

The Evolution of Meta-Learning in AI: Concepts, Taxonomies, and Implications

Muhammad Nagy
Department of Architecture
Ain Shams University
Cairo, Egypt.

Yasser Mansour
Department of Architecture
Ain Shams University
Cairo, Egypt.

Ahmed Iraqi
Department of Architecture
Ain Shams University
Cairo, Egypt.

Abstract - Meta-learning, broadly defined as learning to learn, has evolved from a niche optimization strategy into a foundational paradigm reshaping how artificial intelligence systems acquire, transfer, and generalize knowledge across tasks and domains. This paper presents a narrative and conceptual review tracing the evolution of meta-learning through its theoretical foundations, algorithmic paradigms, and taxonomic developments. Beginning with the bilevel optimization formulation and episodic training framework, the review examines the emergence of three paradigmatic families—metric-based, optimization-based, and model-based methods—and documents how the scope of what is learned progressively widened from model parameters to learning rates, loss functions, and neural architectures. The paper analyzes the progressive dissolution of boundaries between meta-learning, transfer learning, and multi-task learning, and examines how in-context learning within foundation models represents a conceptual reconvergence of classical meta-learning principles at unprecedented scale. Cross-domain implications are explored across knowledge-based systems, predictive modeling, personalized assessment, and user-centered adaptive systems. The review identifies persistent open challenges — including scalability constraints, task distribution assumptions, theoretical gaps, and ethical considerations — that arise directly from the evolutionary trajectory documented herein.

Keywords—*Meta-Learning, Few-Shot Learning, Taxonomy, Foundation Models, Bilevel Optimization, Transfer Learning, Neural Architecture Search*

I. INTRODUCTION

The trajectory of artificial intelligence research reveals a recurring tension between generality and specialization. Early AI systems pursued generality through hand-crafted knowledge representations [1], while machine learning shifted intelligence from explicit rules to statistical regularities extracted from data [2]. Deep learning further amplified this paradigm through hierarchical representations [3]. Yet each advance reinforced a fundamental limitation: the assumption that learning occurs in isolation, from scratch, for each new task. Meta-learning—the systematic study of how learning systems can acquire the capacity to learn more efficiently across tasks—represents a principled attempt to transcend this limitation.

This paper provides an evolutionary analysis of meta-learning within the broader AI landscape: tracing its conceptual genealogy, examining its theoretical and algorithmic maturation, mapping its taxonomic boundaries, and articulating its cross-domain implications. Unlike technical catalogs that inventory algorithms, this review adopts a narrative-analytical

approach seeking to understand why meta-learning emerged, how its core ideas have evolved, and what this evolution implies for adaptive intelligence.

A. Conceptual Genealogy and Motivation

Russell and Norvig [1] established the canonical framework in which AI encompasses computational systems designed to perceive, reason, and act. Within this framework, machine learning constitutes a subfield concerned with algorithms that improve through data exposure, and deep learning represents a methodological specialization employing multi-layered neural networks [2]. Chua et al. [2] demonstrated that conflation of AI, ML, DL, and data mining creates methodological inconsistencies, and their systematic taxonomy clarified that AI functions as the encompassing domain, ML as its learning-capable subset, and DL as a representation-specialized stratum. This hierarchical clarification is essential for meta-learning, which operates at a fundamentally different level of abstraction: rather than learning input-to-output mappings within a fixed task, it learns to optimize the learning process itself across a distribution of tasks.

The motivation for meta-learning emerges from the empirical observation that no single learning algorithm dominates across all tasks. Abdullah et al. [3] demonstrated that neither standalone nor hybrid ML techniques can satisfy all evaluated metrics simultaneously, echoing the classical No Free Lunch theorems. If no universal algorithm exists, then the capacity to select, configure, or construct appropriate algorithms for novel tasks becomes a form of intelligence. In deployment scenarios—medical diagnostics with limited records, robotics in unpredictable environments, NLP for low-resource languages—the assumption that sufficient labeled data exists for training each new model is frequently violated [4], [5]. Meta-learning addresses this by optimizing a learning procedure across a distribution of tasks, each potentially characterized by minimal data.

B. Scope and Differentiation

The meta-learning literature has been served by significant surveys, each adopting distinct analytical lenses. Vettoruzzo et al. [4] provided the most comprehensive recent technical review in IEEE TPAMI, covering state-of-the-art approaches and relationships with adjacent fields. Bahranifard and Ghaffari [5] offered a complementary paradigm survey organizing the field around four core paradigms. Despite their contributions, these surveys treat meta-learning primarily as a technical domain. The present review addresses this gap through three

differentiating commitments summarized in Table 1: (a) foregrounding conceptual evolution over algorithmic inventory, (b) integrating cross-domain implications as a co-equal analytical component, and (c) establishing causal linkages between evolutionary analysis and open problems.

TABLE I. DIFFERENTIATION OF THE PRESENT REVIEW FROM EXISTING SURVEYS

Analytical Dimension	Vettoruzzo et al. [4]	Bahrnifard & Ghaffari [5]	Present Review
Primary Focus	Technical methods & benchmarks	Paradigm taxonomy & applications	Conceptual evolution & implications
Analytical Lens	Algorithmic inventory	Paradigm classification	Evolutionary narrative
Cross-Domain Coverage	Limited to ML domains	Selected application areas	Thematic implications across domains
Open Problems Framing	Listed as future directions	Identified as challenges	Causally linked to evolutionary shifts
Foundation Model Analysis	Mentioned briefly	Emerging direction	Dedicated section (Sec. 6)

This paper adopts a narrative review methodology, selected over systematic review protocols for reasons intrinsic to its analytical objectives. The reference corpus comprises 44 sources spanning foundational AI texts [1] to cutting-edge contributions published in 2025 [2], [3], organized thematically rather than by domain.

II. THEORETICAL FOUNDATIONS

A. The Bilevel Optimization Formulation

The formal distinction between meta-learning and conventional machine learning is expressed through their respective optimization structures. In standard supervised learning, parameters θ minimize a loss $L(\theta, D)$ on a single task. Meta-learning introduces a second level: rather than treating the learning configuration as given, it treats it as learnable—what Bouchattaoui [7] terms the "meta-knowledge" ω . At the inner level, a base learner optimizes task-specific parameters given ω and a task's training data. At the outer level, the meta-learner optimizes ω across a distribution of tasks to minimize expected loss on held-out data:

$$\omega^* = \arg \min_{\omega} \sum_{i=1}^M \mathcal{L}_{\text{meta}} \left(\theta_i^*(\omega), \omega, D_{\text{src}}^{\text{val}}(i) \right) \quad (1)$$

subject to:

$$\theta_i^*(\omega) = \arg \min_{\theta} \mathcal{L}_{\text{task}} \left(\theta, \omega, D_{\text{src}}^{\text{train}}(i) \right) \quad (2)$$

This bilevel structure encodes the fundamental epistemological shift defining meta-learning: in conventional learning, the practitioner occupies the outer loop—manually selecting architectures, learning rates—while the algorithm occupies only the inner loop. Meta-learning automates the outer loop. The mathematical prerequisites draw upon optimization theory foundations: Rameshkumar [6] articulated the underpinnings of gradient-based optimization from basic gradient descent through adaptive methods such as Adam, establishing convergence properties dependent on loss surface geometry. Mohammadi et al. [8] observed that meta-learning differs from classical ML "with respect to the level of adaptation," distinguishing between the fixed bias of base learning and the learnable bias of meta-learning.

B. Representation Learning and Generalization Bounds

A central theoretical insight is that meta-knowledge ω can be understood as a shared representation—a mapping from raw inputs to an intermediate feature space capturing structure common across tasks. Bouchattaoui [7] decomposed the hypothesis space $H = G \circ F$, where F constitutes the shared representation space and G the task-specific head space. The meta-learning objective reduces to finding the representation that minimizes empirical loss averaged across tasks. The theoretical power lies in generalization guarantees: Theorem 3.1 bounds per-task examples n required for good within-task generalization, and Theorem 3.2 extends this to bound both tasks n and examples m for across-task generalization.

The key structural insight is that the shared representation compresses the learning problem's complexity: instead of requiring each task to independently learn features, the representation learner amortizes this cost across tasks. Per-task sample complexity depends only on G 's capacity, while the representation cost is distributed across n tasks. This is the formal basis for meta-learning's few-shot capability (Table 2).

TABLE II. COMPARISON OF GENERALIZATION FRAMEWORKS

Dimension	Standard Learning	Meta-Learning
Data Structure	Single dataset D from one task T	Meta-sample: n tasks \times m examples per task
Optimization	Single-level: $\min_{\theta} L(\theta, D)$	Bilevel: outer (ω) and inner (θ)
What Is Learned	Task-specific parameters θ	Shared representation f + task-specific heads g_i
Generalization Bound	Depends on capacity of H and $ D $	Depends on capacity of G and F , plus n and m [7]
Sample Efficiency	Requires large $ D $ per task	Small m per task; amortized across n tasks
No Free Lunch	Fully applicable	Mitigated by non-uniform task distribution [7]

C. Task Distributions and Episodic Training

Both the bilevel formulation and representation learning framework presuppose a concept absent from classical ML: a distribution over tasks. Bouchattaoui [7] formalized this as an environment E defined over task distributions, from which individual tasks are sampled. Performance is measured by the transfer risk: $R(A, E) = E_{D \sim E} [E_{S \sim D \otimes m} [R(A(S), D)]]$, evaluating the expected risk when confronted with a new task. A critical observation is that the No Free Lunch theorem does not constrain meta-learning in the same way—because tasks are sampled from a non-uniform distribution E , there exists exploitable structure that a meta-learner can leverage [7].

The episodic training paradigm operationalizes these constructs: each episode samples a task, splits it into support and query sets, and optimizes meta-parameters to minimize query loss after adaptation on the support set. This directly implements bilevel optimization within the stochastic task-sampling framework. The episodic structure explicitly optimizes for rapid adaptation—evaluating performance after adaptation, not merely on training data—enabling few-shot generalization [4], [8].

D. Evolution of the Formal Framework

The theoretical foundations emerged progressively through four phases (Table 3): (I) conceptual foundations articulating "learning to learn" as informal principle [8]; (II) bilevel optimization formalization establishing mathematical objectives [7]; (III) representation learning theory with generalization bounds via hypothesis space decomposition [7]; and (IV) task-distribution formalism with environmental measures and transfer risk [7]. Each phase enabled new algorithmic classes: bilevel formulation motivated MAML; representation theory motivated metric-based methods; task-distribution formalism provided the episodic training paradigm.

TABLE III. EVOLUTION OF META-LEARNING'S THEORETICAL FRAMEWORK

Phase	Theoretical Contribution	Key Formalism	Algorithmic Enablement
I: Conceptual	"Learning to learn" as informal principle [8]	Descriptive vocabulary; no unified objective	Heuristic algorithm configuration
II: Bilevel Opt.	Meta-learning as nested optimization [7]	$\min_{\theta} \sum \mathcal{A}^{\text{meta}}(\theta^*(\omega))$ s.t. $\theta^* = \operatorname{argmin} \mathcal{A}^{\text{task}}$	Gradient-based meta-learning (MAML)
III: Representation	Hypothesis decomposition $H = G \circ F$ [7]	Theorems on m (per-task) and n (tasks) bounds	Metric-based and embedding methods
IV: Task-Distribution	Environmental measure E ; transfer risk [7]	$R(A,E) = E_{\omega} [D \sim E] [E_{\omega} \{S \sim D \otimes m\} [R(A(S),D)]]$	Episodic training; task-sampling

III. PARADIGMS AND ALGORITHMIC EVOLUTION

A. Metric-Based Methods

Metric-based meta-learning operationalizes the representation learning framework by learning an embedding function under which semantic similarity corresponds to geometric proximity, reducing classification to nearest-neighbor retrieval. He et al. [10] traced the lineage from Siamese networks through Matching Networks (introducing episodic training to metric learning), Prototypical Networks (computing class prototypes as mean embeddings), and Relation Networks (replacing fixed distance metrics with learned similarity functions). Gharoun et al. [13] clarified that the paradigm's principal advantage lies in computational efficiency—adaptation requires no gradient computation—but at the cost of constraining task-specific adaptation to fixed distance functions. He et al. [10] reported competitive performance on standard benchmarks (MiniImageNet, TieredImageNet), particularly in very few-shot settings, while revealing struggles with tasks requiring complex decision boundaries.

B. Optimization-Based Methods and MAML

Model-Agnostic Meta-Learning (MAML), introduced by Finn et al. [11], trains initial parameters θ such that few gradient steps on a new task's support data yield parameters θ' that generalize well on query data. The meta-objective computes $\theta \leftarrow \theta - \beta \nabla_{\theta} \sum \mathcal{L}_{T_i}(f_{\theta,i})$, where $\theta'_i = \theta - \alpha \nabla_{\theta} \mathcal{L}_{T_i}(f_{\theta})$ Finn et al. [11] demonstrated model-agnosticism and problem-agnosticism—applicable to

classification, regression, and reinforcement learning. Finn and Levine [12] subsequently proved that for sufficiently deep networks, MAML combined with gradient descent has the same representational power as any arbitrary learning algorithm, resolving whether MAML's simplicity came at a representational cost. Alom [9] noted that MAML catalyzed extensions including first-order approximations (FOMAML, Reptile) and variants that meta-learn inner-loop learning rates, adaptation steps, or loss functions.

C. Model-Based Methods

Model-based methods learn an entire learning algorithm implemented as a neural network, typically a recurrent or memory-augmented architecture that ingests the support set and produces predictions in a single forward pass. He et al. [10] identified their defining characteristic as external or internal memory for accumulating task-specific knowledge. Finn and Levine's [12] universality analysis established that recurrent meta-learners are universal learning procedure approximators, but that MAML achieves the same universality. The critical difference lies not in expressive capacity but in inductive bias and statistical efficiency: model-based methods make no assumption about learning algorithm structure, offering flexibility at the cost of requiring more meta-training data [9], [13].

D. Widening Scope: From Parameters to Architectures

The preceding paradigms share a significant constraint: the neural architecture is fixed during meta-learning. Elsken et al. [14] introduced MetaNAS, integrating gradient-based NAS with MAML—both methods optimize nested objectives using gradient descent and can be combined into a single bilevel procedure that jointly meta-learns weights and architecture. Zhao et al. [15] extended this with H-Meta-NAS, addressing hardware heterogeneity by integrating MAML into a hardware-aware NAS flow, reducing search complexity from $O(T \times H \times C)$ to $O(1)$. This trajectory reflects meta-learning's maturation from a technique for few-shot classification into a comprehensive framework for adaptive system design.

E. Comparative Analysis

The paradigms represent fundamentally different answers to what meta-knowledge should consist of. The evolutionary trajectory reveals a consistent pattern: each successive paradigm broadens the scope of what is meta-learned (Table 4). This expansion has not rendered earlier paradigms obsolete—metric-based methods remain preferred in latency-critical applications, and MAML's simplicity continues to make it a dominant baseline.

TABLE IV. COMPARATIVE ANALYSIS OF META-LEARNING PARADIGMS AND THEIR THEORETICAL FOUNDATIONS

Paradigm	Meta-Knowledge	Inner-Loop	Expressiveness	Key Innovation	Ref.
Metric-Based	Embedding space	Non-parametric distance	Limited by fixed metric	Learned similarity with non-parametric classification	[10], [13]
Optimization (MAML)	Initialization θ	Gradient descent	Universal (proven [12])	Model-agnostic initialization for rapid	[11], [12]

				adaptation	
Model-Based	Learned learning algorithm	Forward pass	Universal (by construction)	External memory for task knowledge accumulation	[10], [13]
Architecture-Level	Architecture + weights	Gradient on both	Universal structure	+ Joint meta-learning via DARTS + hardware awareness	[14], [15]

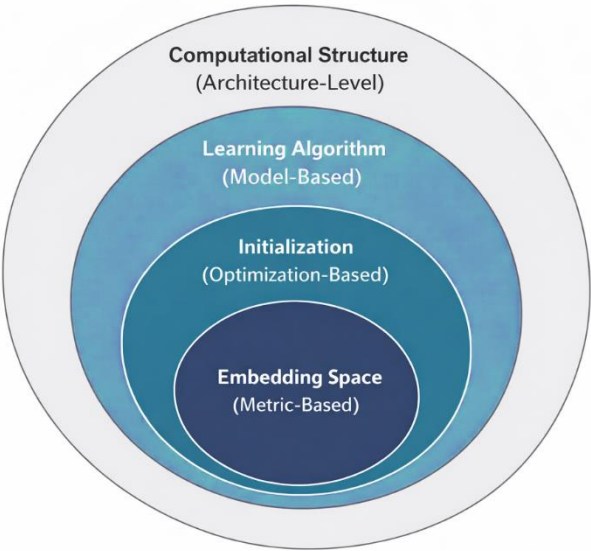


Fig. 1. Hierarchical scope of meta-learning paradigms. Concentric regions represent progressively broader scopes: embedding space (metric) → initialization (optimization) → learning algorithm (model-based) → computational structure (architecture-level).

IV. TAXONOMIC EVOLUTION AND PARADIGM BOUNDARIES

A. Meta-Learning vs. Transfer Learning vs. Multi-Task Learning

Upadhyay et al. [16] provided the most systematic comparative analysis of these three paradigms. All involve tasks defined over domains $\mathcal{D} = \{\mathcal{X}, p(\mathbf{x}, \mathbf{y})\}$ with loss functions, but differ in how knowledge-sharing is structured: transfer learning operates sequentially (source trained first, knowledge transferred to target); multi-task learning operates simultaneously (joint training with shared representations); meta-learning operates episodically (bilevel optimization extracting meta-knowledge ϕ enabling rapid adaptation to unseen tasks). The formal distinction becomes blurred in practice because multi-task and meta-training objectives are structurally similar—the distinction lies in the presence of the outer-loop meta-objective. Upadhyay et al. [16] revealed that the taxonomy is inherently multi-axis, with each paradigm occupying a distinct region in a space defined by training structure, knowledge type, and optimization level. Bahranifard and Ghaffari [5] independently corroborated this multi-axis view.

Multimodal, Continual, and Self-Supervised Extensions

Ma et al. [17] established multimodal meta-learning as a distinct research area, identifying two fundamental problems:

enrichment of task inputs through complementary multimodal information, and generalization across heterogeneous task distributions with different modality combinations. Their taxonomy organized algorithms by meta-knowledge type: learning the optimization (multimodal bilevel parameters), learning the embedding (multimodal prototypes and attention kernels), and learning the generation (cross-modal data augmentation). Bahranifard and Ghaffari [5] identified continual meta-learning (integrating rapid adaptation with catastrophic forgetting mitigation for non-stationary task distributions), self-supervised meta-learning (constructing pseudo-tasks from unlabeled data using contrastive learning or pretext tasks), and causal meta-learning (learning interventional meta-knowledge for robust out-of-distribution generalization) as emerging extensions. These expand meta-learning along three orthogonal dimensions: temporal dynamics, supervision requirements, and reasoning structure [5].

B. Cross-Domain Taxonomic Positioning

Cross-domain taxonomies reveal how meta-learning is perceived outside its native context. Vissers-Similon et al. [22] evaluated AI techniques for early architectural design across seven categories; notably, meta-learning does not appear as an independent category—its adaptive capabilities are distributed across Classic ML and Transformer categories. Castro Pena et al. [23] corroborated this in their review of AI for conceptual architectural design, where taxonomic organization follows application function rather than algorithmic lineage. Li et al. [24] demonstrated that design applications require combinations of generative, discriminative, and adaptive capabilities—a functional decomposition that crosscuts algorithmic taxonomy. This evidence reveals meta-learning's taxonomic transformation: from a specific algorithmic family to a general-purpose paradigm to a fundamental computational capability transcending any single category (Table 5).

TABLE V. CROSS-DOMAIN TAXONOMIC POSITIONING OF META-LEARNING

Study	Domain	Taxonomic Approach	Meta-Learning's Role
Vissers-Similon et al. [22]	Architectural design	Seven AI categories across four potential levels	Implicitly distributed; not independent category
Castro Pena et al. [23]	Conceptual design	Application-function-centered taxonomy	Subsumed under adaptive AI capabilities
Li et al. [24]	AI for design efficiency	Functional: generative, discriminative, adaptive	Positioned as the adaptive capability
Upadhyay et al. [16]	ML (general)	Multi-axis paradigm comparison	Distinct paradigm; converging toward hybrids

V. CONCEPTUAL SHIFTS AND PARADIGM MATURATION

A. Evolutionary Phases and Technique-to-Paradigm Transition

Meta-learning's evolution reveals four distinct phases. The foundational period (pre-2017) articulated "learning to learn" as programmatic rather than algorithmically precise [8], recognizing that single-task learning's limitations were structural [4]. The algorithmic crystallization (2017–2020),

catalyzed by MAML [11], generated the paradigmatic diversity documented in Section 3 and enabled the field to identify its structural dimensions. The taxonomic expansion (2020–2023) witnessed continual, self-supervised, causal, and multimodal extensions [5] alongside progressive convergence with transfer and multi-task learning [16]. The capability transformation (2023+) sees meta-learning transition from a paradigm defined by specific mechanisms to a fundamental computational capability—as evidenced by cross-domain taxonomies [22], [23], [24].

This trajectory constitutes a genuine technique-to-paradigm transition, evidenced by three structural markers: (a) convergence on shared theoretical vocabulary (bilevel optimization, task distributions) rather than ad hoc formalisms [4]; (b) development of systematic internal taxonomy with recognized trade-offs [5]; and (c) capacity to subsume adjacent fields—the bilevel structure generalizes multi-task learning and recovers transfer learning as special cases [16]. A technique has variants; a paradigm has schools of thought (Table 6).

B. Inductive Bias: From Hand-Crafted to Learned

The foundational conceptual shift unifying meta-learning's entire trajectory is the transition from hand-crafted to learned inductive biases. In classical ML, inductive bias is determined a priori: hypothesis class, loss function, regularization, optimization procedure [8]. Meta-learning's insight is that these choices can be parameterized and optimized through experience across tasks. Each paradigm implements this differently: metric-based methods learn the embedding space (similarity bias), optimization-based learn the initialization (starting-point bias), model-based learn the entire algorithm (procedural bias), and architecture-level learn computational structure (structural bias). Vettoruzzo et al. [4] documented a characteristic pattern: each generation expands the set of learnable components while preserving some fixed ones, and the next generation then makes those fixed components learnable. Finn et al.'s [11] universality theorem establishes that this shift does not sacrifice representational power—it is a strict generalization.

TABLE VI. PARADIGM MATURATION AND PROGRESSIVE EXPANSION OF LEARNABLE INDUCTIVE BIASES

Stage	Learnable Component	Fixed Components	Evidence
Classical ML	Parameters (weights) only	Architecture, init, learning rate, loss, hypothesis class	Standard supervised learning [8]
Metric-based Meta-Learning	Embedding space (similarity metric)	Architecture, adaptation procedure, task structure	Prototypical Matching Networks [4]
MAML (Optimization-based)	Initialization θ for gradient descent	Architecture, adaptation steps, LR schedule	Finn et al. [11]; Universality [12]
Model-based Meta-Learning	Entire learning algorithm	Meta-architecture, memory access patterns	MANNs, recurrent meta-learners [4], [5]
Architecture-level	Structure, initialization, weights	+NAS search space, hardware constraints	MetaNAS [14], H-Meta-NAS [15]
Frontier Extensions	Task distributions, supervision, causality	Core learning-to-learn objective	Continual, self-supervised, causal ML [5]
Paradigm Status: Technique →	Shared formalism, internal	N/A	[4], [5], [11], [16]

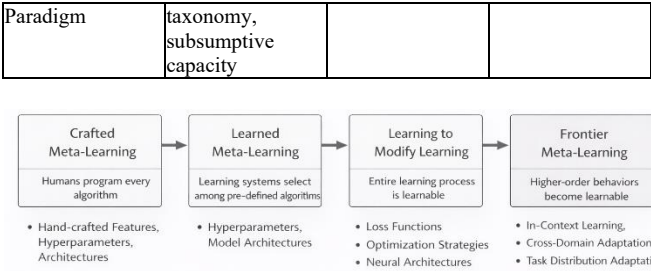


Fig. 2. Four evolutionary phases of meta-learning and the progressive liberation of inductive biases from hand-crafted to learned. Each generation expands what is learnable; the trajectory continues through frontier extensions.

VI. META-LEARNING IN THE FOUNDATION MODEL ERA

A. In-Context Learning as Implicit Meta-Learning

The rise of foundation models intersects with meta-learning's evolution at a critical juncture. Kılınç and Keçecioglu [18] traced generative AI's development from Shannon's communication theory through GANs (2014) to the Transformer architecture (2017)—the same year MAML catalyzed meta-learning's algorithmic crystallization. Both developments represented moves from domain-specific to general-purpose frameworks. The Transformer's attention mechanism enabled training on vastly larger datasets, producing models capable of learning transferable representations.

The most significant intersection lies in in-context learning (ICL)—the capacity of large language models to adapt behavior to new tasks based on demonstration examples within the input prompt, without parameter updates. In the standard meta-learning framework, a meta-learner optimizes initialization through an outer loop such that few gradient steps yield good performance. In ICL, a pretrained model achieves an analogous outcome through a fundamentally different mechanism: the forward pass processes demonstrations (analogous to the support set) and generates predictions (analogous to the query set), with adaptation occurring within activation patterns rather than through parameter optimization. This functional equivalence suggests that meta-learning's core functional capability—few-shot task adaptation—has become an emergent property of systems trained at sufficient scale on sufficiently diverse data [18], [19].

The distinction between explicit and implicit meta-learning carries profound implications. Explicit meta-learning treats the learning-to-learn objective as a distinct optimization problem requiring specialized procedures. Implicit meta-learning achieves functionally equivalent outcomes as an emergent property of large-scale pretraining. This suggests the learned inductive bias framework may be more general than its original bilevel optimization instantiation, and challenges taxonomic boundaries by suggesting meta-learning's identity may reside in functional capability rather than specific optimization structure (Table 7).

TABLE VII. COMPARISON OF EXPLICIT AND IMPLICIT META-LEARNING FRAMEWORKS

Dimension	Explicit Meta-Learning	Implicit Meta-Learning (ICL)
-----------	------------------------	------------------------------

Adaptation Mechanism	Gradient-based parameter updates via bilevel optimization	Attention-mediated context processing; no parameter updates
Training Paradigm	Episodic training over task distribution	Autoregressive pretraining on diverse corpora; task structure implicit
Task Specification	Formal: $T = (D_{train}, D_{test}, L)$ with class structure	Natural language demonstrations; task boundaries implicit
Inductive Bias Source	Learned initialization/embedding optimized over tasks	Learned attention patterns from massive pretraining
Data Requirements	Curated episodic datasets with task boundaries	Massive unstructured corpora; few-shot at inference only

Personalization in the Foundation Model Era

The convergence finds concrete expression in personalization—adapting general-purpose models to individual users with minimal data. Zhu et al. [21] demonstrated this rigorously for personalized image aesthetics assessment (PIAA), reconceptualizing each user's aesthetic preferences as a meta-learning task. Their BLG-PIAA approach directly instantiates bilevel optimization: the meta-training phase learns an aesthetic meta-learner through bilevel gradient updates, extracting shared prior knowledge about how people judge aesthetics—not what any individual judges, but the shared capacity for aesthetic judgment. Newton [25] documented that generative models in specialized domains face a persistent data curation bottleneck that meta-learning's few-shot capabilities directly address. Soares Koshiyama et al. [19] anticipated multi-capability foundation model systems combining generative, sequential, and adaptive components for capital markets. This cross-domain consistency suggests personalization via meta-learning represents a fundamental operational mode for the foundation model era.

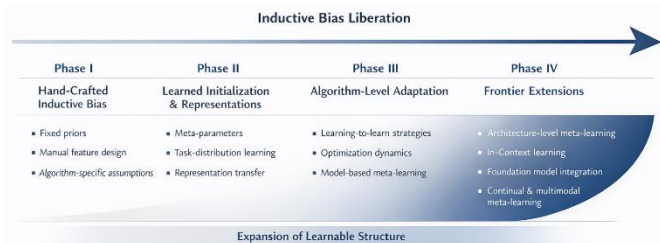


Fig. 3. Parallel and converging trajectories. Upper: generative AI milestones [18]. Lower: meta-learning milestones. Convergence at 2017 (Transformers + MAML) and the foundation model era (ICL operationalizes meta-learning).

VII. CROSS-DOMAIN IMPLICATIONS

This section demonstrates how meta-learning's core principles manifest in applied contexts extending beyond canonical few-shot benchmarks. Rather than organizing by domain, the analysis follows four thematic trajectories revealing progressive deepening of meta-learning's relevance.

A. From Knowledge-Based to Learning-Based Systems

The trajectory from rule-based AI to data-driven learning constitutes the foundational arc upon which meta-learning's cross-domain implications are built. El-Attar [26] developed a computational framework encoding architectural design expertise as explicit rules and case-based prototypes,

epitomizing the knowledge-based paradigm's bounded generalization. Mahmoodi [27] proposed cognitive meta-strategies for design education that anticipated meta-learning's formal objective of task-general inductive biases. Eastman [28] automated evaluation of preliminary courthouse designs using BIM, prefiguring the task-distribution framework where a common evaluation mechanism operates across diverse instances. Sönmez [29] traced the evolution from hand-operated Shape Grammars through Case-Based Design to ML approaches, noting that few-shot learning from precedent examples is precisely the operational mode meta-learning formalizes. Bhatt et al. [30] demonstrated predictive, evidence-based architecture design integrating spatial reasoning with empirical behavioral data, exemplifying the multi-modal cross-task transfer that meta-learning enables. This trajectory reveals progressive expansion of computational autonomy—from executing rules to learning from data to learning how to learn—driven precisely by the limitations meta-learning was designed to overcome.

B. Predictive and Performance-Based Systems

In domains characterized by data scarcity and cross-context variation, meta-learning's few-shot capabilities become operationally significant. Alotaibi [31] integrated multiple ML algorithms with explainable AI for residential energy prediction; the multi-model evaluation framework constitutes an informal approximation of meta-learning's model selection objective. Elbeltagi et al. [32] exposed the interoperability gap between parametric design tools and simulation engines—a bottleneck meta-learning's transferable prediction functions directly address. Runge and Zmeureanu [33] documented persistent challenges in model generalization across building types, mapping onto meta-learning's distinction between within-task and cross-task transfer. Panchalingam and Chan [34] found building AI heavily skewed toward domain-specific models with little cross-context transfer, underscoring meta-learning's unrealized potential. Krishnan et al. [35] introduced ArchGym, finding that with sufficiently tuned hyperparameters, no single ML algorithm consistently dominated—the "hyperparameter lottery" directly motivating meta-learning's task-adaptive strategy.

C. Personalized Assessment and Aesthetic Judgment

Personalized aesthetics is a domain where meta-learning is structurally necessary: each user constitutes a distinct task with limited data, while shared perceptual structure provides cross-task regularity. Zhu et al. [21] established the bilevel framework (Section 6.2). Zhu et al. [36] extended this to multi-attribute interactive reasoning, modeling interactions between objective image attributes and subjective user attributes. Yang et al. [37] provided the PARA dataset (31,220 images, 438 subjects, 13 dimensions), demonstrating that personalized preferences exhibit structured patterns correlated with measurable user characteristics—enabling meta-learning to leverage user metadata for more informative task representations. Zhang and Ban [38] represented the pre-meta-learning baseline (GIAA)—population-level prediction without personalization. Hartanto et al. [39] validated using EEG and eye-tracking that aesthetic responses have consistent physiological correlates, grounding meta-learning's assumption that individual tasks share underlying structure.

D. User-Centered Adaptive Systems

The most forward-looking implications emerge where meta-learning intersects with user-centered design leveraging physiological and behavioral data. Abdelmohsen et al. [40] introduced an affective computing framework generating emotionally responsive environments through multi-modal sensing—each user's emotional profile constitutes a personalization task operationalizing meta-learning's few-shot adaptation in real-time. Ma et al. [41] proposed IVR-based discrete choice modeling predicting design preferences across 162 alternatives; meta-learning could learn a shared preference model rapidly adaptable to new users. Tu and Nagakura [42] demonstrated measurable correlations between multi-modal physiological data and spatial parameters, providing the sensing infrastructure meta-learning personalization requires. Cho et al. [43] developed CNN-LSTM models for EEG-based architectural preference prediction; meta-learning could resolve their precision-recall tradeoff through task-adaptive ERP feature weighting. Zhang et al. [44] found expertise level moderates emotional perception of AI-generated architecture, constituting a task-defining variable in the meta-learning sense (Table 8).

TABLE VIII. CROSS-DOMAIN EVIDENCE FOR META-LEARNING IMPLICATIONS

Trajectory	Representative Studies	Key Finding	Meta-Learning Connection
Knowledge Learning	El-Attar [26]; Sönmez [29]; Bhatt et al. [30]	Progressive computational autonomy from rules to evidence-based prediction	Few-shot learning from precedents; cross-task transfer via learned representations
Predictive Systems	Alotaibi [31]; Runge & Zmeureanu [33]; Krishnan et al. [35]	No single algorithm dominates; generalization across contexts limited	Task-adaptive strategy; transferable prediction functions; hyperparameter lottery
Personalized Assessment	Zhu et al. [21],[36]; Yang et al. [37]; Hartanto et al. [39]	Individual aesthetics structured by measurable attributes; physiological correlates	Bilevel user-as-task formulation; structured task distribution via user metadata
User-Centered Adaptive	Abdelmohsen et al. [40]; Tu & Nagakura [42]; Cho et al. [43]	Multi-modal & physiological responses correlate with design parameters	Real-time few-shot emotion adaptation; task-adaptive neural feature weighting

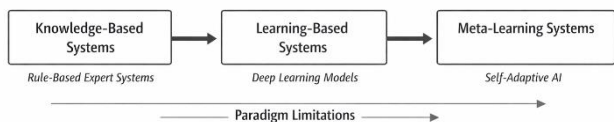


Fig. 4. Three-stage trajectory in applied AI: knowledge-based systems [26–28] → learning-based systems [29–30] → meta-learning systems, driven by successive paradigm limitations.

VIII. OPEN PROBLEMS ARISING FROM THE EVOLUTION

The evolutionary analysis reveals that meta-learning's open problems are logical consequences of its developmental trajectory, not arbitrary gaps. Each conceptual advance

simultaneously opened new capabilities and exposed new limitations.

Scalability is a direct consequence of the field's widening scope. The bilevel optimization structure imposes costs that scale multiplicatively with inner-loop steps, model size, and task count [4]. First-order approximations sacrifice precise meta-gradient information, creating a fundamental trade-off [5]. The architecture-level integration compounds this: MetaNAS [14] requires trilevel optimization, and H-Meta-NAS [15] demonstrated that naive multi-task multi-hardware deployment creates $O(T \times H \times C)$ complexity. However, Kuszczak et al. [20] showed meta-learned initializations reduced optimization iterations by 33.6% in neural topology optimization, with effective cross-resolution transfer, suggesting domain-informed strategies can manage the complexity-performance trade-off.

Task distribution assumptions were exposed by cross-domain transfer. Standard methods learn a single globally shared meta-parameter set, but performance degrades as task dissimilarity increases [4]. Meta-learning's generalization guarantees depend on test tasks being drawn from the same distribution as training tasks—systematically violated in cross-domain scenarios (Section 7). In building energy prediction, geographic and climatic factors create distribution shifts; in personalized aesthetics, user populations vary across cultures; in physiological systems, measurement artifacts create unpredictable shifts. The fundamental question is what constitutes a "task" when the task space itself is ill-defined [4], [5].

Theoretical gaps trace to boundary dissolution. The intersections between meta-learning and adjacent paradigms—mapped by Vettoruzzo et al. [4] across multi-task learning, transfer learning, domain adaptation, self-supervised learning, federated learning, and continual learning—create regions where existing formal guarantees do not apply. Causal meta-learning exposes gaps between statistical learning and structural causal models [5]. The NAS–meta-learning integration [14] lacks convergence analysis for the joint architecture–weight space. These gaps represent a widening disconnect between empirical capabilities and theoretical understanding.

Foundation model integration generates a distinct class of problems. The conceptual recognition that ICL constitutes implicit meta-learning raises unresolved questions: under what conditions does explicit bilevel optimization provide benefits beyond ICL? How can meta-learning's data-efficiency principles reduce foundation model training requirements? The bidirectional integration—where generative capabilities enhance meta-learning through synthetic task generation—creates feedback loops whose theoretical properties are unexplored [4], [5].

Ethical implications arise from scope expansion into human-centered domains. Meta-learning's few-shot personalization creates a structural tension between utility and privacy: the paradigm builds individualized models from minimal user-specific data, potentially making users more identifiable than in aggregate approaches. The EEG-based systems of Cho et al. [43] and physiological sensing of Tu and Nagakura [42] collect intimate neurophysiological data that, processed through few-shot personalization, could enable granular profiling. Abdelmohsen et al.'s [40] environmentally

responsive framework raises consent and autonomy questions. Addressing these challenges requires both technical solutions (differential privacy, consent-aware protocols) and normative frameworks for determining when few-shot user modeling is appropriate (Table 9).

TABLE IX. OPEN PROBLEMS: EVOLUTIONARY ORIGINS AND RESEARCH PRIORITIES

Open Problem	Evolutionary Origin	Current Status	Research Priority
Scalability	Widening scope (Sec. 3); cross-domain transfer (Sec. 7)	First-order approximations; implicit differentiation (SAMA)	Domain-informed initialization; modular meta-learning [4],[5],[14],[15],[20]
Task Distribution	Taxonomic expansion (Sec. 4); cross-domain implications (Sec. 7)	Multi-modal distribution methods; clustered initialization	Task space characterization; task-level regularization [4],[5]
Theoretical Gaps	Boundary dissolution (Sec. 4); paradigm shift (Sec. 5)	Paradigm-specific guarantees; no unified framework	Unified theory for boundary paradigms; architecture ML bounds [4],[5],[14]
Foundation Model Integration	Generative AI convergence (Sec. 6)	ICL recognized as implicit ML; bidirectional integration nascent	Hybrid ICL–bilevel frameworks; continual meta-learning [4],[5],[20]
Ethical Implications	Human-centered applications (Sec. 7)	Federated meta-learning; no normative frameworks	Differential privacy; consent-aware adaptation protocols [4],[5]

IX. CONCLUSION

This paper has traced the evolution of meta-learning from its formal inception as a bilevel optimization strategy to its current position as a foundational paradigm within artificial intelligence. The evolutionary narrative reveals a field whose development follows a discernible logic: each conceptual advance simultaneously resolved limitations and generated new challenges, producing the recursive pattern of capability expansion and problem emergence that characterizes genuinely transformative paradigms.

The review established that meta-learning's theoretical architecture rests upon bilevel optimization, representation learning with generalization bounds, and the task-distribution formalism (Section 2). Three paradigmatic families—metric-based, optimization-based, and model-based—represent fundamentally different answers to what constitutes transferable meta-knowledge, while the integration of NAS expanded the scope from parameters to computational structure itself (Section 3). The taxonomic analysis (Section 4) revealed progressive dissolution of boundaries between meta-learning and adjacent paradigms, reflecting a transition from technique to paradigm evidenced by shared formalism, systematic taxonomy, and subsumptive capacity (Section 5). The foundational shift from hand-crafted to learned inductive biases unifies the entire trajectory—a strict generalization that does not sacrifice representational power.

The foundation model era (Section 6) revealed that in-context learning constitutes an implicit instantiation of meta-learning principles at unprecedented scale, validating that

learning to learn is a fundamental computational principle rather than merely an algorithmic technique. The complementarity between foundation models' broad representational capacity and meta-learning's structured adaptation emerges as a defining theme. Cross-domain validation (Section 7) demonstrated meta-learning's implications across knowledge-based systems, predictive modeling, personalized assessment, and user-centered adaptive design—confirming its status as a general paradigm whose principles possess a generality transcending few-shot classification benchmarks.

The five open problems identified (Section 8)—scalability, task distribution, theoretical gaps, foundation model integration, and ethical implications—are not arbitrary gaps but structural consequences of the field's own developmental trajectory. Each solution creates conditions for the next set of problems, a recursive dynamic characteristic of paradigms with genuine intellectual depth. Ultimately, meta-learning's significance lies in its demonstration that the capacity to learn how to learn constitutes a qualitatively distinct level of adaptive intelligence—one bridging narrow task-specific optimization and the flexible, generalizable intelligence that remains the central aspiration of AI research (Table 10).

TABLE X. CONSOLIDATED EVOLUTIONARY SYNTHESIS

Phase	Key Developments	Conceptual Contribution	Paradigmatic Significance
Theoretical Foundations (Sec. 2)	Bilevel optimization; episodic training	Formal separation of learning levels	Mathematical architecture for all subsequent developments
Algorithmic Paradigms (Sec. 3)	Metric → Optimization → Model-based; NAS	Progressive scope widening	What is learned can itself be learned
Taxonomic Evolution (Sec. 4)	Boundary dissolution with adjacent paradigms	From discrete categories to continuous spectrum	Revealed structural commonalities across paradigms
Conceptual Shifts (Sec. 5)	Technique → Paradigm; inductive bias reconceptualization	From isolated strategy to coherent paradigm	General principle of adaptive intelligence
Foundation Model Era (Sec. 6)	ICL as implicit meta-learning; personalization	Reconvergence of classical principles	Principles operate beyond original bilevel framework
Cross-Domain (Sec. 7)	Knowledge systems; prediction; aesthetics; user-centered	Domain-specific instantiation of abstractions	Validated generality across application ontologies
Open Problems (Sec. 8)	Scalability; distributions; theory; integration; ethics	Causal linkage between advances and challenges	Research agenda generated by developmental trajectory

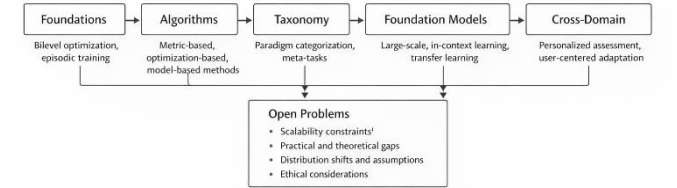


Fig. 5. Consolidated trajectory through seven phases: foundations → algorithms → taxonomy → concepts → foundation models → cross-domain → open problems. Arrows indicate causal relationships between phases and generated problems.

X. REFERENCES

- [1] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Upper Saddle River, NJ, USA: Pearson Education, 2016.
- [2] H. N. Chua, M. B. Jasser, B. Issa, and R. T. K. Wong, "Differentiating artificial intelligence, machine learning, deep learning, and data mining," *IEEE Access*, 2025.
- [3] A. A. Abdullah et al., "In-depth analysis on machine learning approaches: Techniques, applications, and trends," *ARO – The Scientific Journal of Koya University*, vol. 13, no. 1, 2025.
- [4] A. Vettoruzzo, M.-R. Bouguelia, J. Vanschoren, T. Rognvaldsson, and K. C. Santosh, "Advances and challenges in meta-learning: A technical review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 7, pp. 4763–4783, 2024.
- [5] Z. Bahranifard and S. Ghaffari, "A survey of meta-learning: Paradigms, applications, and challenges," *Amirkabir Univ. Technol.*, Tehran, Iran, preprint.
- [6] T. Rameshkumar, "Mathematical foundations of deep learning: Theory, algorithms, and practical applications," 2025.
- [7] M. E. Bouchattaoui, "Meta-learning and representation learner: A short theoretical note," *arXiv:2407.04189v2*, 2024.
- [8] F. Z. Mohammadi et al., "An introduction to advanced machine learning: Meta-learning algorithms, applications, and promises," in *Optimization, Learning, and Control for Interdependent Complex Networks*, AISC 1123, pp. 21–44, Springer, 2020.
- [9] M. Alom, "Meta-learning: Adaptive and fast learning systems," *J. Artificial Intelligence General Science*, vol. 2, no. 1, pp. 91–97, 2024.
- [10] Y. He et al., "Few-shot and meta-learning methods for image understanding: A survey," *Int. J. Multimedia Information Retrieval*, vol. 12, Article 14, 2023.
- [11] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. 34th Int. Conf. Machine Learning (ICML)*, PMLR 70, pp. 1126–1135, 2017.
- [12] C. Finn and S. Levine, "Meta-learning and universality: Deep representations and gradient descent can approximate any learning algorithm," in *Proc. ICLR*, 2018.
- [13] H. Gharoun, F. Momenifar, F. Chen, and A. H. Gandomi, "Meta-learning approaches for few-shot learning: A survey of recent advances," *ACM Computing Surveys*, vol. 56, no. 12, Article 310, 2024.
- [14] T. Elsken, B. Staffler, J. H. Metzner, and F. Hutter, "Meta-learning of neural architectures for few-shot learning," in *Proc. CVPR*, 2020.
- [15] D. Zhao et al., "Rapid model architecture adaption for meta-learning," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.
- [16] R. Upadhyay, R. Phlypo, R. Saini, and M. Liwicki, "Sharing to learn and learning to share: Fitting together meta, multi-task, and transfer learning: A meta review," *IEEE Access*, vol. 12, pp. 148553–148583, 2024.
- [17] Y. Ma, S. Zhao, W. Wang, Y. Li, and I. King, "Multimodality in meta-learning: A comprehensive survey," *Knowledge-Based Systems*, 2022.
- [18] H. K. Kılınc and Ö. F. Keçecioglu, "Generative artificial intelligence: A historical and future perspective," *Academic Platform J. Eng. Smart Systems*, vol. 12, no. 2, pp. 47–58, 2024.
- [19] A. Soares Koshiyama et al., "Algorithms in future capital markets," 2020.
- [20] I. Kuszczak, G. Kuś, F. Bosi, and M. A. Bessa, "Meta-neural topology optimization: Knowledge infusion with meta-learning," *arXiv:2502.01830*, 2025.
- [21] H. Zhu et al., "Personalized image aesthetics assessment via meta-learning with bilevel gradient optimization," *IEEE Trans. Cybernetics*, vol. 52, no. 3, pp. 1798–1811, 2022.
- [22] S. Vissers-Similon et al., "Classification of artificial intelligence techniques for early architectural design stages," 2025.
- [23] M. L. Castro Pena, A. Carballal, N. Rodríguez-Fernández, I. Santos, and J. Romero, "Artificial intelligence applied to conceptual design: A review of its use in architecture," *Automation in Construction*, vol. 124, 103550, 2021.
- [24] C. Li et al., "A review of artificial intelligence in enhancing architectural design efficiency," 2025.
- [25] D. Newton, "Generative deep learning in architectural design," *Technology|Architecture + Design*, vol. 3, no. 2, pp. 176–189, 2019.
- [26] M. S. T. El-Attar, "Application of artificial intelligence in architectural design," *Doctoral dissertation*, Al-Azhar University, Cairo, Egypt, 1997.
- [27] A. S. M. Mahmoodi, "The design process in architecture: A pedagogic approach using interactive thinking," *Doctoral dissertation*, University of Leeds, UK, 2001.
- [28] C. Eastman, "Automated assessment of early concept designs," *Architectural Design*, vol. 79, no. 1, pp. 52–57, 2009.
- [29] N. O. Sönmez, "A review of the use of examples for automating architectural design tasks," *Computer-Aided Design*, vol. 96, pp. 13–30, 2018.
- [30] H. Bhatt et al., "Artificial intelligence for predictive and evidence based architecture design," in *Proc. AAAI-16*, 2016.
- [31] S. Alotaibi, "Advancing energy performance efficiency in residential buildings for sustainable design: Integrating ML approaches," *Wiley*, 2024.
- [32] E. Elbeltagi et al., "Visualized strategy for predicting buildings energy consumption during early design stage using parametric analysis," *J. Building Engineering*, vol. 13, pp. 127–136, 2017.
- [33] J. Runge and R. Zmeureanu, "A review of deep learning techniques for forecasting energy use in buildings," *Energies*, vol. 14, no. 3, p. 608, 2021.
- [34] R. Panchalingam and K. C. Chan, "A state-of-the-art review on artificial intelligence for smart buildings," *Intelligent Buildings International*, 2021.
- [35] A. Krishnan et al., "ArchGym: An open-source gymnasium for machine learning assisted architecture design," 2023.
- [36] H. Zhu, Y. Zhou, Z. Shao, W. Du, G. Wang, and Q. Li, "Personalized image aesthetics assessment via multi-attribute interactive reasoning," *Mathematics*, vol. 10, Article 4181, 2022.
- [37] Y. Yang et al., "Personalized image aesthetics assessment with rich attributes," in *Proc. IEEE/CVF CVPR*, 2022.
- [38] Z. Zhang and J. Ban, "Aesthetic evaluation of interior design based on visual features," *Int. J. Mobile Computing and Multimedia Communications*, vol. 13, no. 2, 2022.
- [39] E. Hartanto, A. Chen, and I. Koh, "Empirical insights into architectural aesthetics: A neuroscientific perspective," in *CAADRIA 2024 Proceedings*, 2024.
- [40] S. Abdelmohsen, F. Farrag, M. Kassas, and A. Ibrahim, "AiMotional ecosystems: An affective computing pedagogical framework for generating emotionally responsive environments," 2024.
- [41] J. Ma, E. Erdogmus, and E. Yang, "A user-centered building design approach using immersive virtual reality and discrete choice modeling," *Building and Environment*, vol. 284, 113400, 2025.
- [42] H. Tu, "Analyzing affective responses to virtual spaces using physiological sensors and verbal descriptions," *M.S. thesis*, MIT, 2023.
- [43] J. Cho et al., "Predicting architectural space preferences using EEG-based emotion analysis: A CNN-LSTM approach," 2025.
- [44] Z. Zhang, J. M. Fort, and L. Giménez Mateu, "Decoding emotional responses to AI-generated architectural imagery," *Frontiers in Psychology*, vol. 15, Article 1348083, 2024.