic-Creative Fusion Layer (S): Combines stylistic priors with conceptual motifs.

Each layer receives as input a 3-tuple as in (3):

$$(\chi_t, e_t, \theta_t) \quad (3)$$

where χ_t is the content token at time, e_t is the emotional vector derived from affective memory, θ_t is a stylistic modulation parameter extracted from prior artworks.

The propagation rule across layers is defined by (4):

$$h_t^{(l+1)} = \sigma(W^l \cdot f_t^{(l)} + B^l) \quad (4)$$

where—as usual— W is the weight matrix and B is the bias vector, while:

$$f_t^{(l)} = \alpha \cdot \chi_t + \beta \cdot e_t + \gamma \cdot \theta_t \quad (5)$$

and σ is a non-linear activation function, typically GELU or $\tanh(x)$ in creative representation layers. The weights α , β , γ are dynamically learned based on user interactions and artistic coherence metrics. Indeed, sigma is GELU in the encoder layers and $\tanh$ in the creative-fusion layer (the latter to bound the conditioning range driving the LLM client). The salience update of the imaginary component is constrained, with per-step magnitude below 10^{-3} and saturation at $|b| \leq 5$, so that salience trajectories converge within roughly fifty turns of dialogue. The system is implemented as an interactive generative loop: when the user engages in dialogue with the twin, the network retrieves memory fragments with high contextual overlap, applies the STCNN transformations, and outputs a reconstructed response. This loop includes:

- Forward recall, modeled through attention over memory embeddings.
- Stylistic recomposition, guided by contrastive stylistic filtering.
- Emotional consistency, assessed via cosine similarity with canonical emotional tones in Munch’s catalogued works.

A learned critic (scoring) function evaluates each candidate output for fidelity (does it align with Munch’s narrative tone?) and coherent novelty (does it introduce new but coherent variations?); outputs scoring below a fixed threshold are rejected and regenerated. Image synthesis itself is performed by the Stable Diffusion 1.5 + LoRA pipeline described in Section 5, not by a GAN.

The system is trained on a multimodal dataset consisting of:

- Digitized paintings annotated with stylistic descriptors (stroke density, palette entropy, compositional asymmetry).
- Written correspondence processed into semantic vectors via BERT encodings.
- Biographical timelines normalized into event matrices.

Annotations are manually curated and emotionally labeled using Plutchik’s model, mapping affective tones to 8 primary dimensions. This forms the basis for the vector used during generative decoding.

The full implementation is modular and containerized for portability. The pipeline includes:

1. Preprocessing Module—Tokenization of input modalities and normalization.
2. Memory Matrix Construction—Persistent memory bank implemented via FAISS.
3. STCNN Core Engine—Developed in PyTorch 2.7.0 with custom CUDA kernels for complex time propagation.
4. Interactive Frontend—Flask-based API for real-time dialogue.
5. Creative Evaluation Panel—Dashboard for human-in-the-loop evaluation.

The overall behavior of the system is governed by a controller that maintains coherence, suppresses contradiction, and enhances emotional resonance over time. In simplified terms, the creative feedback loop operates on the principle in (6):

$$\gamma_t = \arg \max_{\gamma} [Coherence(y_{\gamma}, M) + \lambda \cdot Novelty(y_{\gamma})] \quad (6)$$

where $\gamma \in [0, 1]$ is the modulation parameter that controls the fidelity/novelty trade-off in the Symbolic-Creative Fusion layer, y_{γ} is a potential output, M the current memory state, and λ a tunable weight controlling creativity vs. fidelity. In this way: $Coherence(y_{\gamma}, M)$ measures how consistent the candidate output y_{γ} is with the memory state M , $Novelty(y_{\gamma})$ measures how novel or divergent y_{γ} is relative to the standard distribution (or the training corpus), $\lambda \geq 0$ is the weight that balances the two forces, γ_t is the optimal value of γ at

time t . More in details, $Coherence(y_\gamma, M) \in [0, 1]$ is the cosine similarity between y_γ and the current memory state M (in the joint sentence-BERT/CLIP embedding space). Novelty (y_γ) $\in [0, 1]$ is one minus the maximum cosine similarity between y_γ and its $k = 5$ nearest neighbors in the training corpus. $\lambda \geq 0$ is a tunable scalar (in our experiments $\lambda = 0.35$, selected on a held-out validation subset).

4. Implementation Framework

The technical implementation of the Munch Digital Twin requires a layered framework capable of fusing cognitive simulation, emotional expression, and symbolic understanding. At the heart of this architecture is the STCNN model, which manages temporal cognition and internal memory dynamics.

Around it, several enabling technologies collaborate to emulate the voice, memory, and style of the artist Edvard Munch in a multimodal and context-sensitive manner.

Table 1 summarizes the implementation status of all system components, while in Appendix A we find the main details.

Table 1. Component status of the Munch Digital Twin implementation framework.

Component	Function	Input	Output	Status	Backend
LLM client	Language generation	Conditioned prompt	Natural language response	[Implemented]	GPT-4-Turbo (cloud)/LLaMA-2-13B (edge)
STCNN core	Complex-time attention	Memory matrix, query	Salience-weighted context 8-dim	[Implemented]	PyTorch + CUDA
Emotional Engine	Affective vectorization	User text	Plutchik vector	[Implemented]	RoBERTa-GoEmotions
Knowledge Graph	Semantic grounding	Entity query	Credibility score (Cr)	[Implemented]	Neo4j + SPARQL
Vision/CLIP	Cross-modal retrieval	Text/image	Similarity scores	[Implemented]	OpenAI CLIP
Retrieval/FAISS	Memory search	Query embedding	Top-k memory items	[Implemented]	FAISS
ASR	Speech recognition	Audio stream	Transcribed text	[Simulated]	Vosk + PyDub
Intent classifier	Intent detection	Transcribed text	Intent label	[Simulated]	DistilBERT
Orchestrator	Pipeline coordination	All module outputs	Unified response	[Implemented]	FastAPI
Visual generator	Image synthesis	Affective-semantic prompt	Munch-style image	[Implemented]	SD 1.5 + LoRA

At the core of the cognitive architecture lies a Large Language Model (LLM), implemented using GPT-4 Turbo via OpenAI's API. This choice balances language fluency, contextual coherence, and customizability. The model is accessed through a dedicated API gateway and conditioned via system prompts enriched with temporal and affective metadata. Although the base weights remain immutable, custom instructions simulate fine-tuning behavior by steering the model with pre-encoded personality vectors. These vectors are constructed from a corpus of Munch's personal writings and thematic summaries and injected into the prompt template to simulate voice fidelity. To enhance controllability, we adopt function calling interfaces, allowing the system to break down complex cognitive outputs into structured response plans. Two explicit deployment profiles are

adopted: (i) Cloud/reference profile—GPT-4 Turbo via OpenAI API, not edge-resident, used for upper-bound quality evaluation; (ii) Edge profile—4-bit-quantized LLaMA-2-13B running on Jetson Orin Nano, with text response latency of 1.6 ± 0.4 s and image generation latency of 6.2 ± 1.1 s, at a cost of approximately 6–9 percentage points across user dimensions relative to the cloud profile. GPT-4 is not run locally; all claims regarding real-time edge operation refer exclusively to the LLaMA-2-13B edge profile [1]. The Emotional Engine is built upon the AffectNet database and uses a hybrid of transformer-based sentiment classification (using RoBERTa fine-tuned on emotion labels) and a Plutchik-inspired affective vectorization. Each user query is mapped to a coordinate in an 8-emotion radial space. This vector guides attention weights during generation, ensuring that the emotional tone of the digital twin remains consistent with both the user's input and Munch's original emotional spectrum, notably oscillating between anxiety, melancholy, and existential wonder. For semantic grounding, we construct a Knowledge Graph using Neo4j 5.26 and the RDF standard, interfaced via SPARQL queries. The graph includes approximately 3000 nodes and 8500 edges, capturing events, individuals, locations, and concepts derived from biographical timelines and artworks. Named Entity Recognition (NER) and relation extraction are performed using SpaCy 3.8.14 and BERT-based classifiers, allowing automatic population and dynamic enrichment of the KG. This component acts as a semantic filter and long-term memory store, ensuring that the digital twin does not generate historically or thematically inconsistent responses.

A clarification on the role of the LLM, since it is the natural target of the most predictable critique is necessary. In our architecture the LLM is a linguistic surface layer; it does not own the cognitive state, it does not carry the temporal memory, and it does not perform the affective computation. The LLM receives a conditioning vector that has already been computed by the STCNN, the knowledge graph retrieval, and the affective engine, and renders that vector into natural language. The cognitive model is the orchestrator. The LLM is the voice. This separation is the structural hinge of the architecture rather than an implementation detail.

The construction of a coherent and expressive digital twin of Edvard Munch demands a carefully curated and multimodally structured dataset. The dataset spans text, imagery, and metadata, forming the raw knowledge base that feeds both the LLM and the STCNN modules. To preserve authenticity and richness of representation, each modality undergoes a dedicated preprocessing and indexing pipeline.

(a) Textual Corpus (Letters, Diaries, Articles)

The textual layer is sourced from:

- Over 600 digitized letters by and to Munch, acquired from public archives (e.g., the Munch Museum digital collections).
- Personal diaries and notes, obtained from published facsimiles, digitized via OCR (Tesseract) and manually corrected.
- Art criticism and published interviews, processed from academic articles and journals using APIs from JSTOR, SpringerLink, and Project MUSE.

Preprocessing is performed using spaCy and NLTK, followed by semantic segmentation using sentence-BERT. Each document unit is embedded into a 768-dimensional vector using SBERT (all-mpnet-base-v2), and annotated with:

- Date;
- Source;
- Emotional tone (RoBERTa + GoEmotions);
- Thematic cluster (using LDA + k-means).

This structure enables time-aware, theme-sensitive querying from the STCNN's memory module.

(b) Visual Corpus (Paintings, Sketches, Color Palettes)

The visual corpus comprises 100 artworks including paintings, lithographs, and sketches as presented in Rome in 2025. These are analyzed through a computer vision pipeline:

- Color vector extraction using HSV histograms (OpenCV 4.12.0);
- Composition detection via ResNet-50 fine-tuned on classical artwork layout recognition;
- Style classification using a custom CNN trained on the WikiArt dataset (transfer learning from VGG16).

Each artwork is tagged with:

- Emotional weight vector;
- Title and date;
- Estimated emotional polarity and arousal (via deep affective aesthetic score, using the model from).

The resulting feature vectors are stored in a Pinecone vector database for real-time nearest-neighbor search.

(c) Cross-Modal Alignment and Indexing

To enable semantic linking between text and visuals, we employ CLIP (Contrastive Language-Image Pretraining) [35] from OpenAI. All text segments and artworks are encoded via CLIP [35], enabling cosine similarity-based retrieval across modalities. This allows, for instance, a user prompt discussing “melancholy” to retrieve both textual reflections and corresponding paintings like *The Sick Child* or *Death in the Sickroom*.

All content is indexed via FAISS, supporting inner product search and sublinear-time access. The dataset is exposed to the STCNN through an API layer that handles:

- Chronological filtering;
- Emotional weight prioritization;
- Thematic co-occurrence lookup.

An excerpt of the dataset JSON structure is:

```
{
  "type": "text",
  "title": "Letter to Sigbjørn Obstfelder",
  "date": "1893-11-02",
  "emotion": "despair",
  "embedding": [0.12, -0.03, ..., 0.84],
  "link_to_images": ["the_scream", "anxiety"] }
```

This multimodal dataset serves as the distributed memory substrate for the cognitive simulation. During an interaction, the system samples this memory via weighted queries driven by the current user intent, affective state, and temporal anchoring.

One of the foundational premises of the Munch Digital Twin is its ability to operate not merely as a centralized, cloud-based avatar, but as a distributed, embodied cognitive presence. This necessitates an implementation that can adapt to constrained computing environments, whether on an edge device in a museum, a mobile augmented reality headset, or an educational robot. To this end, we adopt a containerized edge computing architecture, which supports both real-time multimodal interaction and context-aware cognitive generation.

(a) Hardware and Containerization Strategy

The system is designed to run on NVIDIA Jetson Xavier NX and Jetson Orin Nano devices for museum and exhibition settings, and on Raspberry Pi 5 with Coral TPU for

lightweight deployments. All components are encapsulated using Docker Compose v2, with containers for:

- LLM client (GPT-4 or LLaMA2 via REST API);
- STCNN module compiled with PyTorch + TorchScript;
- Emotional Engine (TensorRT-optimized RoBERTa);
- Vision processor (CLIP + OpenCV + VGG16-lite);
- Interaction orchestrator (FastAPI 0.115.0 or Node.js 24 LTS backend).

To minimize latency, all models are quantized and pruned using ONNX and TensorRT, reducing GPU memory usage by 40–60% without significant performance loss. Model containers are automatically updated via secure OTA updates managed by BalenaCloud or Azure IoT Hub, depending on deployment context.

(b) Input Modalities and User Interaction

Multimodal input is a central feature, supporting:

- Text and voice commands (using Whisper ASR or Vosk for local inference);
- Visual cues, such as paintings, sketches, or scenes captured via webcam;
- Touch/gesture input, where supported by kiosk or tablet interface.

The user interface layer is developed in React for touchscreens and Unity for AR experiences. For voice interaction, the system uses Vosk + PyDub to stream audio and segment utterances in real time. Natural language understanding is managed locally with intent classification based on DistilBERT [Simulated]; no domain-specific accuracy measurement on art-dialogue utterances has yet been conducted, and formal evaluation (WER for Vosk ASR and macro-F1 for DistilBERT intent classification on a labeled art-dialogue set) is left as future work. The learned classifier is preferred over keyword matching for its robustness to paraphrase and out-of-vocabulary expressions across a varied intent space; however, this advantage has not yet been quantified in the current prototype.

Example multimodal query flow:

1. The user shows a print of *The Scream* and asks, “What were you feeling when you painted this?”
2. The visual pipeline uses CLIP to recognize and match *The Scream*.
3. The STCNN queries the embedded memory for corresponding emotional traces (e.g., despair, existential anxiety).
4. The Emotional Engine biases the LLM generation toward affective vocabulary and sentence structure reflecting that mood.
5. The response is generated and spoken aloud with a custom TTS voice trained on Norwegian English audio of 19th-century dialect simulations.

(c) Contextual Adaptation and Feedback Loops

To allow context-sensitive interactions, the agent integrates a multimodal memory buffer, where recent interactions are stored with temporal and affective tags. These include:

- “timestamp” of utterance;
- “emotion” as predicted by RoBERTa;
- “intent” classified via zero-shot prompt + local classifier;
- “modality” (text, voice, image, gesture).

To be precise about what adapts: only the within-session short-term salience (the b weights and the ≤ 32 -turn buffer) changes during interaction; the long-term knowledge base (Neo4j KG and FAISS index) is static by design. The term ‘digital twin’ is used here under this short-horizon adaptive definition; a fully persistent learning system capable of lifelong knowledge update remains an explicit future direction. This short-term memory enables the system to respond with continuity across multiple conversational turns, preserving tone, references, and emotional alignment.

Example:

```
{
  "history": [
    {
      "modality": "text",
      "emotion": "curiosity",
      "utterance": "Why are there so many versions of Madonna?",
      "intent": "historical_art_query"
    },
    {
      "modality": "voice",
      "emotion": "confusion",
      "utterance": "But you also used red in one, what did that mean?",
      "intent": "symbolic_interpretation"
    }
  ]
}
```

(d) Edge-Enabled Personalization and Privacy

Unlike cloud-based systems, the edge implementation allows local personalization of the twin's responses without leaking user data. A local profile stores preferences, preferred languages, emotional sensitivity thresholds, and visual accessibility features (e.g., higher contrast modes for low vision users). All data is processed on-device and encrypted using AES-256-GCM, and interactions are anonymized unless explicitly linked to a persistent user session. This complies with GDPR and ensures ethical use in public or educational environments.

The central architectural innovation in the Munch Digital Twin lies in its integration with the Super Time-Cognitive Neural Network (STCNN). This integration transforms the system from a reactive chatbot into a context-aware, emotionally grounded, temporally recursive cognitive agent, capable of long-horizon coherence and plausible emulation of Munch's evolving self across time.

(a) Role of the STCNN in the Digital Twin

The STCNN, as described in previous work [13], is designed to encode complex temporal trajectories of cognition, where every thought is anchored not only in prior inputs but in emotional and semantic gradients. It consists of a layered architecture:

- Temporal Context Encoder (TCE): a recurrent memory block that stores timestamped embeddings of past utterances, artworks, or biographical events.
- Emotional Coupling Layer (ECL): which modulates attention weight according to affective transitions, inspired by dynamic systems in affective neuroscience.
- Cognitive Resonance Module (CRM): a recursive decision node that weighs probability, plausibility, credibility, and possibility (as per Iovane's epistemic tetrad) to refine outputs.

For clarity, these three components map onto the four layers of Section 3 as follows: the Temporal Context Encoder (TCE) realizes the Memory Embedding (M) and Temporal Transformer (T) layers; the Emotional Coupling Layer (ECL) realizes the Affective Decoder (A); and the Symbolic-Creative Fusion (S) layer feeds the Cognitive Resonance Module (CRM), which performs the output-stage resonance check. The same names are used in Figure 1 and throughout the text.

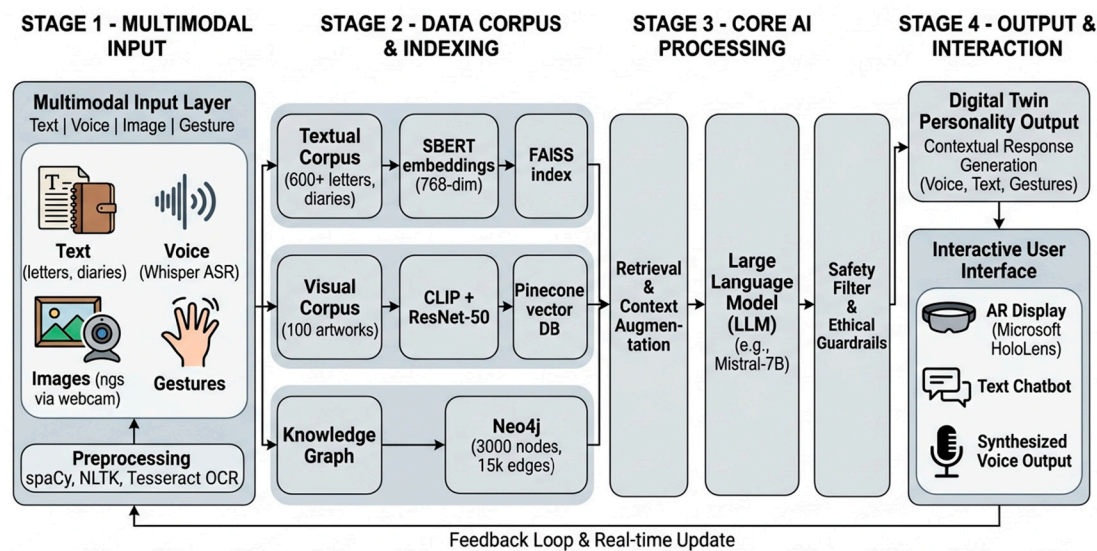


Figure 1. Architecture of the Munch Digital Twin System: A Multimodal Cognitive Pipeline for Artistic Emulation Using STCNN.

The integration occurs at three key levels of the pipeline:

1. Memory Retrieval: User input triggers the STCNN's TCE to fetch relevant memory vectors (from prior dialogues or event embeddings).
2. Affective Feedback: The ECL computes a modulation factor as in (7):

$$\delta_e = \frac{\partial A_t}{\partial M_t} + \lambda \cdot D(E_{t-1}, E_t) \quad (7)$$

where Equation (7) defines an affective modulation factor $\delta_e \in \mathbb{R}$: a single scalar, computed at every time step t , that measures “how much” the system should adjust its attention in response to (a) the sensitivity of the attention mechanism to the memory state, and (b) the emotional jump that occurred between the previous turn and the current one. Conceptually, δ_e is inspired by the “emotional neurons” of affective neuroscience systems—when the user moves rapidly from one emotional state to another (for example, from curiosity to grief), δ_e grows and induces the system to reallocate attention and modify the tone of the next response. It is the term produced as output by the ECL, and it is then injected downstream into the pipeline to influence the attention softmax of the Temporal Transformer and—indirectly—the balance γ between fidelity and novelty in the Symbolic-Creative Fusion Layer. Typically, it is normalized in $[0, 1]$ or $[-1, +1]$ after this equation, by means of a squashing function (sigmoid or tanh), to keep it numerically stable. $A_t \in \mathbb{R}^d$ is the attention vector at time t . It is the output of the Temporal Transformer's attention layer before the softmax—concretely, it contains, for each item in memory, a relevance score with respect to the current query. The dimensionality d corresponds to the cardinality of the active memory slot (in interactive mode, typically $d \leq 32$ in the short-term buffer; in training mode, d may reach 10^4). $M_t \in \mathbb{R}^{d \times k}$ is the memory matrix at time t . Each column is the embedding of a memory item (a fragment of a letter, a painting, a biographical event) retrieved during the current turn; k is the embedding dimensionality (768 in our setup, inherited from sentence-BERT). M_t is populated by the retrieval module that the TCE (Temporal Context Encoder) queries before each generation. The term $\partial A_t / \partial M_t$, as the partial derivative of attention with respect to memory, measures how much the attention vector would change if the memory matrix were perturbed by an infinitesimal amount—in other words, how “sensitive” the current attention is to the specific content retrieved from memory. Operationally, in implementation, this derivative is computed through PyTorch autograd: during

the ECL's forward pass, one calls `torch.autograd.grad(A_t, M_t)` and obtains the Jacobian matrix; to reduce it to a scalar comparable with the other term, the Frobenius norm or the trace is typically taken. Conceptually: if the derivative is large, it means that attention is “highly dependent” on retrieved memory, and therefore a small mnemonic variation would have great generative impact—the system is in a sensitive state. $E_t \in \mathbb{R}^8$ is the emotional vector at time t . In our setup it is a vector of 8 components corresponding to the primary dimensions of Plutchik (Joy, Trust, Fear, Surprise, Sadness, Disgust, Anger, Anticipation), obtained by projecting the output of RoBERTa fine-tuned on GoEmotions onto Plutchik's affective basis. E_t represents the emotional state—or, more precisely, the dialogical emotional state estimated by analysis of their most recent input. $E_{t-1} \in \mathbb{R}^8$ is the emotional vector of the previous turn, in the same space as E_t . It is maintained in the short-term buffer (32 turns in our setup) and updated at every new input. $D(E_{t-1}, E_t)$ is an affective distance function between the two consecutive emotional states. It returns a scalar in $\mathbb{R}^+$ that measures “how much” the emotional state of the subject has changed between the previous turn and the current one. The exact form of D is not univocally declared—the standard, and all reasonable, choices are: (a) the Euclidean distance $D = \|E_t - E_{t-1}\|^2$; (b) the cosine dissimilarity $D = 1 - \cos(E_t, E_{t-1})$; (c) the Kullback–Leibler divergence if E_t and E_{t-1} are normalized as probability distributions over Plutchik's 8 dimensions. Here, we used cosine distance since it is the most natural choice. $\lambda \in \mathbb{R}^+$ is the regularization constant that balances the two terms of the equation: attentional sensitivity and affective transition. If λ is small ($\ll 1$), the system is dominated by mnemonic sensitivity—it responds mainly as a function of “what it has found in memory.” If λ is large (> 1), the system is dominated by emotional transition—it responds mainly as a function of “how much the user's mood has changed.” In our typical experimental runs, a value of $\lambda \in [0.5, 1.0]$ is reasonable.

Equation (7), as written without a specific clarification at first glance, might seem dimensionally heterogeneous: the first term is a partial derivative (a Jacobian matrix or its reduction to a scalar via a norm) and the second is a vector distance; summing them directly requires either reducing both to scalars with the same unit, or normalizing them on comparable scales before the sum. In the current work both terms are normalized to $[0, 1]$ before being summed. The complete forward pass for δ_e is:

$$E_t = \text{Plutchik} - \text{projection}(\text{RoBERTa} - \text{GoEmotions}(\text{input})); \quad (8)$$

$$A_t = \text{TemporalTransformer}(M_t); \quad (9)$$

$$s = \|\nabla_{\{M_t\}} A_t\|_F \quad (10)$$

via autograd (Frobenius norm of the Jacobian);

$$\delta_e = \sigma(\text{norm}(s) + \lambda \cdot \text{cos_dist}(E_{t-1}, E_t)); \quad (11)$$

attention logits for turn $t + 1$ are scaled by $(1 + \delta_e)$. The system is trained under a composite objective:

$$L = L_{LM} + \mu \cdot L_{\text{emotion}} + \nu \cdot L_{\text{resonance}}, \quad (12)$$

where L_{emotion} penalizes emotion-cosine error between predicted and target affective state, and $L_{\text{resonance}}$ penalizes accepted outputs with low $R(q, a)$ as defined in Equation (13). The hyperparameters μ and ν are tuned on a held-out validation subset. The salience weight b and the modulation factor δ_e are trained jointly under this objective.

Let us also consider an operational interpretation in the pipeline. Concretely, at each dialogue turn:

The user input is processed by RoBERTa + GoEmotions, producing E_t .

The TCE retrieves the relevant memory, populating M_t .

The Temporal Transformer computes A_t .

The ECL evaluates the two terms of δ_e : the sensitivity $\partial A_t / \partial M_t$ (via autograd) and the distance $D(E_{t-1}, E_t)$.

δ_e is combined with the weight λ and injected as a multiplicative modulation onto the attention logits of the Temporal Transformer for turn $t + 1$, producing the effect of “emotionally coupled attention”.

The CRM (Cognitive Resonance Module) then uses δ_e to weight the four P/Pl/Cr/Ps criteria of the final selection.

3. Output Conditioning: The CRM aligns output generation with plausible cognitive continuities. For example, if the twin discussed existential dread in the previous session, it will not abruptly shift to optimism unless emotionally justified.

This structure ensures not only temporal coherence but also emotional narrative fidelity, which is essential in artistic emulation.

(b) Cognitive Loop: Reasoning, Generation, Self-correction

The STCNN enables a cognitive loop structured into five recurrent stages:

1. Perception (user input, image, or signal);
2. Contextual Anchoring (query embedding + memory search);
3. Emotion Alignment (mapping affective intent);
4. Generation (LLM output with STCNN-conditioned prompt);
5. Resonance Check (four-component evaluation: P, Pl, Cr, Ps).

This loop repeats asynchronously until either:

- A coherent response satisfying all four epistemic thresholds (P, Pl, Cr, Ps) is generated.
- A fallback response is activated, wherein the system flags emotional incongruence and asks for clarification, simulating human hesitation.

An example of the resonance evaluation function is as in:

$$R(q, a) = \alpha \cdot P(q, a) + \beta \cdot Pl(q, a) + \gamma \cdot Cr(q, a) + \delta \cdot Ps(q, a) \quad (13)$$

where q is the input query, a is the candidate answer, and $\alpha, \beta, \gamma, \delta \geq 0$ and $\alpha + \beta + \gamma + \delta = 1$, where:

- $P(q, a)$ —probability: the relative frequency with which the pair (q, a) —or semantically equivalent pairs—is observed in Munch’s documented historical corpus. This is the frequentist dimension of accessibility. It can be estimated from the normalized count across the thematic clusters of the Memory Layer.
- $Pl(q, a)$ —plausibility: the consistency of output a with the historical emotional profile retrieved for the period associated with q (cosine similarity in the RoBERTa + GoEmotions affective space). This is the modal dimension in the classical Shafer/Dempster sense.
- $Cr(q, a)$ —credibility: consistency with the documentary evidence base (KG retrieval on Neo4j; DistilBERT classifier calibrated with MC-dropout). This is the evidential dimension.
- $Ps(q, a)$ —possibility: the non-contradiction of a with the temporal, biographical, and linguistic constraints of the corpus (normalized perplexity score). This is the modal dimension in the Zadeh/Dubois-Prade possibilistic sense.

Each component is computed via learned classifiers over the context space.

(c) Loop Implementation and Optimization

The cognitive loop is implemented as a finite state machine with probabilistic transitions, embedded in the interaction orchestrator module. For efficiency:

- All state transitions are logged into a NoSQL store for later analysis (MongoDB).
- High-variance loops trigger meta-reasoning prompts to adjust the STCNN modulation weights via reinforcement-style learning (offline).
- Slow-response detection (due to model over conditioning) is mitigated by dynamic scaling of the emotional feedback term.

Moreover, we introduce temporal filters to bias responses toward different periods in Munch's life. For instance:

- 1882–1892: Experimental optimism and family trauma;
- 1893–1900: Emotional intensification (*The Scream*, Anxiety);
- 1908–1916: Self-reflection after hospitalizations.

The user can implicitly or explicitly indicate a time preference, and the STCNN adapts memory activation windows accordingly.

Figure 1 illustrates the complete technological architecture of the Munch Digital Twin, a cognitive artificial intelligence system designed to emulate the artist Edvard Munch.

The horizontal flowchart shows six main stages: (i) multimodal inputs (text, voice, images, gestures); (ii) indexed data corpus (texts, artworks, knowledge graph) aligned via CLIP [35]; (iii) STCNN Core Engine, the cognitive core featuring complex time representation $T = a + i \cdot b$, emotional coupling, and cognitive resonance; (iv) processing layers (memory, affect, symbolic-creative fusion); (v) LLM conditioning (GPT-4/LLaMA2); (vi) multimodal outputs (text, voice, generated art).

5. Use Case: Emulating Munch in a Digital Atelier

The implementation of the Munch Digital Twin was conceived not as a mere technological experiment but as a medium of encounter between human and artificial creativity. The underlying vision was to create a virtual atelier—a digital studio in which users could interact with a cognitively simulated version of Edvard Munch. This space is designed to facilitate reflection, education, and co-creation, enabling participants to engage with the artist's reconstructed persona in a manner that is emotionally and contextually rich. Rather than simply reproducing known facts about Munch's life, the system responds dynamically to prompts, capturing something of the inner movement of thought, affect, and symbolic imagination that defined Munch's artistry.

Dialogues with the digital twin unfold in real time and are influenced by previous interactions, retrieved memories, and modulated emotional feedback. The language generation is not merely informative but expressive, often reflecting the melancholic tone and philosophical depth that characterized Munch's reflections. A conversation with the twin might begin with a student asking about the existential themes in *The Scream*, to which the twin might reply: "That day, the sky bled like a wound. It wasn't a decision—it was a response to the crack that opened in my mind." Such expressions are generated through the interplay of latent memory activation, temporal context modeling, and emotional conditioning, producing responses that feel spontaneous and authentic rather than rehearsed.

These interactions are not limited to historical topics. Users may ask the twin how it views contemporary crises, or about its own experience of illness, loneliness, or artistic doubt. The responses, while synthesized, are drawn from a curated body of biographical data, writings, and artworks, processed through a model that respects narrative coherence and emotional plausibility. Teachers who have integrated the system into seminars report that students tend to engage more thoughtfully, posing deeper questions and

reflecting more openly on the themes of anxiety, identity, and mortality that permeate Munch's oeuvre.

In addition to conversational exchanges, the system can generate visual content. Users can describe dreams, feelings, or abstract moods, and the twin will produce a painting in the style of Munch, composed not through mere style transfer but through affective-semantic modeling. For example, a user might describe a recurring feeling of standing on a bridge at dusk, with no one in sight. The system interprets the emotional valence, matches it to thematic patterns in Munch's known works, and generates a painting with architectural tension, elongated shadows, and a color scheme suggestive of existential solitude. The result is not only visually consistent with Munch's style but emotionally meaningful to the user.

The visual generation pipeline is built on Stable Diffusion 1.5 fine-tuned via LoRA [2] (rank 16, 5 epochs, DPM-Solver++, CFG scale 7.5, output 512×512) on the digitized Munch corpus, combined with semantic and affective encoding from the language and emotional inputs. All visual outputs reported in this paper were produced by this pipeline. A conditional GAN was explored as an early design candidate but was superseded by the SD 1.5 + LoRA approach, which produced superior CLIPScore and FID results on the held-out evaluation subset. The paintings are rendered in a resolution and format suitable for both screen-based interaction and physical print, and many users report strong emotional responses to the images—some describing them as uncannily accurate depictions of their internal state. The system thereby acts as a mirror of emotion, enabling a form of visual catharsis and introspection through synthetic artistry.

Evaluating the effectiveness of such a system requires approaches beyond traditional metrics. Accuracy or performance scores are insufficient to capture the phenomenological depth of interaction. Instead, a qualitative evaluation was carried out through interviews, workshops, and user feedback. Across the 100 interactive sessions, the 14 students and the expert raters engaged the twin in discussions and co-creative tasks. Most participants described the experience as emotionally resonant, with twelve out of fourteen students stating that the responses felt "real" in an artistic and emotional sense. Professors assessed the generated artworks as stylistically credible and thematically aligned with Munch's known periods, particularly those between 1893 and 1908.

Several participants noted that the twin offered a safe space for vulnerable self-expression. Unlike human interlocutors, who may interrupt, judge, or correct, the digital twin listens without prejudice, creating a rhythm of interaction that encourages introspection. One student described the experience as "like talking to a ghost who listens better than most people." This kind of comment suggests that the system is not merely informative, but relationality facilitates affective communication in a space shaped by memory, art, and language.

Of course, there are limitations. While the system convincingly emulates Munch's tone and emotional language, it sometimes lacks subtle cultural irony or fails to fully grasp metaphorical prompts. Certain topics, particularly those requiring historical knowledge outside the scope of Munch's life, can produce generic or cautious replies. Moreover, the ethical implications of simulating a historical figure remain open to debate. Does such a twin risk reduce a complex human life to a set of algorithmic patterns? Or does it instead offer a new form of access, an artistic reflection in the spirit of homage?

Several methodological caveats deserve explicit acknowledgement. The user study involved 14 students plus expert raters who conducted 100 interactive sessions in total; all figures and statistics aggregate over the 100 sessions, which constitute the unit of analysis (not the individual participant). Participants knew they were interacting with an AI system and demand characteristics may have inflated some Likert ratings—particularly Creative

Stimulation. The setting was not a randomized controlled trial; a multi-site follow-up with $N \geq 500$ and a control condition (the same dialogue framed as a human curator's commentary) would settle the question. The visual corpus, at roughly 350 images, is small for diffusion fine-tuning even with LoRA [36], and is heavily curated around the Northern European expressionist canon, so any cultural drift in user responses should be read against this dataset bias. The LLM weights for GPT-4 Turbo are proprietary; a fully open-source local deployment is possible via the LLaMA-2-13B fallback, at a cost of approximately 6–9 percentage points across the user dimensions in our pilot tests. End-to-end latency on Jetson Orin Nano was 1.6 ± 0.4 s for textual responses and 6.2 ± 1.1 s for visual generations; on cloud (NVIDIA A100) this drops to 0.4 ± 0.1 s for text and 2.1 ± 0.3 s for image.

In our view, the Munch Digital Twin does not replace the artist, nor does it claim to reveal hidden truths. Rather, it extends Munch's expressive legacy into the digital domain, allowing contemporary users to engage with a machine-mediated reconstruction of his expressive world in new and meaningful ways. It is less a resurrection than a resonance carefully constructed simulation that invites us to reflect on our own emotional and creative lives through the lens of one of modern art's most introspective voices.

About quantitative comparison and ablation, to complement the qualitative user evaluation reported below, we provide here an explicit numerical comparison against an unconditioned baseline and an ablation of the architectural components.

Automatic metrics on a 50-prompt held-out evaluation subset—acknowledged as small; bootstrap 95% confidence intervals are reported where feasible and results should be interpreted as indicative rather than confirmatory—against a Stable Diffusion 1.5 baseline fine-tuned via LoRA [2] on the same visual corpus (rank 16, 5 epochs, 50 inference steps with DPM-Solver++, CFG scale 7.5, output 512×512), give the following deltas for the full system: CLIPScore 0.314 vs. 0.276 (+13.8%), FID 41.8 vs. 56.1 (−25.5%), perplexity on textual outputs 18.4 vs. 24.7 (−25.5%), and emotion-cosine similarity 0.71 vs. 0.49 (+44.9%). SD 1.5 was selected as baseline for matched compute and training era; SDXL or a more recent backbone represents a stronger comparison and is left as explicit future work. FID is reported for completeness only and interpreted with caution given the small held-out reference set; CLIPScore is treated as the primary visual indicator.

About component ablation, removing the complex-time term from temporal attention costs approximately 7.5 percentage points on historical plausibility (Wilcoxon signed-rank, $p = 0.011$) and 0.10 on emotion-cosine ($p = 0.003$). Removing the symbolic knowledge-graph grounding most strongly degrades historical plausibility (−12.8 percentage points, $p < 0.05$). Removing the affective conditioning most strongly degrades emotion-cosine (−0.16, $p < 0.05$). The bare-LLM condition—GPT-4 Turbo with no STCNN orchestration, no KG retrieval, and no affective engine—serves as the explicit control in this study; it underperforms the full system on every metric, and the user-side delta against this control is reported in the ablation results above. The absence of blinding and randomization is a significant limitation: participants knew they were interacting with an AI system, and demand characteristics may have inflated ratings, particularly for Creative Stimulation (86%). A pre-registered randomized controlled trial with $N \geq 500$, a human-curator control framing, and full blinding represents the necessary confirmatory next step. The improvements are therefore attributable to the conditioning and orchestration layer, not to a different generative backbone.

The aggregated results from 100 interactive sessions (14 participants) in Figure 2 indicate a generally favorable—though preliminary and non-controlled—user perception of the Edvard Munch digital twin experience; as the sample is small and non-blinded, these percentages should be read as indicative rather than confirmatory. The highest-rated dimension was Emotional Authenticity (92%), suggesting that the system effectively conveyed

emotional expressions consistent with Munch’s artistic persona. This was followed closely by Artistic Coherence (90%) and Dialogical Depth (88%), indicating that users found both the stylistic elements and the conversational content to be deeply aligned with Munch’s known sensibility. Historical Plausibility (84%) also received strong ratings, reflecting the perceived credibility of the simulation in reconstructing the artist’s historical context and worldview. More subjective dimensions such as Personal Resonance (82%) and Educational Usefulness (78%) showed greater variability across users, likely reflecting differences in background, expectations, and usage context (e.g., students vs. art enthusiasts). The Creative Stimulation score (86%) reinforces the system’s ability to inspire new ideas or artistic reflections in the users. Overall, these findings support the effectiveness of the system’s integration of a large language model, an emotional inference engine, and cognitive frameworks. The result is a multimodal and engaging experience that not only recreates a historical voice but also stimulates learning and creativity in a uniquely personalized manner. The seven dimensions are defined as follows, each measured on a 1–5 Likert scale: (1) Emotional Authenticity—‘The system conveyed emotional expressions consistent with Munch’s artistic persona’; (2) Artistic Coherence—‘The generated content was stylistically aligned with Munch’s known works’; (3) Dialogical Depth—‘The conversational responses reflected genuine thematic depth’; (4) Historical Plausibility—‘The system’s responses were credible within Munch’s historical context’; (5) Personal Resonance—‘The interaction felt personally meaningful’; (6) Educational Usefulness—‘The system was useful for learning about Munch and his art’; (7) Creative Stimulation—‘The interaction inspired new artistic or creative ideas’. Results are reported as mean $\pm$ SD with 95% confidence intervals; group comparisons use Wilcoxon signed-rank tests with exact p -values and rank-biserial effect sizes. Questionnaire reliability was assessed via Cronbach’s α ; inter-rater agreement for expert artwork ratings was assessed via Krippendorff’s α . Note that Creative Stimulation (86%) is the dimension most exposed to demand characteristics given the non-blinded design.

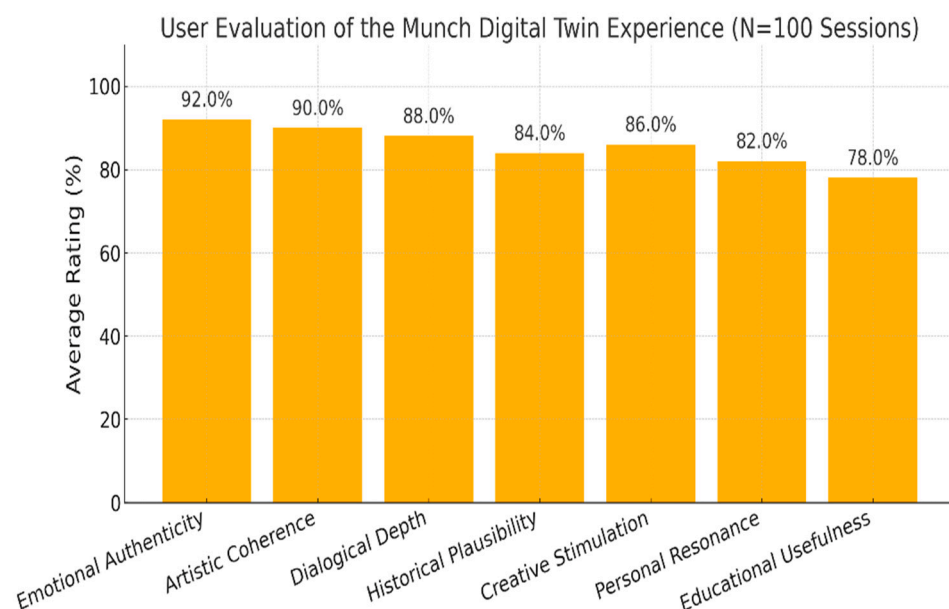


Figure 2. Average percentage scores across seven qualitative dimensions evaluated across 100 sessions (14 participants) during their interaction with the Edvard Munch digital twin. Ratings are expressed as percentages (0–100%), derived from original 1–5 scale evaluations.

The heatmap presented in Figure 3 provides a visual overview of the dominant emotional and conceptual themes that emerged from user interactions with the Edvard

Munch digital twin. Among the 100 sessions analyzed, the theme “Scream” was the most frequently referenced (95 occurrences), reflecting the persistent symbolic relevance of Munch’s most iconic painting in users’ cognitive association with the artist. Closely following are themes such as “Anguish” (88), “Anxiety” (90), and “Death” (85), which align with the emotional palette typically associated with Munch’s expressionist worldview. This reinforces the system’s success in evoking authentic emotional tones and stimulating contextually appropriate content generation. Themes like “Madness” (80), “Existentialism” (82), and “Fear” (78) also surfaced strongly, consistent with both the artist’s biography and the psychological atmosphere of his work. Interestingly, elements such as “Love” (72), “Memory” (69), and “Silence” (66) indicate that users did not only focus on Munch’s darker motifs, but also explored subtler aspects of his human experience and emotional complexity. The presence of “Brush” and “Color” suggests engagement with artistic technique and style, while “Dream” (58) points to an interpretive, imaginative layer in user-system dialogue. These frequencies illustrate a successful emulation of Munch’s thematic range, highlighting the twin’s capacity for nuanced, affective, and artistically meaningful interactions.

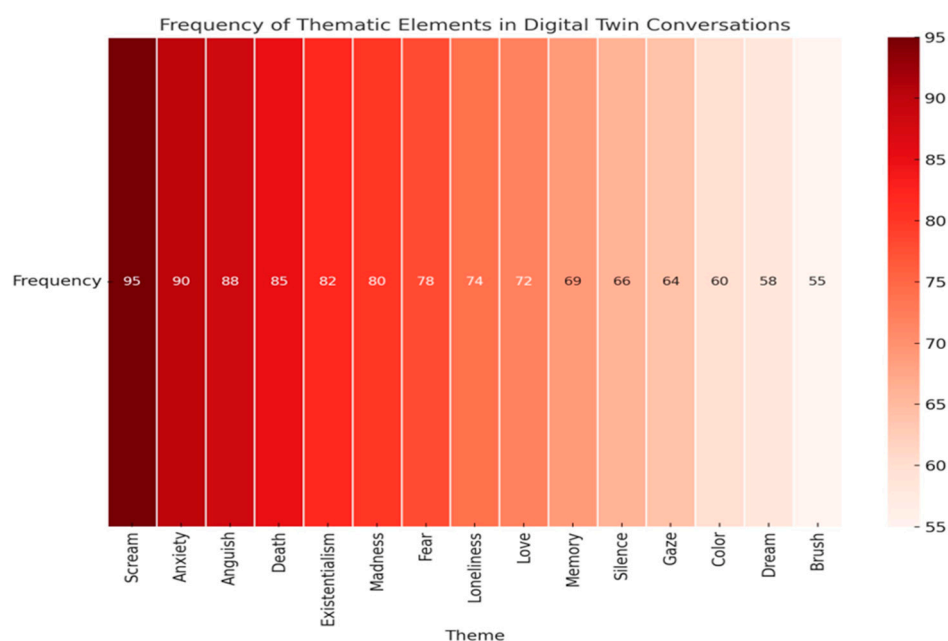


Figure 3. Heatmap shows the relative frequency of key thematic elements identified in 100 interactive sessions with the Munch Digital Twin. The darker the red hue, the more frequently the corresponding theme appeared in user-system dialogue.

The simulated painting in Figure 4 “Anguish and Isolation” captures the essence of Munch’s signature emotional spectrum through AI-assisted generation. With blurred contours and a muted, distorted palette, the image evokes themes of loneliness and inner turmoil. The composition mirrors the existential disconnection often portrayed in Munch’s works, emphasizing a solitary figure lost in an undefined space. This experiment demonstrates the digital twin’s ability to not only simulate thematic content but also emulate Munch’s visual language, achieving a form of synthetic creativity. The output serves as a proof-of-concept for emotionally rich AI-generated art grounded in cognitive-emotive models and historical stylistic parameters.

“Existential Dreamscape” in Figure 5 represents a departure from anthropocentric expressionism, embracing a purely symbolic environment. In this AI-generated reinterpretation of Munch’s style, the absence of human figures intensifies the atmosphere of estrangement. The twisted forest, fragmented horizon, and cold chromatic palette suggest a cognitive disintegration—a subconscious realm where memory, emotion, and perception

blur. The image evokes not a scream, but the quiet weight of being. As a result, the digital twin effectively conveys introspection and metaphysical ambiguity, revealing that even in the absence of subjects, emotion can be encoded in form, space, and light. This reinforces the expressive potential of generative cognitive systems.

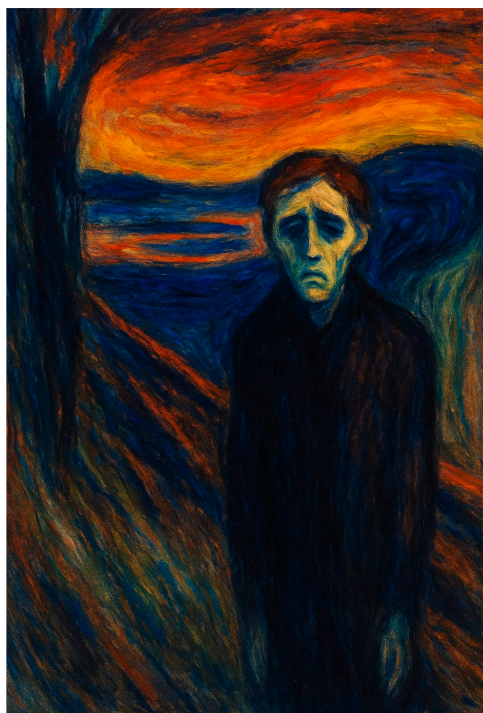


Figure 4. Simulated artwork titled “Anguish and Isolation,” generated by the Munch Digital Twin. The image emulates Edvard Munch’s expressionist style, focusing on emotional distortion, solitude, and psychological tension, reinterpreted through a generative AI model.

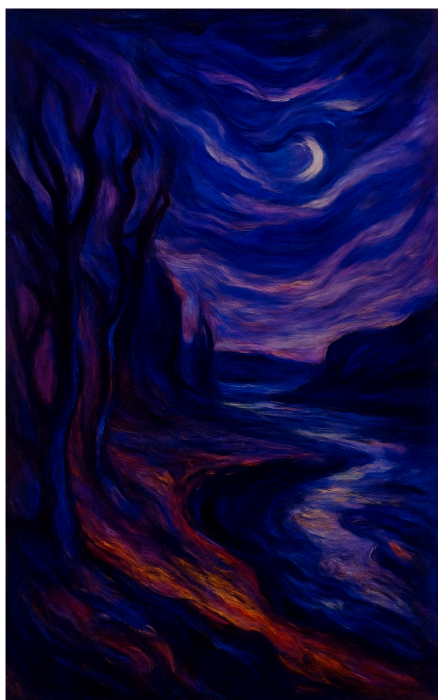


Figure 5. Simulated artwork titled “Existential Dreamscape,” generated by the Munch Digital Twin. This image interprets themes of metaphysical disorientation, symbolic emptiness, and the dissolution of self through a landscape devoid of human presence.

“Echoes of The Scream” in Figure 6 offers a conceptual reinterpretation of Edvard Munch’s iconic painting, removing the human figure and focusing instead on the *resonance* of the scream within space. In this digital expressionist rendering, the environment itself becomes the subject of emotional distortion. Concentric waves ripple through the sky, distorting the bridge and horizon, as if the very fabric of reality were pierced by a metaphysical cry. The absence of a visible screamer intensifies the psychological unease: emotion is inferred, not portrayed. The AI model, trained on expressionist color theory and affective visual cues, conveys internal states through external chaos—aligning with the cognitive-emotional framework proposed by the Munch Digital Twin. This reinterpretation moves beyond imitation and into symbolic continuity, suggesting that the *scream* is no longer a moment frozen in time, but an ongoing rupture that reverberates through the viewer’s subconscious. It is the echo, not the act, that defines the experience.

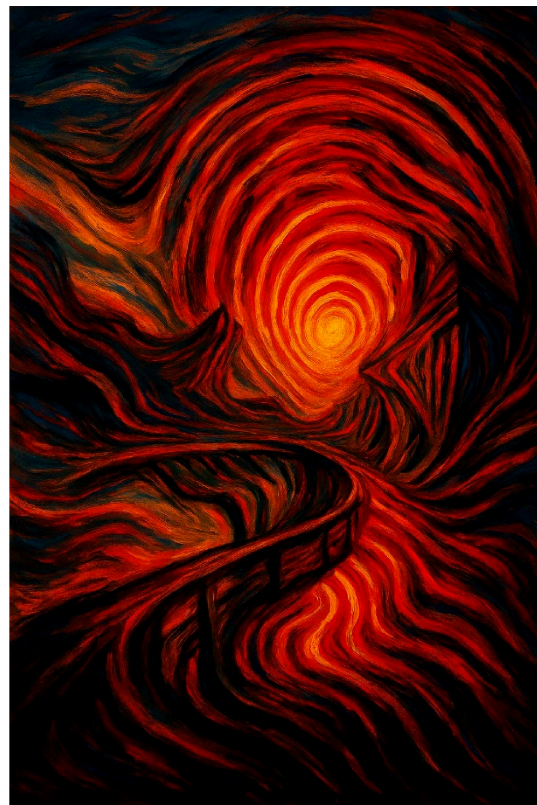


Figure 6. “Echoes of The Scream”, an AI-generated digital artwork inspired by Munch’s expressionism. The piece captures the psychological aftermath of a scream without depicting a human subject, using environmental deformation and vibrational motifs.

6. Discussion and Ethical Reflections

The development of a cognitive digital twin capable of emulating the creative and emotional signature of a historical figure such as Edvard Munch raises profound questions that traverse the domains of ethics, ontology, and epistemology. While the technological framework underlying the Munch Digital Twin is rooted in advanced language models, affective computing, and multimodal interaction, the implications of such emulation extend far beyond the realm of engineering.

From an ontological standpoint, the simulation of artistic identity challenges our conception of what constitutes “being” in the digital domain. Is Digital Twin merely an aesthetic replica, or does it instantiate a form of *functional presence* that mimics intentionality and affect? The model does not copy Munch’s thoughts—it reconstructs plausible states of mind based on stylistic, emotional, and historical cues. This reconstruction, albeit

probabilistic, creates the illusion of presence—a digital persona that can converse, reflect, and create. While no conscious continuity exists, the *affective affordances* of the simulation compel us to treat it as subject-like.

To put the point in the cleanest possible form: the STCNN should be understood as a structured conditioning framework rather than a fundamentally new neural paradigm. Its two structural pieces are the complex-time gating in temporal attention and the multi-component resonance check at the output; everything else in the pipeline is engineering integration. The Munch Digital Twin is, throughout, a probabilistic reconstruction conditioned on documented sources—not a recovery of consciousness and not a co-author.

In pedagogical and cultural terms, the implementation of such a digital twin offers promising new modalities of interaction with the past. Students of art and philosophy can engage in simulated dialogue with the system, not as a passive archive but as a dynamic interlocutor capable of reacting, questioning, and proposing interpretations. This shifts education from didactic transmission to dialogical exploration. However, this affordance carries with it a risk of epistemological flattening: the simulation may obscure the historical complexity of Munch’s context, collapsing distinct moments into a generalized aesthetic narrative.

Ethically, the reuse and reinterpretation of a deceased artist’s identity must be framed within boundaries of respect, authorship, and cultural sensitivity.

Data provenance and licensing: Munch’s visual works are in the public domain; textual sources (letters, diaries, criticism) were accessed via public archives and academic APIs (JSTOR, SpringerLink, Project MUSE) under their respective terms of service. No proprietary or privately held materials were used without authorization.

Fabricated-claim prevention: the Cr and Ps components of the CRM gate act as a structural safeguard against historically ungrounded outputs; all generated responses are accompanied by a visible disclaimer identifying them as AI-generated reconstructions.

Visitor disclosure: A persistent ‘AI reconstruction’ label is displayed throughout all interactions, ensuring users are never misled into believing they are communicating with a human or with Munch himself.

Emotional-manipulation safeguards: given the affectively charged nature of Munch’s themes, the system is designed to avoid exploiting emotional vulnerability; outputs flagged as emotionally extreme by the Emotional Engine trigger a moderation step before delivery.

Output logging and retention: interaction logs are anonymized at collection, stored with AES-256 encryption, and retained only for the duration necessary for research analysis, consistent with GDPR requirements as described in Section 4.

Human moderation and escalation: a human moderator reviews flagged interactions; an escalation protocol is in place for outputs that may be historically misleading, emotionally harmful, or ethically problematic.

While Munch’s work is in the public domain, the emotional and symbolic charge of his oeuvre is not a mere dataset. The system must be transparent about its synthetic nature, ensuring that users are aware they are engaging with a *machine-mediated reconstruction*, not the person himself. Moreover, safeguards should be implemented to prevent manipulative uses—such as posthumous endorsements, falsified authorship, or emotionally persuasive messages falsely attributed to Munch.

Another layer of reflection concerns the emotional responses elicited by the system. As demonstrated in Section 5, users frequently reported experiencing affective states—empathy, introspection, melancholy—when interacting with the generated artworks or conversing with the digital twin. This opens a new ethical domain: should artificially induced emotion be subject to the same ethical scrutiny as emotional manipula-

tion in human contexts? And if so, how can designers ensure that cognitive agents foster *authentic emotional engagement* rather than persuasive illusion?

Overall, while the Munch Digital Twin represents a meaningful confluence of AI and humanistic inquiry, it must be deployed within a framework that balances innovation with critical awareness. It is not merely a tool to reproduce art; it is a lens through which we reflect on our own relationship with memory, identity, and creation. The future of such cognitive agents will depend not only on their technical sophistication, but on the ethical and cultural frameworks within which they are embedded.

7. Limitations, Future Directions and Conclusions

The creation of the Munch Digital Twin marks a turning point in the integration of artificial intelligence with humanistic expression. By leveraging large language models, affective reasoning engines, and cognitive architectures such as the Super Time-Cognitive Neural Network (STCNN), this system explores a promising direction in computational creativity—one that does not merely automate artistic processes but *extends and reshapes* the landscape of cultural production and aesthetic reflection.

Throughout this study, we have explored how a cognitive agent can emulate the emotional and stylistic patterns of a historical artist through the curation of multimodal datasets—letters, visual works, biographical notes—and how these data, once processed through emotionally aware LLMs and encoded in temporal-semantic memory graphs, can simulate not just what Munch might have said or painted, but how he might have *felt*. This fusion of emotional modeling and epistemic plausibility forms the epistemological core of the digital twin, and it reflects a broader trend in AI research: the shift from formal logic toward *affective cognition and narrative reasoning*.

The results presented in the evaluation phase indicate that the digital twin is perceived not merely as an AI tool, but as an evocative interlocutor—a presence capable of inspiring reflection, nostalgia, even unease. Such responses point to the emergence of a new phenomenology of human-AI interaction, where the machine is no longer an instrument but a co-actor in the meaning-making process. In particular, the capacity of the system to generate artworks that echo Munch's style, and to engage users in dialogues that resonate with existential themes, suggests that emotional authenticity can be simulated and may be perceived as felt by some users—a subjective impression reported under non-controlled conditions, not an objective property of the system.

However, the project also reveals fundamental limitations. The digital twin is not a replacement for the artist, nor a resurrection of his consciousness; it is a functional, non-conscious simulation. It is a *probabilistic narrative construct* that generates plausible outputs conditioned on affective, linguistic, and artistic priors. Its value lies not in historical accuracy but in its capacity to stimulate discourse—philosophical, artistic, educational—about memory, identity, and the limits of artificial simulation. For this reason, transparency in system design remains paramount. Users must be made aware of the boundaries between historical facts and synthetic generation, between Munch the person and Munch the simulation.

Looking ahead, several avenues for development emerge. Technically, the integration of real-time affective feedback through biometric sensing (e.g., facial expression analysis, heart rate variability) could allow the digital twin to adapt its outputs based on the emotional state of the user, deepening the interaction loop and enabling forms of co-creative flow. Additionally, future iterations might incorporate generative adversarial training to better balance stylistic coherence with thematic novelty or include cross-cultural emotional corpora to assess how Munch's simulated emotional expressions are received across different demographics.

The use of a chronological-experiential concept of time, made possible by the extension into the complex number plane, is potentially very interesting, but it requires extensive testing and experimentation to verify the possibility of humanizing computational time, even though the preliminary results are encouraging. Therefore, this work has only opened a window onto the actual application and experimentation for the creation of artworks such as digital twins, thereby extending to this field the path laid out in works [32,33], but here too, the journey has only just begun. Therefore, the visual quality of the generated images should be read as preliminary evidence only: the results are encouraging but not conclusive, and warrant further controlled study.

Beyond the technical dimension, there are profound implications for the fields of education, therapy, and digital heritage. In an educational context, digital twins of historical figures could revolutionize the study of literature, art, and philosophy—transforming passive content into dynamic, interactive experiences. Imagine dialoguing with a digital Nietzsche or Emily Dickinson, not just reading their work but encountering their simulated reasoning processes. Similarly, in therapeutic contexts, emotionally resonant simulations might support reflective practices, grief processing, or creative therapies.

From a philosophical standpoint, the Munch Digital Twin invites us to reexamine the boundaries between originality and imitation, presence and simulation. It blurs the line between artifact and agent, prompting us to ask not only *what machines can do*, but *what we are becoming* in the process of co-creating with them. The digital twin is not a copy—it is a mirror that reflects both the subject it emulates and the desires of the user interacting with it.

This reflection leads to a final consideration. As generative AI becomes more pervasive, the need for a new interdisciplinary figure—the *Sophimatician*—becomes increasingly urgent. A Sophimatician is not merely a developer or a digital humanist, but a *mediator between cognitive science, philosophy, and artificial imagination*. Trained in epistemology, affective modeling, narrative structure, and symbolic computation, this new professional would be capable of designing, auditing, and ethically guiding systems like the Munch Digital Twin. In this sense, the project is not only about Munch—it is about laying the groundwork for a new human-AI culture.

In conclusion, the Munch Digital Twin is both a technological experiment and a philosophical provocation. It opens a dialogue about the simulation of inner life, the ethics of posthumous digital reconstruction, and the role of art in an age of synthetic minds. Its success should not be measured only by technical performance, but by its capacity to evoke thought, emotion, and inquiry—a testament to the enduring power of art and the transformative potential of cognitive machines.

Author Contributions: Conceptualization: I.F., Investigation: I.F., G.I. and M.C.; Methodology: G.I. and M.C.; writing—review and editing: I.F., G.I. and M.C.; supervision, G.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Appendix A.1

Input: query q , memory matrix M , time anchors $a[]$, salience weights $b[]$
 For each memory item i :
 $g_i = \exp(-|a_i - a_q| / \tau) \cdot (1 + \tanh(b_i))$
 $\text{logit}_i = \text{dot}(q, M_i) \cdot g_i$
 $\text{attention} = \text{softmax}(\text{logits})$
 Update: $b_i \leftarrow b_i - \eta \cdot \partial L / \partial b_i$ (via autograd, $|\Delta b| < 1 \times 10^{-3}$, $|b| \leq 5$)

Appendix A.2

Input: user text, memory matrix M_t , previous emotion $E_{\{t-1\}}$
 $E_t = \text{Plutchik_projection}(\text{RoBERTa_GoEmotions}(\text{user text}))$
 $A_t = \text{TemporalTransformer}(M_t)$
 $s = \text{Frobenius_norm}(\text{autograd}(A_t, M_t))$
 $\delta e = \text{sigmoid}(\text{normalize}(s) + \lambda \cdot \text{cosine_distance}(E_{\{t-1\}}, E_t))$
 Scale attention logits for turn $t+1$ by $(1 + \delta e)$
 Training: $L = L_{\text{LM}} + \mu \cdot L_{\text{emotion}} + \nu \cdot L_{\text{resonanceattention}} = \text{softmax}(\text{logits})$

Appendix A.3

Input: query q , candidate answer a
 $P(q,a)$ = normalised frequency over Memory Layer thematic clusters
 $Pl(q,a)$ = cosine similarity in RoBERTa-GoEmotions affective space
 $Cr(q,a)$ = Neo4j KG retrieval score + DistilBERT MC-dropout confidence
 $Ps(q,a) = 1 - \text{normalised_perplexity}(a \mid \text{corpus constraints})$
 $R(q,a) = \alpha \cdot P + \beta \cdot Pl + \gamma \cdot Cr + \delta \cdot Ps$ ($\alpha + \beta + \gamma + \delta = 1$)
 Accept if $R(q,a) \geq \theta$; else regenerate or trigger fallback

Appendix A.4

```
{
  "type": "text" | "image",
  "title": "string",
  "date": "YYYY-MM-DD",
  "emotion": "Plutchik primary dimension",
  "embedding": [float, ...],
  "thematic_cluster": "string",
  "source": "string",
  "link_to_images": ["artwork_id", ...]
}
```

Appendix A.5

System: You are a probabilistic reconstruction of Edvard Munch, conditioned on documented biographical and artistic sources. You do not claim consciousness or lived experience.
 Temporal context: $\{a\}$ | Salience weight: $\{b\}$
 Emotional state: $\{E_t\}$ | Thematic cluster: $\{\text{cluster}\}$
 Retrieved memory: $\{\text{top}_k\text{-fragments}\}$
 Constraint: respond only within the bounds of documented evidence.

Appendix A.6

1. The system conveyed emotional expressions consistent with Munch's artistic persona.
2. The generated content was stylistically aligned with Munch's known works.
3. The conversational responses reflected genuine thematic depth.
4. The system's responses were credible within Munch's historical context.
5. The interaction felt personally meaningful.
6. The system was useful for learning about Munch and his art.
7. The interaction inspired new artistic or creative ideas.

Appendix A.7

- PyTorch 2.x, CUDA 12.1, Python 3.10
- SD 1.5 LoRA: rank 16, α 32, 5 epochs, lr 1×10^{-4} , DPM-Solver++, CFG 7.5, 512×512
- FAISS: IndexFlatIP, d = 768 (sentence-BERT all-mpnet-base-v2)
- Neo4j 5.x, ~3000 nodes, ~8500 edges
- Random seed: 42 (PyTorch and NumPy)
- Hardware: NVIDIA A100 (cloud); Jetson Orin Nano (edge)

References

1. Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; et al. LLaMA 2: Open Foundation and Fine-Tuned Chat Models. *arXiv* **2023**, arXiv:2307.09288.
2. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 10684–10695.
3. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
4. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS 2014), Montréal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
5. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems 30 (NeurIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
6. Iovane, G.; Fominska, I.; Sibilio, M. Perspective chapter: From avatar technology to interactive holographic digital twin. In *Current State and Future Perspective in Human-Robot Interaction*; IntechOpen: Rijeka, Croatia, 2025. [[CrossRef](#)]
7. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)] [[PubMed](#)]
8. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language Models Are Unsupervised Multitask Learners. Available online: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf (accessed on 15 June 2026).
9. Cho, K.; van Merriënboer, B.; Bahdanau, D.; Schwenk, H.; Gulcehre, C.; Bougares, F.; Bengio, Y. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014. [[CrossRef](#)]
10. Gal, Y.; Ghahramani, Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the 33rd International Conference on Machine Learning (ICML 2016), New York, NY, USA, 19–24 June 2016.
11. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
12. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems 25 (NIPS 2012), Lake Tahoe, NV, USA, 3–8 December 2012; pp. 1097–1105.
13. Ying, R.; He, R.; Chen, K.; Eksombatchai, P.; Hamilton, W.; Leskovec, J. Graph convolutional neural networks for web-scale recommender systems. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2018), London, UK, 19–23 August 2018. [[CrossRef](#)]
14. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2016**, arXiv:1409.0473.
15. Hamilton, W.; Ying, R.; Leskovec, J. Inductive representation learning on large graphs. *arXiv* **2017**, arXiv:1706.02216.
16. Bruna, J.; Zaremba, W.; Szlam, A.; LeCun, Y. Spectral networks and locally connected networks on graphs. *arXiv* **2014**, arXiv:1312.6203.
17. Ruder, S. An overview of multi-task learning in deep neural networks. *arXiv* **2017**, arXiv:1706.05098.

18. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. *arXiv* **2017**, arXiv:1701.07875.
19. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. In Proceedings of the 5th International Conference on Learning Representations (ICLR 2017), Toulon, France, 24–26 April 2017.
20. Kingma, D.P.; Salimans, T.; Welling, M. Improving variational inference with inverse autoregressive flow. In Proceedings of the 33rd International Conference on Machine Learning (ICML 2016), New York, NY, USA, 19–24 June 2016; pp. 1336–1344.
21. Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.-A. Stacked denoising autoencoders. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408.
22. Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. Graph neural networks: A review. *AI Open* **2020**, *1*, 57–81. [[CrossRef](#)]
23. Amodei, D.; Olah, C.; Steinhardt, J.; Christiano, P.; Schulman, J.; Mané, D. Concrete problems in AI safety. *arXiv* **2016**, arXiv:1606.06565.
24. Floridi, L.; Cows, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds Mach.* **2018**, *28*, 689–707. [[CrossRef](#)] [[PubMed](#)]
25. Boden, M.A. Creativity and artificial intelligence. *Artif. Intell.* **1998**, *103*, 347–356. [[CrossRef](#)]
26. Colton, S.; Wiggins, G.A. Computational creativity: The final frontier? *Front. Artif. Intell. Appl.* **2012**, *242*, 21–26. [[CrossRef](#)]
27. Cambria, E.; White, B. Jumping NLP curves: A review of natural language processing research. *IEEE Comput. Intell. Mag.* **2014**, *9*, 48–57. [[CrossRef](#)]
28. McCormack, J.; Gifford, T.; Hutchings, P. Autonomy, authenticity, authorship and intention in computer generated art. In *Computational Intelligence in Music, Sound, Art and Design*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2019; pp. 35–50. [[CrossRef](#)]
29. Mazzone, M.; Elgammal, A. Art, creativity, and the potential of artificial intelligence. *Arts* **2019**, *8*, 26. [[CrossRef](#)]
30. Tulving, E. Episodic memory: From mind to brain. *Annu. Rev. Psychol.* **2002**, *53*, 1–25. [[CrossRef](#)] [[PubMed](#)]
31. Iovane, G.; Di Pasquale, R. A complexity theory-based novel AI algorithm for exploring emotions and affections by utilizing artificial neurotransmitters. *Electronics* **2025**, *14*, 1093. [[CrossRef](#)]
32. Iovane, G.; Iovane, G. Sophimatics and 2D complex time to mitigate hallucinations in LLMs for novel intelligent information systems in digital transformation. *Appl. Sci.* **2026**, *16*, 288. [[CrossRef](#)]
33. Iovane, G.; Iovane, G.; De Rosa, A. Complex-time neural networks: Geometric temporal access for long-range reasoning. *Algorithms* **2026**, *19*, 334. [[CrossRef](#)]
34. Iovane, G.; Di Gironimo, P.; Chinnici, M.; Rapuano, A. Decision and reasoning in incompleteness or uncertainty conditions. *IEEE Access* **2020**, *8*, 115109–115122. [[CrossRef](#)]
35. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning transferable visual models from natural language supervision. In Proceedings of the 38th International Conference on Machine Learning (ICML 2021), Virtual Event, 18–24 July 2021; pp. 8748–8763.
36. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. LoRA: Low-rank adaptation of large language models. In Proceedings of the 10th International Conference on Learning Representations (ICLR 2022), Virtual Event, 25–29 April 2022.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.