Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Intelligent Product Conceptual Design: A Comparative Review of Paradigms, Challenges, and Future Trends

ALI Mohammed Adam Jamea

ali@qzuie.edu.cn

Quanzhou University of Information Engineering https://orcid.org/0000-0003-3707-5877

Systematic Review

Keywords: Intelligent product conceptual design, Human-AI collaboration, Computational design synthesis, Systematic literature review, Comparative review

Posted Date: August 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10690714/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Intelligent product conceptual design (IPCD) is evolving from predominantly symbolic reasoning systems toward data-driven, generative, and hybrid AI architectures, yet the field lacks a coherent understanding of how these paradigms redistribute design knowledge, reasoning, automation, and human agency. This review examines 247 studies published between 1990 and 2026 using a PRISMA-informed selection protocol and a five-dimensional comparative framework covering knowledge representation, reasoning mechanism, automation level, scalability, and human–AI interaction. The findings reveal not a simple replacement of earlier paradigms by machine learning, but persistent structural trade-offs. Symbolic methods provide explicit and interpretable reasoning but scale poorly; evolutionary approaches expand design-space exploration but remain dependent on representation and evaluation; and ML/DL methods offer strong scalability and generative capacity but lack reliable explainability and physical-feasibility guarantees. Hybrid architectures are increasingly prominent, although current evidence remains insufficient to establish their general superiority over well-designed single-paradigm baselines. Only 5.7% of the reviewed studies include human-designer baseline comparisons, while reproducibility and CAD/PLM integration remain limited. The review argues that IPCD is moving toward distributed design intelligence in which human judgment, symbolic knowledge, generative search, and physically grounded reasoning perform complementary roles. A research agenda is proposed for open FBS-oriented datasets, feasibility-aware generative methods, standardized human–AI benchmarks, reproducible experimentation, and industrial workflow integration.

1. Introduction

Conceptual design constitutes the foundational phase of product development, during which fundamental decisions regarding functionality, working principles, and system architecture are established in the absence of detailed geometric or material specifications (Beitz et al., 1996; Ulrich & KarlT, 2008). Despite its relatively brief duration compared to detailed design and manufacturing preparation, conceptual design commits approximately 70–80% of the final product cost and largely determines long-term performance and market success (Cross, 2021; Ullman, 2010).

Traditionally, conceptual design is a human-centered, creative activity that relies heavily on designer experience, analogical reasoning, and iterative sketching. However, increasing product complexity, shorter development cycles, and the growing need to explore larger design spaces have motivated the development of computational methods to support or automate conceptual design activities. This emerging field is known as intelligent product conceptual design (IPCD) (Chakrabarti, 2002; Goel et al., 2012). In this context, the term “intelligent” refers to the use of computational intelligence techniques including knowledge-based systems, case-based reasoning, evolutionary algorithms, machine learning, and more recently deep learning and large language models to perform or assist design tasks such as function generation, concept synthesis, evaluation, and selection (Gero & Kannengiesser, 2004; S. Kumar et al., 2023).

Despite substantial progress, the literature remains fragmented. Prior reviews are often either paradigm-specific or application-specific; to the best of our knowledge, no prior review has jointly combined a PRISMA-informed protocol with a cross-paradigm comparison of the four major IPCD paradigms within a unified analytical framework. Furthermore, rapid advancements in deep learning and generative AI since 2018 have rendered earlier reviews incomplete. Given the breadth of prior research, the foundational review that most directly motivates the present work examined collaborative conceptual design at the dawn of the Web era, identifying agent technology, Web-based frameworks, and conflict resolution as key enablers; that review analyzed approximately 80 papers and 20 projects, concluding with eight research opportunities including Web-based architectures, product-centric design methodology, and knowledge management (L. Wang et al., 2002). The present review extends this foundation by: (1) incorporating two decades of subsequent advances in ML/DL; (2) systematically comparing four paradigms using a five-dimensional framework; and (3) addressing challenges such as data scarcity and explainability that have emerged since 2002.

Several reviews have addressed intelligent design in general, but a focused, comparative review of conceptual design specifically and one that systematically compares multiple paradigms along consistent dimensions has not previously been conducted (Figoli et al., 2025; Reinpold et al., 2025; Sreenivasan & Suresh, 2024). A further underreported gap is the limited use of direct human-baseline comparisons. Fewer than 6% of publications in the reviewed corpus evaluate IPCD systems against human designers, and no standardized benchmarking protocol currently exists. This absence limits the generalizability of reported performance gains and weakens the practical interpretation of cross-paradigm comparisons. Accordingly, this review documents the gap and proposes a minimum benchmarking protocol in Section 5.6.

This work addresses these gaps by presenting a comparative review of the major computational paradigms, persistent challenges, and emerging research directions in intelligent product conceptual design through the following research questions:

RQ1: What computational paradigms dominate IPCD, and how do they differ across key design dimensions?

RQ2: What is the current state-of-the-art (2018–2026), and what are the persistent challenges?

RQ3: What are the emerging trends and future directions for the field?

Drawing on a systematically curated corpus of 247 publications (primarily peer-reviewed journal articles and conference papers, supplemented by selected recent early-access studies for qualitative discussion) selected through a PRISMA-informed protocol, we develop a five-dimensional analytical framework organized into three layers: epistemic (knowledge representation, reasoning mechanism), computational (automation level, scalability), and socio-technical (human–designer interaction). This layered structure avoids single-metric bias and enables explicit comparison of how agency is distributed between human designers and computational systems.

The paper makes three contributions. First, it advances the argument that IPCD evolution is not a linear progression toward full automation, but a redistribution of design reasoning across human, symbolic, data-driven, and physically grounded computational agents. Second, it identifies five structural tensions that constitute the intellectual core of cross-paradigm analysis. Third, it proposes a research agenda centered on distributed design intelligence,

prioritizing open function–behavior–structure datasets, explainable generative methods, standardized human–AI benchmarking, and CAD/PLM integration, and domain-specific foundation models as summarized in Table 1.

Table 1
Positioning of the present review against prior related reviews

Review	Publication year	Scope	Paradigms covered	Conceptual design focus	Systematic protocol	Comparative framework
Collaborative conceptual design—state of the art and future trends (L. Wang et al., 2002).	2002	Web-based collaborative design	KBS, agents	Yes (partial)	Narrative only	No unified framework
Design thinking and artificial intelligence: A systematic literature review exploring synergies (Sreenivasan & Suresh, 2024).	2024	Design thinking + AI	ML, LLM	No	No	No
AI-against-design map: A systematic review of 20 years of AI-focused studies in design research (Figoli et al., 2025).	2025	20 years AI in design research	Multiple AI methods	Partial	Bibliometric	No 5-dim framework
Present review (2026)	2026	IPCD across 4 paradigms, 1990–2026	KBS, CBR, Evolutionary, ML/DL, Hybrid	Yes, primary focus	PRISMA-informed, 247 papers, k = 0.87	5-dimensional analytical framework

In general, this comparative review contributes to the design science agenda by organizing fragmented evidence into a comparative framework that supports future theory-building, benchmark development, and practice-oriented research in intelligent product conceptual design. Specifically, the comparative analysis informs design science by identifying which paradigms best support early-stage design theory-building through explainable reasoning; characterizing human–AI co-creation dynamics that are central to the field's emerging socio-technical research agenda; and proposing standardized benchmarking criteria that are prerequisites for cumulative, evidence-based progress in computational design support.

2. Review scope and methodology

This section defines the scope of IPCD and explains the methodological procedure. Detailed search strings, borderline classification rules, sensitivity analyses, and full coding sheets are provided in Appendices A–D.

2.1 Definition of intelligent product conceptual design

IPCD refers to the application of computational intelligence techniques that support or automate the generation, representation, evaluation, and selection of product concepts before detailed embodiment. These activities are driven by functional and behavioral specifications, not fully defined geometries.

This definition has two main emphases. First, IPCD focuses on early-stage abstraction, in which product purpose, functional principles, and conceptual structures are established while detailed geometry, materials, and manufacturing processes remain intentionally unspecified. Second, it explicitly excludes activities that belong to later design stages or that lack reasoning. Concretely, parametric optimization of existing geometries, manufacturing process planning, and pure geometric modelling that lacks design reasoning have been excluded, Fig. 1 provides a graphical summary of this scope.

2.2 Search strategy and selection

A systematic literature search was conducted across three databases: Web of Science, Scopus, and Google Scholar. The search period spanned January 1990 to March 2026, with particular attention to 2018–2026 for the state of the art. Quantitative trend analyses (Section 5.1) extend through December 2025; papers retrieved in January–March 2026 were included for qualitative coverage but excluded from longitudinal trend calculations, as this three-month window was insufficient for stable bibliometric analysis. The three-month window for 2026 papers may over-represent early-year submissions and preprints; accordingly, these studies were excluded from growth-rate estimation, and readers should treat qualitative findings from 2026 as preliminary. Papers from 2025–2026 that were preprints or early-access were included qualitatively but may not represent final peer-reviewed versions; readers should interpret recent trends cautiously.

The 18 papers from January–March 2026 are included in the 247-paper corpus for descriptive and qualitative purposes only; they were excluded from paradigm-trend percentages, regression analyses, and comparative performance claims. All percentage-based trend figures (e.g., the 62% ML/DL raw corpus share for 2025; rescaled primary-paradigm share: 75.6%) are based on the 2018–2025 corpus and should be interpreted as provisional. Raw percentages describe corpus share; rescaled percentages are used only for within-row comparison across mutually exclusive primary categories (see Appendix C, Table C4).

This stopping rule may underrepresent low-citation papers by approximately 8–12%, as acknowledged in Section 8.4; forward and backward citation tracking of key papers was applied to partially mitigate this bias. Exact search strings for each database are provided in Appendix A.

The initial search yielded 1847 records. After duplicate removal (n = 412 removed), 1435 records underwent title and abstract screening; 1123 were excluded (reasons: 432 not conceptual design, 351 no computational intelligence, 340 non-English or theoretical only), resulting in 312 full-text articles assessed for

eligibility. Following full-text review, 65 papers were excluded (reasons: 28 no empirical evaluation, 22 detailed design focus only, 15 duplicate findings), leaving 247 papers ultimately included.

Two independent reviewers classified each paper into paradigm categories (knowledge-based systems, case-based reasoning, evolutionary and generative design, machine learning/deep learning, or hybrid). Inter-rater reliability was assessed using Cohen's kappa (overall $\kappa = 0.87$, 95% CI: 0.83–0.91), which indicates substantial to almost perfect agreement. Agreement was highest for knowledge-based systems ($\kappa = 0.92$) and lowest for hybrid papers ($\kappa = 0.78$, still substantial), reflecting the inherently greater ambiguity of multi-paradigm classification. Given the inherent ambiguity in hybrid classification, $\kappa = 0.78$ is acceptable for this field; published benchmarks suggest that values above 0.90 are rare in design research literature (Landis & Koch, 1977; McHugh, 2012). Approximately 8% of papers required third-author adjudication; all disagreements were resolved by consensus. The PRISMA flowchart and full paper list are provided in Appendices A and B.

The complete review corpus comprises 247 studies. Of these, 229 studies published through December 2025 constitute the quantitative temporal-analysis corpus, whereas 18 studies retrieved from January to March 2026 are used only for descriptive and qualitative analysis. Accordingly, all longitudinal trend estimates and annual paradigm-share calculations exclude the 18 studies from 2026.

2.3 Data extraction and analysis

For each included paper, the following attributes were systematically extracted: (1) publication year, (2) computational paradigm, (3) application domain, (4) knowledge representation type, (5) evaluation method, and (6) level of automation.

Quantitative analysis was conducted using descriptive statistics and temporal trend analysis. Paradigm distribution over time was analyzed using log-linear regression to identify growth patterns. Inter-paradigm comparisons were performed using normalized frequency distributions across the five analytical dimensions. Notably, fewer than 12% of included studies provided executable code or publicly accessible full datasets, a reproducibility limitation that is discussed further in Section 8.5.

2.4 Quality assessment and coding reliability

To ensure methodological transparency and rigor, a formal quality appraisal was conducted on all 247 included papers using a 3-item scale. Each paper was scored on: (1) clarity of research aims and scope (0–2 points); (2) empirical validation or evaluation method (0–2 points); and (3) reproducibility of reported results (0–2 points), yielding a maximum quality score of 6 per paper. Scoring was performed by two independent reviewers, with discrepancies resolved by the third author.

Mean quality score across the corpus was 4.1/6 (SD 1.3). ML/DL papers scored higher (4.7) than KBS (3.4). Hybrid papers (4.9) showed the highest quality but also the largest variance, reflecting the wide heterogeneity of hybrid system designs. Distribution by quality tier: high (score 5–6): 38% of papers; medium (score 3–4): 44%; low (score 0–2): 18%. These quality distributions inform the interpretation of paradigm comparisons throughout this review. Findings from low-quality studies (score ≤ 2) are reported descriptively and are not used as the basis for quantitative performance claims. Unweighted aggregation is acknowledged as a limitation for studies where quality scores were not available at the time of initial coding.

Coding reliability was assessed as described in Section 2.2 (Cohen's $\kappa = 0.87$). The full coding sheet, including quality scores for all 247 papers, is available as Appendix B (supplementary material). Each paper was assigned to one dominant primary paradigm; a separate binary hybrid flag identifies papers integrating two or more paradigms architecturally. Primary-paradigm percentages therefore sum to 100%, while hybrid prevalence is reported separately (Appendix C, Table C4).

2.5 Inclusion and exclusion criteria

2.5.1 Inclusion criteria: (1) Peer-reviewed journal articles or conference proceedings; (2) explicit focus on conceptual design (not detailed or embodiment design); (3) use of computational intelligence methods; (4) English language.

2.5.2 Exclusion criteria: (1) Purely theoretical or philosophical papers without implementation or empirical evaluation; methodological foundation studies were retained only when their frameworks were directly operationalized by later IPCD studies and were not used as empirical performance evidence; (2) papers focused exclusively on detailed design or manufacturing; (3) papers where artificial intelligence is used only for literature review or citation analysis.

2.6 Comparative framework

To enable structured cross-paradigm comparison, this review introduces a five-dimensional analytical framework, Table 2. The dimensions were synthesized from automation level and problem-solving scalability with human-designer interaction characteristics (Cagan et al., 2005; Singh, n.d.). The selected dimensions are grounded in established design research literature and integrate both technical characteristics, such as knowledge representation and reasoning mechanisms with socio-technical characteristics, such as automation level, scalability, and human–designer interaction (Chandrasekaran, 1990). As a result, this framework was developed specifically to avoid single-metric bias, which has undermined prior comparative reviews that relied exclusively on performance metrics without accounting for explainability, scalability, or human-designer interaction requirements.

Table 2
Five-dimensional comparative framework

Dimension	Description	Example Metrics/Indicators
Knowledge representation	The manner in which design knowledge and information are structured and encoded within the system	Rules, cases, parametric models, embeddings, semantic graphs
Reasoning mechanism	The computational process through which new design solutions are generated, inferred, or adapted	Rule chaining, retrieval and adaptation, mutation and selection, neural inference
Automation level	The extent to which the design process can operate independently of human intervention	Fully automatic, human-in-the-loop, interactive, human-driven
Scalability	The capability of the approach to manage increasing design complexity, data volume, or problem dimensionality	Polynomial growth, exponential growth, computational tractability
Human–designer interaction	The degree of transparency, explainability, controllability, and collaborative support provided to designers	Explainability level, controllability, interpretability, creative synergy, fixation risk

Throughout this review, scalability refers to the practical capacity of a method to accommodate increasing problem dimensionality, design-space size, dataset volume, model complexity, or case-base size without prohibitive growth in computational cost, memory requirements, or knowledge-engineering effort. Assessments of low, medium, and high scalability are therefore comparative and application-dependent rather than strict asymptotic complexity classes.

3. Historical overview and conceptual framework

This section situates IPCD within its historical development, showing how the field has progressed from rule-based knowledge systems to case-based, evolutionary, and data-driven approaches. It also introduces the function–behavior–structure perspective as a conceptual basis for comparing how different paradigms represent, reason about, and generate product concepts.

3.1 Three waves of intelligent conceptual design

The evolution of intelligent product conceptual design (IPCD) can be understood as three broads, overlapping developmental phases, each characterized by the ascendancy of a dominant computational paradigm rather than by abrupt displacement of its predecessors, Table 3. Critically, these “waves” denote periods of paradigm dominance, not first appearance. Although machine learning methods appeared sporadically from the early 1990s—with neural network approaches to engineering design explored as early as the late 1980s—they did not achieve paradigm dominance until after 2015–2018. Deep learning gained meaningful traction in design research by 2015–2016, following the landmark ImageNet results of 2012. Accordingly, the “third wave” label denotes a period of paradigm dominance rather than the moment of first methodological introduction. A potential fourth wave, centered on generative AI and foundation models (2022–present), is discussed in subsequent sections.

The field has progressively shifted from rule-based systems (i.e., expert knowledge encoded explicitly) to example-based approaches (i.e., reusing past designs) and, most recently, to data-driven methods (i.e., learning patterns from large datasets). Crucially, knowledge-based systems (KBS) and case-based reasoning (CBR) tools continue to be deployed in both industrial practice and hybrid system architectures; they should not be regarded as superseded, but rather as complementary components within increasingly integrated design intelligence frameworks.

Table 3
Three historical waves of IPCD

Wave	Period	Dominant Paradigm	Representative Work	Key limitation
First	1980s–1990s	Knowledge-based systems (KBS)	(Brown & Chandrasekaran, 2014; Tong & Sriram, 1992)	Knowledge acquisition bottleneck
Second	1990s–2010s	Case-based reasoning (CBR) + Evolutionary	(Bentley, 1999; Maher & Pu, 2014)	Adaptation knowledge; computation cost
Third (dominance from ~ 2018)	2018–present	Machine learning / Deep learning	(Camburn et al., 2020; Xiao et al., 2023)	High data requirements; limited interpretability

3.2 A Unified framework for comparison

The function–behavior–structure (FBS) framework provides a useful ontological perspective for comparing IPCD paradigms (Gero & Kannengiesser, 2014). Function describes what an artefact is intended to achieve, behavior describes how its performance is expected or derived, and structure describes its constituent elements and relationships. The four paradigms differ not because each is restricted to a single FBS transformation, but because they operationalize these relationships differently. KBS provides FBS knowledge explicitly through rules and ontologies; CBR retrieves and adapts previously instantiated function–behavior–structure configurations; evolutionary methods search structural alternatives according to behavioral and functional evaluation criteria; and ML/DL approaches learn statistical mappings among functional, behavioral, structural, geometric, and semantic representations.

This review therefore proposes that the historical evolution of IPCD should be interpreted not as successive paradigm replacement, but as the progressive redistribution of design intelligence across symbolic reasoning, experiential knowledge, optimization, statistical learning, and collaborative human–AI interaction.

Based on the historical evolution and comparative analyses presented throughout this review, we propose that the development of intelligent product conceptual design should be interpreted not as a sequence of paradigm replacement but as a progressive redistribution of design intelligence across symbolic reasoning, experiential knowledge, optimization, statistical learning, and collaborative human–AI interaction. Each paradigm contributes distinct computational capabilities, and advances in IPCD increasingly arise from their complementary integration rather than from the dominance of any single methodology.

4. Comparative analysis of four paradigms

This section examines four key paradigms in intelligent product conceptual design, knowledge-based systems, case-based reasoning, evolutionary/generative design, and machine learning/deep learning—using a five-dimensional framework (knowledge representation, reasoning, automation, scalability, human–designer interaction). It compares strengths, limitations, and trade-offs, showing the evolution from symbolic and experiential methods toward data-driven and hybrid approaches, and highlights opportunities for integrating paradigms to enhance adaptability, creativity, and practical design performance.

4.1 Knowledge-based systems (KBS)

4.1.1 Description

Knowledge-based systems are artificial intelligence systems that leverage encoded domain expertise typically represented through symbolic reasoning, rule-based inference, or structured ontologies to solve complex problems that traditionally require human specialist knowledge. These systems emerged as a cornerstone of early AI research. They are distinguished by their explicit separation of knowledge repositories from inference mechanisms, which enables modular updates to domain understanding without restructuring the underlying reasoning architecture (Hayes-Roth & Jacobstein, 1994).

4.1.2 Comparative assessment

Table 4 summarizes the main characteristics of KBS. Their principal advantage is explicit and inspectable reasoning, whereas their main limitations arise from knowledge-acquisition costs, rule-base growth, and brittleness outside the encoded domain.

Table 4
KBS comparative assessment

Dimension	Assessment
Knowledge representation	Explicit, symbolic, human-readable rules
Reasoning mechanism	Logical inference (forward/backward chaining)
Automation level	Medium to high (can be fully automatic if rules are complete)
Scalability	Poor (rule explosion; knowledge acquisition bottleneck)
Human interaction	High transparency; low creativity support

4.1.3 Strengths

Explainable and verifiable; captures deep domain knowledge; benefits from a mature methodology with established tools.

4.1.4 Weaknesses

Suffers from a knowledge acquisition bottleneck; brittle outside the scope of encoded rules; handles novel or unexpected requirements poorly.

4.2 Case-based reasoning (CBR)

4.2.1 Description:

Case-based reasoning solves novel problems by retrieving and adapting prior experiential episodes, or “cases”, stored in a structured memory. Rather than relying on abstract rules, CBR emulates human analogical thinking. When encountering a new situation, the system identifies similar past cases, reuses their solutions with appropriate modifications, evaluates the outcomes, and retains the new experience for future retrieval. CBR excels particularly in domains rich with experiential knowledge that resists complete formalization such as medical diagnosis, legal reasoning, and design configuration where the contextual nuances of specific precedents often prove more actionable than generalized heuristics (Kolodner, 1992).

4.2.2 Comparative assessment

Table 5 summarizes the CBR approach across five dimensions, highlighting practical trade-offs. Case-based retrieval is largely automatic, but adaptation typically requires expert input, and scalability is constrained more by the adaptation process than by retrieval. This assessment supports a balanced interpretation of system performance across all five dimensions.

Table 5
CBR comparative assessment

Dimension	Assessment
Knowledge representation	Cases (problem–solution pairs) + similarity metrics
Reasoning mechanism	Retrieve – Reuse – Revise - Retain cycle
Automation level	Medium (retrieval automatic; adaptation often manual)
Scalability	Medium (retrieval scales but adaptation is challenging)
Human interaction	Moderate transparency; adaptation requires expertise

4.2.3 Strengths

Learns incrementally from experience; handles ill-structured problems effectively; supports continuous learning through addition of new cases.

4.2.4 Weaknesses

Adaptation knowledge is often missing; case retrieval can become computationally expensive for large, high-dimensional case bases; similarity metric design is problem-specific and often heuristic. Regarding constraint handling, constraints are incorporated during case adaptation; however, similarity measures may not fully capture complex engineering constraints, and adaptation remains largely manual for constraint-rich problems.

4.3 Evolutionary design and search-based generative methods

In this review, evolutionary design refers to search-based generative approaches based on population evolution, variation, selection, and objective-driven evaluation. Neural generative models, including GANs, VAEs, diffusion models, and LLM-based systems, are classified under ML/DL unless they form part of an explicitly hybrid architecture.

4.3.1 Description

Evolutionary and generative design have become important paradigms in intelligent product conceptual design, particularly for complex, multi-objective engineering problems. These approaches use computational algorithms to automatically generate, evaluate, and optimize large numbers of design alternatives based on predefined objectives, constraints, and performance criteria. Evolutionary design draws inspiration from biological principles such as selection, mutation, and crossover. Generative design, in turn, integrates artificial intelligence, topology optimization, and parametric modeling to explore innovative solutions that may lie beyond conventional human intuition. It bears emphasis, however, that evolutionary and generative methods are not inherently creative in the human sense: they explore combinatorial spaces defined by their genotype representations and objective functions, and can only discover solutions that are expressible within the encoding chosen by the designer. Compared with traditional methods, these approaches can expand the explored design space and may support creative exploration, although their outputs remain dependent on the quality of objective functions, constraints, and human evaluation. They offer improved design efficiency and solution diversity, making them well suited for smart product development, lightweight structures, and adaptive engineering systems (Goldberg, 1989; Holland, 1992).

4.3.2 Comparative assessment

Table 6 describes the evolutionary design approach across five dimensions, highlighting practical trade-offs. Variation operators and selection drive adaptation, while genotype–phenotype mappings enable structural search. Automation is high, although fitness evaluation may still require human guidance.

Table 6
Evolutionary design comparative assessment

Dimension	Assessment
Knowledge representation	Parametric or grammatical (genotype–phenotype mapping)
Reasoning mechanism	Variation (mutation, crossover) + selection
Automation level	High–Very high (can be fully automatic)
Scalability	Medium (population size grows with complexity)
Human interaction	Low transparency (limited interpretability); interactive evolution possible

4.3.3 Strengths

Explores large, discontinuous, and non-convex design spaces; naturally handles multiple conflicting objectives (multi-objective evolutionary algorithms, MOEAs — see Nomenclature); does not require gradient information.

4.3.4 Weaknesses

Computationally expensive (requires many function evaluations); no guarantee of global optimality; handles mixed discrete–continuous variables poorly; results are difficult to interpret analytically.

4.4 Machine learning and deep learning approaches

4.4.1 Description

Machine learning (ML) and deep learning (DL) approaches have become increasingly important in intelligent product conceptual design because of their ability to learn patterns, predict design outcomes, and support automated decision-making using large-scale datasets (J. Wang et al., 2018; Xu et al., 2022). ML techniques enable intelligent systems to analyze user requirements, optimize design parameters, and improve overall design efficiency, while DL methods provide advanced capabilities in feature extraction, image recognition, semantic understanding, and complex pattern analysis (Guo, 2017). These approaches play a critical role in enabling intelligent, adaptive, and data-driven product design and manufacturing processes (Tao et al., 2018).

4.4.2 Comparative assessment

Table 7 compares the ML/DL approach across five dimensions, highlighting practical trade-offs. Neural inference and learned embeddings drive reasoning with very high automation and scalability. However, low transparency and black-box behavior limit human interaction.

Table 7
ML/DL comparative assessment

Dimension	Assessment
Knowledge representation	Learned embeddings (latent vectors, neural weights, feature maps)
Reasoning mechanism	Neural inference (forward pass-through learned weights; autoregressive generation).
Automation level	High–very high (conditional on data quality; end-to-end automation typically requires domain-specific preprocessing, curated training data, and fine-tuning).
Scalability	High (with sufficient data and compute; sub-linear inference time)
Human interaction	Very low transparency (black-box); limited controllability without specialized methods.

4.4.3 Strengths

ML/DL approaches are highly effective in handling high-dimensional and unstructured data, including images, point clouds, and textual information. They can automatically learn complex patterns without requiring explicit rule definitions and are capable of generating innovative and unexpected design solutions. Once trained, these models also provide rapid inference and efficient large-scale processing.

4.4.4 Weaknesses

ML/DL approaches remain weak in constraint handling because they do not inherently guarantee physical feasibility. For safety-critical or high-cost design domains, future systems should embed feasibility constraints directly into the model architecture, for example through physics-informed networks, neuro-symbolic layers, or constraint-aware decoding. Table 8 summarizes the technical characteristics of each paradigm across the five analytical dimensions.

Table 8
Cross-paradigm summary: Technical characteristics

Paradigm	Knowledge representation	Reasoning mechanism	Automation level	Scalability	Explainability	Data requirements	Typical output	Constraint handling
KBS	Explicit symbolic rules, ontologies, and expert knowledge	Logical inference using forward and backward chaining	Medium–High	Low	High	Low to moderate, primarily expert-derived knowledge	Rule-based discrete solutions and recommendations	Strong explicit constraint representation through encoded rules; however, performance becomes brittle when rules are incomplete or conflicting
CBR	Historical cases and problem–solution pairs with similarity metrics	Retrieval, adaptation, revision, and retention cycle	Medium	Medium	Moderate	Moderate, requiring representative case libraries	Adapted solutions based on prior experiences	Constraints are incorporated during case adaptation, although similarity measures may not fully capture complex engineering constraints
Evolutionary and Generative Design	Parametric models, chromosomes, or genotype–phenotype mappings	Evolutionary variation, crossover, mutation, and selection	High	Medium	Low	Relatively low, mainly objective and fitness function definitions	Multiple optimized design alternatives and solution populations	Effective handling of multi-objective and constrained optimization through penalty functions and constraint-based fitness evaluation
ML/DL	Learned embeddings, latent representations, neural weights, and feature maps	Neural inference, pattern learning, and autoregressive generation	Very high	High	Very Low	Very high, requiring large-scale training datasets	Predictive models, generated concepts, and continuous design mappings	Constraint handling remains challenging and often relies on specialized architectures or post-processing feasibility checks
Hybrid approaches	Combined symbolic, experiential, and data-driven representations	Integrated reasoning mechanisms combining multiple paradigms	Medium–to very high	Medium–High	Moderate	Varies depending on the integrated paradigms	Context-dependent hybrid outputs combining prediction, adaptation, and generation	Improved flexibility in constraint handling by leveraging complementary strengths of multiple paradigms

Figure 2. Evolution of design intelligence across IPCD paradigms, illustrating the progressive redistribution of design intelligence from symbolic reasoning toward collaborative human–AI systems rather than a simple replacement of earlier paradigms.

5. Contemporary State-of-the-Art (2018–2026)

This section reviews contemporary IPCD developments from 2018 to 2026, with particular attention to the growth of ML/DL, generative AI, and hybrid systems. It examines publication trends, application domains, representative state-of-the-art systems, geographic and institutional patterns, human-comparison studies, and adjacent frameworks, while identifying persistent challenges in benchmarking, explainability, feasibility assessment, and industrial integration.

5.1 Publication and paradigm trends

A bibliometric analysis based on the Scopus dataset and the PRISMA-filtered corpus of 247 included studies demonstrates sustained and statistically significant growth in IPCD research since 2018. Using the search strategy defined in Section 2.2, annual publication counts were analyzed through log-linear Poisson regression. The analysis yielded an incidence rate ratio (IRR) of 1.18 per year (95% CI: 1.14–1.22, $p < 0.001$), indicating an approximate compound annual growth rate of 18%. The Poisson model demonstrated acceptable fit (Pearson $\chi^2/df = 1.12$), confirming no significant overdispersion. Statistical analyses were performed in R v4.3.1 using glm(family = Poisson).

Sensitivity analyses (full results in Appendix C) excluding pre-2000 papers, 2025 publications, conference proceedings, and alternative hybrid classifications consistently yielded IRR values between 1.15 and 1.19 (all $p < 0.001$), confirming robustness of the primary estimate.

The publication landscape also reveals a substantial paradigm transition. ML/DL-based approaches rose from approximately 15% of the corpus in 2018 to dominate by 2025 (see Appendix C, Table C4 for full detail), while rule-based and CBR approaches declined from 45% to 18%. Papers carrying a secondary hybrid flag expanded from 8% to 34% in the same period.

Within the PRISMA-filtered corpus, annual output grew from fewer than 10 papers before 2010 to 25–40 papers by 2025, reflecting both AI advances and growing interdisciplinary interest in intelligent conceptual design.

5.2 Dominant application domains

The reviewed literature indicates that IPCD applications between 2018 and 2025 were concentrated primarily in mechanical engineering and product architecture design, although increasing diversification toward multimodal and semantic design tasks is evident, Table 9.

Table 9
Dominant application domains in IPCD (2018–2025)

Application domain	Share (%)	Common methods	Example output
Mechanical component concept generation	35%	GANs, GNNs, evolutionary algorithms, diffusion models	Brackets, linkages, gearbox concepts
Product architecture and layout design	22%	CBR, graph-based ML, reinforcement learning	Circuit-board layouts, building floorplans
Sketch-based conceptual design	18%	CNNs, VAEs, Sketch-RNN	3D concepts generated from 2D sketches
Function–structure mapping	15%	LLMs, knowledge graphs, neuro-symbolic systems	Function trees, component mappings
Concept evaluation and selection	10%	AHP, TOPSIS, ML-based ranking systems	Ranked and scored concept sets

Mechanical component generation remains the dominant application area, largely because geometric representations and CAD-compatible datasets are more readily available than functional or behavioral design data. Recent work increasingly employs diffusion models, graph neural networks (GNNs), and multimodal generative systems capable of synthesizing structurally feasible design alternatives.

Another important trend is the emergence of semantic and language-driven conceptual design systems. In particular, LLM-based systems have enabled natural-language interaction with design tools, supporting early-stage ideation, requirement interpretation, and concept refinement. This shift reflects a broader movement from geometry-centric design support toward cognitively informed design assistance.

Research attention toward electronics, service systems, biomedical products, and socio-technical domains remains limited, reflecting a persistent mechanical-engineering bias in the corpus.

5.3 Representative state-of-the-art systems

The reviewed literature indicates that contemporary IPCD systems can be organized into four major state-of-the-art clusters: LLM-assisted design systems, graph-based design intelligence, generative aesthetic design systems, and surrogate-assisted evolutionary or inverse-design frameworks, as summarized in Table 10. These clusters reflect the field's movement from isolated algorithmic optimization toward interactive, data-driven, and context-aware design intelligence. On this basis, representative examples from each cluster are summarized below:

1. **LLM-based design assistants.** Recent systems employ multi-agent LLM architectures in which specialized agents take on distinct roles such as ideation, evaluation, critique, and refinement. These agents collaborate to support natural-language-driven conceptual design workflows and enable human–AI co-creation in early-stage design (V. Kumar & Karniadakis, 2026; S. Liu et al., 2026).
2. **Graph neural network (GNN)-based design reasoning systems.** GNNs are increasingly used to learn spatial and functional dependencies from existing design repositories. For example, they have been applied to support decision-making in shared mobility systems and to generate architectural layouts. Such approaches are particularly effective for topology-sensitive design problems (Liang et al., 2025; Peng et al., 2026).
3. **Generative adversarial and diffusion-based aesthetic design systems**: Researchers have integrated generative AI models with consumer preference learning to produce visually optimized product forms, including automotive exteriors and footwear concepts. These systems demonstrate how data-driven generation can align with subjective aesthetic criteria (Bing et al., 2026; Gaikwad et al., 2025).
4. **Surrogate-assisted evolutionary design systems**: In this family of methods, expensive simulation-based fitness evaluations are replaced by surrogate models such as Gaussian processes or neural networks dramatically reducing computational cost while maintaining acceptable optimization accuracy. These hybrid frameworks are especially valuable for multi-objective and computationally intensive design tasks (M.-T. Chen et al., 2026; Z. Liu et al., 2018; Tanriover et al., 2022; Wu et al., 2024).

Collectively, these systems signal a shift from static rule-based automation toward adaptive, multimodal, and human-centered design intelligence — integrated environments combining language reasoning, graph-based representations, generative modeling, and human feedback.

Representative systems discussed in this section include both studies contained in the formal review corpus and selected recent publications used to contextualize emerging technological developments. Only studies satisfying the eligibility criteria described in Section 2 contributed to the quantitative

analyses.

Table 10
Analytical comparison of representative state-of-the-art IPCD system clusters

System cluster	Typical input	Core method	Typical output	Evaluation approach	Main limitation
LLM-based design assistants	Natural-language requirements, design briefs	Multi-agent LLM, RAG	Concept descriptions, design rationale	Expert evaluation, user studies	Hallucination; limited physical feasibility checking
GNN-based design reasoning	Graph-structured design repositories	Graph neural networks (GNN)	Topology-sensitive design layouts	Benchmark datasets, expert assessment	Requires large graph-structured training data
Generative / diffusion aesthetic design	Product images, consumer preference data	GAN / diffusion models	Visually optimized product forms	Preference surveys, Likert ratings	Subjective evaluation; limited manufacturability checking
Surrogate-assisted evolutionary design	Parametric design variables, simulation results	Evolutionary algorithm + surrogate (GP/NN)	Pareto-optimal design alternatives	Paired comparison vs. full simulation; statistical tests	Surrogate accuracy degrades in high-dimensional spaces

**Hallucination refers to the generation of factually incorrect, physically infeasible, or internally inconsistent outputs that are not grounded in training data or constraints.*

5.4 Hybrid and integrated approaches

A notable development in contemporary IPCD research is the emergence of hybrid systems that integrate complementary computational paradigms to overcome the limitations of individual methods, Table 11.

Table 11
Common hybrid IPCD archetypes

Hybrid Type	Example	Mechanism	Benefit
CBR + ML	(Hudson & Parmee, 1996)	ML learns similarity metrics from case libraries	Reduces manual similarity engineering
Evolutionary + Surrogate models	(Q. Wang et al., 2026)	Neural surrogates replace expensive fitness evaluations	Reduces optimization cost by 10–100×
LLM + Knowledge graph	(P. Chen et al., 2026)	Knowledge graphs validate LLM-generated concepts	Improves factual consistency and feasibility
Generative + Interactive system	(Burnap et al., 2023; Takagi, 2002)	Human feedback guides generative exploration	Aligns outputs with subjective preferences

Hybrid systems are increasingly attractive because they combine the interpretability of symbolic systems with the adaptability and scalability of data-driven learning. Detailed hybrid classification criteria and borderline case adjudication are provided in Appendix A.6 and Appendix B.5. Hybrid approaches are associated with reported performance gains in the literature, yet fewer than half of hybrid studies reported statistical comparisons against non-hybrid baselines, and none met pre-specified criteria for both statistical testing and independent replication. Therefore, hybrid systems appear promising because they combine complementary strengths across paradigms; however, current evidence remains insufficient to establish their general superiority over well-designed single-paradigm baselines. Table 12 summarizes the integration depth and reported performance improvement for representative hybrid architectures.

Table 12
Illustrative performance outcomes reported by selected hybrid IPCD studies

Hybrid type	Integration depth	reported performance improvement	Statistically tested?	Example reference
CBR + ML	Tight (ML replaces core similarity metric)	+ 23% retrieval precision vs. hand-tuned CBR	Partial (ablation only)	(Nkisi-Orji et al., 2022)
Evolutionary + Surrogate	Tight (surrogate replaces fitness function)	10–100x evaluation cost reduction	Yes (paired comparison)	(B. Liu et al., 2025)
LLM + Knowledge Graph	Loose (KG post-filters LLM output)	Hallucination rate reduced ~ 40%	Partial (manual inspection)	(Jin et al., 2024; X. Wang et al., 2026)
Generative + Interactive	Loose (human selects; model adapts)	Higher preference scores (Likert + 1.2)	No (subjective ratings only)	(Jacobs et al., 2025)

**Performance gains are reported from individual studies with varying experimental conditions, task domains, and baseline definitions. Cross-study comparisons are not statistically valid. See* Section 5.4 *for discussion of the absence of standardized benchmarks.*

5.5 Geographic and institutional distribution

The geographic distribution of IPCD research demonstrates substantial concentration within a relatively small number of technologically advanced regions. Between 2018 and 2025, institutions affiliated with China contributed approximately 38% of publications within the reviewed corpus, followed by the United

States (24%), Europe (21%), and other regions (17%). It should be noted that, because this review applied an English-language inclusion criterion (Section 2.5.1), Chinese-language and other non-English publications are likely systematically underrepresented; the true Chinese share may therefore be higher than these figures indicate. Within Europe, Germany, the United Kingdom, and France were the dominant contributors.

Leading institutions include MIT, Carnegie Mellon University, Beihang University, TU Delft, and the University of Cambridge. Industry-affiliated publications (Siemens, Autodesk, Dassault Systèmes) represent only approximately 12% of the reviewed corpus. Meanwhile, geographic contributions were classified according to the corresponding author's institutional affiliation to avoid multiple counting of internationally co-authored publications.

5.6 Comparison with human designers

An important trent of this review is the limited number of studies directly comparing IPCD systems with human designers performing equivalent conceptual design tasks. Of the 247 reviewed papers, only 14 studies (5.7%) included any form of human baseline evaluation. The available evidence suggests several consistent patterns:

1. ML/DL-based concept generation systems can achieve levels of concept diversity comparable to junior human designers but generally underperform experienced experts in feasibility assessment and contextual appropriateness.
2. LLM-based systems produce broader and more structurally diverse concept spaces than individual designers working independently; however, these systems also generate higher rates of physically infeasible or internally inconsistent solutions. This hallucination phenomenon is qualitatively more consequential in engineering design than in general NLP: an incorrect functional principle propagated through conceptual design can cascade into erroneous embodiment specifications, misdirected simulation campaigns, and ultimately costly physical prototype failures.
3. Interactive evolutionary systems often outperform fully automated optimization systems when guided by expert human preferences, particularly in multi-objective conceptual design tasks involving subjective trade-offs.

The lack of standardized human-comparison benchmarks significantly limits the interpretability of current “state-of-the-art” claims. This study suggests a minimal benchmark comprising three standardized design problems, such as a robotic gripper, a heat sink, and a handheld consumer enclosure which are evaluated on three performance metrics: novelty (i.e., pairwise distance in design space), feasibility (i.e., satisfaction of physical constraints), and user preference (i.e., paired comparison).

5.6.1 Critical synthesis of paradigm evolution

The transition from KBS and CBR toward ML/DL should not be interpreted as a simple technological replacement. Instead, it reflects a redistribution of computational intelligence across complementary mechanisms. Knowledge-based systems externalize expert reasoning through explicit symbolic representations, whereas machine-learning approaches internalize statistical regularities within model parameters. Evolutionary approaches contribute systematic search capabilities, while large language models increasingly facilitate semantic interaction between human designers and computational tools. Consequently, paradigm evolution represents increasing functional specialization rather than convergence toward a universally superior methodology. The next generation of IPCD systems are therefore expected to integrate complementary reasoning paradigms rather than replace them, enabling adaptive and context-dependent design intelligence across diverse engineering applications.

5.7 Relationship to adjacent frameworks

Two emerging frameworks increasingly intersect with IPCD research: Cognitive intelligence-enabled product design (CIPD) and digital twin–knowledge graph (DT-KG) integration. CIPD emphasizes semantic understanding, cognitive reasoning, and multimodal knowledge integration as central capabilities for intelligent design systems. DT–KG integration combines real-time operational data from digital twins with structured semantic representations provided by knowledge graphs. Rather than competing paradigms, these frameworks represent complementary infrastructural directions, and their future implications are discussed further in Section 7.6.

6 Persistent challenges and gaps

Despite rapid advances in computational methods, several fundamental barriers continue to limit the practical adoption, reproducibility, and industrial impact of IPCD systems. These challenges are not independent; rather, they interact to constrain progress across the entire design lifecycle. The following subsections examine these interconnected barriers from technical, organizational, and human-centered perspectives.

6.1 Data scarcity and quality

Engineering design data remains fundamentally constrained by issues of scarcity, heterogeneity, and proprietary ownership. Unlike domains such as computer vision, where large-scale labeled datasets (e.g., ImageNet) have catalyzed rapid progress, IPCD lacks standardized, richly annotated datasets capturing function–behavior–structure relationships. Public datasets (ABC, ShapeNet, Fusion 360 Gallery) currently focus primarily on geometry, not the function–behavior–structure triplets required for conceptual design.

To contextualize the severity: the largest public engineering design dataset — Fusion 360 Gallery — contains only 8,625 CAD models with minimal functional annotations, far below the scale that enabled the deep learning revolution in computer vision. Only 3 of 47 IPCD systems in a sub-corpus reporting complete system descriptions integrated with commercial CAD platforms. The field urgently needs a community-curated open benchmark containing annotated F–B–S triplets in open exchange formats (STEP, IGES) compatible with established engineering taxonomies.

6.2 Explainability and trust

Conceptual design decisions have high downstream costs. Engineers are reluctant to trust black-box ML recommendations without explanations. Current explainable AI (XAI) methods for design are immature: saliency maps are often uninformative for structural outputs, and natural language explanations from LLMs may be plausible but incorrect. Of 247 included papers, fewer than 38 (approximately 15%) include any form of user study with professional designers a critical gap given the human-centered nature of conceptual design practice. Subsequent research should follow established human-factors protocols, use industrial design tasks (not toy problems), and recruit experienced practitioners (≥ 5 years), reporting both process metrics (time-to-concept, design-space coverage) and outcome metrics (novelty, feasibility, preference).

6.3 Integration with existing workflows

Most IPCD systems are standalone prototypes. Integration with commercial CAD (SolidWorks, CATIA, Fusion 360, NX) and PLM systems remains limited. The lack of standardized APIs or interchange formats between IPCD tools and CAD platforms is a practical barrier that the research community has not adequately addressed. Within the sub-corpus of 47 IPCD papers reporting complete system descriptions (see Section 6.1), only 3 (6.4%) reported any form of integration with commercial CAD platforms.

6.4 Evaluation metrics

There is no consensus on how to evaluate an IPCD system. Metrics used in literature include: (1) success rate, (2) novelty, (3) diversity, (4) designer preference, (5) computational cost, and (6) task completion time. These metrics are often subjective, task-dependent, and mutually conflicting. Therefore, establishing a standardized evaluation suite is an important requirement for ensuring fair, reproducible, and comparable assessment of IPCD systems.

6.5 Creativity vs. automation

A fundamental tension exists between automation (efficiency) and creativity (novelty). Fully automatic systems risk producing only incremental variations; highly interactive systems are slow. This tension is not merely technological but also cognitive: designers may experience creative fixation when AI suggestions are too dominant (Youmans & Arciszewski, 2014). Both failure modes — fixation and complacency — must be actively managed through structured evaluation checkpoints and explicit designer-agency reinforcement mechanisms.

6.6 Semantic gap

The gap between natural language requirements (e.g., “a lightweight gripper for delicate objects”) and structured design representations (forces, kinematics, materials) remains wide. LLMs have narrowed this gap but still produce factually incorrect, physically infeasible, or internally contradictory suggestions. This ‘hallucination’ problem in engineering design is qualitatively different from hallucinations in general NLP: an incorrect suggestion can propagate through downstream embodiment and manufacturing decisions with significant cost implications. Systematic benchmarking of hallucination rates for engineering design tasks remains absent from the literature.

Addressing this gap requires controlled studies across mechanical, aerospace, and consumer domains benchmarking hallucination rates, downstream cost implications, and the effectiveness of mitigation strategies such as retrieval-augmented grounding. A further challenge is the potential for multilingual performance disparities, as most pre-training data remains English-dominant. Continued advances should benchmark LLM-based design reasoning across languages such as Chinese, German, and Arabic. As a concrete, deployable benchmark protocol, the field should adopt the following minimum standard: (1) define two standard design problems specifically, a lightweight cantilever bracket and a handheld device enclosure; (2) require independent reporting of hallucination rate (%), physical infeasibility rate (%), and a simple F_1 score for engineering constraint satisfaction.

6.7 What does not work: Failure analysis and negative results

The IPCD literature overwhelmingly reports successes, reinforcing publication bias and obscuring empirical reality. From our corpus, three failure patterns merit explicit documentation as first-class findings. First, across the reviewed evidence, unconstrained LLM-based approaches frequently produced geometrically invalid, physically infeasible, or internally inconsistent design outputs, particularly when language generation was not coupled with CAD validation, simulation, or explicit engineering constraints, because standard language models lack geometric reasoning and physical feasibility guarantees. Second, fully automated evolutionary search can also produce infeasible or difficult-to-manufacture concepts when objective functions and constraint formulations inadequately represent manufacturing realities, particularly in multi-material topology optimization with complex manufacturing constraints. Third, knowledge-based systems become brittle outside their encoded domain, generating confidently incorrect recommendations when facing novel requirements. These patterns suggest that automation without domain-aware constraint handling, human feedback, or scope management is a recurring failure mode. Future IPCD publications should treat failure reporting as a standard, not optional, element of empirical studies.

6.8 Emerging cross-cutting considerations

The following subsections address broader cross-cutting considerations that are distinct from the core technical barriers (Sections 6.1–6.6) but will increasingly shape IPCD development and adoption.

6.8.1 Sustainability and environmental impact

A critical gap in the current IPCD literature is the near-complete absence of sustainability considerations. Of the 247 reviewed papers, only 2 mentioned environmental impact or life-cycle considerations. This omission is increasingly untenable for three reasons. First, training large generative design models consumes substantial computational resources. Second, EU regulations including the Ecodesign for sustainable products regulation (ESPR, EU 2024/1781) and the EU product environmental footprint framework increasingly mandate life-cycle assessment (LCA) integration at the conceptual design stage; IPCD tools that cannot interface with LCA databases will face adoption barriers. Third, the “Green AI” movement in machine learning research has not yet substantially influenced IPCD.

Longer-term progress should incorporate: (1) energy-efficiency as an explicit design objective alongside performance and manufacturability; (2) cradle-to-gate LCA estimates integrated into generative concept evaluation pipelines; and (3) model-efficiency metrics as standard reporting requirements. The computational cost and carbon footprint of large-scale model training can be substantial and vary considerably with model architecture, hardware efficiency, training duration, and the carbon intensity of the electricity supply. Accordingly, future IPCD research should report computational-resource use and, where feasible, estimated energy or carbon impacts alongside conventional performance metrics.

6.8.2 Regulatory and ethical frameworks

Beyond the EU, additional jurisdictions introduce distinct compliance requirements that IPCD researchers must anticipate. China's Interim Measures for the Management of Generative AI Services impose content review and data transparency obligations, while the U.S. NIST AI Risk Management Framework adopts a voluntary, principles-based approach. Such jurisdictional fragmentation complicates global deployment and demands adaptable system architectures. Researchers building cross-jurisdictional IPCD systems should anticipate regulatory divergence and design systems with jurisdiction-configurable compliance modules. Ethical considerations — including IP ownership of AI-generated concepts, displacement of entry-level engineering roles, and training-data bias — will increasingly shape the pace and geography of IPCD industrial adoption.

6.8.3 Cognitive paradoxes in GenAI–human collaborative conceptual design

Beyond the technical challenges discussed in Sections 6.1–6.8.2, emerging evidence suggests that human–AI collaboration introduces a new class of cognitive tensions that influence creativity, trust, responsibility, and decision-making during conceptual design. These tensions represent structural characteristics of collaborative intelligence rather than deficiencies of either human designers or AI systems. Table 13 synthesizes five such structural paradoxes from the reviewed literature and established HCI design theory.

Table 13
Five structurally recurrent paradoxes in GenAI–human collaborative design (identified from the reviewed literature and design theory)

Paradox name	Core tension	Design manifestation	Proposed mitigation
Ambiguity vs. Precision	Humans work with vague, evolving intent; AI requires precise, stable inputs.	Designers struggle to articulate requirements as prompts; AI misinterprets fuzzy goals, producing structurally irrelevant outputs.	Mixed-initiative interfaces that infer intent incrementally from partial specifications.
Control vs. Serendipity	Designers want agency but also unexpected inspiration.	AI suggestions that are too predictable bore the designer; suggestions that are too surprising disrupt workflow and induce fixation.	Adjustable novelty controls and selective suggestion timing (e.g., suggestions offered only at impasse points).
Speed vs. Reflection	AI generates rapidly; creative ideation requires slow, deliberate incubation.	Fast iteration bypasses critical evaluation; designers accept the first adequate solution (satisficing), foreclosing better alternatives.	Mandated reflection intervals, structured evaluation prompts, and staged review protocols.
Individual vs. Collective	Single-designer intent vs. team dynamics, with AI as a third agent of uncertain role.	Role ambiguity arises (tool, partner, or supervisor?); conflicting expectations emerge across team members regarding AI authority.	Explicit role negotiation at session start; persistent team–AI interaction protocols and shared accountability frameworks.
Originality vs. Remix	Novelty is valued; AI inherently recombines and interpolates training data.	Legal and ethical concerns about IP ownership of AI-generated concepts; risk of unintentional plagiarism of source designs.	Attribution tracing, provenance tracking for generated concepts, and training-data disclosure policies.

7. Future Trends and Directions

The persistent challenges identified in Section 6 indicate that future progress in IPCD will depend not only on algorithmic innovation but also on advances in data infrastructure, human–AI collaboration, industrial integration, and trustworthy AI. Consequently, the future directions discussed below extend beyond individual computational methods toward broader technological ecosystems. On this basis, the future trajectory of IPCD can be usefully partitioned into two distinct categories: incremental advances that extend existing paradigms such as LLMs and hybrid systems and transformative paradigms including design-specific foundation models and physical AI that could fundamentally restructure how intelligent design is conducted. Evidence-based projections for each are presented below.

7.1 LLM-assisted design

Large-scale transformer-based language models and their successors are already being adapted for conceptual design. Near-term research directions include: (1) LLMs as creative ideation partners generating diverse function–structure mappings; (2) LLM-driven requirement elicitation and refinement; (3) retrieval-augmented generation (RAG) over design knowledge bases; and (4) LLM agents that call CAD APIs or simulation tools. Rather than full automation, the future points toward collaborative intelligence: interactive generative systems that learn designer preferences in real time, and cognitive design assistants that augment rather than replace human creativity.

7.2 Human-AI co-creation and self-improving design systems

Self-improving design systems learn continuously from each design iteration without full retraining. Key mechanisms include: (1) online learning from designer feedback and design outcomes; (2) meta-learning architectures that adapt model parameters from few design examples; (3) reinforcement learning from human preferences expressed through pairwise concept comparisons; and (4) federated learning across distributed design teams, enabling knowledge sharing without centralizing proprietary design data. These capabilities would shift IPCD tools from static, pre-trained artifacts to adaptive partners that improve with deployment.

7.3 Design-specific foundation models

A design foundation model pre-trained on massive, diverse design data—including text, images, CAD assemblies, simulation results, and material databases—could be fine-tuned for specific tasks. Unlike general-purpose models (GPT-4), a design foundation model would encode domain-specific priors about physical feasibility, manufacturing constraints (e.g., draft angles, clearances), and functional relationships. Early, limited efforts include concept demonstrations for CAD language models. Developing a genuinely general-purpose design foundation model remains a long-term scientific challenge requiring unprecedented data sharing, standardized design representations, and sustained collaboration among academia, industry, and software vendors. An achievable intermediate milestone—a domain-adapted foundation model pre-trained on existing public engineering datasets (e.g., Fusion 360 Gallery, ABC dataset) and fine-tuned for conceptual design tasks appears technically achievable within an approximate three- to five-year research horizon, provided that sufficient community-scale datasets and collaborative infrastructures become available. This is identified as the highest-priority research direction in Table 14.

Table 14
Prioritized research directions

Priority	Research direction	Expected impact	Estimated effort	Risk
1	Open, large-scale design datasets (F–B–S + CAD)	Enable foundation models	Very high (consortium)	High (IP barriers)
2	Explainable AI methods for generative design	Build trust and adoption	High	Medium
3	Standardized evaluation benchmarks	Compare methods fairly	Medium	Low
4	Integration APIs with commercial CAD/PLM	Real-world deployment	Medium (industry collaboration)	Medium
5	Human–AI co-creation frameworks	Balance automation and creativity	Medium	Medium
6	LLM grounding with design ontologies	Reduce hallucinations	Medium	Low

Among these priorities, open function–behavior–structure (F–B–S) design datasets (Priority 1) represent the single most critical enabling condition for the field: without such datasets, design-specific foundation models, explainable generative design methods, and standardized benchmarks will remain aspirational rather than achievable research goals.

7.4 Physical AI and embodied design reasoning

Physical AI refers to AI systems that reason about the physical world—including forces, kinematics, material behavior, and manufacturability—rather than operating purely in symbolic or linguistic spaces. For conceptual design, this convergence manifests as: (1) physics-informed neural networks that enforce thermodynamic and structural feasibility during generative synthesis; (2) robot learning approaches applied to manipulation task-driven design; and (3) embodied simulation environments where AI design agents receive physical feedback from virtual prototyping. The integration of IPCD with robotics and physically grounded simulation represents a fundamental transition from designing isolated artifacts to designing intelligent systems capable of reasoning about real-world behavior and operational context. Critically, explainability for structural design outputs remains qualitatively different from textual XAI: spatial and functional interpretability methods require domain-adapted approaches that current saliency-based techniques do not adequately provide.

7.5 Explainable and trustworthy AI for design

Advances in XAI for engineering design include concept attribution maps, natural-language explanations linked to physical principles, and uncertainty quantification for ML recommendations. Trustworthy AI for design will also require alignment with designer cognitive workflows and institutional accountability requirements.

7.6 Digital twins, IoT, and lifecycle-closed-loop design tools

Real-world product performance data from digital twins (sensors, usage logs) can be fed back into conceptual design models, creating a closed-loop design optimization cycle. This would shift IPCD from static, pre-production support to continuous, lifecycle-long design refinement. Integration with IoT platforms and DT-KG architectures (Section 5.7) is central to realizing this vision.

7.7 Democratized and low-code IPCD Tools

The democratization of IPCD through user-friendly, low-code or no-code platforms will lower the barrier for small and medium enterprises (SMEs). If successful, IPCD could shift from an elite research capability to a standard engineering tool, analogous to finite element analysis (FEA) becoming accessible in the 1990s. Key enablers include cloud-based deployment, visual workflow editors, and pre-trained design model libraries accessible without deep ML

expertise. However, without careful interface design, low-code tools may propagate poor design practice or over-rely on AI-generated concepts without adequate human judgment. Controlled studies comparing outcomes from expert vs. novice users of low-code IPCD tools are urgently needed.

7.8 Long-term convergent trends

The maturity horizons and evidence levels presented in Table 15 should be interpreted as qualitative research assessments rather than deterministic forecasts. They are derived from the convergence of technological maturity, empirical evidence, deployment barriers, and research trajectories identified across the reviewed literature. Their purpose is to organize potential development pathways according to both expected time horizon and current evidential support.

Table 15
Eight convergent trends in cognitive intelligence for product design (beyond 2026)

No.	Trend name	Est. maturity	Evidence level	Description
T1	Cognition–Decision Bimodal Synergy	Long-term (6–10 years)	High	RL frameworks that explicitly model both cognitive (understanding, perception) and decision-making (selection, trade-off) capabilities, enabling AI to simulate human-like design reasoning.
T2	Emotionally Intelligent Design Assistants	Mid-term (3–6 years)	Moderate	Systems that recognize designer affective states via biometric or interaction-pattern sensing, adapting suggestion strategies to maintain creative engagement.
T3	Agentic UX and Intent-Based Interaction	Long-term (6–10 years)	Emerging	Shift from command-driven interfaces to intent-based systems where designers express high-level goals and AI agents autonomously decompose, plan, and execute.
T4	Multilingual Design Reasoning Parity	Long-term (6–10 years)	Speculative	LLM-based design assistants performing technical reasoning equivalently across Chinese, German, Arabic, and English—critical for global manufacturing enterprises.
T5	Collective Intelligence Systems	Long-term (6–10 years)	Emerging	Multi-agent platforms where multiple human designers and multiple AI agents collaborate as heterogeneous teams, with AI agents specializing in distinct domains.
T6	Lifecycle-Closed-Loop Design	Mid-term (3–6 years)	High	Integration of IPCD with digital twins and IoT enabling design systems to learn from product performance in operation, automatically updating design rules throughout the product lifecycle.
T7	Regulatory-Compliant AI by Design	Mid-to-long-term (3–10 years)	Moderate	Systems built with traceability-by-design to satisfy EU AI Act and equivalent high-risk requirements: auditable decision logs, human-override mechanisms, and liability attribution.
T8	Democratized Cognitive Design Tools	Long-term (6–10 years)	Moderate	Low-code/no-code platforms embedding IPCD capabilities, enabling SMEs and non-experts to access advanced generative design without ML expertise.

Note: *Evidence level reflects the degree of support available in the reviewed literature rather than the likelihood of future realization. "High" indicates substantial empirical evidence from multiple independent studies; "Moderate" indicates consistent but still developing evidence; "Emerging" indicates promising research with limited validation; and "Speculative" denotes concepts that remain largely visionary or are supported primarily by conceptual or early-stage studies*

8 Conclusion and research agenda

8.1 Summary of findings

This review demonstrates that IPCD has undergone a fundamental transition from knowledge-driven to data-driven paradigms, with hybrid approaches emerging as a consistently promising design strategy. Hybrid approaches are associated with reported performance gains in the literature, but fewer than half of hybrid studies included statistical comparisons against non-hybrid baselines, and none met pre-specified criteria for both statistical testing and independent replication. Therefore, hybridization remains promising but not yet empirically verified as superior. However, the field remains constrained by systemic limitations in data availability, evaluation standardization, and integration with industrial workflows. A central insight is that no single paradigm dominates universally; instead, performance depends on the interplay between data availability, interpretability requirements, and design complexity. This comparative review answers the three research questions posed in the introduction as follows:

- **RQ1: Distribution of design intelligence across paradigms**: The review demonstrates that IPCD paradigms distribute knowledge representation, reasoning mechanisms, automation, scalability, and human involvement in fundamentally different ways. Rather than forming a hierarchy of increasingly superior methods, they represent complementary computational strategies suited to different conceptual design contexts.
- **RQ2: Structural trade-offs and recurring limitations**: Each paradigm exhibits characteristic strengths and limitations. Symbolic systems provide transparency but scale poorly; CBR depends on effective adaptation; evolutionary approaches excel at exploration but require expensive evaluation; and ML/DL methods offer scalability while remaining limited in explainability and physical feasibility.
- **RQ3: Emerging architecture of intelligent conceptual design**: The evidence indicates that IPCD is evolving toward distributed design intelligence in which symbolic reasoning, experiential knowledge, optimization, data-driven learning, and human creativity operate collaboratively rather than competitively. As a result, the key conclusions are as follows:

1. **Paradigm shift**: The center of gravity has shifted from rule-based (KBS) and case-based methods to data-driven ML/DL. Hybrid systems show consistent promise in the corpus; however, their superiority over single-paradigm baselines has not yet been established through rigorous statistical comparison and independent replication.
2. **No universally optimal paradigm exists**: Each paradigm has fundamental trade-offs, as summarized in Table 8. The choice depends on data availability, need for explainability, and desired automation level.
3. **Benchmarking gap**: The absence of open, annotated design datasets and standardized metrics severely hampers progress and fair comparison.
4. **Integration is the bottleneck**: Technical IPCD capabilities have outpaced practical deployment in commercial CAD/PLM workflows.
5. **Human-AI co-creation, not automation, is the likely future**: The most successful systems will augment, not replace, human designers.

8.2 Research agenda for the next five years

The findings synthesized in this review indicate that future progress in intelligent product conceptual design will depend not only on continued algorithmic advances but also on coordinated improvements in data infrastructure, evaluation standards, human–AI collaboration, and industrial deployment. Building upon the limitations and opportunities identified throughout the preceding sections, the following research agenda outlines the most strategically important directions for advancing the field over the next five years.

Accordingly, the six priority directions summarized in Table 14 define a tractable five-year research agenda. First, the field requires as a high priority open, large-scale function–behavior–structure annotated datasets; without shared training and evaluation corpora, cross-paradigm comparison and foundation model development remains extremely difficult. Concrete milestones include: a community dataset of at least 10,000 annotated design cases within the next three to five years, and an agreed annotation schema released for public comment during the initial phase of community standardization. Second, explainable generative design methods should move beyond post-hoc attribution toward architecturally embedded interpretability, enabling designers to interrogate why a concept was generated, not merely what it looks like.

This distinction between post-hoc and ante-hoc explainability is not merely definitional: post-hoc methods can be inconsistent, unfaithful to the underlying model, and sensitive to input perturbations, making them unreliable as the basis for high-stakes design decisions.

Measurable targets include methods that achieve ≥ 70% designer-comprehension ratings in controlled studies by 2028. Third, standardized human–AI benchmarking protocols, specifying reference design problems, evaluation metrics, and human baseline conditions, are needed before any comparative claim of system superiority can be scientifically validated. Fourth, CAD/PLM integration APIs should be co-designed with industry partners to ensure real-world deployability; Developing a minimum viable plug-in for at least one major CAD platform appears achievable within the next three to five years through close collaboration between academia, software vendors, and industrial partners. Fifth, human–AI co-creation frameworks should formalize designer agency models, drawing on the cognitive paradoxes identified in Table 13. Finally, LLM grounding with design ontologies constitutes the most tractable near-term path to reducing hallucination rates in engineering contexts. Achieving these directions will require coordinated collaboration among design researchers, AI scientists, industry practitioners, and standards bodies.

8.3 Practical recommendations

The implications of this review extend beyond academic research to industrial practice and software development. Accordingly, the following recommendations are organized for three principal stakeholder groups: design practitioners, researchers, and CAD/PLM software developers.

8.3.1 For design practitioners:

Design practitioners should adopt hybrid AI approaches that combine complementary computational paradigms rather than relying solely on data-driven methods. Explainable and transparent AI techniques should be prioritized for high-stakes engineering decisions to enhance trust and support effective validation. In addition, organizations should invest in the systematic collection, curation, and management of high-quality design data, as robust data infrastructure is fundamental to the successful deployment and continuous improvement of intelligent product conceptual design systems.

8.3.2 For researchers:

Researchers should prioritize the development and public release of high-quality benchmark datasets that enable rigorous and reproducible comparison of IPCD methods. Future studies must adopt standardized evaluation protocols, report negative results where appropriate, and promote the open sharing of datasets, models, and experimental configurations to strengthen reproducibility and accelerate cumulative scientific progress.

8.3.3 For software developers (CAD/PLM vendors):

CAD/PLM software developers should prioritize standardized plugin architectures and interoperable APIs that facilitate seamless integration between intelligent conceptual design methods and industrial engineering workflows. Particular emphasis should be placed on explainability, traceability, and compatibility with existing engineering software ecosystems to support practical industrial adoption.

8.4 Limitations of this review

As with all systematic reviews, the present work is subject to several limitations which are summarized in Table 16: (1) Publication bias: Conference proceedings may underrepresent negative results, potentially overstating ML/DL maturity. (2) Language bias: English-only papers may miss significant Chinese, German, or Japanese contributions. (3) Time lag: Publications from 2025–2026 may not yet have accumulated stable citation records. (4) Domain bias: Mechanical components dominate the corpus (35%), while electronics, software, and service design are underrepresented. (5) Preprint bias: The inclusion of 2025–2026 papers increase the risk of incorporating studies whose findings have not yet undergone extensive scholarly validation. (6) Quality

heterogeneity: Significant variance in methodological rigor across included papers means paradigm performance comparisons aggregate studies of widely varying evidential strength. (7) Google Scholar stopping-rule bias: The non-standard 'two consecutive pages' stopping criterion may underrepresent less-cited but relevant work, biasing the corpus toward high-citation papers, as described in Section 2.2.

Table 16
Summary of limitations of this review

Limitation	Severity	Mitigation
Publication bias (negative results underreported)	Moderate	Acknowledged; failure analysis section (6.7) added
English-language bias	Moderate	Chinese contributions estimated as undercounted; acknowledged in Section 5.5
Google Scholar stopping rule bias (~ 8–12% for low-citation papers)	Low	Sensitivity analysis (Appendix C, Table C5) confirms robustness; WoS + Scopus-only IRR = 1.16
Preprint inclusion (2025–2026)	Low	Qualitative only; not used in trend analysis; readers advised to interpret 2026 findings cautiously
Quality heterogeneity across included studies	Moderate	3-item quality scale applied; low-quality studies excluded from quantitative claims (Section 2.4)

8.4.1 Deliberate exclusions

Deliberate exclusions include sub-paradigm comparisons, meta-analysis, cost–benefit analysis of IPCD adoption, longitudinal tracking of research groups, and non-English literature (Section 8.4). Readers seeking implementation details would benefit from consult method-specific reviews cited herein.

8.5 Reproducibility and open science

A systemic reproducibility concern affects the IPCD field. Of the 247 papers reviewed, fewer than 12% provided executable code or publicly accessible full datasets; this severely limits independent validation and reuse of reported results. Broken down by paradigm: KBS 5%, CBR 10%, EA/GD 15%, ML/DL 8%, and hybrid architectures 20%. An additional 18% of studies provided at least pseudocode or algorithmic descriptions sufficient for partial replication. This code-sharing rate is substantially lower than comparable ML fields (where code sharing has risen to 30–40% following NeurIPS and ICML open-science requirements). In addition to the overall methodological quality assessment, reproducibility was independently evaluated using a five-point ordinal scale assessing code availability, data availability, methodological transparency, parameter reporting, and execution documentation. A reproducibility badge system, analogous to those adopted by Nature and ACM, is recommended for the next generation of IPCD publications. Among studies providing code/data, only 4% included sufficient documentation to reproduce results without contacting authors; hybrid systems scored highest (2.9/5) and KBS lowest (1.6/5).

A further cross-cutting concern is dual-use risk: the same generative capability that produces novel product concepts could, under different intent, produce harmful configurations. Consistent with the NASEM framework, journals publishing IPCD research should require dual-use impact statements and establish editorial review protocols for high-autonomy generative design systems (Sciences et al., 2017).

8.6 Final remarks

Intelligent product conceptual design has matured from a niche research topic to a practical engineering capability. The convergence of large language models, generative AI, and human–AI collaboration promises to transform conceptual design from a purely human creative activity into a synergistic partnership between designer and machine.

The evidence of this review suggests that the most successful systems of the coming decade will not be those that fully automate design, but those that empower human designers to explore more and better concepts, faster and with greater confidence. Achieving that future requires not only better algorithms but also open data, standardized benchmarks, and deep integration with the tools that designers already use. Critically, this agenda cannot be pursued by the research community in isolation: its realization requires active partnership with industrial CAD/PLM vendors (for integration APIs and workflow adoption), standards bodies (for evaluation metric consensus), and policy makers (for open data governance and IP frameworks). The six priority directions in Table 14 define a tractable and coordinated roadmap for this interdisciplinary collaboration.

Ultimately, the future of intelligent product conceptual design will not be determined by increasingly autonomous algorithms alone, but by the effective integration of human creativity, symbolic knowledge, computational intelligence, optimization, and physically grounded reasoning within collaborative design ecosystems. By synthesizing more than three decades of research through a unified comparative framework, this review demonstrates that the evolution of IPCD is best understood as the progressive redistribution of design intelligence across complementary paradigms, thereby providing both a conceptual foundation for future research and a practical roadmap for next-generation intelligent engineering design systems.

Abbreviations

Abbreviation	Definition
IPCD	Intelligent Product Conceptual Design
KBS	Knowledge-Based Systems
CBR	Case-Based Reasoning
ML	Machine Learning
DL	Deep Learning
LLM	Large Language Model
GNN	Graph Neural Network
GAN	Generative Adversarial Network
VAE	Variational Autoencoder
MOEA	Multi-Objective Evolutionary Algorithm
FBS	Function-Behavior-Structure
XAI	Explainable Artificial Intelligence
RAG	Retrieval-Augmented Generation
CAD	Computer-Aided Design
PLM	Product Lifecycle Management
SME	Small and Medium Enterprise
CAGR	Compound Annual Growth Rate
GLM	Generalized Linear Model
ESPR	Ecodesign for Sustainable Products Regulation
LCA	Life Cycle Assessment
FLOPs	Floating-Point Operations
RQ1–RQ3	Research Questions 1–3
IRR	Incidence Rate Ratio
CI	Confidence Interval

Declarations

Acknowledgements

The authors gratefully acknowledge the support provided by Quanzhou University of Information Technology for this research. No external funding was received for this study.

Declaration of Generative AI Use

During the preparation of this work, the authors used DeepSeek exclusively for language refinement, grammar correction, and initial structuring of the literature review. After using this tool, the author(s) have reviewed and edited the content as necessary and take full responsibility for the content of the published article.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors acknowledge the institutional support provided by Quanzhou University of Information Technology.

Ethics Statement

This review does not involve human participants, animal subjects, or primary data collection and therefore does not require ethics committee approval. All data analyzed in this review are from publicly available peer-reviewed publications.

Conflict of Interest

The authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability Statement

The complete dataset supporting this review—including the coding sheet for all 247 included studies, bibliographic metadata, paradigm classifications, application domains, hybrid-system status, quality scores, reproducibility indicators, database-specific search strings, the PRISMA-informed selection flowchart, annual publication counts, and the R/Python analysis scripts—is archived in the Open Science Framework repository specified in the supplementary materials.

References

1. Beitz W, Pahl G, Grote K (1996) Engineering design: a systematic approach. MRS Bull, *71*
2. Bentley P (1999) Evolutionary design by computers. Morgan Kaufmann
3. Bing Y, Yu L, Li S, Cho Y, Li C (2026) A novel product shape innovation design method based on Kansei Engineering and GAN model with limited sample data. J Eng Des 37(3):981–1006
4. Brown DC, Chandrasekaran B (2014) Design problem solving: knowledge structures and control strategies. Morgan Kaufmann
5. Burnap A, Hauser JR, Timoshenko A (2023) Product aesthetic design: A machine learning augmentation. Mark Sci 42(6):1029–1056
6. Cagan J, Campbell MI, Finger S, Tomiyama T (2005) *A framework for computational design synthesis: model and applications*
7. Camburn B, He Y, Raviselvam S, Luo J, Wood K (2020) Machine learning-based design concept evaluation. J Mech Des 142(3):31113
8. Chakrabarti A (2002) Engineering design synthesis: understanding, approaches and tools. Springer Science & Business Media
9. Chandrasekaran B (1990) Design problem solving: A task analysis. AI Magazine 11(4):59
10. Chen M-T, Pan Y, Zuo W, Zhao O, Gardner L (2026) Generative inverse design of steel gridshell joints with multi-objective optimisation. Adv Eng Inform 72:104483
11. Chen P, Cai Y, Zhou Z, Yao J, Li J, You W, Sun L (2026) An LLM-based multi-agent system to assist early-stage product design and evaluation. J Eng Des, 1–36
12. Cross N (2021) Engineering design methods: strategies for product design. Wiley
13. Figoli FA, Bruggeman R, Rampino L, Ciuccarelli P (2025) AI-against-design map: A systematic review of 20 years of AI-focused studies in design research. Des Stud 96:101279
14. Gaikwad S, Daase C, Haertel C, Staegemann D, Turowski K (2025) *Modular Generative Adversarial Networks for Support in Product Design*
15. Gero JS, Kannengiesser U (2004) The situated function–behaviour–structure framework. Des Stud 25(4):373–391
16. Gero JS, Kannengiesser U (2014) The function-behaviour-structure ontology of design. An anthology of theories and models of design: Philosophy, approaches and empirical explorations. Springer, pp 263–283
17. Goel AK, Vattam S, Wiltgen B, Helms M (2012) Cognitive, collaborative, conceptual and creative—Four characteristics of the next generation of knowledge-based CAD systems: A study in biologically inspired design. Comput Aided Des 44(10):879–900
18. Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison. Reading
19. Guo Y (2017) *Deep learning for visual understanding*
20. Hayes-Roth F, Jacobstein N (1994) The state of knowledge-based systems. Commun ACM 37(3):26–39
21. Holland JH (1992) Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT Press
22. Hudson MG, Parmee IC (1996) The application of genetic algorithms to conceptual design. *AI System Support for Conceptual Design: Proceedings of the 1995 Lancaster International Workshop on Engineering Design, 27–29 March 1995*, 17–36
23. Jacobs O, Pazhoohi F, Mullen G, Kingstone A (2025) Comparative designs reveal preferences for human-generated rather than AI-generated art. Empir Stud Arts, 02762374251360129
24. Jin B, Liu G, Han C, Jiang M, Ji H, Han J (2024) Large language models on graphs: A comprehensive survey. IEEE Trans Knowl Data Eng 36(12):8622–8642
25. Kolodner JL (1992) An introduction to case-based reasoning. Artif Intell Rev 6(1):3–34
26. Kumar S, Gopi T, Harikeerthana N, Gupta MK, Gaur V, Krolczyk GM, Wu C (2023) Machine learning techniques in additive manufacturing: a state of the art review on design, processes and production control. J Intell Manuf 34(1):21–55
27. Kumar V, Karniadakis GE (2026) Agentic Risk-Aware Set-Based Engineering Design. *ArXiv Preprint ArXiv:2604.16687*
28. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics, 159–174
29. Liang J, Zhong X, Zhao S, Fricker P, Koh I (2025) Building-graph-AI: Graph neural networks learning and generating 3D detailed and layered building models. Int J Architectural Comput 23(3):640–654
30. Liu B, Tan J, Lyu H, Zhang Z, Liu Z, Huang H (2025) Surrogate-assisted evolutionary optimization and microgravity experimental validation of a planar deployable mechanism. Aerosp Sci Technol, 111299
31. Liu S, Shen Y, Zhang Y, Hou Z, Wang X, Luo J, Zhang Z (2026) iDesignGPT enhances conceptual design via large language model agentic workflows. Nat Commun
32. Liu Z, Zhu D, Rodrigues SP, Lee K-T, Cai W (2018) Generative model for the inverse design of metasurfaces. Nano Lett 18(10):6570–6576
33. Maher M, Lou, Pu P (2014) Issues and applications of case-based reasoning to design. Psychology
34. McHugh ML (2012) Interrater reliability: the kappa statistic. Biochemia Med 22(3):276–282

35. Nkisi-Orji I, Palihawadana C, Wiratunga N, Corsar D, Wijekoon A (2022) Adapting semantic similarity methods for case-based reasoning in the cloud. *International Conference on Case-Based Reasoning*, 125–139
36. Peng S, Feng R, Huang S, Cui L (2026) Physics-aware edge-centric graph neural networks for full-field mechanics simulation of reticulated shells. Eng Struct 360:122759
37. Reinpold LM, Wagner LP, Gehlhoff F, Ramonat M, Kilthau M, Gill MS, Reif JT, Henkel V, Scholz L, Fay A (2025) Systematic comparison of software agents and Digital Twins: differences, similarities, and synergies in industrial production. J Intell Manuf 36(2):765–800
38. Sciences NA, of, Medicine P, Affairs G, Science Con (2017) Law, Concern, C. on D. U. R. of, & Management, O. for F. *Dual use research of concern in the life sciences: current issues and controversies*
39. Singh V (n.d.). Revisiting Gero's Function-Behaviour-Structure (FBS) Framework to Establish Fundamental Principles, Novel Insights, and Propositions. *Novel Insights, and Propositions*
40. Sreenivasan A, Suresh M (2024) Design thinking and artificial intelligence: A systematic literature review exploring synergies. Int J Innov Stud 8(3):297–312
41. Takagi H (2002) Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation. *Proceedings of the IEEE*, *89*(9), 1275–1296
42. Tanriover I, Lee D, Chen W, Aydin K (2022) Deep generative modeling and inverse design of manufacturable free-form dielectric metasurfaces. ACS Photonics 10(4):875–883
43. Tao F, Zhang H, Liu A, Nee AYC (2018) Digital twin in industry: State-of-the-art. IEEE Trans Industr Inf 15(4):2405–2415
44. Tong C, Sriram D (1992) Artificial intelligence in engineering design (Volume I) design representation and models of routine design. Academic Press Professional, Inc
45. Ullman DG (2010) The mechanical design process: Part 1. McGraw-Hill
46. Ulrich, KarlT (2008) &. *Product design and development /-4th ed*. Product design and development /-4th ed
47. Wang J, Ma Y, Zhang L, Gao RX, Wu D (2018) Deep learning for smart manufacturing: Methods and applications. J Manuf Syst 48:144–156
48. Wang L, Shen W, Xie H, Neelamkavil J, Pardasani A (2002) Collaborative conceptual design—state of the art and future trends. Comput Aided Des 34(13):981–996
49. Wang Q, Zhang Y, Gong D, Song X, He C, Ji X, Chu F (2026) A Review of Surrogate-Assisted Evolutionary Multi-Objective Optimization. IEEE Trans Emerg Top Comput Intell
50. Wang X, Hu Y, Wang Y, Shi C, Zhang M (2026) Position: How can Graphs Help Large Language Models? *ArXiv Preprint ArXiv:2605.02452*
51. Wu T, Maruyama T, Wei L, Zhang T, Du Y, Iaccarino G, Leskovec J (2024) Compositional generative inverse design. *ArXiv Preprint ArXiv:2401.13171*
52. Xiao Y, Ahmed F, Sha Z (2023) Graph neural network-based design decision support for shared mobility systems. J Mech Des 145(9):91703
53. Xu J, Kovatsch M, Mattern D, Mazza F, Harasic M, Paschke A, Lucia S (2022) A review on AI for smart manufacturing: Deep learning challenges and solutions. Appl Sci 12(16):8239
54. Youmans RJ, Arciszewski T (2014) Design fixation: Classifications and modern methods of prevention. AI EDAM 28(2):129–137

Figures

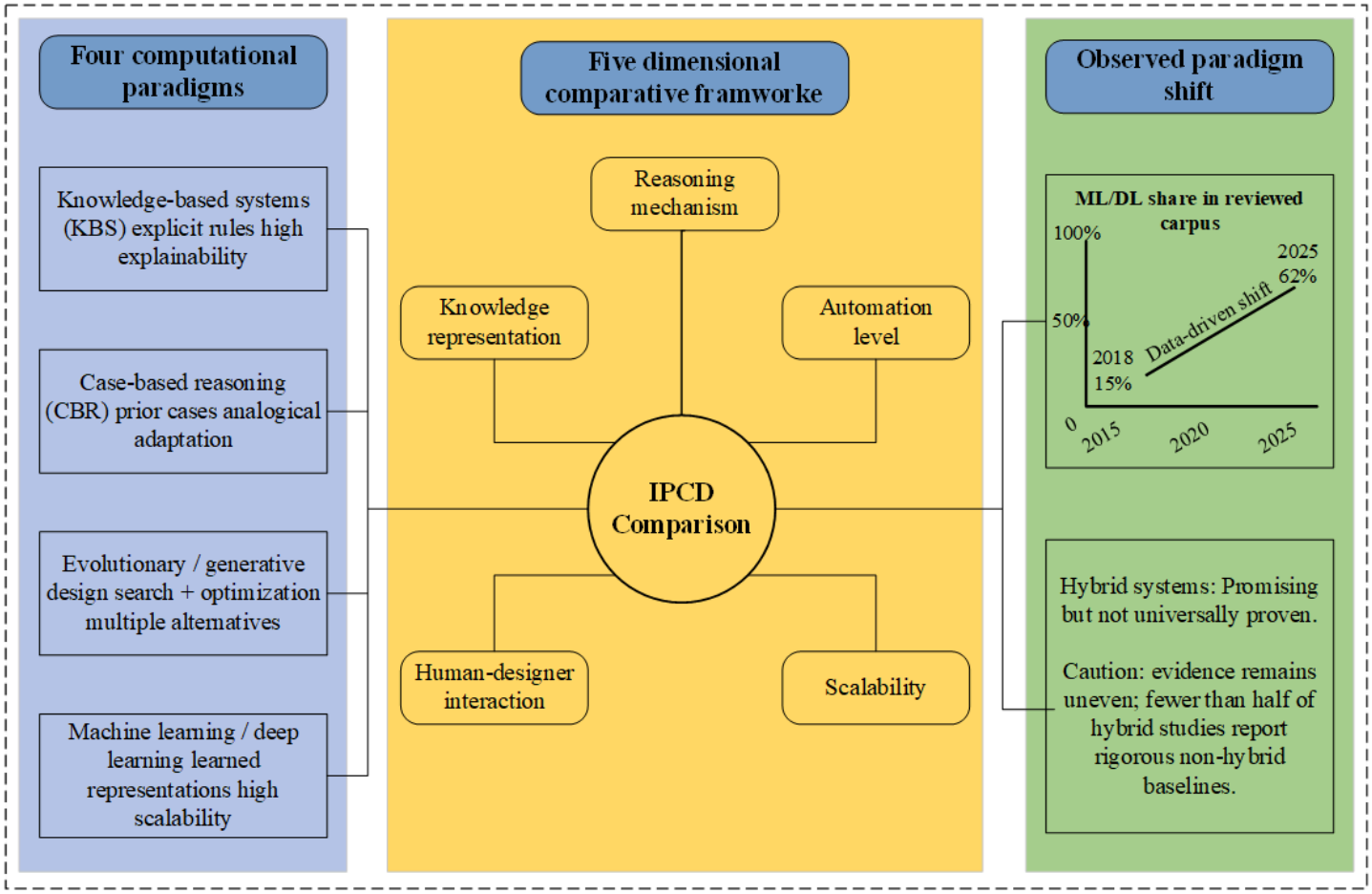


Figure 1

Graphical overview of the IPCD paradigms, comparative dimensions, and observed paradigm shift

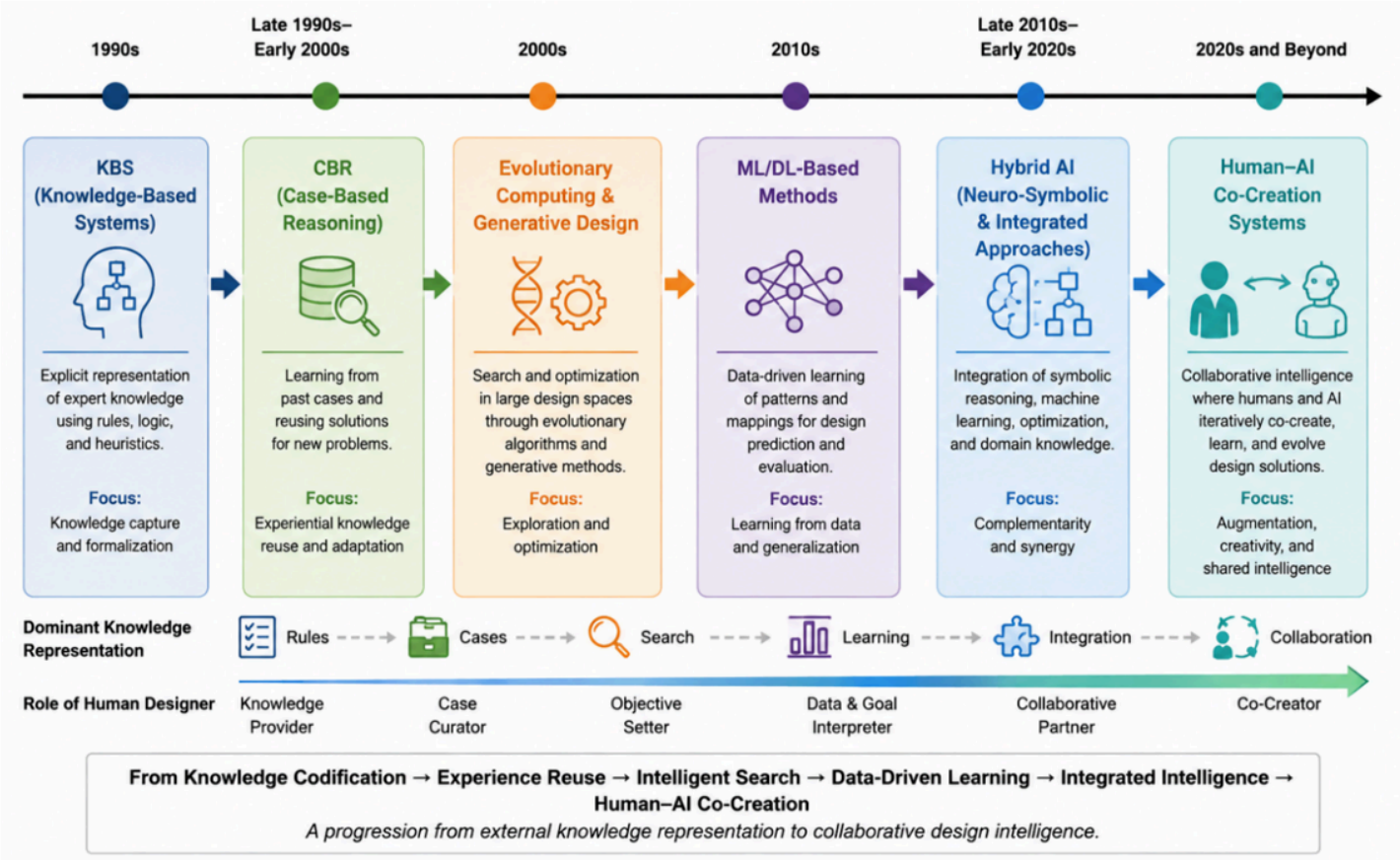


Figure 2

Evolution of design intelligence across IPCD paradigms