Asian Journal of Computer Science and Technology
ISSN: 2249-0701 (Print) | 2583-7907 (Online)
Journal Homepage: http://ajcst.co
DOI: https://doi.org/10.70112/ajcst-2026.15.1.4443

Review Article

From Diffusion Models to Instruction-Tuned LLMs: A Unified Taxonomy and Trend Analysis of Generative AI

Stephen Akobre[1]*, Japheth Kodua Wiredu[2], Iven Aabaah[3] and Nelson Seidu Abuba[4]
[1]Department of Cyber Security and Computer Engineering Technology, [3]Department of Information Systems and Technology, University of Technology and Applied Sciences, Navrongo, Ghana
[2,4]Department of Computer Science, Regentropfen University College, Bolgatanga, Ghana

Article Information

Article History
Received: 1 March 2026
Revised: 5 April 2026
Accepted: 12 April 2026
Published online: 17 April 2026

Keywords
Generative Artificial Intelligence
Large Language Models
Diffusion Models
Instruction Tuning
Taxonomy and Trends

Correspondence*
sakobre@cktutas.edu.gh

ORCID
Stephen Akobre
https://orcid.org/0000-0003-3320-212X
Japheth Kodua Wiredu
https://orcid.org/0009-0008-0313-5011
Iven Aabaah
https://orcid.org/0009-0008-1241-1153
Nelson Seidu Abuba
https://orcid.org/0009-0003-7674-8808

Abstract

Generative AI has brought about swifter and more profound changes in the computing environment, with machines capable of creating high-quality text, images, audio, video, code, and multimodal data. This paper presents a detailed taxonomy and trend overview of generative AI models and products, which follow the historical development of the earlier diffusion-based models all the way to instruction-tuned large language models (LLM). The study is a systematic taxonomy of models based on architecture, modality, training paradigm, deployment type, and scale, and map research innovations to commercial products, such as chat assistants, image and video generators, code assistants, and multimodal systems. The analysis illustrates common patterns of model scaling, benchmark performance, patterns of adoption, and the increased convergence of the vision and language modalities. The present study also points to the important gaps and challenges such as alignment, hallucination, ethical issues, and limitations in evaluation and suggests possible research and industrial implementation directions. This work offers a visionary reference to future scholars, practitioners, and policymakers who want to grasp the development, abilities, and potentials of generative AI by merging technical, commercial, and societal attitudes.

I. INTRODUCTION

Generative Artificial Intelligence (AI) has quickly become one of the most disruptive fields of study in the field of computer science, redefining the human-machine interaction and driving innovation in creative sectors, in education, in healthcare, and in business [65][66][8][47][63]. Generative models generate new content, such as textual, visual (images), audio (music), video (both live-action and animation), and multimodal content (such as art), unlike discriminative models that classify, or predict, such content, and are usually of high quality and diversity, akin to that of human authors [67][61][8]. Its significance is not only its technical complexity but also the fact that it has been widely adopted by society which is why it is the center of the current AI studies and applications [61][47][8]. The development of generative AI in the last ten years is full of both paradigm shifts and increased magnitude. One of the milestones was the Generative Adversarial Networks (GANs) introduced by [21] which put the generative process into the context of a minimax game, where one of the parties (a generator G) generates samples, and the other party (a discriminator D) differentiates between genuine and generated samples [21]. GANs allowed the synthesis of realistic images, audio and data and inspired a lot of research on adversarial learning, but issues like mode collapse and unstable training were still present [54].

Although GANs used to rule the early days of generative modelling, the paradigm shift occurred with the introduction of diffusion models. Diffusion models are trained to de-noise a signal with noise added to it, rather than competing against each other. [26] Popularised the formulation as denoising diffusion probabilistic models (DDPMs) and demonstrated that it is possible to learn to reverse a diffusion (noise) process in order to produce high-quality samples [26]. Subsequent models such as Imagen [53] and Stable Diffusion [51] offered better photorealism and diversity (and more stable training and controllability by use of guidance methods e.g. classifier guidance, classifier-free guidance). Surveys like Diffusion Models: A Comprehensive Survey divide the accelerated methodological growth of diffusion methods (efficient sampling, likelihood estimation, structured data handling) [64] and observe that they have wide-ranging application in regard to image, video, and molecule generation.

The Transformer architecture [58] along with the visual generative modelling transformed language generation because it supports attention-based autoregressive sequence modelling [58]. Large language models (LLM), including GPT-2 [45] and GPT-3 [7] based on transformers demonstrated emergent capabilities: models with increased parameters, data and compute deliver greater reasoning, coherence and generalization capabilities including zero-shot and few-shot learning [28]. These models quickly were applied to various natural language processes (summarization, translation, QA, code generation) with no need to redesign the architecture.

Nevertheless, pre-training in itself does not ensure synchronized and manageable conduct. One step was the introduction of instruction tuning, whereby LLMs are also further refined to adhere to user instructions. Reinforcement Learning through Human Feedback (RLHF) emerged as a key method (e.g. Instruct GPT, the deployment implementation of OpenAI) where human preferences are used to update the models [41]. The methods of incorporating human feedback have been further optimized by preference tuning and direct preference optimization (DPO) [62]. Instruction tuning models like ChatGPT and Claude have been turned into a safer, more controllable, and usable conversation assistants effectively turning generative models into instruction-following agents. This generation-to-alignment shift of instruction compliance has increased the pace of generative AI adoption in physical products. Modern commercial systems hide their architectural specifics, although they are based on the same modelling primitives. Intelligent assistants, creative tools, copilots and synthesis APIs dominate the digital ecosystems. Even with this fast development, the generative AI market is disjointed. Outputs in research can be in the fields of vision, NLP or multimodal subfields; model internals are commonly hidden in industrial products. Existing pre-surveys are usually focused on individual families e.g. GANs [14], diffusion models [64], or transformer/LLM methods [38] but do not cross-paradigmatically integrate, or connect them to commercial deployment strategies. In addition, scaling laws, benchmark improvement and compute trends have been studied individually, but there is no consolidated taxonomy or mapping between architectures, modalities, benchmarks, and product applications. The paper aims to address these gaps, including building a single taxonomy of generative AI models (since the early GANs up to the modern instruction-tuned LLM), a trend analysis of scaling, performance, and adoption, and a product-model crystallization of mapping the real-world AI tools to the techniques behind them. The key research questions are:

1. What has changed since diffusion-based approaches to generative AI models? Parameterized advice to instruction-tuned LLMs.
2. How can existing architectures and uses of generative AI be most taxonomically classified?
3. What is the model scale and benchmark performance and product deployment trends?
4. What are current limitations and future opportunities of research and application of generative AI?

Our contributions are:

1. Possible taxonomy: to organize the generative AI models into architecture, modality, training paradigm and product deployment.
2. Trend analysis: measuring the time-scale, benchmarks and adoption of evolution in academia and industry.
3. Mapping of model products: the connection between commercial products based on generative AI and other architecture paradigms.
4. Challenge identification: Be able to state the gaps in evaluation, alignment risks, costs in the environment, and implications on policy to guide the future research.

This paper is a comprehensive and futuristic exploration of generative AI by combining technical underpinnings, historical developments, commercial applications, and a prognosis. The resulting taxonomy and analyses are to become a resource to researchers, practitioners, policymakers and educators to gain insight about not only where generative AI is at the current stage but also where the future lies.

II. BACKGROUND AND THEORETICAL FOUNDATIONS

The swift rise of generative artificial intelligence has been influenced by the demanding theoretical frameworks, correspondence strategies, and usage sectors, which will be measured using the newest benchmarks and metrics. Such understanding of these foundations is essential to place the modern models, including instruction-tuned large language models (LLMs), in the context of the overall research directions of generative AI [21], [58], [26], [6]. This part of the paper is a review of the major generative modelling paradigms, the development of instruction tuning and alignment methods, the growing number of applications across modalities, and the assessment models that still dominate the field.

A. Generative Modelling Paradigms

The concept of generative modelling has been used throughout history to integrate the distinct aspects of various art forms via its paradigm, a characteristic that can be regarded as a constitutional convention. The paradigm of generative modelling has been applied at all times and used to accommodate the unique features of different art forms, which can be considered as a constitutional convention. Eclectic choice of methods of generative modelling is indicative of the complexity of the task of modelling high-dimensional data distributions. The first breakthroughs were Generative Adversarial Networks (GANs) introduced by [21] which used a competition principle between a generator and a discriminator model to create high aesthetic samples. Although GANs have introduced new benchmarks in photorealism image generation [44], [29], their adversarial architecture was too unstable and mode collapsing, which could not be scaled up [4]. Parallel to that, Variational Autoencoders (VAEs) provided a probabilistic method which encodes data in latent variables and then reconstructs them using learned distributions [31] [50] VAEs were offered with stable training and interpretability, however, tended to lack sample sharpness, which limited them to high-resolution generation tasks [49]. Autoregressive transformers emerged and this was a paradigm shift. Autoregressive models, including GPT-2 [45] and GPT-3 [7] that are based on the Transformer architecture, showed the ability to learn long-range sequential dependencies in text using scalable self-attention patterns. These models also generalized effectively independently of various natural language processing tasks [16][47], making the foundation model era a reality [6]. Diffusion models have taken the place of other generative models as the canonical generative architecture of the 2020s, in the visual world.

A diffusion architecture, including DDPM [26], Stable Diffusion [51] and Imagen [53] by modelling the process of noise-injection reversal, achieved record-breaking fidelity, diversity, and controllability. Compared to GANs, which tended to have unsteady convergence, diffusion models had steady likelihood-based education and more intelligible sampling algorithms [55], [17]. Since then, they have grown to be more than just capable of producing vision to speech, music, and cross-modal applications [27], [32]. Energy-based models (EBMs) are not so popular yet still receive some theoretical attention [33], [18].

EBMs are a model of data configurations which gives a model of complex distributions based on the data configuration, without explicitly normalizing the probabilities. Their super modelling nature has been investigated as controllable generation [20], where practical implementation is limited by the problem of training. Collectively, these paradigms depict the intellectual heritage of adversarial training and latent-variable inference to transformers and diffusion, which currently dominate the academic and industrial generative AI.

B. Instruction Tuning and Alignment

The incredible generative models scaling introduced new issues: albeit being fluent, even large pretrained generative models were not compatible with human intent [60]. The transition of raw language modelling to instruction-following systems is one of the most radical changes in the study of AI in the recent past. The initial alignment processes involved using supervised fine-tuning (SFT) on manual collections of instructions and responses [59][13]. An even more radical innovation was the adaptation of Reinforcement Learning including Human Feedback (RLHF), that is, the operationalization of alignment as a reward maximization objective [12], [41]. This was done by gathering human preference data, training a reward model that models the preferences and refining the base model to optimize the alignment. This approach was the baseline in systems like ChatGPT and it made them interactive and context-sensitive agents [5]. The effectiveness of RLHF highlighted the role of the large-scale pretraining in association with structured human interventions to reach the usability and trustworthiness.

The current studies have widened the alignment of alignment beyond RLHF. Direct Preference Optimization (DPO) and its variants do not use the concept of reinforcement learning; instead, they directly optimize against human preference data at a lower level of complexity [46]. Meanwhile, more advanced datasets of instructions, including FLAN [59], Alpaca [57], and Dolly [15], have added coverage to models in reasoning, coding, and multimodal tasks. These advances in combination are indicative of a shift to instruction-tuned assistants and not only powerful, but responsive, controllable, and general to different domains.

C. Applications of Generative AI

Generative AI is currently available in virtually any form of human expression and knowledge work [67], [61] Instruction-conditioned LLMs can summarize, discuss, answer questions and make inferences adaptively and competitively to domain-specific systems in natural language processing [7], [11]. Education and legal services, industries that require knowledge have also been transformed by these models as conversational interfaces and collaborative creators [63][63]. GANs and diffusion models have revolutionized image synthesis in computer vision, making it possible to create content that is photorealistic as well as gain a wider range of applications in image editing, restoring, and transferring artistic styles [30][53]. Diffusion-based systems like Stable Diffusion have made such diffusion architecture more mainstream, with open-source releases [52], leading to new industry and arts and design works [64]. Generation of audio has also been assisted by autoregressive and diffusion-based models generating realistic speech and generative music as well as restoring audio of high quality [32], [27]. The technologies form the basis of voice aids, accessibility machines, and innovative audio platforms. The most ambitious frontier is perhaps the emergence of the

multimodal systems, which combine text, vision, and audio into cohesive generative systems. Cross-modal reasoning has been shown by such models as DALL·E, which takes pictures as inputs and generates descriptive text [48] and GPT-4 with visual capabilities that generate images when presented with a text input [1]. This multimodal grounding capability broadens the scope of generative AI beyond the field of specialization to human-machine interaction in general. Lastly, the code generation domain has also come up as a revolutionary application. Such systems as Codex [9], AlphaCode [35], and StarCoder [34] apply the capabilities of LLMs to software engineering, where they can be used to synthesize and debug code and produce documentation. Such models have a great potential to increase the productivity of the developers and cast doubts of reliability, intellectual property, and the future of programming as a human skill [43].

D. Benchmarks and Metrics

One of the most problematic and disputed fields of generative AI is evaluation. Conventional measures of translation and summarization like BLEU and ROUGE, which are quantitative measures of n-gram overlap, do not tend to reflect semantic faithfulness or human fluency [42][37]. Image generation uses quality metrics, such as the Inception Score (IS) and Frechet Inception Distance (FID), to measure visual quality, and visual diversity [54][25], but these metrics as well do not align with human judgments of quality or diversity. With the increase in scale and capability of models, benchmark suites have developed to test more competencies. Datasets used in language to test reasoning, knowledge and ethical alignment include MMLU [24], BIG-bench [56] and HELM [36]. Cross-modal standards such as those of COCO Captions [10] and VQAv2 [22] may be used in the multimodal domains, but they are still narrow in scope compared to the real-life complexity. Human evaluation is more often considered the gold standard, especially when it comes to subjective work, e.g. quality of dialogue, creativity, or ethical correctness [2]. Human preference studies in large scale, typically with model evaluation pipelines, are now central to model development [5]. However, there are still major issues of high costs of evaluation, cultural bias, and inability to identify long-term or emergent behaviours [60]. These constraints will have to be dealt with to create fair, reliable, and trustworthy generative systems.

III. METHODOLOGY

The following section presents a taxonomy to order the rather fast-evolving world of generative AI. It incorporates five fundamental dimensions, including architecture, modality, training paradigm, deployment type and scale, which combine both the technical diversity and practical usability.

The taxonomy was constructed in a systematic five stages approach (Figure 1; Algorithm 1) that aims to make the taxonomy consistent, reproducible, and objectively classified.

A. Taxonomy Construction Process

Five sequential stages are used in constructing it:

1. *Corpus Construction:* Gathering of pertinent literature of key academic sources utilizing preset search terms.
2. *Screening and Filtering:* Inclusion criteria are used to screen and filter.
3. *Dimensional Extraction:* These are key technical attributes that are identified and clustered with the help of open coding and clustering.
4. *Operationalization:* Categorization of dimensions and rules of classification.
5. *Validation:* Testing of the taxonomy by using expert judgment and inter-rater reliability.

B. Data Collection and Filtering

It was searched in the largest repositories (e.g., arXiv, IEEE Xplore, ACL Anthology) and the best conference proceedings with keywords related to generative AI (e.g., large language models, diffusion models, GANs, multimodal AI). The original list of 2,847 articles published in January 2022 to March 2024 was narrowed down to 247 following the following inclusion criteria: Generative model contribution, Empirical validation, Impact of publication in a good publication or adequate citation and Availability of full text.

C. Dimensional Extraction and Operationalization

Each paper was mined in terms of technical terminology, which was further clustered together under semantically related techniques. The resulting clusters were studied and grouped into five fundamental dimensions upon which the basis of the taxonomy is based as shown in Figure 1.

1. Architecture: Generative models may be subdivided into structural foundations. Vision synthesis is dominated by diffusion models whereas autoregressive transformers form the basis of large language models. Hybrid systems are becoming more integrated between the two paradigms and provide better quality and controllability of the samples. GANs and VAEs have not been pushed out yet but are increasingly being replaced by diffusion transformer pipelines in niche tasks.
2. Modality: Generative AI falls under text, vision, audio, video, and code. GPT-series and the diffusion-based Stable Diffusion and Imagen models were massively used and transformed the way images are generated respectively. Multimodal systems (e.g. GPT-4V, Gemini) are embracing a trend of converting foundational systems to multimodal integration (i.e. combining language, vision, and audio) at the foundation level.
3. Training Paradigm: Training previously involved only the use of supervised methods but currently involves self-supervised pretraining using huge corpora. The most popular approaches to identifying a match between the output of the models and the expectations of humans are Reinforcement Learning with Human Feedback (RLHF), preference

modelling, and direct preference optimization (DPO). In downstream application usability and reliability has become the norm of instruction tuning.

4. Deployment Type: Deployment can be split into open-source (e.g. LLaMA, Stable Diffusion), API-based (e.g. OpenAI GPT-4, Anthropic Claude) and proprietary embedded product (e.g. Microsoft Copilot, Adobe Firefly). The open-closed dichotomy determines the accessibility, reproducibility and the forces of competition within the AI ecosystem.

5. Scale: Scale is a vital axis, which consists of parameters, data size, and computing resources. Trends show that scaling is still able to provide performance benefits, but emerging methods that are based on efficiency, like parameter efficient fine-tuning, quantization, and sparse architectures, are becoming increasingly popular to find solutions to sustainability and accessibility aspects.

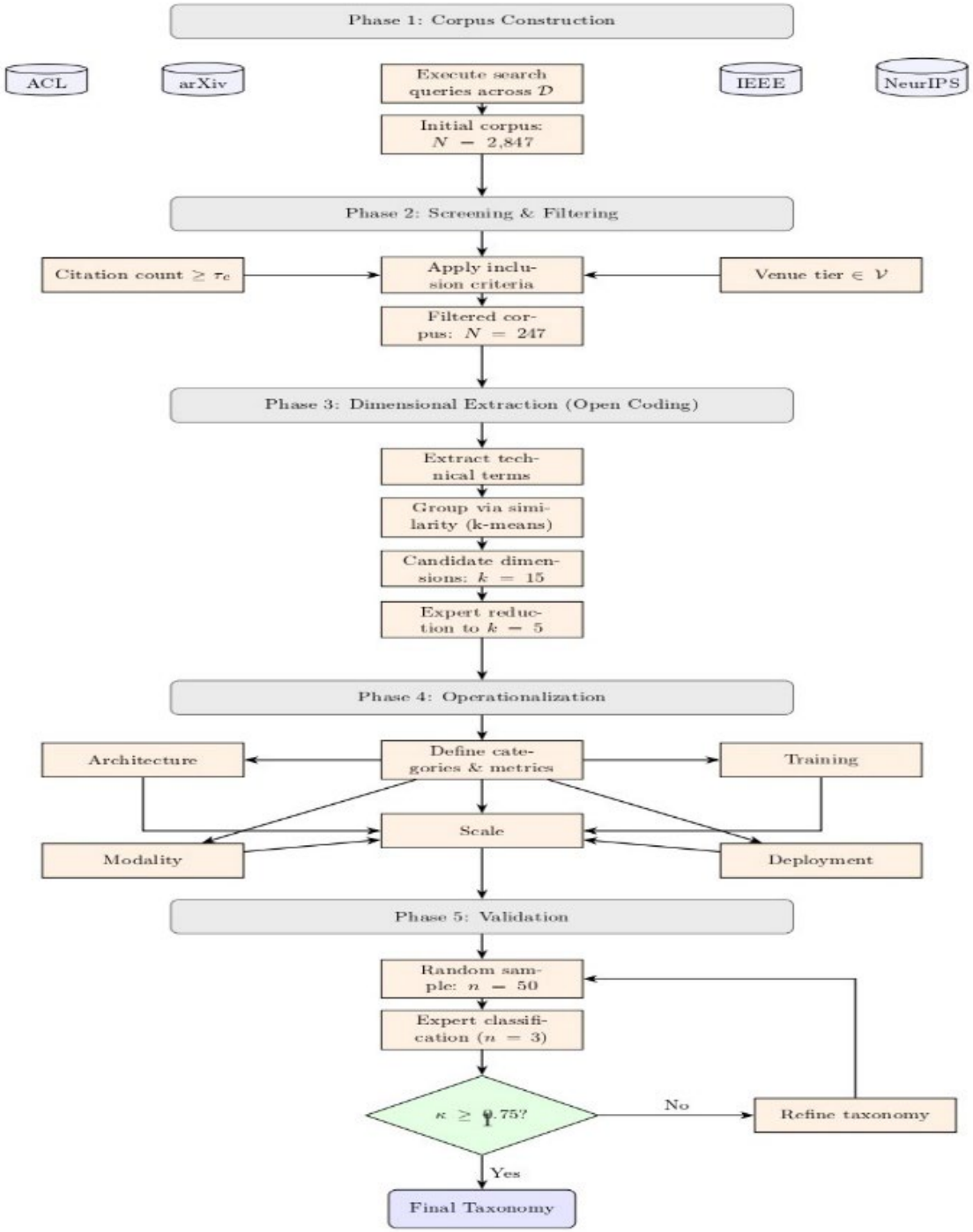

Fig.1 Proposed Taxonomy Architecture

Algorithm 1: Systematic Taxonomy Construction

Input: Search query Q, databases $\mathcal{D}$, time window T, citation threshold τ_c, venue list $\mathcal{V}$
Output: Taxonomy $\mathcal{T}$ with dimensions $\mathcal{D}_1 \dots \mathcal{D}_5$

Function *ConstructTaxonomy*(Q, $\mathcal{D}$, T, τ_c, $\mathcal{V}$):
  // Phase 1: Corpus Collection
  $\mathcal{P} \leftarrow \emptyset$;
  foreach $d \in \mathcal{D}$ **do**
    $results \leftarrow \textsc{Search}(d, Q, T)$;
    $\mathcal{P} \leftarrow \mathcal{P} \cup results$;
  end
  // Phase 2: Screening & Filtering
  $\mathcal{F} \leftarrow \emptyset$;
  foreach $p \in \mathcal{P}$ **do**
    if $p.citations \geq \tau_c$ ***and*** $p.venue \in \mathcal{V}$ **then**
      $\mathcal{F} \leftarrow \mathcal{F} \cup \{p\}$;
    end
  end
  // Phase 3: Dimensional Extraction (Open Coding)
  $\mathcal{T}|\nabla\Updownarrow\int \leftarrow \emptyset$;
  foreach $p \in \mathcal{F}$ **do**
    $terms \leftarrow \textsc{ExtractTechnicalTerms}(p)$;
    $\mathcal{T}|\nabla\Updownarrow\int \leftarrow \mathcal{T}|\nabla\Updownarrow\int \cup terms$;
  end
  $\mathcal{C} \leftarrow \textsc{GroupBySimilarity}(\mathcal{T}|\nabla\Updownarrow\int, k = 15)$;
  $\mathcal{D}\rangle\Updownarrow|\backslash\int\rangle\wr\backslash\int \leftarrow \textsc{ReduceToK}(\mathcal{C}, k = 5)$;
  // Phase 4: Operationalization
  foreach $dim \in \mathcal{D}\rangle\Updownarrow|\backslash\int\rangle\wr\backslash\int$ **do**
    $dim.categories \leftarrow \textsc{DefineCategories}(dim)$;
    $dim.metrics \leftarrow \textsc{DefineMetrics}(dim)$;
  end
  // Phase 5: Validation
  $\mathcal{S} \leftarrow \textsc{RandomSample}(\mathcal{F}, n = 50)$;
  $\mathcal{C}\ddagger\dashv\int\int\rangle\{\}\rfloor\dashv\sqcup\rangle\wr\backslash\int \leftarrow \emptyset$;
  foreach $c \in \mathcal{E}$ **do**
    $\mathcal{C}\ddagger\dashv\int\int\rangle\{\}\rfloor\dashv\sqcup\rangle\wr\backslash\int \leftarrow \mathcal{C}\ddagger\dashv\int\int\rangle\{\}\rfloor\dashv\sqcup\rangle\wr\backslash\int \cup \{\textsc{Classify}(c, \mathcal{S})\}$;
  end
  $\kappa \leftarrow \textsc{FleissKappa}(\mathcal{C}\ddagger\dashv\int\int\rangle\{\}\rfloor\dashv\sqcup\rangle\wr\backslash\int)$;
  if $\kappa < 0.75$ **then**
    $\mathcal{D}\rangle\Updownarrow|\backslash\int\rangle\wr\backslash\int \leftarrow \textsc{RefineDimensions}(\mathcal{D}\rangle\Updownarrow|\backslash\int\rangle\wr\backslash\int, \kappa, \mathcal{E}.feedback)$;
  end
  $\mathcal{T} \leftarrow \textsc{BuildTaxonomy}(\mathcal{D}\rangle\Updownarrow|\backslash\int\rangle\wr\backslash\int)$;
  return $\mathcal{T}$;
end

Fig.2 Algorithm Systematic Taxonomy Construction

D. Validation

Taxonomy was justified by expertise assessment. An independent sample containing 50 papers was independently rated by three domain experts and inter-rater agreement was assessed with Fleiss k. The findings showed significant to close agreement (k=0.86) which validates the reliability of the taxonomy. The differences were solved by consensus and the definitions were also narrowed down.

IV. RESULTS AND ANALYSIS

A. Evolution and Trend Analysis

The development of generative AI from 2014 to 2025 reflects a trajectory defined by architectural innovation, scaling dynamics, benchmark performance, and adoption patterns. This section synthesizes these trends, highlighting the milestones that shaped the field and the structural forces driving its rapid transformation.

1. Timeline of Major Milestones (2014–2025): The development of generative AI can be followed in different stages. In 2014, Generative Adversarial Networks (GANs) were published, the first step in realistic data generation, and the field of image generation was dominated by them during the next few years. Diffusion models became a more stable and higher-fidelity alternative and have been fast supplanted by diffusion models as the state-of-the-art on vision tasks, by 2020. Simultaneously, the emergence of large language models (LLMs) was enabled by the introduction of the Transformer architecture (2017) and models like GPT-3 (2020) and PaLM (2022) showed scaling-based emergent behaviour. Generative modelling was extended to text, vision, and audio by the introduction of multimodal systems like DALL·E 2, Stable Diffusion, Gemini, and GPT-4V, and further advances featured such things as instruction tuning and Reinforcement Learning with Human Feedback (2022 onward), making LLMs extremely useful as conversational agents. By 2025, the field has reached a convergence phase with transformer based multimodal models and diffusion transformer hybrids becoming the dominant models.

Takeaway: The field has evolved from adversarial and latent-variable methods toward transformer- and diffusion-based multimodal systems, with alignment and instruction tuning marking the most recent paradigm shift.

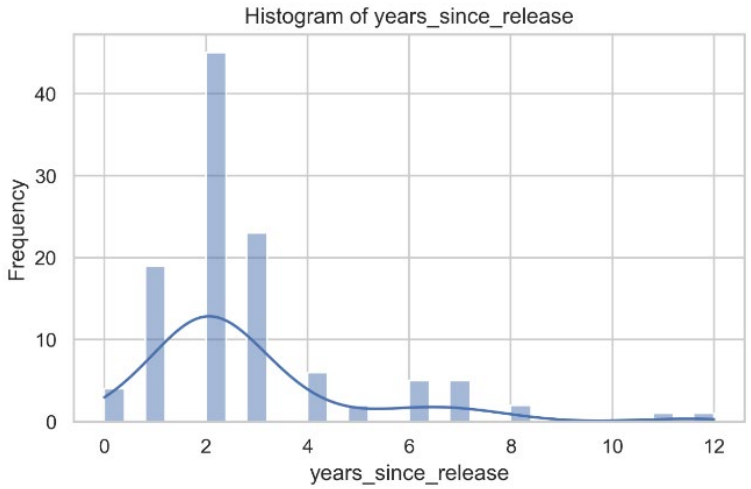

Fig.3 Distribution of AI Model Age (Years Since Release)

2. Model Scaling Trends: The key agent of performance increase in generative AI has been scaling. Count of leading models has increased to hundreds of billion in frontier models like GPT-4 and Gemini, up to billions in transformers during early phases of LSTMs and GANs and millions in early transformers. Equally, the size of datasets has increased compared to curated corpora in the gigabyte scale to gigantic web-scale datasets in the trillions of tokens. Computer training has grown exponentially, with the developments of the GPU/TPU hardware and massive distributed training. Empirical scaling laws postulate that loss is a predictable power law dependency on parameters, data and compute. But efficiency-oriented innovations have developed by 2024-2025, including parameter-efficient fine-tuning, sparse attention mechanisms, quantization, and distillation have been proposed to help reduce the unsustainable increase in compute and energy demands.

Takeaway: Scaling remains the engine of capability growth, but efficiency innovations are increasingly necessary to balance performance with sustainability.

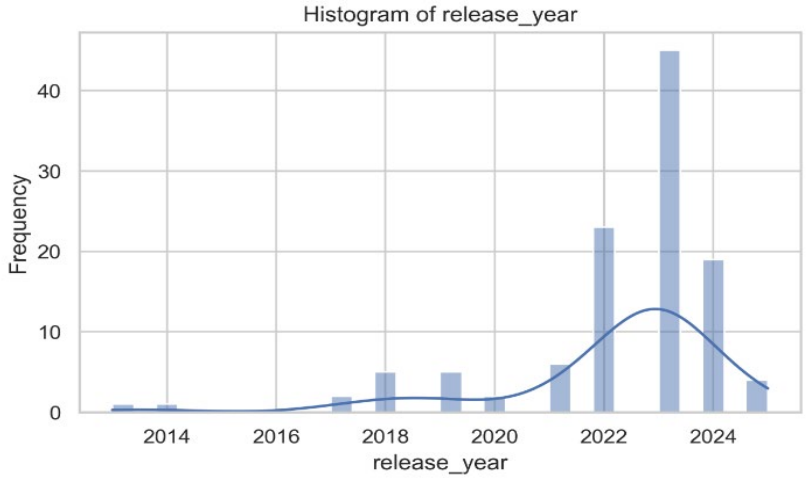

Fig.4 Histogram of AI Model Release Years

3. Performance Trends: Long term benchmark results indicate the trend of performance of generative models. BLEU and ROUGE scores were steadily increasing until being outperformed by more general benchmarks like GLUE, SuperGLUE, MMLU and BIG-bench which found emergent reasoning capabilities in large LLMs. Inception Score (IS) and Frechet Inception Distance (FID) showed a radical decrease in the development of diffusion models to generate images. Assessment of quality, safety and creativity of conversations has shifted to human judgments that are progressively becoming part of large-scale preference studies.

Lesson: Scaling improvements and methodological improvements mean that the benchmarks themselves have been extended to new dimensions of performance, including reasoning, alignment as well as multimodality.

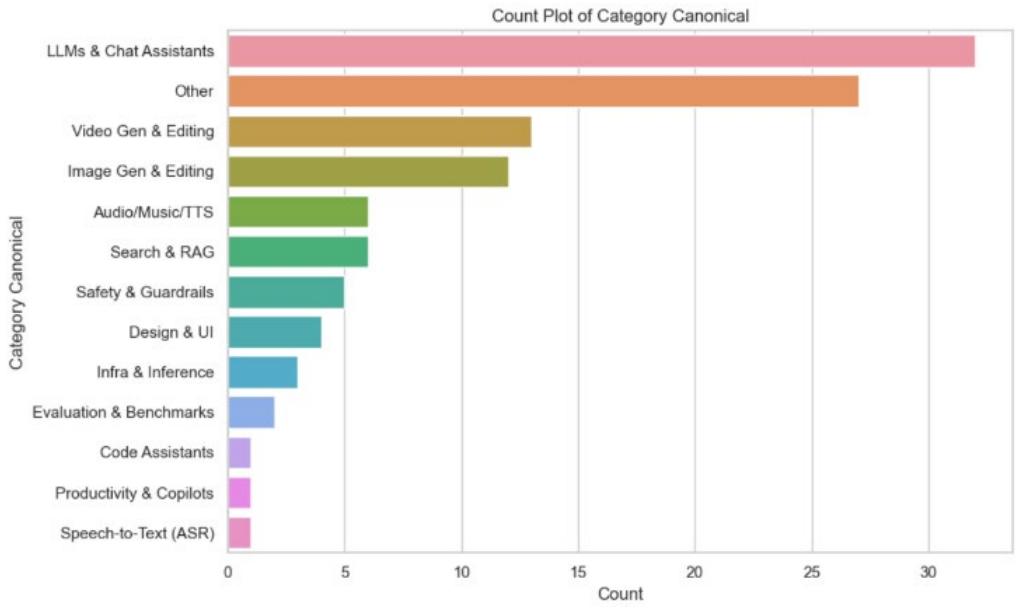

Fig.5 Frequency of AI Use Cases by Canonical Category

4.Adoption Trends: Adoption patterns portray the dynamic between scholarly studies and its application in industry. Academic communities were primarily responsible for the creation of GANs and VAEs, whereas diffusion models became more popular due to open-source releases (e.g., Stable Diffusion) that further accelerated their usage. Such proprietary deployments, to which the GPT family of OpenAI and Claude of Anthropic belong, have dominated LLMs and open-source alternatives (e.g., LLaMA, Falcon, Mistral) have been announced as demand is forced to offer transparency and reproducibility. This binary has demonstrated a rising conflict: as the diffusion models went open source and made the world of creativity more democratic, the proprietary versions of the LLM API made access and control centralized. Meanwhile, industry has replaced academia as frontier models development leader because the necessary compute and data resources are beyond the ability of most research institutions. Implication: The trends of adoption show a dichotomous open-source innovation prevails in the vision, and proprietary hegemony characterizes language as a reflection of deeper accessibility, governance, and competition issues.

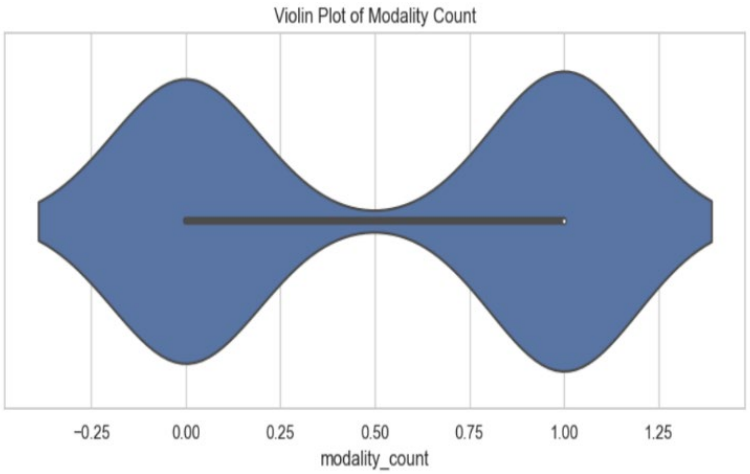

Fig.6 Distribution of AI Model Modality Count

B. Generative AI Products and Market Mapping

The generative AI curve has not stayed within the four walls of the research laboratories, and it has been transformed into commercial products within a short timeframe that dictate the creative industries, business operations, learning, and everyday life. In order to interpret this translation, the study draws a relationship between basic models and applications that can be seen in the market, in which case we consider five types: chat assistants, image generators, video/audio generators, code assistants, and multimodal systems.

1. Categories of Products:

a. *Chat Assistants:* Conversational agents are the most commonly used category of generative AI products. ChatGPT (OpenAI), Claude (Anthropic), and Pi

(Inflection AI) are systems that are driven by instruction-tuned large language models fine-tuned with RLHF. They are used in personal productivity, as well as the education, customer support, and healthcare information services.

b. *Image Generators:* Image generating products like DALL·E 2, Midjourney, and Stable Diffusion have democratized the generation of images so that both professional designers and non-professional users can generate images of high quality. Whereas proprietary systems such as MidJourney focus on the artistic style and closed ecosystems, open-source models such as Stable Diffusion have triggered the emergence of an active developer community.

c. *Video and Audio Generators:* Video generation products like Runway Gen-2 and OpenAI Sora apply the same concept of diffusion to time, which makes it possible to generate realistic motion. Products that involve audio generation, such as the Jukebox by OpenAI and voice cloning services by ElevenLabs, are based on either an autoregressive or diffusion architecture that generates human-like voice and music.

d. *Code Assistants:* GitHub Copilot (with OpenAI Codex) and Replit Ghostwriter as well as AlphaCode are examples of software development that use LLM. The products have made the coding work faster and enhanced the work of the developer, as well as they are used to attract an educational support, yet issues of accuracy, reliability, and intellectual property are still present.

e. *Multimodal Systems:* The latest generation of products combines more than one modality. Google's Gemini is an example of GPT-4V, cross-modal reasoning, including image interpretation, text generation, and multimodal dialogue support, along with Anthropic's Claude 3 Opus. These systems bring out the overlap between language, vision, and audio at the foundation level models.

C. Product-to-Model Mapping

A systematic mapping illustrates how research innovations translate into commercial offerings:

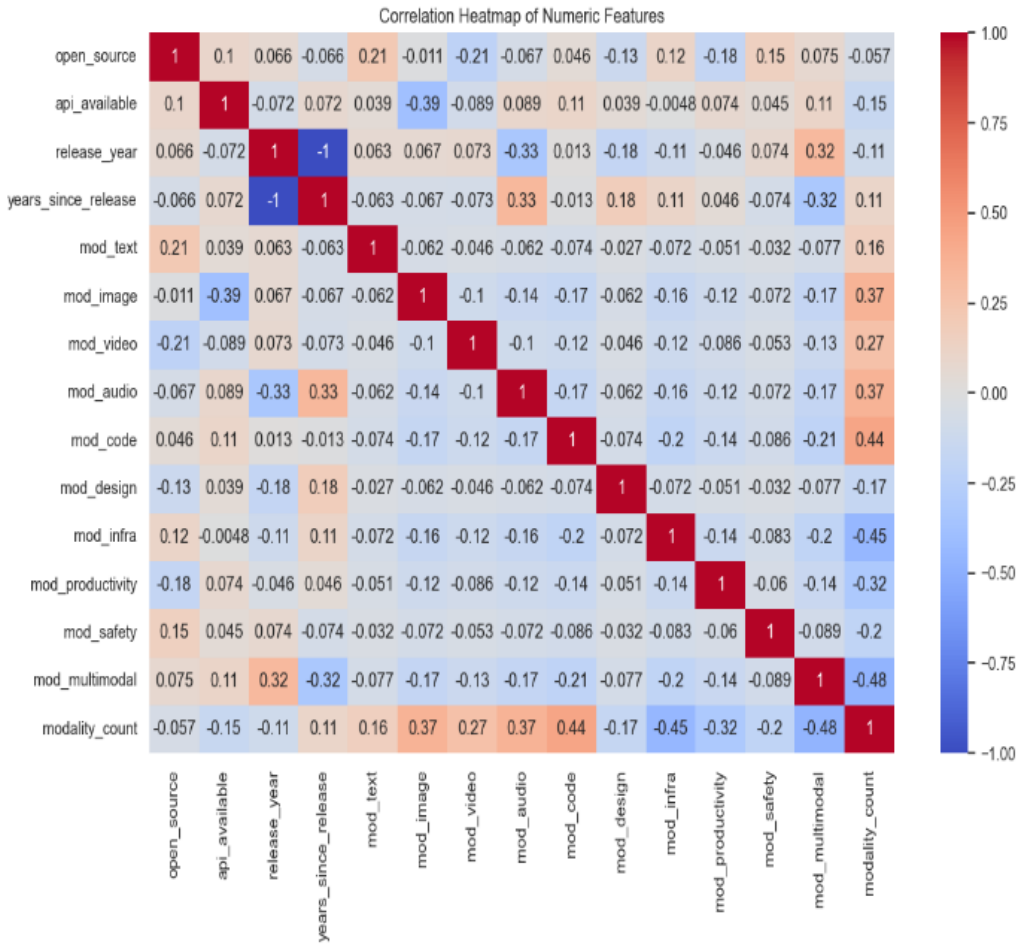

Fig.7 AI Model Feature Correlation Heatmap

TABLE I OVERVIEW OF GENERATIVE AI PRODUCTS, MODALITIES, AND TECHNICAL FOUNDATIONS

Product/ Tool	Modality/ Output	Special Features/ Notes	Likely Underlying Model/ Tech	Open/ Proprietary/ Mixed
ChatGPT	Text, chat, image (with Vision)	Multimodal inputs, broad domain; strong instruction-tuning and RLHF etc.	GPT-4 / GPT-4V etc.	Proprietary API
Claude (Anthropic)	Text, long-document processing etc.	Large context windows; safety-oriented, preference or constitutional AI	Claude models (v2, v3…)	Proprietary API
Pi (Inflection AI)	Text / chat	Focused on human alignment	Inflection-1 etc.	Proprietary
LLaMA / LLaMA 3 Chat	Text / chat (open weight)	SFT + preference tuning	LLaMA family	Open source (weights)
Stable Diffusion	Image generation	Latent diffusion; many derivatives and style-variants	Stability models	Open source
MidJourney	Image generation	Strong style, artist-like outputs	Proprietary diffusion	Proprietary
DALL·E 3	Image generation	Instruction tuned, complex prompt handling	OpenAI diffusion / hybrid model	Proprietary API
Adobe Firefly	Text-to-image, Text-to-video etc.	“Commercially safe” content, integrated into Adobe tools	Adobe proprietary diffusion + training	Proprietary
Runway Gen-2	Video & multimodal	Video generation from text or image + editing tools	Diffusion / hybrid	Proprietary
Sora (OpenAI)	Video	Combines diffusion + autoregressive mix	OpenAI video-capable models	Proprietary API
Pika Labs	Video	Fine-tuned for short video generation	Proprietary diffusion	Proprietary
ElevenLabs	Audio / Speech (TTS etc.)	High voice quality, voice cloning etc.	Transformer / autoregressive audio models	Proprietary API
Suno AI	Music / Audio	Text-to-music generation; creative audio	Diffusion / autoregressive audio	Proprietary
GitHub Copilot	Code / Programming assistance	Completion, suggestions, integration in IDE	Codex / GPT-3.5 base etc.	Proprietary API
Replit Ghostwriter	Code / Dev tools	Code assistance within dev environment	Code-specialized LLMs	Proprietary
AlphaCode (DeepMind)	Code generation / problem solving	Research prototype; challenges and competition style tasks	Transformer-based code LLM	Research / Proprietary
StarCoder	Code LLM	Open-source code generation + dev community use	StarCoder-15B etc.	Open source
Gemini (Google)	Multimodal (text + vision + etc.)	Integrated with Google ecosystem; multiple variants (workspace, assistant etc.)	Gemini family	Proprietary
GPT-4V	Multimodal (vision + text)	Can see images, interpret visual content	Extension of GPT-4	Proprietary API
Claude 3 Opus	Multimodal	Vision extension + text, etc.	Claude multmodal version	Proprietary API
Kosmos-2	Multimodal (vision-language)	Research prototype	Microsoft model	Research / Proprietary
Jasper	Text / Marketing content	Templates, SEO tools; specialized features for brand voice etc.	Likely fine-tuned LLMs	Proprietary
Notion AI	Text / Notes / Workspace	Tools for brainstorming, summarization etc.	LLMs + Prompt engineering	Proprietary
Anyword	Text / Marketing content	Predictive copywriting, A/B testing etc.	LLM specialized for marketing	Proprietary
Shortwave	Email management + text generation	Helps manage writing and responses etc.	LLM text generation	Proprietary
Mem	Note-taking / knowledge management	Search, organize, generate notes etc.	LLMs	Proprietary / Mixed
Beautiful.ai	Presentation design automation	Auto layout, design based on content	Likely diffusion or generative layout models	Proprietary

Pitch (tool)	Sales / Slide decks	Automated design, maybe generative templates	Hybrid models	Proprietary
Wix AI / Framer etc.	Web / UI generation	Helps build websites, design etc.	Likely hybrid / diffusion + layout models	Proprietary
Wondershare Filmora	Video editing + generative effects	Combines traditional editing + generative features	Proprietary	Proprietary
ImagineArt (Vyro.ai)	Images, Video, Audio, Voice	Creative suite bundling many modalities under one app	Mix of in-house + third-party models	Proprietary
Google AI Studio	IDE / Prototyping + multimodal tools	Access to Gemini, APIs, free tiers	Google's Gemini etc.	Proprietary with free tiers
Adobe Firefly Boards etc.	Image / Video / Creative collaboration	Whiteboard / collaborative creative workflows	Firefly + partner models	Proprietary
Qodo (formerly Codium)	Code generation / integrity checking	Focus on code quality as well as generation	Code LLMs + static analysis etc.	Proprietary
Tencent Hunyuan3D-2.0	3D generation	Open-source models for text/image→3D visuals; "turbo" versions for faster output	3D diffusion / generative models	Open source (for Tencent's release)

D. From Research to Adoption

Generative AI research to product translation has an easily traceable path. The breakthroughs in the field of GANs in image synthesis, diffusion in high-fidelity generative modelling, transformers in language and RLHF in alignment quickly find their way into commercial systems. The open-source models are faster to adopt as they allow the community to experiment and customize, whereas proprietary systems are more used with large, enterprise deployment as a result of the resource needs and management of intellectual property. Another important move is the separation of modalities: image generation was democratized by open-source releases, whereas language-based systems are mostly proprietary since there is a risk of misuse, safety, and competitive advantage. The multimodal integration is also an indication of the convergence of the architectures surrounding the transformer-based systems and this is an indication of meeting the research ends with the market needs.

Lesson: The market mapping indicates the product types are supported by architectural and training innovations, whereas the deployment options create accessibility, innovation directions and competitive forces.

E. Case Studies

This segment is a detailed discussion of some of the latest state-of-the-art AI systems, their architecture, training approaches, performance aspects, deployment models and their effects to the market. Examples Diffusion models Stable Diffusion Stable Diffusion is a Latent Diffusion Model (LDM)-based model, built on a U-Net backbone and attention to efficiently generate high-quality images in a condensed latent space. These models are conditioned on large-scale datasets with high optimization methods and their performance is usually measured in terms of Frechet Inception Distance (FID) and Inception Score (IS). The access strategies and integration into creative tools are frequently based on API and implemented in the area of deployment contributing to its popularity in open-source and commercial sectors [51].

GPT-4 and LLaMA-3 are examples of instruction-tuned large language models, which are a significant breakthrough in natural language processing. GPT-4 is a large multimodal model trained by the OpenAI that can be fed with image and text inputs to generate highly coherent texts as outputs. It performs at the human level on the professional and academic test, such as passing a simulated bar exam in the top 10 percent of test takers [40]. The LLaMA-3 models offered by Meta in various parameter sizes such as 8B to 70B are efficient and versatile. They use such innovations like rotary positional embeddings and grouped multi-query attention to improve performance. These models are trained on high-performance infrastructure and optimized to various tasks, with the state-of-the-art results [23].

Multimodal systems, like Gemini, Claude 3.5 Sonnet and Sora multimodal systems combine several input modalities like text, images and in a few cases videos to allow complex reasoning and content generation. The Gemini models offered by Google DeepMind, scaled to Ultra, Pro and Nano, are a high-quality benchmark with a score of approximately 18.8% on the Humanity Last Exam, and a pass rate of above 84% on GPQA. Google Cloud and Workspace applications allow integrating all tasks to create a smooth process of writing documents and analysing data and automating tasks [19]. The Claude 3.5 Sonnet by Anthropic also leads the way in terms of both intelligence and coding, as well as reasoning and vision tasks, which are also improved and cost-effective and can be considered effective in a broad spectrum of applications [3]. The Sora text-to-video generative Artificial Intelligence of OpenAI creates videos of realistic or imaginative scenes based on text instructions. It is shown to bring a high potential of simulating the physical world and improving the creative workflow, which is quite a progressive advancement in video generation technologies [39].

Combined with each other, diffusion models, instruction-tuned large language models, and multimodal systems exemplify the fast development and decentralization of technologies based on artificial intelligence. Although a

diffusion model, such as Stable Diffusion, has been used to make significant advancements in the generation of creative and visual content, instruction-tuned models, like GPT-4 and LLaMA-3, have been developed to better generate natural language understanding and reasoning in the workplace and academia. Multimodal systems such as Gemini, Claude 3.5 Sonnet, and Sora can be considered a meeting of modalities, and allow AI to process and generate complex information in text, images, and video. Not only do each category reflect technical innovations in the field of architecture and training, but they also reflect different deployment approaches and market forces, such as open-source use and enterprise integration, content creation and next-generation AI services. All these systems contribute to the importance of the impending potential of AI to transform various industries and the present-day challenges in scaling, alignment, and ethical application [23][40][23][19][3][39].

F. Benchmarking, Gaps, and Challenges

Although this has been adopted at a very high rate, there are critical limitations. Current benchmarks might not be able to reflect hallucinations, biases, explainability problems, and other real-world performance fits. There are loopholes in the transformation of research into safe and regulated products, especially in copyright, intellectual property and standards.

TABLE II BENCHMARKS, GAPS, AND CHALLENGES OF GENERATIVE AI MODELS AND PRODUCTS

Product / Model	Benchmark / Evaluation	Key Strengths	Gaps / Challenges	Notes
ChatGPT (OpenAI)	HumanEval, MMLU, user studies	Multimodal, strong instruction-tuning, broad domain	Hallucination, bias, proprietary, expensive API	Text, chat, image (vision)
Claude (Anthropic)	Internal QA benchmarks, coding tasks	Large context windows, safety-oriented	Proprietary, limited research access	Text, long-document processing
Pi (Inflection AI)	Human alignment tests	Human-alignment focused	Proprietary, limited deployment	Chat/text only
LLaMA / LLaMA 3 Chat	MMLU, reasoning benchmarks	Open-weight, flexible, efficient	Requires high-performance infrastructure	Text/chat, open-source weights
Stable Diffusion	FID, Inception Score	High-quality image gen, open-source, flexible	Style consistency issues, hallucinations in complex scenes	Image generation
MidJourney	User evaluation, aesthetic quality	Strong artistic style	Closed ecosystem, limited technical transparency	Image generation
DALL·E 3	FID, human evaluation	Instruction-tuned, complex prompt handling	Proprietary, copyright/safety concerns	Image generation
Adobe Firefly	Human evaluation	Commercially safe content, integrated with Adobe tools	Proprietary, limited community experimentation	Text-to-image/video
Runway Gen-2	Video quality metrics, human evaluation	Text-to-video + editing tools	Proprietary, limited real-world testing	Video/multimodal
Sora (OpenAI)	Video realism, human evaluation	Text-to-video, creative workflows	Temporal coherence issues, expensive computation	Video
Pika Labs	Human eval, video metrics	Short video generation	Proprietary, limited feature set	Video
ElevenLabs	MOS (Mean Opinion Score), speech clarity	High-quality voice, voice cloning	Proprietary, limited customizability	Audio/TTS
Suno AI	User studies, audio quality	Text-to-music generation	Proprietary, limited dataset transparency	Music/audio
GitHub Copilot	Code correctness, functional tests	Productivity, code completion	Accuracy issues, IP concerns	Code/IDE integration
Replit Ghostwriter	Code correctness, dev tests	Integrated into IDE, code assistance	Proprietary, dependent on LLM updates	Code
AlphaCode (DeepMind)	Programming contest benchmarks	Code reasoning, problem-solving	Research prototype, limited adoption	Code generation
StarCoder	Code correctness tests	Open-source, community support	Hardware/resource requirements	Code LLM
Gemini (Google)	GPQA, Humanity's Last Exam	Multimodal reasoning, integration with Google tools	Proprietary, limited open research	Text + vision + other modalities
GPT-4V	MMLU + vision benchmarks	Vision + text, interprets images	Proprietary API, cost	Multimodal

Claude 3 Opus	Internal QA, vision+text tasks	Vision extension + reasoning	Proprietary API	Multimodal
Kosmos-2	Vision-language benchmarks	Research prototype	Proprietary/research only	Multimodal
Jasper	Human evaluation, marketing content success	Templates, brand-focused	Proprietary, limited transparency	Text/marketing
Notion AI	User studies, productivity metrics	Brainstorming, summarization	Proprietary	Text/notes/workspace
Anyword	Marketing metrics, A/B testing	Predictive copywriting	Proprietary	Text/marketing
Shortwave	Email efficiency, user studies	Email management + generation	Proprietary	Text/email
Mem	Knowledge management metrics	Note-taking, search, generate	Proprietary/mixed	Text/notes
Beautiful.ai	Presentation quality metrics	Automated design	Proprietary	Presentation design
Pitch	Productivity/user studies	Sales deck automation	Proprietary	Slide decks
Wix AI / Framer	User testing, UI metrics	Website/UI generation	Proprietary	Web/UI generation
Wondershare Filmora	Video quality metrics	Editing + generative effects	Proprietary	Video
ImagineArt (Vyro.ai)	User evaluation	Creative suite, multiple modalities	Proprietary	Images, video, audio, voice
Google AI Studio	Productivity, prototyping metrics	Access to Gemini, APIs, free tiers	Proprietary	IDE/prototyping multimodal
Adobe Firefly Boards	Collaboration metrics	Creative collaboration	Proprietary	Image/video/creative collaboration
Qodo (Codium)	Code correctness	Code generation + integrity checking	Proprietary	Code
Tencent Hunyuan3D-2.0	3D rendering metrics	Text/image → 3D, fast output	Open-source but limited documentation	3D generation

V. DISCUSSION

The paper introduces a single taxonomy of generative AI models and products, starting with the development of diffusion-based models and moving on to instruction-tuned large language models. The results indicate an industry with a convergence of architecture, an increasing scaling as well as an increasing dichotomy of open and proprietary ecosystem. This part is a synthesis of these findings, implications, and limitations, and future directions.

A. Synthesis of Key Findings

We have identified three trends that are dominant. To begin with, the field is being remodelled by architectural convergence: transformer-based architecture is now used to solve language tasks, and diffusion-based architecture is used to generate visual synthesis. Nevertheless, hybrid systems like diffusion transformers, and autoregressive-diffusion hybrids are an indication of a path toward combined multimodal systems that will capture the best of both paradigms.

Second, instruction tuning and alignment is another paradigm shift, which is independent in importance to the invention of transformers or diffusion models. Even competent raw pretrained models are not immediately usable and controllable. These models were transformed into deployable interactivity agents with the development of Reinforcement Learning from Human Feedback (RLHF), Direct Preference Optimization (DPO) and instruction tuning. This change characterizes a third generative AI development wave: generative capability (wave one), scale (wave two), and alignment and controllability (wave three).

Third, there is one stark contrast between open-source and proprietary ecosystems. Open-source releases (such as Stable Diffusion) have made diffusion models very popular with the community, allowing them to innovate. On the contrary, the best large language models are proprietary and accessible through an API. This divergence indicates the variation in risk profile language models have more risks of misuse and it has serious consequences to the accessibility, reproducibility, and governance.

B. Implications

To a researcher, the intersection of architectures demands cross-paradigmatic activities bridging the worlds of vision and language. Alignment research needs to be more methodologically rigorous, i.e. it should have standardized reports on training methods. Assessment systems should no longer be focused on limited measures such as BLEU and FID but comprehensive methods to address safety, fairness, and strength. To a practitioner, there is no decision to make but a significant trade-off between flexibility and transparency and convenience and scalability in the issues of open-source or

proprietary deployment. Capability is just as important as usability, instruction tuning and preference optimization must be considered high priority in development pipelines. The efficiency-oriented methods include quantization, distillation, parameter-efficient fine-tuning, which makes it possible to operate under resource-constrained settings.

To policy makers, there should be open-source versus proprietary dichotomy that is context sensitive to governance. The practice of democratized innovation is only possible through open-source models that need to be responsibly released. Proprietary APIs are very concentrated in terms of power and risk and they require transparency and accountability. Embedded products create a concern of user understanding and acceptance. Sustainability of the environment also needs to be considered because computing resources needed to run frontier models have lots of carbon footprint.

C. Limitations

There are a number of limitations that should be mentioned. The taxonomy is based on the state of generative AI in January 2022 to March 2024; the new architectures might not be represented exhaustively. The selection of the corpus could be biased to industrial research and non-English works. Although validation attained high inter-rater agreement. Although validation was associated with high inter-rater agreement (k = 0.86), dimensional extraction was done subjectively and other organizational schemes can be done. These constraints imply that the taxonomy must be considered as a dynamic structure that needs to be updated on a regular basis.

D. Future Directions

The frontier of architecture should be advanced through diffusion transformers and state-space models being developed as an efficient alternative to transformers, and neuro-symbolic hybrids that make use of generative models to provide explicit reasoning. The alignment research should go beyond post-instruction alignment to appreciation alignment, interactive learning through continuous feedback, and balancing the preferences of various stakeholders. The high compression methods required by efficiency and sustainability, the standardized Green AI metrics of carbon footprint reporting, and hardware-sensitive co-design are all extreme. The next horizon is Multimodal and embodied systems, where world models have to learn physical dynamics, using embodied instruction following robotics, and cross-modal transfer learning. Evaluation frameworks must have holistic metrics that measure capability, alignment, efficiency and robustness as one; dynamic evaluation to avoid overfitting; and human-AI joint evaluation.

VI. CONCLUSION

It has presented an in-depth taxonomy and trend analysis of generative AI, followed by tracing its development since the early days of diffusion models to instruction-tuned large language models and presenting a mapping of innovations in research to commercial offerings. The results indicate a discipline with convergence of architecture, scaling forces, and an expanding dichotomy among open and proprietary ecosystems. The revolution of alignment has changed raw generative ability into deployable, controllable systems and efficiency-oriented innovations are coming up to handle the scalability sustainability of challenges. Generative AI is changing rapidly more than ever, and this is due to the combination of research innovation, commercial implementation, and social influence. The presented taxonomy and analyses can serve as a source of information to both researchers, practitioners, policymakers and educators that want to know not just where generative AI is today, but where it is going. A way ahead includes interdisciplinary and intersectoral cooperation - cutting across the gulfs between vision and language study, between open and proprietary development, between technical competence and ethical conformity. When used in a cautious manner, generative AI can be used to enhance the creativity of people, democratize knowledge, and solve complex problems in science, industry, and society. This task of fulfilling that vow, but reducing the dangers, is yet one of the characteristics of our generation.

Declaration of Conflicting Interests
The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The authors received no financial support for the research, authorship, and/or publication of this article.

Use of Artificial Intelligence (AI)-Assisted Technology for Manuscript Preparation
The authors confirm that no AI-assisted technologies were used in the preparation or writing of the manuscript, and no images were altered using AI.

REFERENCES

[1] J. B. Alayrac, J. Donahue, P. Luc, A. Miech, I. Barr, Y. Hasson, A. M. Gordon, Y. M. Song, S. Sukthankar, C. Schmid, and K. Simonyan, "Flamingo: A visual language model for few-shot learning," *Advances in Neural Information Processing Systems*, vol. 35, pp. 23716–23736, 2022.

[2] K. Li, Y. He, Y. Wang, Y. Li, W. Wang, P. Luo, X. Wang, Z. Li, and Y. Qiao, "VideoChat: Chat-centric video understanding," *Science China Information Sciences*, vol. 68, no. 10, p. 200102, 2025.

[3] Anthropic, "Introducing Claude 3.5 Sonnet," Anthropic, 2024.

[4] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the 34th International Conference on Machine Learning (ICML)*, PMLR, 2017, pp. 214–223.

[5] Y. Bai, A. Jones, K. Ndousse, A. Askell, A. Chen, N. DasSarma, *et al.*, "Training a helpful and harmless assistant with reinforcement learning from human feedback," *arXiv preprint arXiv:2204.05862*, 2022.

[6] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.

[7] M. Pawelczyk, S. Neel, and H. Lakkaraju, "In-context unlearning: Language models as few-shot unlearners," *arXiv preprint arXiv:2310.07579*, 2023.

[8] V. Capraro, A. Lentsch, D. Acemoglu, S. Akgun, A. Akhmedova, E. Bilancini, *et al.*, "The impact of generative artificial intelligence on socioeconomic inequalities and policy making," *PNAS Nexus*, vol. 3, no. 6, p. pgae191, 2024.

[9] M. Chen, J. Tworek, H. Jun, Q. Yuan, H. P. D. O. Pinto, J. Kaplan, *et al.*, "Evaluating large language models trained on code," *arXiv preprint arXiv:2107.03374*, 2021.

[10] X. Chen, H. Fang, T.-Y. Lin, R. Vedantam, S. Gupta, P. Dollár, and C. L. Zitnick, "Microsoft COCO captions: Data collection and evaluation server," *arXiv preprint arXiv:1504.00325*, 2015.

[11] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, et al., "PaLM: Scaling language modeling with Pathways," Journal of Machine Learning Research, vol. 24, no. 240, pp. 1–113, 2023.

[12] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei, "Deep reinforcement learning from human preferences," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[13] H. W. Chung, L. Hou, S. Longpre, B. Zoph, Y. Tay, W. Fedus, *et al.*, "Scaling instruction-finetuned language models," *Journal of Machine Learning Research*, vol. 25, no. 70, pp. 1–53, 2024.

[14] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018.

[15] M. Conover, M. Hayes, A. Mathur, X. Meng, J. Xie, J. Wan, A. Ghodsi, P. Wendell, and M. Zaharia, "Hello Dolly: Democratizing the magic of ChatGPT with open models," *Databricks Blog*, Mar. 24, 2023.

[16] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2019, pp. 4171–4186.

[17] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," Advances in Neural Information Processing Systems, vol. 34, pp. 8780–8794, 2021.

[18] Y. Du and I. Mordatch, "Implicit generation and modeling with energy-based models," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[19] G. Team, R. Anil, S. Borgeaud, J.-B. Alayrac, J. Yu, R. Soricut, *et al.*, "Gemini: A family of highly capable multimodal models," *arXiv preprint arXiv:2312.11805*, 2023.

[20] R. Gao, Y. Song, B. Poole, Y. N. Wu, and D. P. Kingma, "Learning energy-based models by diffusion recovery likelihood," *arXiv preprint arXiv:2012.08125*, 2020.

[21] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, *et al.*, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[22] Y. Goyal, T. Khot, D. Summers-Stay, D. Batra, and D. Parikh, "Making the V in VQA matter: Elevating the role of image understanding in visual question answering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[23] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, *et al.*, "The LLaMA 3 herd of models," *arXiv preprint arXiv:2407.21783*, 2024.

[24] D. Hendrycks, C. Burns, S. Basart, A. Zou, M. Mazeika, D. Song, and J. Steinhardt, "Measuring massive multitask language understanding," *arXiv preprint arXiv:2009.03300*, 2020.

[25] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[26] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

[27] Z. Shi, Y. Wang, F. Yin, X. Chen, K. W. Chang, and C. J. Hsieh, "Red teaming language model detectors with language models," *Transactions of the Association for Computational Linguistics*, vol. 12, pp. 174–189, 2024.

[28] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, *et al.*, "Scaling laws for neural language models," *arXiv preprint arXiv:2001.08361*, 2020.

[29] T. Kynkäänniemi, T. Karras, S. Laine, J. Lehtinen, and T. Aila, "Improved precision and recall metric for assessing generative models," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[30] P. Sharma, M. Kumar, H. K. Sharma, and S. M. Biju, "Generative adversarial networks (GANs): Introduction, taxonomy, variants, limitations, and applications," *Multimedia Tools and Applications*, vol. 83, no. 41, pp. 88811–88858, 2024.

[31] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," arXiv preprint arXiv:1312.6114, 2013.

[32] Z. Kong, W. Ping, J. Huang, K. Zhao, and B. Catanzaro, "DiffWave: A versatile diffusion model for audio synthesis," arXiv preprint arXiv:2009.09761, 2020.

[33] Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, and F. Huang, "A tutorial on energy-based learning," in *Predicting Structured Data*, 2006.

[34] R. Li, L. B. Allal, Y. Zi, N. Muennighoff, D. Kocetkov, C. Mou, *et al.*, "StarCoder: May the source be with you!," *arXiv preprint arXiv:2305.06161*, 2023.

[35] Y. Li, D. Choi, J. Chung, N. Kushman, J. Schrittwieser, R. Leblond, *et al.*, "Competition-level code generation with AlphaCode," *Science*, vol. 378, no. 6624, pp. 1092–1097, 2022.

[36] P. Liang, R. Bommasani, T. Lee, D. Tsipras, D. Soylu, M. Yasunaga, *et al.*, "Holistic evaluation of language models," *arXiv preprint arXiv:2211.09110*, 2022.

[37] S. K. Bharti and K. S. Babu, "Automatic keyword extraction for text summarization: A survey," *arXiv preprint arXiv:1704.03242*, 2017.

[38] Y. Tian, X. Li, H. Zhang, C. Zhao, B. Li, X. Wang, and F. Y. Wang, "VistaGPT: Generative parallel transformers for vehicles with intelligent systems for transport automation," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 9, pp. 4198–4207, 2023.

[39] Y. Liu, K. Zhang, Y. Li, Z. Yan, C. Gao, R. Chen, *et al.*, "Sora: A review on background, technology, limitations, and opportunities of large vision models," *arXiv preprint arXiv:2402.17177*, 2024.

[40] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, *et al.*, "GPT-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[41] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, *et al.*, "Training language models to follow instructions with human feedback," *Advances in Neural Information Processing Systems*, vol. 35, pp. 27730–27744, 2022.

[42] K. Papineni, S. Roukos, T. Ward, and W. J. Zhu, "BLEU: a method for automatic evaluation of machine translation," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics*, Philadelphia, PA, USA, pp. 311–318, Jul. 2002.

[43] H. Pearce, B. Tan, B. Ahmad, R. Karri, and B. Dolan-Gavitt, "Examining zero-shot vulnerability repair with large language models," in *2023 IEEE Symposium on Security and Privacy (SP)*, San Francisco, CA, USA, pp. 2339–2356, May 2023.

[44] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[45] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI Blog*, 2019.

[46] R. Rafailov, A. Sharma, E. Mitchell, C. D. Manning, S. Ermon, and C. Finn, "Direct preference optimization: Your language model is secretly a reward model," *Advances in Neural Information Processing Systems*, vol. 36, pp. 53728–53741, 2023.

[47] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, vol. 21, no. 140, pp. 1–67, 2020.

[48] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with CLIP latents," *arXiv preprint arXiv:2204.06125*, 2022.

[49] A. Razavi, A. van den Oord, and O. Vinyals, "Generating diverse high-fidelity images with VQ-VAE-2," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[50] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 10684–10695, 2022.

[51] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. L. Denton, *et al.*, "Photorealistic text-to-image diffusion models with deep language understanding," *Advances in Neural Information Processing Systems*, vol. 35, pp. 36479–36494, 2022.

[52] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[53] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv preprint arXiv:2010.02502*, 2020.

[54] A. Srivastava, A. Rastogi, A. Rao, A. A. M. Shoeb, A. Abid, A. Fisch, *et al.*, "Beyond the imitation game: Quantifying and extrapolating the capabilities of language models," *Transactions on Machine Learning Research*, 2023.

[55] Y. Dubois, C. X. Li, R. Taori, T. Zhang, I. Gulrajani, J. Ba, *et al.*, "AlpacaFarm: A simulation framework for methods that learn from human feedback," *Advances in Neural Information Processing Systems*, vol. 36, pp. 30039–30069, 2023.

[56] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, *et al.*, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[57] J. Wei, M. Bosma, V. Y. Zhao, K. Guu, A. W. Yu, B. Lester, et al., "Finetuned language models are zero-shot learners," arXiv preprint arXiv:2109.01652, 2021.

[58] L. Weidinger, J. Mellor, M. Rauh, C. Griffin, J. Uesato, P. S. Huang, *et al.*, "Ethical and social risks of harm from language models," *arXiv preprint arXiv:2112.04359*, 2021.

[59] M. Wessel, M. Adam, A. Benlian, A. Majchrzak, and F. Thies, "Generative AI and its transformative value for digital platforms," *Journal of Management Information Systems*, vol. 42, no. 2, pp. 346–369, 2025.

[60] G. I. Winata, H. Zhao, A. Das, W. Tang, D. D. Yao, S. X. Zhang, and S. Sahu, "Preference tuning with human feedback on language, speech, and vision tasks: A survey," *Journal of Artificial Intelligence Research*, vol. 82, pp. 2595–2661, 2025.

[61] J. K. Wiredu, N. Seidu Abuba, and H. Zakaria, "Impact of generative AI in academic integrity and learning outcomes: A case study in the Upper East Region," *Asian Journal of Research in Computer Science*, vol. 17, no. 8, pp. 10–9734, 2024.

[62] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, *et al.*, "Diffusion models: A comprehensive survey of methods and applications," *ACM Computing Surveys*, vol. 56, no. 4, pp. 1–39, 2023.

[63] U. Mittal, S. Sai, V. Chamola, and D. Sangwan, "A comprehensive review on generative AI for education," *IEEE Access*, vol. 12, pp. 142733–142759, 2024.

[64] N. Cheng, S. Wu, X. Wang, Z. Yin, C. Li, W. Chen, and F. Chen, "AI for UAV-assisted IoT applications: A comprehensive review," IEEE Internet of Things Journal, vol. 10, no. 16, pp. 14438–14461, 2023.

[65] Z. Huang, X. Zhang, Z. Tang, F. Xu, M. Datcu, and J. Han, "Generative artificial intelligence meets synthetic aperture radar: A survey," *IEEE Geoscience and Remote Sensing Magazine*, 2024.