IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Beyond Visual Persuasion: A Methodological Framework for Evaluating Empathetic Architecture in AI-Generated Design Imagery

[1]Khaled Omara, [2]Ayman Assem, [3]Mohamed Ezzeldin, [4]Abdelrahman Ayman

[1]Research assistant, [2]Professor of Architecture Engineering, [3]Associate Professor of Architecture Engineering,
[4]Assistant Professor of Architecture Engineering,
[1]Department of Architecture,
[1]Faculty of Engineering, Ain Shams University, Cairo, Egypt

Abstract: Generative artificial intelligence now produces architectural imagery faster than the discipline can evaluate it. As AI-generated images enter concept-stage architectural workflows, the key question is no longer whether AI can generate compelling visuals, but how such images should be judged as architectural propositions. This paper argues that the necessary criterion is fidelity to design intent: the degree to which an AI-generated image remains aligned with the experiential, cultural, material, and temporal intentions that informed it. The paper proposes a conceptual-methodological five-stage AI-assisted pipeline for evaluating fidelity to empathetic design intent in AI-generated architectural imagery. Structured around the TACT framework - Tactility, Atmosphere, Culture, and Temporal Identity - the method embeds evaluation across the full workflow: material ingestion, TACT-informed analysis, prompt construction, image generation, comparative assessment, and human validation. The framework is illustrated through the Ethiopian New International Airport concept, where the design intent centers on Ethiopian identity, highland light, Rift Valley spatiality, tactile materiality, threshold hospitality, and temporal continuity. The contribution is methodological: a structured, auditable, and bias-aware framework for evaluating whether AI-generated architectural imagery remains faithful to empathetic architectural intent.

Index Terms: Generative AI; empathetic architecture; TACT framework; fidelity to design intent; AI-assisted design evaluation; architectural image judgment; cultural specificity; temporal identity.

I. INTRODUCTION

Generative artificial intelligence now produces architectural imagery faster than the discipline can evaluate it. The question is no longer whether AI can generate visually compelling architectural images, but how such images should be judged as architectural propositions. This paper argues that the central evaluative problem in AI-assisted architecture has shifted from image generation to image judgment. Existing vocabularies - visual criticism, phenomenological critique, and technical benchmarking - each provide partial insight, but none fully addresses what is specific to AI-generated architecture: its status as a human-intended yet machine-generated design proposition.

This distinction matters because visual persuasion can easily be mistaken for architectural fidelity. A generated image may appear atmospheric, culturally rich, or materially convincing while remaining weakly connected to the design intent that prompted it. In architectural practice, concept-stage imagery is not merely representational; it helps direct design decisions, client interpretation, and future development. The discipline therefore requires a method for assessing whether AI-generated images remain faithful to the experiential, cultural, material, and temporal intentions embedded in the design brief.

The problem is also methodological. When researchers or designers simultaneously define the design intent, construct prompts, select the most successful outputs, and evaluate the final images, curation risk becomes concentrated at the input and selection stages of the workflow. This does not invalidate AI-assisted design research, but it does make evaluative claims difficult to audit. A more rigorous method must separate these roles, document the movement from intent to prompt to image, and make the basis of evaluation explicit.

This paper proposes a five-stage AI-assisted pipeline for evaluating fidelity to empathetic design intent in AI-generated architecture. The pipeline is structured around the TACT framework: Tactility, Atmosphere, Culture, and Temporal Identity. Rather than treating TACT as a manual scoring instrument applied only after image generation, the revised approach embeds TACT across the workflow: empathetic material ingestion, TACT-informed analytical synthesis, structured prompt construction, image generation, and comparative fidelity evaluation. Human expertise is repositioned from isolated image scoring toward framework validation, interpretive calibration, and design-direction recommendation.

The framework is illustrated through the Ethiopian New International Airport (ENIA) concept, where the design intent centres on bringing Ethiopian identity, highland light, Rift Valley spatiality, tactile materiality, threshold hospitality, and temporal continuity into a contemporary airport environment. Accordingly, the paper should be read as a methodological framework paper rather than an empirical validation study. Its aim is to define a repeatable evaluative structure that can later be tested through image generation, expert assessment, and cross-case validation.

The contribution of the paper is methodological. It proposes a structured, auditable, and bias-aware framework for evaluating whether AI-generated architectural imagery remains faithful to empathetic architectural intent. Section II reviews the theoretical and technical literature supporting TACT and AI-mediated design evaluation. Section III presents the revised TACT framework. Section IV describes the five-stage AI pipeline. Section V discusses the methodological contribution. Section VI outlines theoretical implications, limitations, and future directions. Section VII concludes the paper.

II. LITERATURE REVIEW AND THEORETICAL BACKGROUND

2.1 Tactility: Embodied and Haptic Architectural Experience

The tactile dimension of architectural experience has been addressed most fully by Pallasmaa, whose account of the eyes of the skin argues that architectural encounter is fundamentally multisensory [1]. The body does not experience space through vision alone; it reads depth, temperature, texture, weight, and proximity through an embodied field of perception. Holl, Pallasmaa, and Perez-Gomez similarly argue that architecture gains experiential intensity through the interaction of light, material, sound, and bodily movement [2]. These accounts challenge any evaluation method that treats architectural imagery as purely visual.

Zumthor offers a practitioner’s account of the same problem. For Zumthor, atmosphere is inseparable from the tactile presence of material: the weight of stone, the resonance of timber, the density of concrete, and the temperature of surfaces [3]. Leatherbarrow extends this argument by linking material specificity to the ground, topography, and conditions of place [4]. Materiality is therefore not a decorative layer applied to form; it is one of the ways architecture communicates belonging, permanence, and bodily orientation.

For AI-generated architectural images, tactility raises a specific evaluative challenge. The image cannot be touched, but it can suggest whether a material has weight, grain, depth, and environmental credibility. TACT operationalises this question through Material Realism, Textural Detail, and Multisensory Comfort. Together, these indicators ask whether an image offers more than surface texture - whether it implies an architecture that could be physically encountered.

2.2 Atmosphere: Affective Presence and Cultural Specificity

Atmosphere is central to architectural experience but difficult to evaluate precisely. Bohme defines atmosphere as an affective quality arising between subject and environment; it is neither purely subjective nor purely objective, but emerges through encounter [5]. Zumthor similarly describes atmosphere through material presence, light, sound, temperature, intimacy, and the relation between interior and exterior [3]. These qualities are partly visible in architectural images, yet they are not reducible to visual mood.

This distinction is especially important for AI-generated imagery. Contemporary text-to-image models are highly capable of producing coherent mood, dramatic light, and emotionally persuasive spatial scenes.

However, atmospheric coherence does not necessarily indicate fidelity to design intent. An AI-generated airport interior may appear warm, cinematic, and immersive while failing to communicate the atmospheric conditions required by a specific project.

This limitation motivates the revision of the Atmosphere dimension in TACT. The original indicator of Emotional Tone is replaced with Cultural Atmosphere Specificity. In the ENIA context, this means evaluating whether the image communicates Ethiopian highland light, Rift Valley spatial generosity, threshold hospitality, and climatic groundedness, rather than generic airport ambience. Atmosphere is therefore repositioned from a measure of visual quality to a measure of cultural-environmental fidelity.

2.3 Culture: Identity, Place, and Symbolic Legibility

The cultural grounding of architectural meaning has long been central to architectural theory. Norberg-Schulz's concept of genius loci argues that architecture should respond to the character of place: its light, materiality, topography, climate, and patterns of human inhabitation [6]. Frampton's theory of critical regionalism gives this position a more explicitly political dimension by arguing for practices that resist placeless global uniformity through attention to local tectonics, climate, and cultural specificity [7]. Rapoport provides an anthropological foundation for this view, showing that built form is shaped by social customs, symbolic systems, and everyday life as much as by environment and technology [8].

For AI-generated architecture, culture presents both an opportunity and a risk. Text-to-image models can generate culturally suggestive forms with speed and visual richness, but they may also rely on generic, stereotyped, or geographically imprecise visual associations. Prior work on generative systems has shown that model outputs can reproduce cultural and demographic biases embedded in training data, particularly when prompts are underspecified or culturally ambiguous [9]-[11].

TACT addresses this problem through Cultural Specificity, Vernacular Sensitivity, and Ritual/Symbolic Layering. These indicators ask whether an image communicates a situated cultural logic, engages local tectonic and spatial principles beyond surface motifs, and embeds symbolic meaning through sequence, hierarchy, orientation, or ritual use.

2.4 Temporal Identity: Time, Memory, and the Limits of the Still Image

Temporal Identity is the TACT dimension most resistant to still-image evaluation. Ruskin's lamp of memory frames architecture as a bearer of accumulated time, where meaning deepens through age, weathering, repair, and historical continuity [12]. Mostafavi and Leatherbarrow develop this argument through weathering, showing how materials can be selected and detailed to anticipate transformation over time [13]. Brand describes buildings as layered systems that change through occupation and adaptation [14], while Lynch argues that environments gain meaning through their capacity to encode multiple temporal layers [15].

AI-generated still images struggle with these temporal conditions. They present architecture at a single moment and often simulate age as visual style rather than material or cultural process. Patina, wear, accumulated use, and seasonal change may appear as surface effects without deeper constructional or narrative logic.

For this reason, the revised TACT framework separates Temporal Identity into more precise indicators. Surface Memory asks whether materials appear capable of carrying patina or accumulated traces. Constructional Logic asks whether the building appears assembled through credible tectonic processes. Narrative Time Depth asks whether the image suggests a before and after - prior occupation, ritual use, seasonal change, or future adaptation.

2.5 Generative AI in Architecture: From Production to Judgment

AI-assisted architectural design has moved from rule-based and parametric systems toward probabilistic image synthesis and multimodal generative workflows. Earlier computational design research focused on parametric schemata, procedural variation, and design automation [16], [17]. More recent developments in generative adversarial networks and diffusion models have enabled rapid production of high-resolution images from textual prompts [18]-[20]. In architecture, this has encouraged experimentation with image-based ideation, rapid concept exploration, and speculative visualisation.

However, the architectural literature increasingly recognises that visual production is not equivalent to architectural reasoning. AI-generated images may be compositionally impressive while remaining detached from site, culture, programme, construction, or lived experience [21], [22]. The discipline already has powerful image generators; what it lacks is a structured method for judging whether those images remain aligned with the design intentions they should represent.

The issue is particularly acute at concept stage, where images can shape decisions before technical or spatial feasibility has been tested. A visually persuasive image can prematurely stabilise a design direction even when its relationship to the project's deeper intent is weak.

2.6 Text-to-Image Fidelity Evaluation Beyond Architecture

Machine learning research has developed several methods for evaluating text-to-image fidelity. TIFA decomposes prompts into question-answer pairs that can be checked against generated images through visual question answering [23]. DALL-Eval evaluates compositional and reasoning capacities in text-to-image generation [24]. HEIM proposes a broader multi-dimensional evaluation framework for text-to-image models, including alignment, quality, aesthetics, originality, fairness, and bias [25].

These approaches are valuable, but they are not sufficient for architectural evaluation. Most text-to-image benchmarks assess whether an image corresponds to prompt content. Architecture requires a different question: whether an image corresponds to design intent. A prompt may faithfully describe a large airport terminal with warm light, stone textures, and cultural patterns, and an image may satisfy those tokens while still failing to communicate the intended spatial, cultural, or experiential logic of a specific project.

TACT differs from general text-to-image evaluation frameworks in three ways. First, it evaluates alignment with architectural intent rather than prompt wording alone. Second, it requires culturally and professionally informed interpretation. Third, it treats disagreement not merely as a reliability problem, but as evidence of interpretive ambiguity.

2.7 Empathic AI, Multimodality, and Human-AI Interpretation

The idea that computational systems can engage with affective and experiential content has its roots in Picard's work on affective computing [26]. Recent multimodal AI systems have expanded the scope of machine interpretation by combining language, image, audio, and video analysis within a single reasoning environment. This has direct relevance for architectural evaluation, where meaning often emerges through the relationship between visual form, spatial sequence, cultural reference, and experiential description.

Recent multimodal models suggest that AI systems can support structured interpretation of images and spatial scenes, but also reveal the limits of automated cultural and experiential judgment. The GPT-4o system card describes a model capable of reasoning across text, image, and audio inputs [27], while the Gemini 2.5 technical report emphasises advanced reasoning, multimodality, and long-context processing [28]. In architectural design research, Cheung et al. demonstrate the potential of agentic multimodal LLMs to support iterative conversational design processes across visual and textual modalities [29].

Research on cultural sensitivity in multimodal models shows that cultural interpretation remains highly context-dependent. Mukherjee and Ghosh examine cultural competence in multimodal story generation [30], while Yadav et al. evaluate cultural value sensitivity in multimodal systems [31]. These studies support a cautious position: AI can assist cultural and spatial interpretation, but its outputs require expert calibration.

2.8 Analytical Frameworks for AI-Mediated Design Evaluation

Schön's concept of the reflective practitioner identifies the gap between explicit professional knowledge and tacit design judgment [33]. Designers do not simply apply rules; they engage in a reflective conversation with materials, representations, constraints, and emerging possibilities. AI-mediated design evaluation operates within this same gap. It must make design intent explicit enough to guide machine analysis without reducing architectural judgment to mechanical scoring.

TACT functions as the analytical vocabulary through which this translation can occur. Its dimensions convert broad empathetic concerns - material presence, affective atmosphere, cultural identity, and temporal continuity - into structured criteria that can guide both AI and human interpretation. Chung et al. show that generative systems can support creative work more effectively when guided by both a creative brief and an analytical framework [34]. Cross similarly describes design cognition as iterative movement between problem framing and solution development [35].

2.9 The Gap This Paper Addresses

Five gaps converge in the current literature. Empathetic architectural theory offers rich qualitative criteria for embodied, atmospheric, cultural, and temporal experience, but does not provide a protocol for evaluating AI-generated architectural images. AI-in-architecture literature has focused primarily on generation and visual production rather than judgment. Text-to-image fidelity benchmarks evaluate alignment with prompts, but not alignment with architectural meaning or design intent. Multimodal and empathic AI research demonstrates relevant technical capabilities, but these have not yet been systematically adapted to architectural fidelity evaluation. Finally, methodological work on AI-assisted design has not yet produced a framework that integrates empathetic theory, AI-mediated analysis, and human validation into a coherent evaluative pipeline.

This paper addresses that gap by restructuring TACT as the analytical infrastructure of a five-stage AI-assisted evaluation pipeline. Its novelty lies not in any single dimension, but in the integration of architectural empathy, multimodal AI analysis, prompt construction, image generation, and human validation into a method that is structured, auditable, and bias-aware.

III. THEORETICAL FRAMEWORK: REVISED TACT FRAMEWORK

3.1 Fidelity to Empathetic Design Intent

This paper defines fidelity to empathetic design intent as the degree to which an AI-generated architectural image preserves the experiential, cultural, material, and temporal intentions embedded in the original design brief or reference project. The term does not refer only to visual similarity, stylistic resemblance, or prompt-token accuracy. It refers to whether the generated image remains aligned with the deeper architectural intentions that give the project meaning.

This distinction is necessary because AI-generated images can be visually persuasive while remaining weak as architectural propositions. A generated airport interior may appear atmospheric, photorealistic, and culturally suggestive, yet still fail to communicate the intended material logic, spatial experience, cultural specificity, or temporal continuity of the project. In such cases, the image succeeds visually but fails architecturally.

Fidelity to design intent therefore provides a more rigorous criterion for judging AI-generated architecture. It asks whether the image carries forward the intended relationship between body, material, place, culture, memory, and use. In this paper, that criterion is structured through the revised TACT framework: Tactility, Atmosphere, Culture, and Temporal Identity.

3.2 TACT as a Framework for Architectural Image Judgment

TACT is proposed as a framework for judging AI-generated architectural images as design propositions rather than as isolated visual outputs. Its purpose is to translate empathetic architectural intent into a set of evaluative dimensions that can guide interpretation, prompting, image assessment, and expert validation.

The framework is based on four dimensions. Tactility evaluates material presence and bodily encounter. Atmosphere evaluates environmental tone, spatial immersion, and cultural-environmental mood. Culture evaluates situated identity, vernacular intelligence, and symbolic depth. Temporal Identity evaluates memory, continuity, use, aging, and the image's capacity to suggest time.

Together, these dimensions provide a vocabulary for distinguishing visual appeal from architectural fidelity. A visually strong image may still fail if it lacks tactile credibility, produces generic atmosphere, relies on unstable cultural references, or appears temporally static. TACT is therefore not a style guide or aesthetic checklist; it is an evaluative framework for asking whether an AI-generated image remains faithful to the empathetic intentions that informed it.

3.3 Tactility: Material Presence and Haptic Legibility

Tactility concerns the bodily and material dimension of architectural experience. Architecture is not encountered through vision alone; it is sensed through scale, texture, temperature, weight, sound, proximity, and bodily orientation. Even in an image, the viewer reads whether a material appears heavy or light, warm or cold, crafted or synthetic, grounded or superficial.

For AI-generated architectural images, tactility is especially important because models can easily produce surface texture without material depth. A wall may appear richly textured but lack believable thickness, jointing, gravity, or constructional logic. The Tactility dimension therefore asks whether the image suggests an architecture that could be physically encountered, not merely viewed.

TACT evaluates Tactility through three indicators. Material Realism assesses whether materials behave believably in relation to light, scale, gravity, and construction. Textural Detail assesses whether surface articulation supports haptic reading through grain, layering, joints, edges, and variation. Multisensory Comfort assesses whether the image implies thermal, acoustic, ergonomic, and bodily comfort.

3.4 Atmosphere: From Generic Mood to Cultural Atmosphere Specificity

Atmosphere is one of the most immediately persuasive qualities of AI-generated imagery. Contemporary models can produce dramatic lighting, soft shadows, immersive interiors, and coherent emotional tone with remarkable ease. This strength creates a methodological problem: atmospheric quality may be mistaken for architectural fidelity.

The original Atmosphere dimension included Emotional Tone as one of its indicators. In the revised TACT framework, this is replaced by Cultural Atmosphere Specificity. Emotional tone alone is too generic: an image may feel calm, warm, monumental, or welcoming while remaining detached from the specific cultural and environmental conditions of the project.

The revised Atmosphere dimension retains Mood Consistency and Spatial Immersion, but adds Cultural Atmosphere Specificity as the key diagnostic indicator. In the ENIA case, the question is not simply whether the image feels welcoming or impressive. The question is whether the atmosphere communicates Ethiopian highland light, Rift Valley spatial generosity, threshold hospitality, climatic groundedness, and national gateway identity.

3.5 Culture: Specificity, Vernacular Intelligence, and Symbolic Depth

Culture evaluates whether an AI-generated architectural image communicates situated meaning rather than generic regional style. This is especially important because cultural prompts can easily produce visual stereotypes: patterns, arches, colours, ornaments, or motifs that appear culturally expressive but are geographically unstable or symbolically shallow.

The Culture dimension is not concerned with literal representation alone. It asks whether identity is embedded in spatial organisation, material logic, threshold, hierarchy, orientation, gathering, and use. Culture is therefore treated not as surface decoration, but as a deeper architectural logic.

TACT evaluates Culture through Cultural Specificity, Vernacular Sensitivity, and Ritual/Symbolic Layering. In the ENIA case, cultural fidelity requires more than Ethiopian-inspired motifs. It requires the image to communicate the broader ambition to bring Ethiopia in through spatial generosity, material grounding, communal identity, landscape reference, and the symbolic transition between homeland and global gateway.

3.6 Temporal Identity: Memory, Use, and the Limits of the Still Image

Temporal Identity is the most difficult TACT dimension to evaluate in still AI-generated images. Architecture carries time through patina, weathering, repair, repeated use, ritual, seasonal change, and memory. A still image, however, often presents architecture as a perfected instant. AI-generated images intensify this problem because they may simulate age or historicity as visual effect rather than evidence of lived time.

The revised TACT framework therefore makes Temporal Identity more visually tractable. Historical Continuity assesses whether the image suggests a lineage from past architectures or inherited spatial forms. Surface Memory assesses whether materials appear capable of carrying patina, wear, or accumulated traces. Constructional Logic assesses whether the building appears assembled through credible tectonic or craft processes rather than as a seamless synthetic surface. Narrative Time Depth assesses whether the image suggests a before and after: prior occupation, repeated ritual, seasonal transformation, or future adaptation.

These revisions do not claim that still images can fully represent time. Instead, they clarify which temporal qualities can be evaluated visually and which may require sequential images, animation, or video in future applications. Temporal Identity therefore also exposes the limits of the image medium itself.

3.7 Revised TACT Dimensions and Indicators

The revised framework is summarised in Table 1. The table presents the four TACT dimensions, their indicators, and their operational meanings for evaluating fidelity to empathetic design intent.

Table 1. Revised TACT dimensions and indicators

Dimension	Code	Indicator	Operational Meaning
Tactility	T1	Material Realism	Believability of material presence, weight, scale, and light behaviour.
	T2	Textural Detail	Fine-grain articulation suggesting haptic depth, surface variation, and crafted detail.
	T3	Multisensory Comfort	Visual implication of thermal, acoustic, ergonomic, and bodily comfort.
Atmosphere	A1	Mood Consistency	Coherence of environmental and emotional tone across light, material, scale, and enclosure.
	A2	Spatial Immersion	Sense of presence, depth, experiential draw, and bodily orientation within the space.
	A3*	Cultural Atmosphere Specificity	Whether mood is consistent with project-specific atmospheric conditions. For ENIA, this includes highland Ethiopian light, threshold hospitality, and Rift Valley spatial generosity.
Culture	C1	Cultural Specificity	Recognition of situated cultural identity rather than generic regional or globalised imagery.
	C2	Vernacular Sensitivity	Engagement with local tectonic, climatic, and spatial logic beyond surface motif.
	C3	Ritual / Symbolic Layering	Symbolic meaning conveyed through sequence, hierarchy, orientation, gathering, or ritual use.
Temporal Identity	Ti1	Historical Continuity	Suggestion of lineage from past architectures, cultural memory, or inherited spatial forms.
	Ti2*	Surface Memory	Whether materials appear capable of carrying time, patina, wear, or accumulated traces.
	Ti3*	Constructional Logic	Whether the assembly suggests credible making, craft, tectonic layering, or construction over time.
	Ti4*	Narrative Time Depth	Whether the space suggests a before and after: prior occupation, repeated use, seasonal change, or future adaptation.

Table 1. Revised TACT framework. Asterisked indicators indicate revised or newly introduced criteria. A3 replaces the original Emotional Tone indicator. Ti2* to Ti4* revise and expand the Temporal Identity dimension.*

3.8 Visual Appeal as Supporting Quality, Not Primary Criterion

The revised TACT framework places fidelity to empathetic design intent at the top of the evaluative hierarchy. Visual appeal remains important, but it is not the primary measure of success. This distinction is essential in AI-generated architecture because models can produce images that are attractive, atmospheric, and photorealistic while remaining weakly aligned with the project’s architectural intent.

The primary question is therefore not: Is the image impressive? The primary question is: Is the image faithful? Secondary questions then examine where that fidelity succeeds or fails: material realism, haptic detail, cultural specificity, vernacular sensitivity, symbolic layering, surface memory, constructional logic, and narrative time depth.

Figure 1 illustrates this hierarchy. Fidelity to empathetic design intent sits at the top of the evaluative structure. The four TACT dimensions form the second layer, while individual indicators operate as diagnostic tools. Visual appeal is positioned as a supporting quality rather than the primary measure of success.

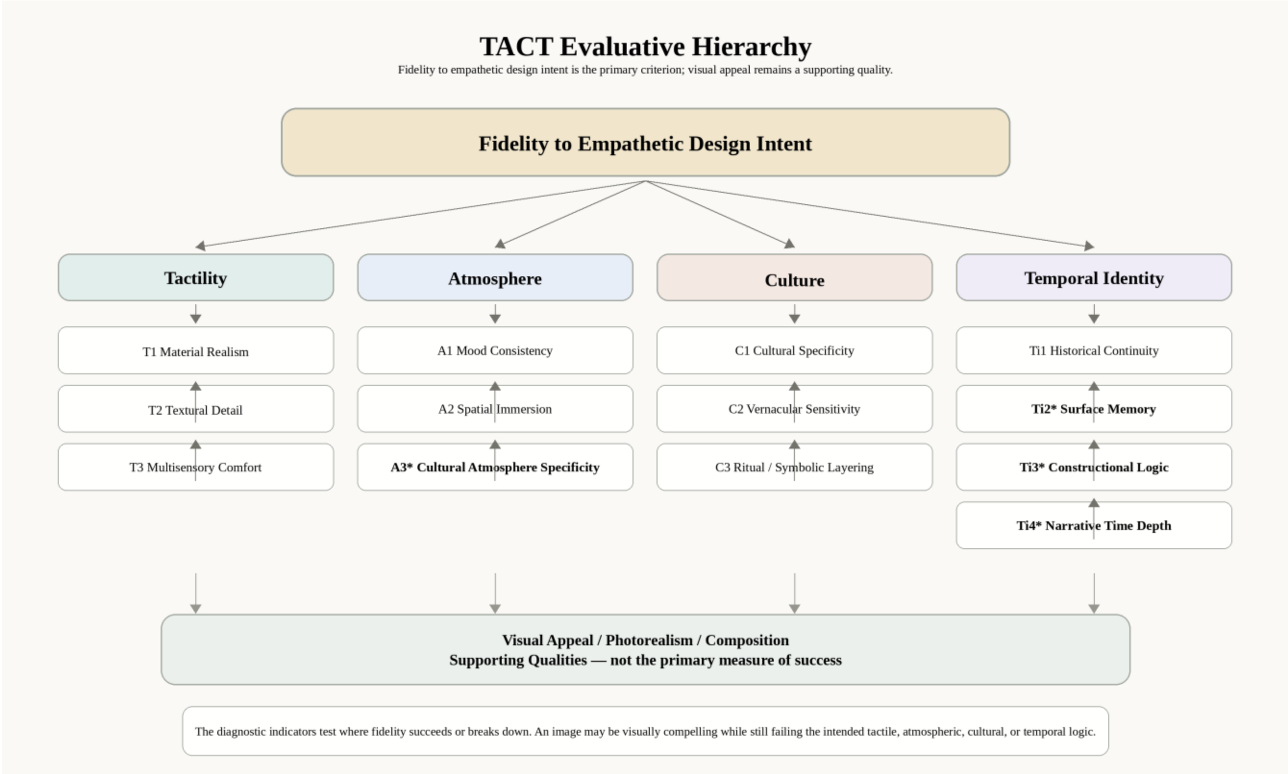

Figure 1. Fidelity to empathetic design intent as the organising principle of TACT evaluation. Source: created by the author.

This hierarchy prevents evaluation from being dominated by the strongest default capacities of image-generation models, such as mood, lighting, realism, or visual drama. Beauty alone does not establish architectural fidelity; visual appeal supports the evaluation only when it reinforces the intended tactile, atmospheric, cultural, and temporal logic of the project.

IV. METHODOLOGICAL IMPLEMENTATION: FIVE-STAGE AI PIPELINE

4.1 Pipeline Logic and Operational Structure

This chapter describes how the revised TACT framework is operationalised as a five-stage AI-assisted pipeline for evaluating fidelity to empathetic design intent. Whereas Section III defined the theoretical and evaluative structure of TACT, this section explains how that structure can be implemented in a research or concept-design workflow.

The pipeline follows a sequential but iterative logic:

Reference Material -> Empathetic Profile -> TACT Analysis -> Structured Prompt -> Generated Image Set -> Fidelity Evaluation -> Human Validation -> Prompt / Design Revision.

The purpose of the pipeline is not to automate architectural judgment. Its purpose is to make the movement from design intent to AI-generated image more explicit, documented, and open to critique. Each stage produces a recordable output that can be reviewed by researchers, designers, or expert validators.

The pipeline is governed by three methodological principles: intent traceability, role separation, and framework continuity. Intent traceability requires each generated image to be connected back to a clearly defined design intention. Role separation distinguishes between design-intent analysis, prompt construction, image generation, evaluation, and human validation. Framework continuity ensures that the same TACT dimensions used to define intent are also used to structure prompts and evaluate outputs.

Figure 2 illustrates this operational sequence. Table 2 summarises the five-stage pipeline, including the function, input, and output of each stage.

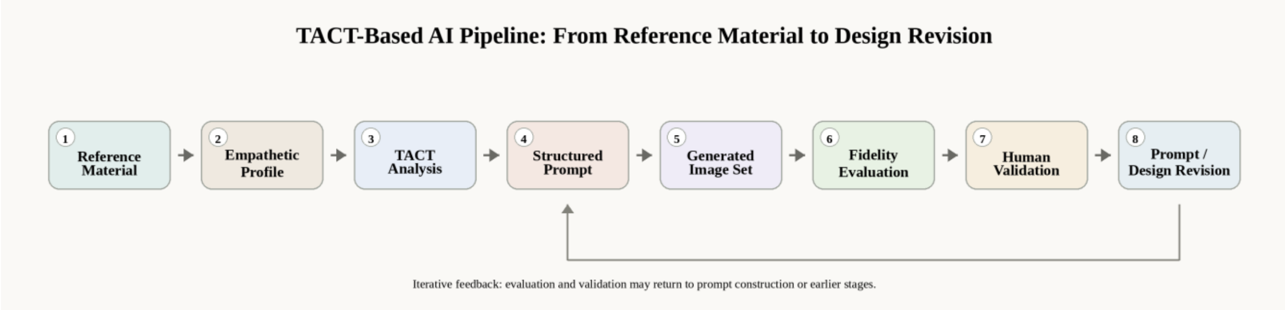

Figure 2. TACT-based AI pipeline from reference material to design revision. Source: created by the author.

Table 2. Five-stage AI pipeline for empathetic fidelity evaluation

Stage	Function	Input	Output
Stage 1	Empathetic Material Ingestion	Project brief, images, videos, cultural references, design narratives, spatial diagrams, material references	Structured empathetic profile
Stage 2	TACT-Informed Analytical Summary	Empathetic profile + revised TACT criteria	Per-space TACT design-intent analysis
Stage 3	Structured Prompt Construction	TACT analysis + spatial intent	Auditable generation prompts
Stage 4	Image Generation and Candidate Pool Formation	Structured prompts + documented model settings	Candidate image set
Stage 5	Comparative Fidelity Evaluation	Generated images + TACT criteria + reference intent	Structured fidelity report and revision recommendations

The ENIA case is used as an illustrative design context based on project design intent, conceptual narratives, and author-held design knowledge. It is not presented as a completed image-generation experiment or as a statistically validated case study. Its role is to demonstrate how TACT structures the translation from design intent to evaluative criteria.

4.2 Stage 1: Empathetic Material Ingestion

Stage 1 establishes the design-intent baseline. The input is not limited to a short text prompt; it includes the broader body of material through which architectural intent is formed. This may include project briefs, concept narratives, site photographs, precedent images, cultural references, material palettes, spatial diagrams, environmental descriptions, stakeholder accounts, and design-development notes.

For the ENIA case, the relevant design intent includes Ethiopian national identity, highland light, Rift Valley spatiality, volcanic material presence, threshold hospitality, communal gathering, and continuity between heritage and future infrastructure. These elements provide the empathetic and cultural ground against which future AI-generated images would be evaluated.

The purpose of Stage 1 is to organise reference material into a structured empathetic profile. This profile describes what each TACT dimension means for the project before any image generation occurs. This prevents the design intent from being reduced too early into stylistic keywords. Table 3 identifies the main categories of reference material required for Stage 1 and the risks associated with their absence.

Table 3. Reference material types for Stage 1 empathetic ingestion

Reference Type	Function in the Pipeline	Risk if Missing
Project brief and design narrative	Defines core intent and conceptual priorities	Image generation becomes stylistic rather than intentional
Cultural and historical references	Grounds identity and symbolic meaning	Cultural output becomes generic or stereotyped
Material and atmospheric references	Defines tactile and environmental qualities	Images become visually attractive but materially superficial
Spatial and typological references	Connects intent to architectural organisation	Outputs lack programme, scale, and spatial logic

Stage 1 is therefore not a mood-board exercise. It is the construction of an evaluative baseline. The quality of later stages depends directly on the richness and clarity of this initial material.

4.3 Stage 2: TACT-Informed Analytical Summary

Stage 2 translates the empathetic profile into explicit architectural criteria. At this stage, the AI model is not asked to generate images or write prompts. It is asked to analyse what each space should communicate through the four TACT dimensions.

For each selected space, the Stage 2 output should define what Tactility, Atmosphere, Culture, and Temporal Identity mean in that context. This creates a bridge between broad design intent and prompt construction. For example, an ENIA arrival plaza may be analysed through grounded stone-like surfaces, highland light, threshold hospitality, public symbolism, and continuity between historical memory and future national infrastructure.

The output of Stage 2 is a written TACT-dimensional analysis for each space. In the ENIA case, the twelve airport spaces listed in Table 4 provide an illustrative spatial sequence through which the framework can be applied. These spaces are not presented as a completed empirical image set, but as a structured design sequence for methodological demonstration.

Table 4. ENIA spaces, primary TACT dimension, and evaluative rationale

Space	Primary TACT Dimension	Rationale
Arrival Plaza	Atmosphere	Gateway mood, first encounter, environmental tone
Façade	Culture	Identity expression, symbolic language, recognisability
Circulation Corridor	Tactility	Material encounter, bodily scale, sensory passage
Waiting Lounge	Atmosphere	Comfort, enclosure, restorative pause
Terminal Forecourt	Culture	Threshold, arrival symbolism, public identity
Central Terminal Hall	Atmosphere	Spatial immersion, light, collective tone
Quiet Area	Tactility	Acoustic and sensory comfort, bodily retreat
Central Courtyard	Temporal Identity	Landscape continuity, time depth
Retail Gallery	Culture	Everyday social life, vernacular cues
Luggage Collection	Temporal Identity	Repetition, cyclical occupation, narrative use
Central Beacon	Temporal Identity	Orientation, symbolic endurance, memory object
Departure Hall	Tactility	Transition, movement, embodied threshold

4.4 Stage 3: Structured Prompt Construction

Stage 3 converts the TACT-informed analysis into structured generation prompts. In this framework, the prompt is treated as a methodological artifact rather than an informal creative instruction. It is part of the research record because it mediates between design intent and generated output.

Each TACT-structured prompt should include six components: spatial description; tactile specification; atmospheric specification; cultural specification; temporal specification; and negative constraints. This structure prevents the prompt from overemphasising visual style at the expense of architectural intent.

The prompt must be traceable back to the Stage 2 TACT analysis. This traceability allows reviewers to examine whether a weak output resulted from model limitation, poor prompt translation, insufficient reference material, or unclear design intent. In this sense, prompt construction becomes equivalent to methodological coding: it is the point where qualitative architectural intent is operationalised into generative language.

4.5 Stage 4: Image Generation and Candidate Pool Formation

Stage 4 generates image candidates from the structured prompts developed in Stage 3. In a full implementation of the framework, each space would be developed through a fixed candidate pool. In a future full implementation, one possible protocol would generate eight images per space using the same prompt version, model, aspect ratio, and generation settings.

The present paper focuses on establishing the methodological structure rather than reporting a completed image-generation experiment. The candidate-pool protocol is therefore presented as an implementation procedure for future application.

The purpose of this stage is not to search indefinitely for the most visually impressive image. It is to create a bounded and comparable set of outputs that can be evaluated consistently. All candidates should be retained in the project record, including unsuccessful outputs, because discarded images may reveal model bias, prompt weakness, cultural ambiguity, or failure of material and temporal fidelity.

Each generation cycle should record the model name and version where available, space name, prompt version, aspect ratio, image parameters, number of candidates generated, date of generation, negative prompts or exclusions, and whether the output is a still image, sequence, animation, or video. Candidate selection should then prioritise fidelity to TACT criteria over visual drama. The selection protocol is summarised in Table 5.

Table 5. Candidate pool and selection protocol

Step	Action	Purpose
1	Generate a fixed candidate pool per space	Create a bounded and comparable image set
2	Preserve all candidates	Prevent retrospective cherry-picking
3	Flag technical failures	Identify unusable or incoherent outputs without deleting them
4	Rank candidates using TACT criteria	Prioritise fidelity to design intent over visual appeal
5	Select one primary image per space	Produce a coherent illustrative or analytical sequence
6	Keep one optional secondary image	Support comparison or appendix discussion

For Temporal Identity, still images may be insufficient. Where the design intent involves time, memory, occupation, seasonal change, or ritual repetition, Stage 4 may use image sequences, animation, or short video in future applications. In such cases, the same rule applies: the candidate set is bounded, documented, preserved, and evaluated according to TACT criteria rather than visual impressiveness alone.

4.6 Stage 5: Comparative Fidelity Evaluation

In a full implementation, Stage 5 would evaluate generated outputs against the same TACT criteria that informed their production. A vision-capable AI model may assist by producing a structured comparative report, but the evaluation remains grounded in the TACT framework.

The evaluation asks four questions. First, does the image preserve the intended tactile, atmospheric, cultural, and temporal qualities? Second, where does the image drift into generic visual effect, superficial ornament, or cultural ambiguity? Third, which aspects of the intent cannot be adequately judged in a still image? Fourth, what should be changed in the prompt, reference material, generation method, or evaluation criteria?

The output is not primarily a numerical score. It is a structured evaluative narrative that identifies the image’s strengths, weaknesses, ambiguities, and revision needs. Quantitative scoring may be added in future empirical validation studies, but the present methodological framework prioritises traceability, interpretive clarity, and design usefulness.

4.7 Human Validation and Calibration

Human validation follows the AI-mediated stages. Expert reviewers would receive the reference material, Stage 2 analysis, Stage 3 prompts, generated outputs, and Stage 5 evaluation report. Their role is not simply to score images, but to audit the quality of the entire interpretive chain.

For ENIA, expert validation should ideally include architectural and cultural knowledge relevant to Ethiopian spatial traditions, vernacular tectonics, contemporary airport design, and symbolic national infrastructure. Validators may examine whether the pipeline correctly interpreted the original design intent, whether images preserve or distort cultural specificity, whether atmosphere is genuinely context-specific, whether materiality feels tactile and climatically appropriate, and whether temporal identity is visible or impossible to judge in the chosen medium.

This structure preserves human expertise where it is most valuable: not in repetitive scoring, but in interpretive calibration and design-direction recommendation. Divergence between AI evaluation and expert judgment should be recorded as meaningful evidence rather than treated as procedural failure.

4.8 Feedback Loops and Iterative Refinement

Although presented as a five-stage sequence, the framework is iterative. Stage 5 evaluation and human validation may require a return to any earlier stage. If cultural ambiguity appears, the process may return to Stage 1 to enrich the reference material or to Stage 2 to refine the cultural analysis. If the image is atmospheric but generic, the process may return to Stage 3 to sharpen the A3* prompt language. If materiality appears superficial, tactile prompt language may need revision. If Temporal Identity is weak, the process may return to Stage 4 and shift from still-image generation to sequential or video-based output.

The feedback logic is summarised in Table 6. TACT does not only judge whether an image succeeds; it identifies where fidelity breaks down and which stage of the pipeline should be revised.

Table 6. Feedback logic for TACT-based pipeline revision

Evaluation Failure	Likely Source of Failure	Pipeline Revision
Generic atmosphere	Weak or underdeveloped A3* specification	Revise atmospheric prompt around site-specific light, climate, air quality, spatial mood, and cultural-environmental conditions.
Cultural ambiguity	Broad cultural descriptors or generic regional references	Enrich reference material and add precise cultural, spatial, symbolic, and material references.
Superficial materiality	Texture treated as visual decoration	Strengthen prompt language around material behaviour, weight, surface depth, assembly, and bodily comfort.
Weak temporal depth	Still-image limitation or vague Temporal Identity criteria	Use sequential images, animation, or video; alternatively, revise Temporal Identity indicators.
Visually strong but intent-poor image	Candidate selection based on aesthetic appeal	Re-rank candidate pool using TACT criteria and select for fidelity to design intent.
AI evaluation conflicts with expert judgment	Cultural, tacit, or embodied knowledge not captured by the model	Use expert feedback to recalibrate Stage 2 analysis, Stage 3 prompt, or Stage 5 criteria.
Repeated prompt failure across candidates	Insufficient Stage 1 reference material or unclear design intent	Return to Stage 1 and enrich the empathetic profile with clearer narratives, references, and precedents.
Culturally specific but architecturally weak image	Motifs present without spatial, tectonic, or programmatic integration	Embed culture through spatial hierarchy, threshold, gathering, orientation, material assembly, and use.

4.9 Framework Assumptions and Implementation Scope

The framework rests on several assumptions that define its applicability and scope. These assumptions are stated explicitly in Table 7 to clarify what the method can and cannot claim.

Table 7. Framework assumptions and justifications

Assumption	Justification	Scope	When to Reconsider
Design intent can be translated into evaluative criteria	Architectural intent is often tacit, but frameworks such as TACT help make it explicit	Suitable for concept-stage evaluation	When the design brief is too vague or undocumented
AI can assist in analysis but not replace judgment	Multimodal AI can organise and compare material, but cultural and embodied meaning require expertise	Useful for structured evaluation	When outputs concern sensitive cultural interpretation
Prompt construction should be auditable	Prompts shape outputs and must be treated as methodological artifacts	Essential for research transparency	When using AI only for informal ideation
Visual fidelity is not the same as empathetic fidelity	Images may be persuasive while failing intent	Central to all AI-generated architectural evaluation	Always retained
Still images have limits	Time, ritual, use, and memory are difficult to represent in single images	Suitable for early-stage image evaluation	Use video or sequence when Temporal Identity is central
Expert disagreement is informative	Divergent readings may reveal ambiguity rather than error	Important for Culture and Temporal Identity	Do not collapse disagreement into forced consensus

The framework is designed for concept-stage architectural imagery and methodological research. It does not evaluate technical feasibility, construction documentation, cost, environmental performance, safety, or post-occupancy experience. Its contribution is specific: it provides a structured, auditable method for judging whether AI-generated architectural images remain faithful to empathetic design intent.

V. METHODOLOGICAL CONTRIBUTION AND OPERATIONAL IMPLICATIONS

5.1 From Image Production to Traceable Image Judgment

The framework addresses a central methodological question in AI-assisted architecture: can AI-generated imagery be evaluated as an architectural proposition rather than as a visually persuasive artifact? In many AI-assisted workflows, image production and architectural judgment remain loosely connected. The designer writes a prompt, the model generates outputs, and the preferred image is selected through aesthetic appeal, visual coherence, or client response. This sequence risks allowing the image to become authoritative before its relationship to design intent has been examined.

The methodological shift is therefore not simply that TACT evaluates images. It is that TACT structures the full chain through which an image is produced and judged. Reference material is organised into an empathetic profile; that profile is translated into TACT-informed criteria; the criteria guide prompt construction; the prompt generates a candidate image set; and the selected image is evaluated against the same criteria that shaped its production. Figure 3 illustrates this shift from visual production to intent-fidelity evaluation.

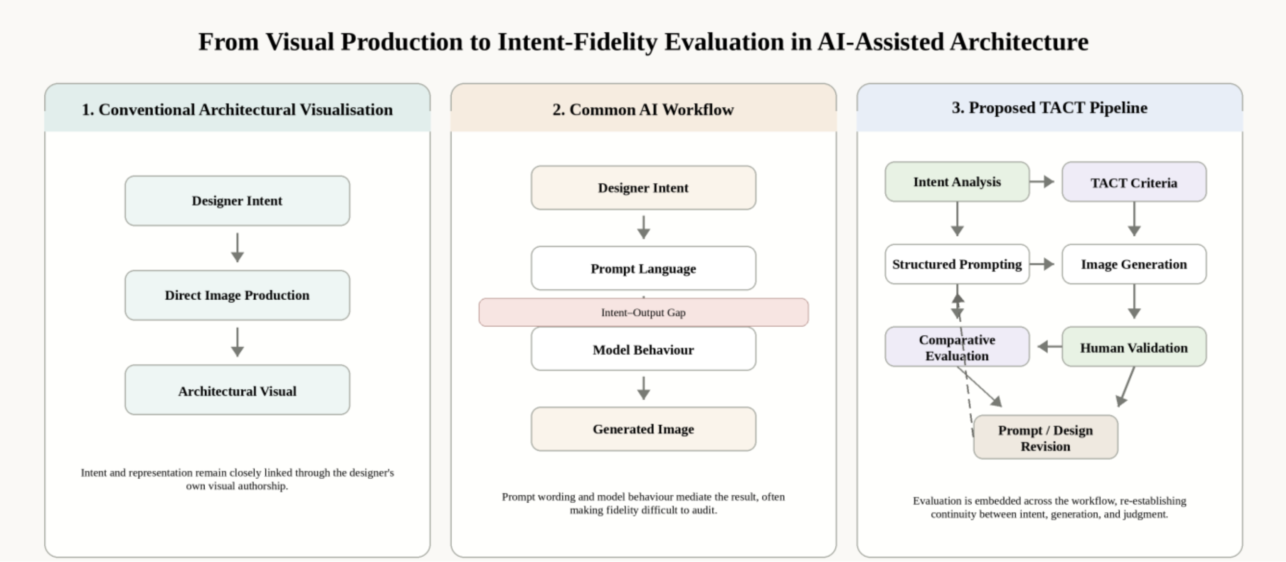

Figure 3. From visual production to intent-fidelity evaluation in AI-assisted architecture. Source: created by the author.

5.2 TACT as an Operational Constraint

A key methodological contribution of the framework is that empathetic fidelity becomes an operational constraint rather than a retrospective interpretive concern. In many AI-image workflows, qualities such as atmosphere, cultural identity, material depth, and temporal memory are judged only after an image has been produced. The proposed method instead positions these qualities as criteria that must guide the image from the beginning.

TACT makes this translation more disciplined. Each prompt is required to carry four categories of intent: Tactility, Atmosphere, Culture, and Temporal Identity. These are not decorative additions to the prompt; they are operational constraints that structure what the image should preserve. In the ENIA case, the ambition to bring Ethiopia in cannot be reduced to surface motifs or warm airport imagery. It must be operationalized through highland light, Rift Valley spatiality, tactile groundedness, threshold hospitality, communal identity, and continuity between heritage and future infrastructure.

5.3 Candidate Selection as a Methodological Step

Candidate selection is a major source of bias in AI-assisted design because multiple outputs are usually generated, while only the most visually convincing image is shown. In the proposed framework, selection is therefore treated as a methodological step, not a neutral curatorial choice.

In a full implementation, each space would use a fixed, pre-declared candidate pool, with all outputs preserved in the research record. The selected image should be chosen according to the Stage 2 TACT criteria, not aesthetic appeal alone. The preferred image is the one that best preserves the intended tactile, atmospheric, cultural, and temporal qualities.

This makes selection auditable. Reviewers can examine whether the chosen image was selected because it was visually impressive, or because it remained most faithful to the design intent.

5.4 Ambiguity as Diagnostic Evidence

The framework does not treat all disagreement or ambiguity as methodological failure. In architectural evaluation, especially in culturally specific contexts, multiple readings may be meaningful. However, the framework distinguishes between productive ambiguity and diagnostic ambiguity.

Productive ambiguity occurs when an image supports layered interpretation while remaining grounded in the intended architectural field. Diagnostic ambiguity occurs when the image cannot be situated clearly within the intended cultural, material, or spatial logic. For example, an ENIA image that is read as Ethiopian, generically African, Moroccan, or Middle Eastern may not have achieved cultural specificity; it may have produced cultural atmosphere without cultural precision.

5.5 From Image Critique to Design Feedback

The final methodological contribution is the conversion of image critique into design feedback. In conventional workflows, critique often stops at judgment: an image works, does not work, or needs improvement. In the proposed framework, evaluation identifies the specific stage at which fidelity breaks down.

If an image is atmospheric but generic, the problem may lie in weak A3* prompt specification. If an image is culturally ambiguous, the Stage 1 reference material may be insufficient or the Stage 3 cultural language too broad. If materiality appears superficial, tactile criteria may need to shift from texture to constructional depth. If Temporal Identity is weak, the limitation may be the still-image medium itself. The framework therefore creates a feedback loop: Evaluation -> Diagnosis -> Prompt Revision -> Regeneration -> Expert Validation.

5.6 Summary of Methodological Contribution

The methodological contribution of the framework lies in converting AI-generated architectural imagery into a traceable design proposition. The image is no longer assessed only as a visual object. It is evaluated as the outcome of a documented chain linking design intent, TACT analysis, prompt construction, generation, selection, and validation.

This gives the framework three practical values. First, it reduces the opacity of AI-assisted image production. Second, it provides a structured method for distinguishing visual appeal from design-intent fidelity. Third, it turns evaluation into a feedback mechanism that supports iterative design development. In this sense, the framework moves beyond visual persuasion.

VI. THEORETICAL IMPLICATIONS, LIMITATIONS, AND FUTURE DIRECTIONS

6.1 AI-Assisted Design as Intent Formalization

The proposed framework has broader theoretical implications for how architectural knowledge is translated into AI-assisted workflows. Architects often work with forms of knowledge that are difficult to state explicitly: atmospheric judgment, cultural sensitivity, material intuition, bodily scale, symbolic resonance, and temporal memory. These forms of knowledge usually operate through sketches, precedents, conversations, site experience, and iterative critique rather than through fixed evaluation rules.

The revised TACT framework can be understood as a formalization of this tacit intent. It does not claim to capture architectural judgment completely. Rather, it provides a structure through which selected aspects of that judgment can be made explicit enough to guide AI-mediated generation and evaluation. The theoretical value of TACT lies in its ability to make architectural intent discussable, traceable, and contestable within an AI workflow.

6.2 Bridging Human Empathy and Machine Generation

The framework clarifies the relationship between human and machine roles in architectural design. Human designers bring situated understanding, cultural awareness, bodily intuition, symbolic interpretation, and ethical responsibility. AI systems contribute speed, variation, pattern recognition, language-image translation, and structured comparison. The challenge is that these capacities are not naturally aligned.

The TACT framework operates as an intermediate language between these two modes. It allows human intent to be expressed in a form that can guide machine generation without pretending that the machine fully understands architectural meaning. AI is not treated as an autonomous judge of architectural value; it is treated as an interpretive partner whose outputs require human calibration.

6.3 Toward an Intent-Fidelity Paradigm

The theoretical position advanced by this paper is that AI-assisted architecture requires an intent-fidelity paradigm. Existing discussions of AI imagery often focus on quality, realism, novelty, speed, or creativity. These are important, but they do not answer the architectural question: does the image remain faithful to the design intent that produced it?

Intent fidelity shifts evaluation from the appearance of the image to its relationship with architectural meaning. A visually impressive image may fail if its materiality is superficial, its atmosphere generic, its cultural references unstable, or its temporality absent. Conversely, a less spectacular image may be more valuable if it better preserves the project's intended spatial and experiential logic.

6.4 Limitations and Scope

The framework's contribution must be understood within explicit limitations. First, ENIA is used as an illustrative application rather than as a completed statistical validation study. The paper demonstrates the methodological logic of the pipeline, but it does not claim generalizable empirical proof across cultures, typologies, or AI systems. The framework depends on author-defined interpretation of ENIA design intent; future work should include independent stakeholder and cultural expert validation of the reference-intent profile.

Second, the framework depends on the richness of the reference material entered at Stage 1. If the design brief is vague, the cultural references are weak, or the material and atmospheric intentions are poorly documented, the resulting empathetic profile and evaluation will be limited. Third, AI-assisted evaluation remains dependent on model capability. Multimodal models vary in their ability to interpret cultural references, spatial relationships, material qualities, and temporal cues.

Fourth, cultural interpretation cannot be fully automated. The framework uses AI to structure analysis and comparison, but human validation remains essential, particularly in non-Western contexts where generic training-data associations may be mistaken for cultural specificity. Fifth, still images are limited as evidence of Temporal Identity. Sequential imagery, animation, or video may be required where temporal fidelity is central. Finally, the framework is developed for concept-stage imagery; it does not evaluate technical feasibility, construction documentation, cost, environmental performance, safety, or post-occupancy user experience.

6.5 Future Directions and Extensions

Future research should test the framework across culturally distinct projects, such as civic buildings, transport hubs, museums, healthcare environments, and educational spaces. This would clarify whether the TACT dimensions remain stable across typologies or require project-specific adjustment.

Future studies should also compare AI-assisted TACT evaluation with expert human assessment, extend the pipeline to sequential and video-based media for Temporal Identity, develop clearer protocols for expert validation, and integrate TACT into professional design workflows as a feedback tool for prompt revision, design review, client presentation, and comparative evaluation of AI-generated concept options.

Finally, future research should connect concept-stage fidelity evaluation to later design outcomes. The key question is whether AI-generated images that perform well in empathetic fidelity lead to stronger design development, better stakeholder understanding, or more culturally grounded architectural decisions.

VII. CONCLUSION

This paper has proposed a methodological framework for evaluating fidelity to empathetic design intent in AI-generated architecture. Its central argument is that the problem facing AI-assisted architectural practice is no longer only image generation, but image judgment. Generative AI can now produce visually persuasive architectural imagery with remarkable speed, but the discipline still lacks robust methods for assessing whether such images remain faithful to the intentions that produced them.

The framework addresses this problem by repositioning TACT - Tactility, Atmosphere, Culture, and Temporal Identity - from a post-generation scoring instrument to an analytical infrastructure embedded across the full AI workflow. The proposed five-stage pipeline links empathetic material ingestion, TACT-informed analysis, structured prompt construction, image generation, comparative evaluation, and human validation within a single traceable process.

The key methodological contributions are fourfold. First, the framework formalizes empathetic design intent into evaluative dimensions that can guide AI-assisted generation. Second, it shifts evaluation from visual persuasion to empathetic intent fidelity. Third, it creates a feedback loop in which evaluation informs prompt revision and design development. Fourth, it preserves human expertise as a process of validation and calibration, rather than reducing architectural judgment to automated scoring.

The ENIA case illustrates how this method can be applied to a culturally specific architectural project. The ambition to bring Ethiopia in requires more than visually attractive airport imagery. It requires attention to Ethiopian identity, highland light, Rift Valley spatiality, tactile materiality, threshold hospitality, communal experience, and temporal continuity.

The fundamental thesis remains that AI-generated architectural images should not be judged primarily by how convincing they appear, but by how faithfully they carry the empathetic intentions they claim to represent. Fidelity to design intent therefore offers a necessary criterion for architectural judgment in the age of generative AI.

REFERENCES

[1] Pallasmaa, J. 2012. The eyes of the skin: Architecture and the senses. 3rd ed. Wiley.

[2] Holl, S., Pallasmaa, J., and Perez-Gomez, A. 1994. Questions of perception: Phenomenology of architecture. A+U Publishing.

[3] Zumthor, P. 2006. Atmospheres: Architectural environments, surrounding objects. Birkhauser.

[4] Leatherbarrow, D. 2000. Uncommon ground: Architecture, technology and topography. MIT Press.

[5] Bohme, G. 2017. Atmospheric architectures: The aesthetics of felt spaces. Bloomsbury Academic.

[6] Norberg-Schulz, C. 1980. Genius loci: Towards a phenomenology of architecture. Rizzoli.

[7] Frampton, K. 1983. Towards a critical regionalism: Six points for an architecture of resistance. In H. Foster, ed., The anti-aesthetic: Essays on postmodern culture, 16-30. Bay Press.

[8] Rapoport, A. 1969. House form and culture. Prentice-Hall.

[9] Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. 2021. On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610-623.

[10] Bianchi, F., Kalluri, P., Durmus, E., Ladhak, F., Cheng, M., Nozza, D., Hashimoto, T., Jurafsky, D., Zou, J., and Caliskan, A. 2023. Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 1493-1504.

[11] Basu, A., Babu, R. V., and Pruthi, D. 2023. Inspecting the geographical representativeness of images from text-to-image models. Proceedings of the IEEE/CVF International Conference on Computer Vision, 5136-5147.

[12] Ruskin, J. 1849. The seven lamps of architecture. Smith, Elder & Co.

[13] Mostafavi, M., and Leatherbarrow, D. 1993. On weathering: The life of buildings in time. MIT Press.

[14] Brand, S. 1994. How buildings learn: What happens after they are built. Viking.

[15] Lynch, K. 1972. What time is this place? MIT Press.

[16] Davis, D., Burry, J., and Burry, M. 2011. Untangling parametric schemata: Enhancing collaboration through modular programming. International Journal of Architectural Computing, 9(2), 187-207.

[17] Kolarevic, B., and Duarte, J. P. 2021. Mass customization and design democratization. Routledge.

[18] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. 2014. Generative adversarial nets. Advances in Neural Information Processing Systems, 27.

[19] Ho, J., Jain, A., and Abbeel, P. 2020. Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840-6851.

[20] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10684-10695.

[21] Burry, M., and Louw, M. 2022. Artificial intelligence and architectural meaning. Architectural Design, 92(4), 24-31.

[22] Duarte, J. P. 2023. Artificial intelligence in architecture: Opportunities and challenges. Design Studies, 85, 101187.

[23] Hu, Y., Liu, B., Kasai, J., Wang, Y., Ostendorf, M., Krishna, R., and Smith, N. A. 2023. TIFA: Accurate and interpretable text-to-image faithfulness evaluation with question answering. Proceedings of the IEEE/CVF International Conference on Computer Vision, 20406-20417.

[24] Cho, J., Zala, A., and Bansal, M. 2023. DALL-Eval: Probing the reasoning skills and social biases of text-to-image generation models. Proceedings of the IEEE/CVF International Conference on Computer Vision, 3043-3054.

[25] Lee, T., Yasunaga, M., Meng, C., Mai, Y., Park, J. S., Gupta, A., Zhang, Y., Narayanan, D., Teufel, H., Bellagente, M., Kang, M., Park, T., Leskovec, J., Zhu, J.-Y., Fei-Fei, L., Wu, J., Ermon, S., and Liang, P. 2023. Holistic evaluation of text-to-image models (HEIM). Advances in Neural Information Processing Systems, 36.

[26] Picard, R. W. 1997. Affective computing. MIT Press.

[27] OpenAI. 2024. GPT-4o system card. arXiv:2410.21276.

[28] Gemini Team. 2025. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. arXiv:2507.06261.

[29] Cheung, L. H., Lombardi, D., and Di Marco, G. 2025. Agentic, multimodal large language model with behavioural parameters and self-iterative capabilities for conversational architectural design processes. Architectural Science Review. https://doi.org/10.1080/00038628.2025.2586655

[30] Mukherjee, A., and Ghosh, S. 2025. Toward socially aware vision-language models: Evaluating cultural competence through multimodal story generation. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops.

[31] Yadav, S., Zhang, Z., Hershcovich, D., and Shutova, E. 2025. Beyond words: Exploring cultural value sensitivity in multimodal models. Findings of the Association for Computational Linguistics: NAACL 2025, 7607-7623.

[32] Schon, D. A. 1983. The reflective practitioner: How professionals think in action. Basic Books.

[33] Chung, J. J. Y., Kim, W., Yoo, K. M., Lee, H., Adar, E., and Chang, M. 2022. TaleBrush: Sketching stories with generative pretrained language models. Proceedings of the CHI Conference on Human Factors in Computing Systems, 1-19.

[34] Cross, N. 2011. Design thinking: Understanding how designers think and work. Berg.