UNIVERSITY OF PLOVDIV
“PAISII HILENDARSKI”

FACULTY OF MATHEMATICS AND INFORMATICS

DEPARTMENT OF COMPUTER TECHNOLOGIES

VIKTOR PLAMENOV MATANSKI

COMPUTATIONAL FRAMEWORK FOR CROSS-MODAL ART TRANSLATION

ABSTRACT

of a dissertation submitted for the award of the educational and scientific degree “Doctor”

field of higher education 4. Natural Sciences, Mathematics and Informatics
professional field 4.6. Informatics and Computer Science
doctoral programme “Informatics”

Scientific supervisors:
Prof. Dr. Anton Iliev
Prof. Dr. Todorka Terzieva

Plovdiv
2026

The dissertation comprises 197 pages and consists of an introduction, seven chapters, a conclusion, a statement of contributions, prospects for future development, approbation of the results, publications related to the dissertation, an appendix, and a bibliography. It contains 54 figures, 49 tables, and 250 bibliographic sources in Bulgarian and English.

The dissertation was discussed and directed for defence at a meeting of the Department of Computer Technologies at the Faculty of Mathematics and Informatics of the University of Plovdiv “Paisii Hilendarski” on [• date •] (Protocol No. [• •]).

Scientific jury:

1. [• academic degree, full name •] – chair

2. [• academic degree, full name •] – reviewer

3. [• academic degree, full name •] – reviewer

4. [• academic degree, full name •] – member

5. [• academic degree, full name •] – member

The defence of the dissertation will take place on [• date •] at [• time •] in [• hall •] of the University of Plovdiv “Paisii Hilendarski”, at an open session of the scientific jury.

The defence materials are available to those interested at the Dean’s Office of the Faculty of Mathematics and Informatics of the University of Plovdiv “Paisii Hilendarski” and on the university’s website.

Author: Viktor Plamenov Matanski

Title: Computational Framework for Cross-Modal Art Translation

CONTENTS OF THE ABSTRACT

INTRODUCTION

Context and relevance of the research

When Homer describes the shield of Achilles, he turns a visual image into words and joins an ancient tradition of carrying artistic meaning from one expressive form into another — from painting to music, from music to poetry, and from poetry back to image. From *ut pictura poesis* and ekphrasis, through Lessing’s distinctions [1] and the colour–tone systems of Newton and Scriabin [2], to Kandinsky’s synaesthetic explorations [3], the different artistic forms have continually entered into relations of analogy, correspondence, and mutual re-expression. In computer science, however, this phenomenon still lacks a sufficiently rigorous computational formalization. Modern generative AI produces text, images, and music from freely specified prompts with striking ease, yet the task in which a system must preserve the relevant properties of a specific input artifact as it crosses into another artistic form remains considerably less explored.

The distinction is one of kind. Most contemporary language models can write a plausible poem about a stormy sea, but a poem that conveys the symbolism, the emotional trajectory, and the compositional emphasis of one specific painting is a task of an entirely different order. The first case is free generation; the second is translation with a commitment to the source. Modern multimodal models [4][5] make the task technically feasible but do not resolve it conceptually. What is missing is a stable formulation that separates the challenge from free generation, style transfer, and descriptive paraphrase; and without explicit invariants and acceptance criteria, evaluation remains weakly controllable and difficult to reproduce.

The nature of the task has a deeper root: the arts do not share a common language of expression. Painting organizes meaning through spatial relations, composition, and visual texture; poetry — through linguistic semantics, imagery, and rhythm; music — through temporal organization, timbre, and harmony. No direct correspondence exists between these expressive systems, which is why the relation between a source work and a target work cannot be established through surface resemblance. What is required is to identify the essential properties of the source work and to represent them in a modality-independent form, as invariants and constraints that the target work must satisfy. A review of existing approaches reveals three systemic deficits for which no comprehensive solution yet exists: controllability, traceability, and reproducible evaluation.

Core concepts and terminology

Art translation is defined [6] as the purposeful re-expression of a specific input artifact from one artistic form into another [7], such that the result is verifiably linked to the input through explicit invariants, constraints, and success criteria. Its defining mark is not thematic similarity but a normatively governed relation between input and output. A translation undertakes a commitment, and that commitment can be checked, unlike free generation, style transfer, and descriptive paraphrase, where no such commitment exists.

Object, subject, and scope of the research

The object of the research is intelligent computational systems for translating artistic content between different forms. The subject is the formal models, knowledge representation models, multi-agent architectures, and evaluation procedures required to preserve and verify relevant invariants. The scope covers painting, poetry, and music, forms with fundamentally different expressive primitives, which throws the problem of modality mismatch into sharp relief. The framework is designed to remain open to extension towards additional art forms.

Aim of the dissertation

The aim of the dissertation is to develop a computational framework for cross-modal art translation, articulated in four sub-goals: to formalize art translation as a computational task with explicit representations, constraints, and criteria for acceptance or rejection; to provide a modality-independent intermediate representation encoding the invariants, the constraints, and the trace of decisions; to design a multi-agent architecture that puts fidelity first, through independent verification, typed repair, and explicit points of human intervention under full traceability; and to provide a fidelity evaluation protocol combining formal checks, automated indicators, and expert judgement.

Research thesis

The working thesis is that the principal deficit of existing cross-modal generative systems stems not from insufficient generative capacity but from the absence of an explicit commitment to the input artifact. Without formalized fidelity, a system can neither establish what has been lost nor justify its result to an external assessor. Hence the central thesis: art translation can be formalized as generation under constraints, in which fidelity to the input artifact is measurable, traceable, and open to diagnosis. This is achievable through a modality-independent intermediate representation, an explicit set of invariants, and an iterative cycle of analysis, generation, verification, and correction. The distinction between principled translation and arbitrary variation can be formalized and measured, without any claim to full cross-modal equivalence.

Research questions and tasks

From the central question (how can art translation be formalized and realized as a computational problem) follow six research questions: its formalization, as distinguished from free generation, style transfer, and paraphrase (RQ1); the modality-independent intermediate space for the invariants (RQ2); a sufficient ontological foundation and the constraints that follow from the ontological commitment (RQ3); a multi-agent architecture for fidelity-first operation, traceability, and correction of deviations (RQ4); reproducible evaluation through formal checks, automated indicators, and expert assessment (RQ5); and the empirical limits of formalizable fidelity and the trade-offs between fidelity and freedom (RQ6).

Five tasks are carried out in pursuit of the aim: a formal definition of art translation (an input–output contract, a taxonomy of failure modes, an intermediate representation, a hierarchy of constraints); an ontological knowledge foundation; a multi-agent architecture with an iterative verify-and-repair mechanism; a prototype system for translation between painting, poetry, and music; and a protocol for evaluating the results.

Methodological framing of the research

The research is positioned within computer science and combines conceptual analysis, formalization, ontological modelling, architectural design, and experimental evaluation. The framework follows four organizing principles: fidelity first (preservation of the invariants is the criterion for all decisions, while interpretative freedom is controlled through an explicit hierarchy of constraints); an iterative process around the intermediate representation (a cycle of analysis, generation, verification, and repair, rather than one-shot generation); the ontology as an active knowledge layer (semantic enrichment, consistency checking, generalization through property chains); and multi-agent execution with traceability (five specialized agents in graph-based orchestration, with a complete trace of decisions).

Practical significance

The practical significance lies wherever controllable, explainable, and measurable work with artistic elements is required: the creative industries, cultural heritage, accessibility, and education; the CDAO ontology itself has independent value as a verifiable, growing corpus of cross-modal correspondences.

The potential extends beyond the initial scope: the prototype has been applied qualitatively to scientific visualizations (phase portraits, bifurcation diagrams, chaotic attractors), and the expert reviews add photography and sculpture as input artifacts tested in practice. Media production and therapeutic applications via the PAD encoding are promising directions. These applications are potential rather than empirically confirmed, and they inherit the limitations set out in Chapter 7.

Limitations of the research

The research has five clearly defined limitations: it works only with those aspects of the artifact that are representable in the intermediate representation; the goal is a manageable optimum, not full aesthetic equivalence; the empirical scope is painting, poetry, and music, with automated indicators mandatorily complemented by formal checks and expert assessment; style is subordinate to semantic and affective fidelity; and the system has the status of a reference research prototype.

Structure of the dissertation

The dissertation comprises 197 pages and consists of an introduction, seven chapters, a conclusion, a statement of contributions, prospects for future development, approbation of the results, publications related to the dissertation, an appendix, and a bibliography of 250 sources. The present abstract follows the structure of the dissertation. The main results are reported in 3 scientific publications, including an article in the scientific journal Future Internet (2025).

CHAPTER 1. LITERATURE REVIEW

The literature review passes through six research traditions (the historico-theoretical foundations of cross-modal correspondences, their cognitive and affective foundations [8], multimodal generative models, multi-agent systems [9][10][11], ontology engineering [12], and computational creativity [13]) and derives from them the two pillars of the framework. First: cross-modal correspondences are not arbitrary. Systematic correspondences between the sensory modalities are observed in the general population [14], and affective content stands out as a more stable intermediate layer than concrete descriptors, a fact on which the framework builds directly. Second: multimodal generative systems make cross-modal generation a technological reality but do not solve the task of faithful, traceable, and assessable translation. Computational creativity is examined through Boden’s taxonomy [15] and the SPECS framework [16], which motivate the evaluation protocol.

A synthesis of the reviewed traditions makes clear what is already established: artistic correspondences are a long-debated problem in aesthetics and the theory of symbol systems; cross-modal correspondences have a cognitive basis and admit empirical study; multimodal models already achieve plausible generation between text, image, and sound; complex generative processes are governed better through distributed roles and iterative cycles; and knowledge-based representation is necessary wherever shared semantics, inference, and provenance are required. From here the problem is no longer whether art translation is conceivable, but why the available tools do not yet realize it in a rigorous and verifiable form.

The analysis shows that monolithic end-to-end approaches do not, by themselves, provide the properties art translation requires: the correspondence between modalities remains in a latent space without an explicit semantic contract, and one can neither specify nor verify which properties of the source are preserved. There is no machine-checkable acceptance criterion beyond human preference, and correction reduces to resampling without a traceable record. Scaling alone does not resolve these problems. The synthesis of the literature identifies five research gaps: the lack of a formal definition of art translation; the lack of a machine-interpretable cross-modal ontological language; semantic opacity in creative multi-agent systems; the lack of a typed taxonomy of failures; and a missing methodology for evaluating cross-modal semantic fidelity.

Conclusions. Cross-modal correspondences have a systematic cognitive basis, affect stands out as a comparatively more stable intermediate layer, and the five research gaps identified justify the integrated approach of the dissertation.

CHAPTER 2. FORMALIZATION OF CROSS-MODAL ART TRANSLATION

Chapter Two is the theoretical core of the dissertation. In it, art translation is formalized as a procedure that transforms an input artifact as into an artifact of a chosen target modality dt, under a given context ctx. The result comprises the target artifact at, the intermediate representation IR, and a status indicating whether the acceptance conditions are met. The central element of the formalization is the intermediate representation IR=⟨SDS,IS,CP,TS,PL⟩. It unites the source descriptors (Source Descriptor Set), the set of invariants (Invariant Set), the correspondence plan (Correspondence Plan), the translation specification (Translation Specification), and the provenance log (Provenance Log). The work thus does not pass directly from canvas to verse: it is first reduced to a modality-independent core and then unfolded from that core into the target artistic form. Figure 1 presents this two-stage process through the hourglass model.

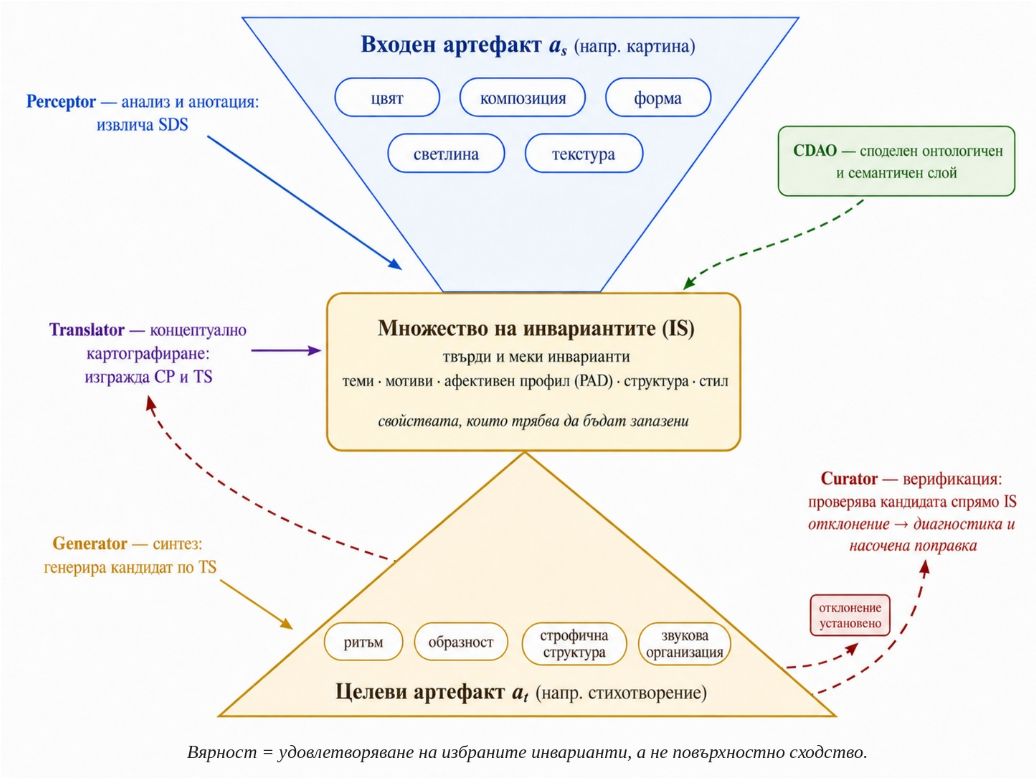

Figure 1. Conceptual model of translation: the input artifact is distilled into the invariant set (IS) and unfolded into the target artifact; fidelity = satisfaction of the invariants, not surface similarity.

An invariant is an explicitly formulated requirement to preserve a property of the input artifact, one judged significant for its interpretation and represented in modality-independent form. It must be stressed that an invariant is not an objective property the system simply discovers, but a revisable normative judgement. It is proposed by the perceptual agent and may be confirmed, amended, or rejected by a human or by the verifying agent. By degree of obligation, invariants divide into hard (mandatory and essential to the artifact's identity) and soft (weighted and open to compensation). By content, they may be semantic, structural, affective, or mixed.

Fidelity is formalized through the function score(·) and a two-stage acceptance condition. A translation is accepted if every hard invariant satisfies score(h)≥θ_hard and the weighted aggregate over the soft criteria reaches AggS(at)≥θ_soft. For hard invariants the condition is non-compensatory: violating a critical invariant leads to rejection of the result regardless of how the remaining criteria fare. For soft criteria, compensation between individual scores is permitted. The thresholds θ are not set as arbitrary constants but as typed, calibrated values for the specific translation profile. Four measures are used for assessment: fidelity, completeness, adequacy, and voice (F, C, A, V). The last acknowledges that the invariants guide and constrain the target realization without exhausting it: a good poem is not merely a description of a painting.

The chapter introduces a typed taxonomy of deviations, classifying mismatches by their causal profile rather than by a low overall score. The class of a deviation is determined unambiguously by the type labels of the violated invariants, so that each type points to a different repair stage. The repair rule is minimal and directed: only the component responsible for the deviation is modified, and at the lowest level of the pipeline at which the cause can be removed without destroying invariants already satisfied, with a terminating condition and a regression check.

Table 1. Typed taxonomy of deviations and directed repair

Deviation type	**Signal**	**Directed repair**
Semantic (Type I)	uncovered mandatory descriptor	Translator – missing descriptors
Affective (Type II)	PAD distance above threshold τ_PAD	Generator – affective constraint
Structural (Type III)	violated structural invariant	Generator – structural constraint
Composite (compositional)	violations across several invariants	full replanning (Translator)

The mechanism is shown through a worked example: an impressionist seascape is translated into a poem. Five invariants are extracted and typed: the sea horizon and the melancholic calm as hard; the sunset, the horizontal layering, and the impressionist diffuseness as soft; while the impasto texture is registered as unrealizable in verse: an explicitly recorded loss, written to the provenance log rather than passed over in silence. The candidate poem clears the hard gate (0.88 and 0.80 against an illustrative threshold of 0.75; the calibrated operational threshold for this profile is 0.62, see Figure 3), the soft aggregate reaches 0.657 against a threshold of 0.60, and fidelity is $F \approx 0.76$. The poem also adds a tidal rhythm and the metaphor of a “closing eye”, properties beyond the invariants, reported as voice: the framework measures not only what is preserved but also what the new work contributes of its own.

All of this is consolidated in the Framework Contract with four clauses (fidelity first, controllability, traceability, and reproducibility;

FC-FF/CT/TR/RP), which are architecturally binding for any realization of the framework. The example also grounds why full cross-modal equivalence is unattainable in principle: losses accumulate both at capture (perception is lossy extraction) and at realization (not every invariant has an analogue in the target modality). This is precisely why fidelity is a graded rather than binary quantity, a matter of measurement, not of declaration.

Conclusions. Art translation is formalizable as generation under constraints, with a verifiable two-stage acceptance condition and directed, typed repair. With this, Chapter 2 answers research questions RQ1 and RQ2.

CHAPTER 3. AN ONTOLOGY FOR CROSS-MODAL ART

Chapter Three builds CDAO, an ontology for art translation. The goal is for it to be part of the translation process, not merely to describe it. CDAO is designed to answer pre-formulated competency questions and encodes modality-independent semantic, affective, and structural descriptors through a class hierarchy, an N-ary Correspondence class for correspondences with typed mapping relations and confidence values, and property-chain axioms. Emotions are represented through the numerical PAD model (pleasure–arousal–dominance) [17][18].

The class hierarchy separates the work level cleanly from the element level, with dedicated properties for affect (the PAD model) and explicit modelling of provenance. Four property chains allow the ontology to generalize on its own: emotion, motif, theme, and style propagate automatically from element level to work level. Since OWL 2 RL is monotonic while the generalization of artistic knowledge is inherently defeasible, consistency is controlled through SHACL rules [19], including a guard shape against unwarranted generalization.

CDAO is aligned with established standards (Getty AAT [20], CIDOC-CRM [21]), which secures interoperability and reuse, and it is conceived as a cumulative cross-modal knowledge base. Through inference and aggregation it reveals knowledge beyond what is explicitly asserted. Ontological grounding also acts as a mechanism for constraining, and detecting part of, unsupported claims. An invariant is grounded not because it points at the meaning, but because behind it stands a traceable chain of assertion, ratification, and verification. The formal evaluation was carried out reproducibly on release 0.9.0: structurally, the ontology comprises 39 classes, 29 object and 27 datatype properties, and 8 SHACL shapes. More than half of the relations are non-taxonomic, reflecting the orientation towards correspondences rather than classification alone. All seven competency questions receive answers, with three of the queries requiring inference, direct confirmation that the ontology derives new statements rather than merely storing them. The chapter openly identifies the weak points as well, above all the lower density of style correspondences, and argues for extensibility.

The ontology's capacity to derive new statements is also demonstrated directly. The affect-mediated bridge artifact–emotion–artifact, executed on the live graph, yields plausible candidate correspondences never explicitly asserted: for example, warm colour palette → (joy) → major tonality between painting and music, or fast tempo → (tension) → high contrast in the opposite direction. The result is illustrative, but it shows that the shared PAD layer generates semantically meaningful, verifiable hypotheses.

Conclusions. CDAO combines decidable inference (OWL 2 RL), generalization through property chains, and SHACL control in a cumulative cross-modal knowledge base. With this, Chapter 3 answers research question RQ3.

CHAPTER 4. ARCHITECTURAL SPECIFICATION OF A MULTI-AGENT SYSTEM FOR ART TRANSLATION

Chapter Four examines the candidate architectures capable of supporting cross-modal art translation. From the Framework Contract, seven architectural requirements (R1–R7) are derived, against which six candidates are assessed: a monolithic agent, a pipeline, centralized/hierarchical, peer-to-peer, blackboard, and market-based/ContractNet. None covers the requirements on its own. A hybrid architecture is therefore justified, with a formal argument for minimal sufficiency. Five agent roles are specified with a clear division of responsibilities.

Table 2. Specialized agent roles in the pipeline.

Agent	Function in the pipeline
Reasoner	Director and orchestrator – governs the flow and replanning
Perceptor	Sensory agent and annotator – extracts descriptors and proposes invariants
Translator	Conceptual cartographer – builds the correspondence plan
Generator	Modality synthesizer – produces a candidate in the target form
Curator	Verifier and critic – checks against the invariants, with veto power

The division of roles is normative, not descriptive. Rights and access are distributed so that the Curator holds an architecturally guaranteed veto, and no path exists from a generated to an accepted result that bypasses independent verification. Memory is two-layered: a semantic layer (the CDAO knowledge graph) and an operational layer (a structured blackboard), with an explicit schema for recording provenance. Control flow is defined through named execution states with admissible transitions (Figure 2): a translation passes through ANALYZED → MAPPED → GENERATED → EVALUATED, from which the paths lead to ACCEPTED, REPAIRED (returned for directed repair), REJECTED, or ESCALATED (to human intervention). The cycle terminates on acceptance, on lack of improvement, or on exhausting iter_max.

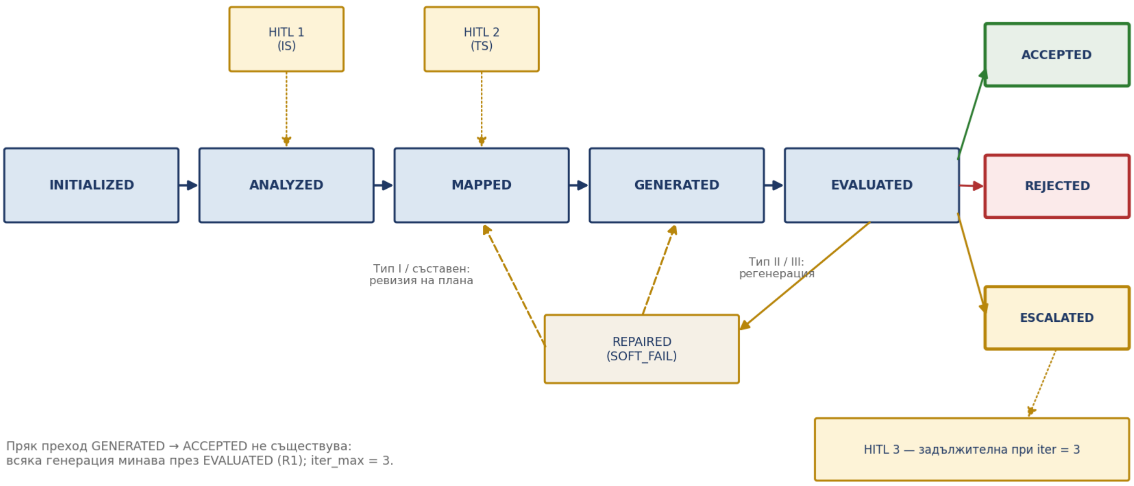

Figure 2. Execution-state diagram: admissible transitions, HITL points, and repair cycles.

The verification mechanism uses a deterministic classifier of the deviation type and routing logic for directed repair. Three formal Human-in-the-Loop intervention points are specified with inputs and outputs as a structural obligation, not a configurable option: the human is part of the architecture, not its fallback. The chapter analyses the prompt-engineering strategy [22] as an architectural interface, and the neuro-symbolic integration: the neural layer proposes, the symbolic layer binds and verifies. The model's limits are also examined: coordination complexity under scaling, dependence on the quality of CDAO, the indeterminism of language models in borderline cases, and dependence on human availability.

Conclusions. The hybrid hierarchical architecture is minimally sufficient with respect to R1–R7, and the Curator's veto together with the formal points of human intervention make independent verification of fidelity a structural obligation. With this, Chapter 4 answers research question RQ4.

CHAPTER 5. PROTOTYPE

Chapter Five describes the reference implementation: a working prototype with a three-layer system architecture: a web application layer (Django), a multi-agent implementation layer (LangGraph) [23], and a knowledge–data layer (PostgreSQL with an on-demand projection of the live knowledge graph). The prototype uses eight locally hosted generative models for perception and synthesis across the three modalities, distributed over the agent roles, and supports multidirectional translation between painting, poetry, and music. Local hosting is justified by three considerations: reproducibility (pinned local versions), data control, and comparability of experimental conditions.

The three layers have a clear division of responsibilities: the web layer (Django/Celery) handles authentication, artwork submission, presentation of results, and the questionnaires; the multi-agent layer (LangGraph) concentrates all AI computation (perception, planning, generation, and evaluation); and the knowledge–data layer links the CDAO graph (RDFlib) with a relational store for artifacts, checkpoints, and provenance, with all exchange between the web layer and the data mediated by the agent layer. Local hosting also carries a deliberate compromise: models of 7–30 billion parameters fall short of the leading proprietary systems, which is why the prototype demonstrates the architectural properties of the framework (fidelity preservation, deviation detection, and ontological planning) at a qualitative level, explicitly acknowledged as a boundary of the proof of concept.

The key design decisions and engineering trade-offs are justified: synchronous access to the ontology against asynchronous agent execution; JSON-schema enforcement against free-form output; the robustness of CDAO against model-invented concepts through a quarantine of user labels; and graph state management under bidirectional translation with session isolation through separate graphs. The chapter describes the correspondence between the specification of Chapters 2 and 4 and the implementation along the main translation path, including the mapping of execution states to graph nodes and of access rights to the tool registry.

The system is publicly available at https://art-translator.com, where every translation, every invariant, and every agent decision can be examined in detail. The prototype is at the same time a research instrument. It provides a component for studying and evaluating translation processes, both with human participants (non-professionals and experts) and with automated criteria and autonomous assessors. The Experiment Runner automates the execution of experiments, the computation of the indicators, and the recording of provenance.

Conclusions. The art-translator.com prototype demonstrates the realizability of the specification under full execution traceability and a configuration that supports reproducibility. With this, Chapter 5 supports the answer to RQ4.

CHAPTER 6. EVALUATION

How is fidelity to be measured when no painting has a single “correct” poem? Chapter Six answers with a five-layer protocol organized as a ladder. The layers are ordered by increasing independence of the assessor: from internal symbolic checks against the system’s own ontology, through automated measures at scale, to external human judgement, which anchors the rest. The evaluation follows the SPECS methodology [16]: first define what success even means for such a system (traceable and transparent preservation of the semantically identified invariants), and only then derive criteria and tests, instead of relying on standard similarity metrics, which are structurally unsuitable in the absence of a single reference. No single layer is sufficient on its own; the evidential force comes from the layers complementing one another, with both their agreements and their divergences carrying diagnostic value.

Table 3. Five-layer evaluation protocol.

Layer	What it checks	Key result
1. Fidelity	symbolic recoverability of the invariants	IS Retrieval: median 0.83; independent semantic measure: AUC 0.72–0.75 (text/visual)
2. Robustness	cycle consistency under composition	563 cycles; at least one strong exemplar in every direction (maxima 0.62–1.00)
3. Assessors	reliability of the automated assessors	reliable at level/discriminative terms, not at fine-grained ranking
4. Users	perception by non-specialists	80 enrolled, 75 raters, 540 ratings;
5. Experts	value for creative practice	6 independent experts: a creative collaborator, not a substitute

Layer 1 establishes the symbolic recoverability of the invariants through IS Retrieval, computed via SPARQL over the materialized OWL 2 RL closure. On the case corpus the median is 0.83, highest for themes (0.89) and affect (0.86), lowest for style (0.71). The difference is not noise but a structural diagnosis: for affect the ontology has a rich infrastructure (PAD properties and chains), while style correspondences are sparser. Discriminative validity is established directly for a parallel semantic measure through permutation negative controls, and remains indirect for IS Retrieval itself. For the two text/visual directions it holds (AUC 0.72–0.75, 95% confidence intervals excluding 0.5), while for the four music directions it is not established on the available sample.

Layer 2 subjects the system to its harshest test: cycle consistency. A work is translated into another modality and back, and the two ends of the chain are compared. Across 563 attempts covering the six directions in the full collection, the key finding concerns existence, not averages. In each of the six directions, at least one highly rated, reproducible cycle was observed by the measure used (maxima 0.62–1.00; the single cycle scoring 1.00 is in the direction music → painting → music). The emotional core is the most durable property under composition, while concrete descriptors and theme are more fragile. The low overall mean is consistent with the difficulty of the task and the limitations of the generative models used, but it does not license a conclusion of impossibility in principle. The very existence of reproducible high-quality cycles establishes feasibility. Eight representative cycles, through poetry and through music, are shown at the end of the chapter.

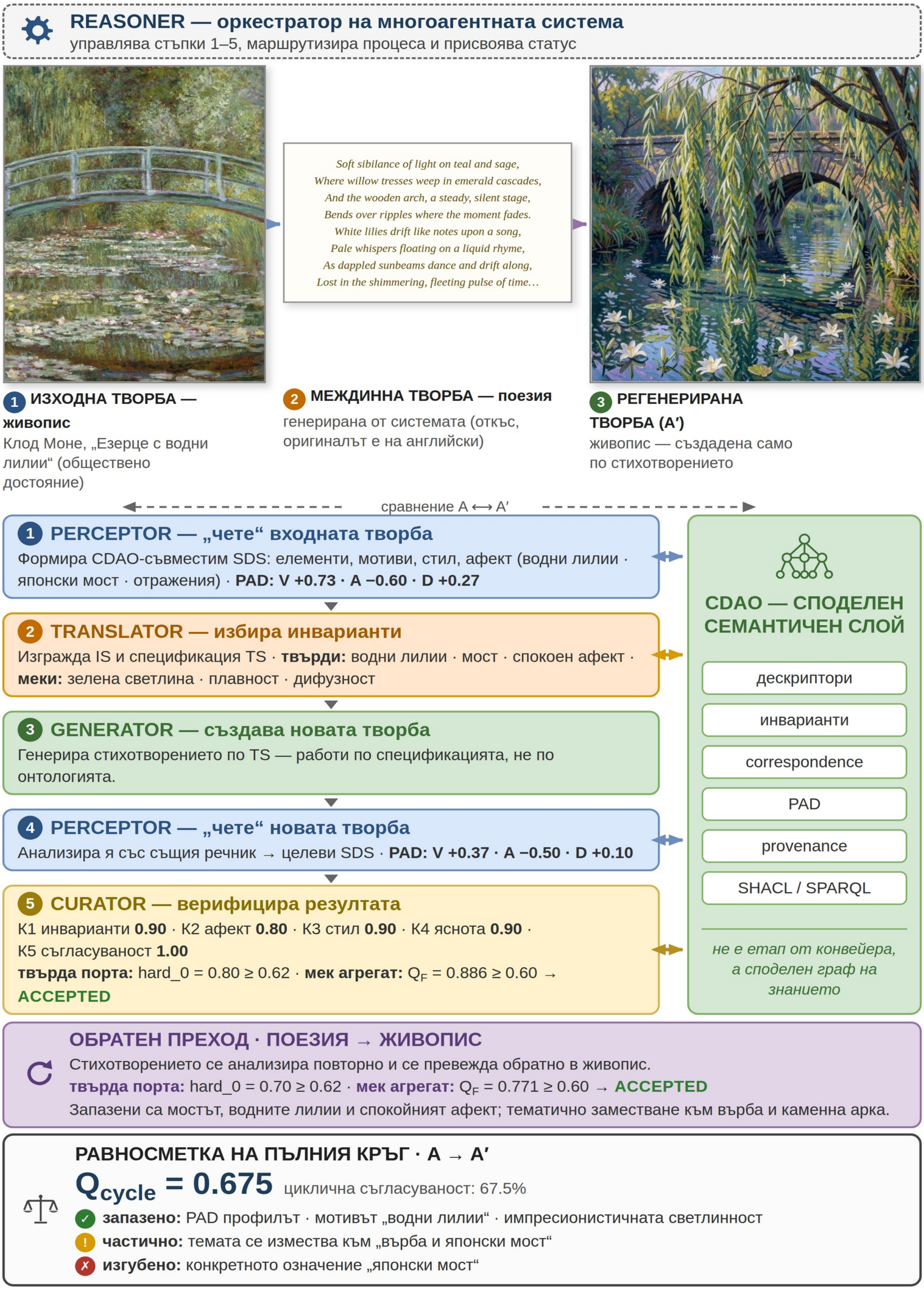

Figure 3. Complete worked example from the prototype (translation No. 368/77): a cross-modal cycle painting → poetry → painting (Claude Monet, “Water Lilies”); the indicators are measured automatically.

Layer 3 assesses the reliability of the automated assessors (Curator, a neutral judge, an agent panel) [24]: they are reliable at the level of scores and discriminatively, but their rank correlation with independent human ratings is low (overall $\rho \approx 0.06$) under a strong ceiling effect. Machine judges serve as a preliminary diagnostic layer, not as a substitute for fine-grained human judgement. Layer 4 (a user study with 80 enrolled participants, 75 of whom contribute 540 ratings across three modalities) establishes high perceived acceptability of the outputs (54–60% maximum ratings across the five dimensions, Figure 4), robust under aggregation. Layer 5 (a qualitative assessment by six independent experts: a painter, a musician, a sculptor, a linguist, a photographer, and an educator) [25] provides a qualitative, interpretative evaluation. The experts predominantly perceive the system as capable of supporting the transfer of the conceptual and affective core, and as functioning as a creative collaborator rather than a substitute.

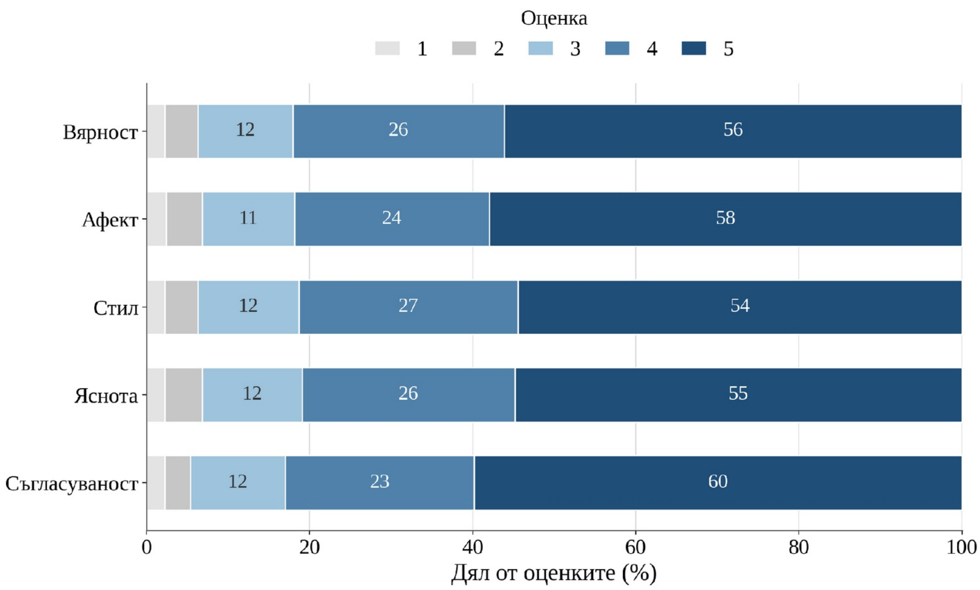

Figure 4. Distribution of ratings by dimension (user study, 80 participants): 54–60% maximum ratings.

The triangulation across the layers shows divergence rather than strict convergence [26], a finding valuable in itself. It points precisely to where machine and human judgement part ways. The systematized threats to validity and the residual risk are summarized in a dedicated table in the dissertation.

Conclusions. Semantic fidelity is operationalizable and measurable, with discriminative validity for the text/visual directions, and the divergence between the layers is diagnostically valuable. With this, Chapter 6 answers research question RQ5 and, in part, RQ6.

Examples of round-trip translations

The sources of the original artworks are listed in order of presentation, with their A2A identifier and cycle-consistency score; the live round-trip translations are available at art-translator.com.

1. A2A #2 · Frederic Edwin Church · The Heart of the Andes, 1859 · cycle consistency 0.845
2. A2A #160 · Paul Gauguin · Where Do We Come From? What Are We? Where Are We Going?, 1897–1898 · cycle consistency 0.448
3. A2A #28 · Claude Monet · Woman with a Parasol – Madame Monet and Her Son, 1875 · cycle consistency 0.644
4. A2A #98 · Odilon Redon · Ophelia among the Flowers, c. 1905–1910 · cycle consistency 0.507
5. A2A #129 · Katsushika Hokusai · Fine Wind, Clear Morning (Red Fuji), c. 1830–1832 · cycle consistency 0.477
6. A2A #48 · Thomas Cole · The Oxbow, 1836 · cycle consistency 0.597
7. A2A #132 · Michelangelo · The Creation of Adam, c. 1508–1512 · cycle consistency 0.475
8. A2A #164 · William Blake · The Ancient of Days, 1794 · cycle consistency 0.444

Table 4. Round-trip translations of paintings through poetry and through music: source work, intermediate artifact, and regenerated work, with cycle-consistency scores.

O, emerald shroud where
ancient shadows sleep,
And thunderous falls through
mossy canyons leap;
From verdant depths to peaks
of crystalline,
Where silence meets the breath
of the divine.
The wandering clouds, in soft
and silver flight…….

Terracotta wraps hold the
weight of a bird,
A child perched soft on a
mother's warm shoulder.
Patterned blues and pale
whites stand in a line,
Silent sentinels where the
island's peace grows.
The jungle breathes emerald,
heavy and lush…….

Emerald silk filters the golden
glare,
As sunlight dances through the
parasol's shade.
A woman in white, with linen
to flare,
Glides through the meadow, a
lighthearted braid.
The wildflowers sway in a
rhythmic sea…..

♪ MUSIC

audio on art-translator.com

♪ MUSIC

audio on art-translator.com

♪ MUSIC

audio on art-translator.com

♪ MUSIC

audio on art-translator.com

CHAPTER 7. DISCUSSION, LIMITATIONS AND FUTURE WORK

Chapter Seven interprets the results in a broader theoretical context. The central argument is this: the framework establishes technical feasibility, not empirical superiority. The chapter answers research questions RQ1–RQ6 with explicit scoping and develops three theoretical directions. The first is the thesis of an ontological interlingua: a finite, machine-interpretable symbolic vocabulary, organized around modality-independent conceptual categories, can serve as a semantic bridge between the modalities, not as symbols true in a descriptive sense, but as binding symbols capable of carrying obligations, provenance, and verification. The second is the taxonomy of typed failure modes as a design principle that generalizes beyond art translation. The third is controlled stochasticity as a creative mechanism: the ontology and the invariants draw the boundary within which the generator explores stochastically, while the verification cycle applies selective pressure in the ontologically defined space.

The limitations fall into three categories. Of the implementation: a more limited correspondence base and corpus for music, CDAO coverage weighted towards Western European and Bulgarian figurative painting, the ceiling of the local models, and restricted access to the points of human intervention. Of the evaluation: the absence of a systematic baseline (ablation) comparison is the most significant gap, since component-level causal attribution requires a controlled ablation design. Structural: the vocabulary ceiling leaves subconceptual properties (brushstroke, rhythm of line, material presence) unrepresented, and the designation of invariants remains a judgement dependent on the framework.

The chapter articulates the framework's guiding perspective: knowledge-based creative AI, in which the neural layer proposes, the symbolic layer binds and verifies, and the human remains the only party who can renegotiate the vocabulary. The resulting programme of future work is systematized in the section “Prospects for Future Development”.

Conclusions. Causal attribution of the components requires an ablation design, and typed failure modes and controlled stochasticity are generalizable design principles. With this, Chapter 7 answers research question RQ6.

CONCLUSION

The dissertation demonstrates that art translation can be formalized as a computational task of generation under constraints, in which fidelity to the input artifact is measurable, traceable, and open to diagnosis. At the architectural level the framework is realized and internally coherent: the Framework Contract, the intermediate representation, the typed taxonomy of deviations, the division of the five agent roles with the Curator's veto, CDAO as a verifiable substrate, and reproducibility through a traceable profile. Empirically, within a stated scope, the operationalizability and verifiability of semantic fidelity are established, with discriminative validity for the text/visual directions. The central scientific contribution is the formalization itself: without an explicit intermediate representation, semantic deviation remains invisible. The system does not know what it has lost, and an invisible loss cannot be repaired in a directed way.

The dissertation situates the research space of cross-modal art translation at the intersection of five traditions and adds to each: to multimodal generative AI it contributes a formal intermediate space that makes semantic preservation measurable and diagnosable, inaccessible to end-to-end mapping; to multi-agent systems it establishes the principle that the taxonomy of failure modes determines the correction strategy; to ontology engineering it offers CDAO, operationally complementing Getty AAT and CIDOC-CRM and resolving links across different art forms, with no direct analogue found in the literature reviewed; to computational creativity it achieves an architectural operationalization of Boden's taxonomy (combinational and exploratory, but not transformational creativity); and to cross-modal studies of art it contributes a modality-independent vocabulary that computationally operationalizes Goodman's insight [27].

The contribution of the dissertation does not lie in the mere idea of carrying meaning between the arts — a practice known at least since the tradition of ekphrasis [28]. As far as can be established from the literature review, the dissertation offers one of the first systematic formalizations of art translation as a distinct and verifiable computational task. The formalization includes explicit conditions for success and failure and is operationalized in a publicly accessible prototype. In this way, intuitions accumulated over millennia of artistic tradition are turned into explicit, machine-interpretable, and procedurally verifiable structures.

STATEMENT OF CONTRIBUTIONS

The dissertation formulates and substantiates six contributions, which form a logical sequence: C1 defines the task, C2 specifies the framework, C3 provides the ontological vocabulary, C4 builds the architecture, C5 realizes it, and C6 defines the evaluation protocol and carries out an empirical demonstration. C1–C2 form the theoretical core, while C3–C6 materialize the framework and present the evidence for its realizability and verifiability.

Table 5. Correspondence between contributions, chapters, research questions, and publications

Contribution	Ch.	Question	Nature	Publ.
C1 – formal problem statement and typology of deviations	2	RQ1	scientific–applied	1, 2
C2 – normative-operational apparatus (Framework Contract, IR)	2	RQ2	scientific–applied	1, 3
C3 – an ontology for art translation	3	RQ3	scientific–applied	1
C4 – multi-agent architecture	4	RQ4	scientific–applied	1
C5 – prototype system	5	RQ4	applied	1
C6 – five-layer evaluation protocol	6	RQ5, RQ6	scientific–applied	1

Contribution 1 (C1). An original formal statement of cross-modal art translation as a distinct computational task is proposed, including its demarcation from related generative processes (free generation, style transfer, descriptive paraphrase) and the introduction of typed failure modes.

Contribution 2 (C2). A formal normative-operational apparatus is developed (the Framework Contract, a modality-independent intermediate representation, a hierarchy of constraints, and a two-stage acceptance condition), which makes fidelity measurable, traceable, and open to diagnosis.

Contribution 3 (C3). The Cross-Domain Art Ontology (CDAO) is proposed not merely as a semantic layer but as a cumulative cross-modal knowledge base which, through inference and aggregation (property chains, N-ary correspondences, SHACL), reveals knowledge beyond what is explicitly asserted.

Contribution 4 (C4). A hierarchical multi-agent architecture is developed, with orchestration, independent verification holding veto power, and typed, directed repair of deviations under full traceability of decisions.

Contribution 5 (C5). A reference proof-of-concept system with a traceable execution cycle is implemented, demonstrating the realizability of the formal framework and publicly available for independent verification.

Contribution 6 (C6). A five-layer evaluation protocol is proposed, and an empirical demonstration of the feasibility of the reference system is carried out across three modalities (80 enrolled participants, 540 ratings, of which 446 cross-modal).

PROSPECTS FOR FUTURE DEVELOPMENT

What remains to be established

The causal attribution of invariant preservation to the individual components (the CDAO ontology, the repair cycle, and the multi-agent architecture) has not been established empirically and defines the immediate priority. An ablation study with four conditions is planned (full system; without ontological grounding; without the repair cycle; a direct query to a single model as baseline), with an independent outcome measure and pre-registered hypotheses. The medium-term programme adds a benchmark suite for cross-modal fidelity, extension of CDAO's coverage and modalities, the inclusion of larger multimodal models as components, and further studies with more participants and experts.

The ontological layer as an open resource

CDAO is published as a traceable open resource (release 0.9.0, https://art-translator.com/cdao/data) and has a life of its own beyond the framework. The frozen research corpus (v0.9.0; n = 701) guarantees the reproducibility of the results of Chapter 6, while the production graph keeps growing: as of July 2026 it covers more than 1,800 translations, 7,400 typed correspondences, and 4,500 assessments, documenting itself with every translation.

The knowledge-based creative AI towards which the framework points is not an alternative to human creativity but an instrument for its extension: a system that can explain what it has preserved, diagnose what it has lost, and propose a correction, leaving the final judgement to the human.

APPROBATION OF THE RESULTS

The results of the dissertation have been validated through a publication in a peer-reviewed journal (Future Internet, 2025) [6], presenting the framework integrally; through papers at scientific conferences; through the publicly accessible working prototype art-translator.com with its live knowledge graph, which allows independent verification of the translations and of the semantic layer; and through the research conducted within three scientific projects, acknowledged below.

Acknowledgements.

The research is supported by the Centre of Excellence in Informatics and ICT (Contract No. BG16RFPR002-1.014-0018-C01, Programme "Research, Innovation and Digitalization for Smart Transformation" 2021–2027, co-funded by the European Union) and by the University of Plovdiv "Paisii Hilendarski" projects No. IT15-FMIIT-004 ("Research in the Domain of Innovative ICT Oriented Towards Business and Education") and No. FP25-FMI-010 ("Innovative interdisciplinary research in Informatics, Mathematics and Pedagogy of Education").

PUBLICATIONS RELATED TO THE DISSERTATION

The main theoretical formulations, architectural decisions, and preliminary experimental results are reported in 3 scientific publications:

1. Matanski, V., A. Iliev, N. Kyurkchiev, and T. Terzieva. "Ontology-Driven Multi-Agent System for Cross-Domain Art Translation." Future Internet 17, no. 11 (2025): 517. https://doi.org/10.3390/fi17110517 (Q2, WoS, IF: 4.6).

2. Matanski, V. "Exploring Synesthesia Utilizing Software Technologies." Scientific Conference "Innovative ICT in Business and Education: Future Trends, Applications and Implementation", Pamporovo, 2016. ISBN 978-954-8852-72-2.

3. Matanski, V. "Evolution of Generative Art Systems: Methods, Milestones and Open Problems." "Computer Technologies and Applications", Pamporovo, 15–17 April 2026, ISBN: 978-619-7768-62-6.

Publication 1 (the central publication) reflects the framework integrally: the formal problem statement (C1), the CDAO ontological layer (C3), the multi-agent architecture (C4), the prototype implementation (C5), the five-layer evaluation protocol (C6), and the intermediate representation as an element of the normative-operational apparatus (C2). Publication 2 is a preliminary study of the cross-modal correspondences between the sensory modalities, which served as the conceptual basis for formalizing the task. Publication 3 systematizes the methods, milestones, and open problems of generative art systems.

BIBLIOGRAPHY

1. Lessing, G.E. (1766/1984). Laocoon: An Essay on the Limits of Painting and Poetry. Johns Hopkins University Press. ISBN 978-0-8018-3139-3.

2. Klemenc, B., Ciuha, P. & Solina, F. (2011). Educational Possibilities of the Project Colour Visualization of Music. Organizacija, 44(3), 67–75. https://doi.org/10.2478/v10051-011-0006-9 ISSN 1318-5454.

3. Kandinsky, W. (1912/1977). Concerning the Spiritual in Art. Dover Publications. ISBN 978-0-486-23411-3.

4. Radford, A., Kim, J.W., Hallacy, C., et al. (2021). Learning transferable visual models from natural language supervision (CLIP). In Proc. ICML 2021 (International Conference on Machine Learning). arXiv:2103.00020.

5. Rombach, R., Blattmann, A., Lorenz, D., et al. (2022). High-resolution image synthesis with latent diffusion models. In Proc. CVPR 2022 (IEEE/CVF Conference on Computer Vision and Pattern Recognition). https://doi.org/10.1109/CVPR52688.2022.01042

6. Matanski, V., Iliev, A., Kyurkchiev, N. & Terzieva, T. (2025). Ontology-driven multi-agent system for cross-domain art translation. Future Internet, 17(11), 517. ISSN 1999-5903. https://doi.org/10.3390/fi17110517

7. Jakobson, R. (1959). On linguistic aspects of translation. In On Translation, pp. 232–239. Harvard University Press. ISBN 978-0-674-73161-5. https://doi.org/10.4159/harvard.9780674731615.c18

8. Ramachandran, V.S. & Hubbard, E.M. (2001). Synaesthesia – A window into perception, thought and language. Journal of Consciousness Studies, 8(12), 3–34. ISSN 1355-8250.

9. Wooldridge, M. (2009). An Introduction to MultiAgent Systems (2nd ed.). John Wiley & Sons. ISBN 978-0-470-51946-2.

10. Rao, A.S. & Georgeff, M.P. (1991). Modeling rational agents within a BDI-architecture. In Proc. KR’91, pp. 473–484.

11. Borghoff, U.M., Bottoni, P. & Pareschi, R. (2025). Beyond Prompt Chaining: The TB-CSPN Architecture for Agentic AI. Future Internet, 17(8), 363. ISSN 1999-5903. https://doi.org/10.3390/fi17080363

12. Noy, N.F. & McGuinness, D.L. (2001). Ontology development 101: A guide to creating your first ontology. Stanford KSL Tech. Report.

13. Wiggins, G.A. (2012). The mind's chorus: Creativity before consciousness. Cognitive Computation, 4(3), 306–319. ISSN 1866-9956. https://doi.org/10.1007/s12559-012-9151-6
14. Spence, C. (2011). Crossmodal correspondences: A tutorial review. Attention, Perception, & Psychophysics, 73(4), 971–995. ISSN 1943-3921. https://doi.org/10.3758/s13414-010-0073-7
15. Boden, M.A. (2004). The Creative Mind: Myths and Mechanisms (2nd ed.). Routledge. ISBN 978-0-415-31453-4.
16. Jordanous, A. (2012). A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. Cognitive Computation, 4(3), 246–279. ISSN 1866-9956. https://doi.org/10.1007/s12559-012-9156-1
17. Mehrabian, A. & Russell, J.A. (1974). An Approach to Environmental Psychology. MIT Press. ISBN 978-0-262-13090-5.
18. Russell, J.A. (1980). A circumplex model of affect. J. of Personality and Social Psychology, 39(6), 1161–1178. ISSN 0022-3514. https://doi.org/10.1037/h0077714
19. Knublauch, H. & Kontokostas, D. (2017). Shapes Constraint Language (SHACL). W3C Recommendation. https://www.w3.org/TR/shacl/
20. Getty Research Institute (2017). Art & Architecture Thesaurus Online. https://www.getty.edu/research/tools/vocabularies/aat/. (accessed 29 May 2026)
21. Doerr, M. (2003). The CIDOC Conceptual Reference Model. AI Magazine, 24(3), 75–92. ISSN 0738-4602. https://doi.org/10.1609/aimag.v24i3.1720
22. Wei, J., Wang, X., Schuurmans, D., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. In Proc. NeurIPS 2022 (Conference on Neural Information Processing Systems). arXiv:2201.11903.
23. Yao, S., Zhao, J., Yu, D., et al. (2023). ReAct: Synergizing reasoning and acting in language models. In Proc. ICLR 2023 (International Conference on Learning Representations). arXiv:2210.03629.
24. Zheng, L., Chiang, W.-L., Sheng, Y., et al. (2023). Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. In Proc. NeurIPS 2023 (Conference on Neural Information Processing Systems), Datasets and Benchmarks Track. arXiv:2306.05685. https://doi.org/10.48550/arXiv.2306.05685
25. Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. ISSN 1478-0887. https://doi.org/10.1191/1478088706qp063oa
26. Creswell, J.W. & Clark, V.L.P. (2018). Designing and Conducting Mixed Methods Research (3rd ed.). SAGE. ISBN 978-1-4833-4437-9.

27. Goodman, N. (1968). Languages of Art: An Approach to a Theory of Symbols. Bobbs-Merrill.
28. Bajohr, H. (2024). Operative ekphrasis: The collapse of the text/image distinction in multimodal AI. Word & Image, 40(2). ISSN 0266-6286. https://doi.org/10.1080/02666286.2024.2330335

DECLARATION OF ORIGINALITY

I declare that the present dissertation is entirely my own work and that no third-party publications have been used in its development improperly or in violation of copyright.