Synthetic Data Generation for Computer Vision: Foundations, Generative Advances, Evaluation Frameworks, and Responsible Deployment

Abstract

Synthetic data has rapidly transitioned from a supplementary tool to a foundational element in computer vision, driven by challenges surrounding data scarcity, privacy, annotation overhead, and the increasing complexity of downstream tasks. This comprehensive survey delineates the evolution, methodologies, and far-reaching implications of synthetic data and generative modeling in vision and beyond. After outlining the motivations—spanning scientific discovery, healthcare, and regulatory compliance—the review systematically covers the theoretical underpinnings and diverse taxonomies of synthetic data generation, encompassing traditional statistical approaches, agent-based simulations, and the transformative rise of deep generative models such as GANs, diffusion models, VAEs, and transformer-based hybrids.

Key contributions include a critical analysis of conditional synthesis, text-to-image and multimodal generation, data augmentation strategies, annotation-efficient learning paradigms, and 3D/multiview synthesis. The survey also presents advances in evaluation, introducing both classical and domain-adapted metrics (e.g., FID, IS, MP-PSNR), as well as context-sensitive benchmarking protocols addressing factuality, fidelity, and fairness. Application domains span computer vision, medical imaging, scientific discovery, 3D simulation, federated and edge learning, and environmental modeling.

Furthermore, the survey synthesizes current methodologies' strengths, limitations, and risks—such as computational demands, domain bias, model memorization, and ethical dilemmas related to privacy, bias propagation, and traceability. It highlights emerging frameworks for responsible and standardized evaluation, as well as guidelines for robust and interpretable deployment, especially in high-stakes contexts. Concluding, the paper identifies critical open challenges—including generalization, adversarial robustness, label efficiency, and scalable evaluation—and offers perspectives on future directions, emphasizing the imperatives of interdisciplinary collaboration, transparent benchmarking, and adaptive, ethically aligned innovation as generative AI continues to reshape both scientific inquiry and societal practice.

ACM Reference Format:
. 2025. Synthetic Data Generation for Computer Vision: Foundations, Generative Advances, Evaluation Frameworks, and Responsible Deployment. In . ACM, New York, NY, USA, 32 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Synthesizing data for machine learning and artificial intelligence (AI) systems has become a critical tool to combat issues such as data scarcity, privacy constraints, annotation costs, and bias. Recent advances have led to a broad spectrum of synthetic data generation paradigms, from classical rule-based algorithms to state-of-the-art generative models. These approaches act as foundational building blocks in developing robust AI systems across domains such as vision, language, and audio.

The domain of synthetic data generation encompasses several core paradigms. For instance, analytical simulation leverages mathematical models to generate synthetic samples, while procedural generation relies on algorithmic rules and stochastic processes to mimic real-world variability. More recently, deep generative models—such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models—have achieved substantial success in generating high-fidelity synthetic data that closely resembles real distributions. Hybrid systems also exist, combining domain-specific knowledge with data-driven techniques to further expand generative capabilities.

Understanding the relationships and workflows among these paradigms is essential for both practitioners and researchers. A taxonomy of generative methods provides a systematic way to survey the field, clarifying distinctions and commonalities among approaches. Methodological pipelines typically involve stages such as data modeling, sampling, and quality evaluation, each with unique challenges depending on the application context.

At the conclusion of this section, it is important to recognize several key takeaways: the landscape of synthetic data generation is diverse, spanning analytical, procedural, and deep learning-based paradigms; the choice of paradigm influences both workflow and applicability; and establishing clear taxonomies and methodological workflows forms the basis for comprehensive analysis in subsequent sections.

This survey proceeds by providing a unified taxonomy, detailed explorations of generative models and their pipelines, and a synthesis of evaluation strategies. These structured overviews aim to equip readers with both foundational knowledge and practical guidance for leveraging synthetic data in contemporary machine learning workflows.

1.1 Motivation for Synthetic Data in Computer Vision

Synthetic data has evolved from a peripheral tool to a foundational resource for advancing computer vision. This transformation is

Conference'17, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

driven by the persistent challenge of acquiring high-quality, annotated datasets: real-world data is frequently scarce, costly, or restricted by privacy and ethical limitations, particularly within sensitive domains such as healthcare and regulated industries [3, 8, 61, 64, 74, 90, 92]. Traditional computer vision systems relied extensively on hand-crafted, manually annotated datasets and basic augmentation techniques—including flipping, rotation, and cropping—creating significant bottlenecks in scalability and generalizability [23, 67, 86, 87]. The advent of deep learning has substantially increased the demand for diverse and richly annotated data, emphasizing the necessity not only for robust object detection but also for enhanced data augmentation and reduced manual annotation burdens [23, 61, 67, 75, 86, 87, 90].

Importantly, the potential of synthetic data transcends that of mere augmentation. Generative models address data scarcity, alleviate class imbalance, and mitigate privacy concerns—often without sacrificing data utility [61, 90, 92]. In object detection, synthetic data facilitates the construction of comprehensive benchmarks, enables research on rare event detection, and substantially reduces dependence on manual labeling [23, 87]. Moreover, synthetic datasets can be engineered to preserve essential statistical properties while permitting controlled attribute manipulation, thereby supporting rigorous benchmarking and enabling explorations of algorithmic fairness and bias [15, 64]. As contemporary computer vision applications require greater adaptability, interpretability, and diversity, the adoption of model-driven synthetic data generation is reshaping both the research agenda and deployment practices in the field.

1.2 Overview of Generative Approaches

The landscape of generative approaches in computer vision has undergone significant transformation, mirroring both technical advancements and evolving theoretical perspectives. Initial strategies were dominated by statistical models and simulation-based frameworks; while effective within specific confines, these early methods were hampered by limitations in fidelity and scalability. A pivotal advancement materialized with the emergence of deep generative frameworks, including Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), and more recently, diffusion models [61, 67, 75, 86, 90, 92]. Among these, GANs have assumed a central role, enabling the synthesis of high-resolution, photorealistic images and supporting sophisticated data augmentation strategies tailored to data-scarce environments [67, 86, 90, 92].

Each generative paradigm presents distinct advantages and concomitant challenges:

- **GANs:** Offer powerful image synthesis capabilities but suffer from training instability, mode collapse, and the delicate balance between image fidelity and diversity [90, 92].
- **Diffusion Models:** Excel at producing detailed textures and complex distributions; however, they pose risks of memorization—particularly in the context of limited or homogeneous training data—raising concerns over ethics and privacy, especially in domains such as medical imaging [86].
- **LLMs and Multimodal Transformers:** Expand synthetic data generation into structured and multimodal domains, bridging data types and supporting richer dataset synthesis strategies [74, 87, 90].

The convergence of these modalities produces a robust and flexible generative toolkit underpinning both supervised learning and emerging self-supervised or contrastive frameworks [67, 87, 90].

Collectively, these innovations foster new paradigms for data generation, accommodating increasingly complex requirements for adaptability, modality fusion, and ethical oversight.

1.3 Scientific Positioning and Societal Relevance

The epistemological and societal ramifications of synthetic data are both profound and multifaceted. Within the research ecosystem, synthetic data underpins the paradigm shift from empiricism to the "Fourth Paradigm" of scientific discovery, wherein data-intensive computational techniques reveal patterns that elude traditional methods [15, 52, 74, 92]. This transition is mirrored in computational social science, where synthetic data frameworks and taxonomies enable the scaling of experimental designs and the simulation of complex social phenomena [3, 15, 74]. Recent surveys emphasize that synthetic data not only expands research capabilities but also generates new challenges, particularly when considering the full spectrum from quantitative to qualitative datasets and diverse applications such as synthetic populations and survey data replacement [74].

Synthetic data's societal utility is particularly salient in the realm of privacy enhancement. Conventional privacy techniques—such as anonymization or aggregation—often prove inadequate in preventing re-identification, which has led to the growing adoption of generative models, frequently augmented by differential privacy mechanisms, to securely share sensitive datasets for domains spanning healthcare analytics and regulatory compliance (in alignment with frameworks such as GDPR and HIPAA) [28, 52, 55, 64]. However, even with these advances, privacy protections remain an unresolved challenge. Generative models, including state-of-the-art GANs, VAEs, and LLM-based methods, may still leak information through memorization or attribute disclosure, especially in the absence of rigorous differential privacy guarantees [28, 52, 64]. Achieving a balance between the utility of synthetic datasets and their privacy assurances involves inherent trade-offs; current research focuses on improving protection strategies (for example, through application of Laplace noise, gradient sanitization, or PATE mechanisms) while systematically auditing and developing clearer guidelines for safe application [52, 64]. Moreover, generative synthetic data methods face technical barriers such as training instability, bias propagation, and difficulty in ensuring true diversity and faithfulness of synthesized samples [15, 28], with new forms of algorithmic discrimination and erosion of trust emerging as possible unintended outcomes [64, 74, 90].

Thus, integrating synthetic data into scientific and societal workflows necessitates rigorous technical validation as well as proactive ethical stewardship. Detailed frameworks for evaluating the "truth, beauty, and justice" of synthetic datasets—considering faithfulness, diversity, and fairness metrics—are progressively becoming indispensable for ensuring responsible advancement in this domain [15, 74].

Table 1: Summary of prominent generative models for synthetic data in computer vision

Model	Strengths	Challenges	Applications
GANs	High-quality, photorealistic generation; wide adoption	Training instability, mode collapse, balancing fidelity/diversity	Object detection, domain adaptation, augmentation
VAEs	Structured latent space; interpretable synthesis	Lower sample fidelity compared to GANs; oversmoothing	Feature learning, anomaly detection, segmentation
Diffusion Models	Excellent texture and distribution fidelity; scalable	Computationally intensive; risk of memorization	Medical imaging, fine-grained synthesis
LLMs/Multimodal Transformers	Structured, text-conditioned or multimodal data synthesis	Interpretability, alignment, scaling	Dataset design, multi-modal learning

1.4 Structure of the Survey

This review systematically investigates the theoretical foundations, methodologies, and applications of synthetic data in computer vision. The manuscript is organized as follows:

Section 2 presents the theoretical and technical foundations, covering data-centric motivations, core generative models, and foundational taxonomies. Section 3 examines principal generative methods, charting progress from classical simulation to state-of-the-art deep generative models, and critically evaluating key merits and limitations. Section 4 surveys important application domains, including object detection, segmentation, domain adaptation, healthcare, and privacy-preserving analytics. Section 5 discusses assessment methodologies for synthetic data, encompassing quality, utility, ethical considerations, current metrics, and ongoing research challenges. Section 6 explores open challenges such as adversarial risks, fairness, scalability, and difficulties in distinguishing synthetic from real data. Section 7 offers perspectives on future directions, emphasizing opportunities for building robust and ethically aligned synthetic data ecosystems.

Through this comprehensive exploration, the survey delineates the contemporary landscape, open questions, and prospective trajectories for synthetic data in computer vision and its broader societal implications.

2 Foundations of Synthetic Data Generation and Image Synthesis

2.1 Definitions and Conceptual Frameworks

Synthetic data has emerged as a cornerstone of modern AI research and deployment, primarily driven by the need to circumvent limitations related to the availability, sensitivity, and expense of acquiring real-world data. In essence, synthetic data refers to algorithmically generated assets that closely mimic the statistical, structural, and semantic characteristics of real datasets, yet do so without directly replicating real-world entities. This attribute supports the mitigation of privacy concerns and circumvents legal-ethical barriers frequently encountered in regulated fields such as healthcare and finance [74]. The conceptual landscape of synthetic data is broad, encompassing both algorithmic and simulation-based methodologies as well as model-driven samples derived from generative architectures, thus reflecting a convergence of computational statistics with recent advances in artificial intelligence [52, 59, 74, 75].

Researchers have established various taxonomies to systematize the rapidly expanding set of synthetic data generation techniques. Approaches can be broadly categorized along several modality-independent dimensions: Instance- versus dataset-level methods: Distinguishing between techniques that augment individual examples and those that synthesize entire datasets. Value-based versus structure-based transformations: Differentiating methods focusing on data values from those manipulating structural properties. Hybrid strategies: Such as domain randomization and multi-step, multimodal processing pipelines [22, 52, 59, 74, 75].

This framework generalizes across diverse data modalities—including images, text, tabular data, time series, and graphs—providing a unified basis for contemporary strategies in data augmentation, simulation, and generative modeling [21, 22, 27, 52, 74, 75].

Simultaneously, usage- and rationale-oriented typologies offer further organizational granularity, delineating areas such as: Quantitative synthetic data: Including simulated measurements and sensor logs. Qualitative assets: Such as synthetic natural language or expert system outputs. Synthetic populations: For demographic simulation and modeling. Highly interactive entities: Including advanced personabots [74].

Application-driven taxonomies underline the versatile utility of synthetic data, which spans privacy protection, alleviation of data scarcity, domain adaptation, and benchmarking for robust scientific discovery [27, 74].

The widespread adoption of synthetic data generation is supported by rigorous theoretical and empirical evaluation frameworks. Central to these are: Data augmentation, leveraging group transformations to enhance generalization and reduce model variance [6, 81]. Empirical Risk Minimization (ERM), which formally connects augmentation processes to robust model training [6]. Bayesian and mathematical approaches, furnishing metrics such as likelihood measures or posterior sampling to quantify distributional alignment between real and synthetic data [31, 46, 81]. Legal-analogy frameworks, like precedential constraint theory, situating synthetic generation within formal systems that balance utility, factuality, fidelity, and fairness—criteria essential for trustworthy AI [14, 27, 74].

2.2 Traditional Generative Methods

Traditional methods of synthetic data generation precede modern deep generative approaches and are grounded in classical statistical modeling and explicit simulation. Statistical models such as Gaussian mixtures, copulas, and Markov models formed the backbone of early systematic efforts, facilitating sampling from assumed or learned data distributions under typical assumptions of independence or stationarity [52, 74, 75]. Explicit simulation—involving agent-based or rule-driven systems—proved especially powerful for generating synthetic environments with precise control over latent parameters, leading to prominent applications in medical imaging phantoms, physics-based simulations, and synthetic population creation [22, 74].

In early pipelines, synthetic data was frequently produced via value-based and structure-based transformations. Value-based techniques included noise injection and pixel or token substitution, while structure-based methods encompassed geometric manipulations such as rotations and translations, as well as other operations affecting the shape or connectivity of data entities [6, 52, 75, 81]. These transformations, as classified in a unified taxonomy [75], offered mathematical tractability and strong interpretability, with clear links to domain-specific invariances (e.g., symmetries in physical or visual data [6]) and theoretical guarantees such as variance reduction and enhanced sampling efficiency (e.g., in Monte Carlo and Bayesian inference settings where calibrated augmentation reduces bias and autocorrelation [81]).

However, these traditional approaches were constrained by their reliance on strong distributional assumptions and limited expressivity. This curtailed their ability to capture multimodal dependencies, represent complex structural or semantic diversity, or address rare-event phenomena present in real-world datasets [22, 46, 74, 81]. Their strength lay in providing tools with high theoretical rigor and interpretability, making them essential precursors to more flexible and automated techniques. Yet, the scale and heterogeneity of modern data analytics—such as the need for large-scale, annotated, and highly variable synthetic datasets—soon exposed their limitations. This realization spurred the evolution toward advanced, data-driven generative frameworks capable of greater representational power and adaptability [22, 52, 75].

2.3 Emergence of GANs and Diffusion Models

Objectives and Unique Contributions: This subsection aims to (i) delineate the technical evolution from classical to state-of-the-art neural generative models for synthetic data, (ii) offer a clear comparative taxonomy of GAN and diffusion paradigms, (iii) synthesize recurring advances and outstanding challenges across techniques, and (iv) explicitly highlight the very latest families and emerging benchmarks (2023–2024) [3, 8, 9, 15, 21, 22, 33, 36, 43, 51, 52, 54, 59, 63, 64, 72, 75, 85, 88]. This survey further distinguishes itself by surfacing not just model architectures but also cross-modal conditioning capabilities, hybrid platforms, and the documented risks of memorization and "AI autophagy" in generative practices.

The introduction of deep generative models, especially Generative Adversarial Networks (GANs) and, more recently, diffusion models, has initiated a transformative shift in synthetic data generation. These advances have enabled the synthesis of high-dimensional, multimodal, and photorealistic data—capabilities that surpass those of traditional paradigms [3, 8, 9, 15, 21, 22, 28, 33, 36, 40, 43, 51, 52, 54, 55, 57, 59, 63, 64, 72, 75, 85, 88]. GANs pioneered the minimax adversarial framework, wherein a generator and discriminator engage in a competitive process, iteratively refining the fidelity of generated samples. Numerous architectural variants—including conditional GANs, auxiliary classifier GANs, and domain-specific extensions—have propelled progress across image synthesis, segmentation, tabular data generation, and audio domains, while facilitating research on domain adaptation and class imbalance [15, 21, 22, 36, 52, 54, 64, 72, 85].

Nevertheless, GANs exhibit inherent challenges, such as mode collapse, training instability, and an inability to directly estimate sample likelihoods or uncertainty [3, 21, 28, 52, 72, 85]. These shortcomings have prompted the ascendance of diffusion models, a class of implicit probabilistic models that gradually transform noise into coherent data via iterative noise-to-signal inversion, often formulated through score-based optimization or stochastic differential equations [3, 7–9, 21, 22, 36, 43, 54, 63, 85, 88]. Diffusion models are characterized by stability in training, controllable output diversity, and state-of-the-art sample quality—frequently outperforming GANs in complex domains such as image, video, medical, and molecular data synthesis [8, 9, 21, 22, 33, 43, 51, 54, 63, 85]. Recent works (e.g., 2024–2025) have advanced theoretical guarantees for synthetic data statistical quality [36, 63], explored diffusion for text, tabular, quantum circuit and large language generation [9, 15, 43, 59], and revealed that although diffusion models set SOTA in sample fidelity, they may also confer heightened memorization risks in sensitive applications, notably in medical imaging, underscoring the need for tailored evaluation protocols and privacy safeguards [8, 85].

Both GANs and diffusion models benefit from flexible conditioning (e.g., class labels, text prompts, or domain attributes), now extended to cross-modal, annotation-efficient or few-shot synthesis tasks through integration with large language models (LLMs) [7, 9, 15, 63, 72, 85, 88]. However, recent literature increasingly notes that unrestrained use of synthetic data in compounding workflows introduces risks of quality degradation or "AI autophagy," motivating careful curation and benchmarking practices [85, 88].

Recent developments have given rise to hybrid architectures (e.g., diffusion-GANs, VAE-diffusion models), reinforcement-augmented and implicit neural generative frameworks, and the incorporation of LLMs as both generators and evaluators in multimodal workflows [3, 7, 9, 33, 40, 43, 52, 55, 72, 85, 88]. These platforms provide flexibility, supporting unsupervised dataset creation and targeted data augmentation in weakly, unsupervised, and few-shot learning settings [7, 22, 54, 64, 75, 89].

As outlined in Table 8, these paradigms differ substantially in generative capabilities, stability, and suitability for complex, multimodal data.

Summary Box—Current Synthesis: GANs and diffusion models represent the twin pillars of modern synthetic data generation, each with distinct algorithmic trade-offs: GANs offer creative, fast sampling but are challenged by instability and privacy utility; diffusion models provide stable, high-fidelity generation, extendable to text and large language tasks with recent state-of-the-art advances and theoretical guarantees [8, 9, 36, 43, 51, 54, 63, 85]. Emerging cross-modal, hybrid, and LLM-integrated architectures, as well as the recent focus on memorization and dataset curation, define a rapidly advancing and multidisciplinary research frontier.

2.4 Importance in Computer Vision and Data Science

This section aims to clearly articulate the main objectives of our survey: (1) to synthesize the motivations, technical families, and broad contributions of synthetic data in computer vision and data science; (2) to integrate and highlight persistent open challenges faced by practitioners; and (3) to present a concise roadmap for actionable improvements in evaluation, fairness, privacy, and responsible use. By clarifying these goals up front, we ensure readers can directly map the reviewed advances and findings to the current landscape and our unique survey perspective.

Table 2: Comparison of Classical, GAN, and Diffusion Approaches to Synthetic Data Generation

Aspect	**Traditional Statistical / Simulation Methods**	**GANs**	**Diffusion Models**
Representative Models	Gaussian Mixture Models, Copulas, Markov Models, Agent-based Simulations	Original GAN, cGAN, ACGAN, StyleGAN	DDPM, Score-based Models, Guided Diffusion, LLM-Diffusion
Key Strengths	Easy to interpret, tractable, controlled latent factors	High-dimensional, realistic sample generation; creative tasks; conditional synthesis	Stable training, high-fidelity outputs, theoretical guarantees (2024+), controllable diversity, effective in multimodal and annotation-efficient tasks
Notable Weaknesses	Limited expressivity; cannot capture high-dimensional or rare-event phenomena; restricted multimodality	Mode collapse; training instability; lack of explicit likelihood estimates; privacy challenges	High computational demand; large sample generation times; memorization risk in small/sensitive domains (e.g., medical); risk of “AI autophagy”
Effective Domains	Simple tabular data, simulations, controlled environments	Images, audio, tabular data, domain adaptation	Images, video, text, quantum circuits, molecular/medical/multimodal data, large language tasks

Synthetic data has become a linchpin in the progress of computer vision and, more broadly, data science. Its adoption is driven by urgent needs for data augmentation, tackling annotation scarcity, and ensuring fairness and privacy. Synthetic data not only expands datasets, but strategically combats class imbalance, enables new learning paradigms, and bolsters model robustness under distributional shifts [2–4, 8–10, 12, 15, 16, 21–23, 25, 26, 28, 29, 33, 36, 41, 43, 45, 50–52, 54–56, 58, 61, 63, 65, 67, 70, 72, 75, 78, 85, 86, 88, 89, 93]. In computer vision, synthetic datasets drive aggressive augmentation—balancing data imbalances and supporting training where labeled data is limited or biased. Cutting-edge techniques such as diffusion models and advanced LLM-augmented pipelines [9, 15, 16, 22, 26, 33, 36, 43, 51, 54, 56, 63, 85] have enabled high-fidelity sample generation across domains like imaging, molecular design, video, and language.

Beyond vision, synthetic data delivers analogous impacts in NLP, healthcare, tabular and time-series data, providing an integrated toolkit that addresses scarcity, bias, and privacy constraints across modalities [22, 52, 64, 75, 86]. In scenarios where annotation is costly or infeasible, synthetic data enables weak supervision, unsupervised, and few-shot learning—empowering progress in domains such as dense segmentation, anomaly detection, and multimodal medical analytics [7, 10, 22, 29, 41, 45, 50, 54, 56, 58, 63–65, 85]. Its use in bridging disparate modalities has catalyzed advances in vision-language integration, foundation model pretraining, and robust cross-domain transfer [15, 22, 33, 63, 72, 75, 85].

Synthetic data generation further addresses pivotal issues in privacy and algorithmic fairness. By decoupling model development from direct reliance on sensitive real-world data, it enables collaborative research and safer industry adoption. Synthetic datasets underpin efforts to reduce bias, prevent data leakage, and perform fair representation analysis [2, 27, 41, 61, 74, 86]. Current research also explores integrating differential privacy, auditing, and principled utility-privacy trade-offs [8, 27, 52, 74, 85]. However, open challenges persist:

- Faithful representation of complex, high-dimensional ground-truth distributions—especially in scientific, medical, and physically or statistically constrained domains—remains difficult [26, 36, 51, 56, 85]. - The lack of standardized, domain-specific evaluation protocols for judging synthetic data quality, utility, and real-world transferability hampers rigorous benchmarking [27, 52, 63, 74, 85, 86]. - Privacy guarantees are incomplete, as risks of memorization or re-identification remain unresolved, especially in sensitive or low-diversity datasets [8, 27, 52, 74, 85]. - Biases and inequities can persist or be amplified if generation pipelines are not critically audited; actionable and scalable guidance is still emerging [27, 74, 85]. - Most synthetic data practice still lacks transparent reporting around data generation steps, provenance, and intended usage, which undermines responsible adoption.

In summary, synthetic data stands as a foundational enabler for AI-driven research and real-world deployments across vision and data science. Its evolving technical landscape—spanning adversarial, diffusion, language-based, and simulation-based generation—is matched by a new wave of benchmarks and evaluation protocols (e.g., FID, precision/recall, TSTR, domain-specific utility metrics) [9, 33, 36, 54, 63, 64, 75, 85]. Yet, the responsible use of synthetic data now demands advances in fidelity, systematic evaluation, privacy preservation, and transparency of process and intent.

For practitioners, actionable priorities now include: - Developing and adopting robust, domain-tailored evaluation benchmarks and reporting standards for synthetic data; - Designing synthesis pipelines with explicit privacy and fairness trade-off measurement and auditing; - Implementing scalable guidelines for bias prevention and mitigation during data generation; - Ensuring provenance and clear documentation of generation processes and usage scope for all synthetic datasets; - Intensifying community-driven synthesis and documentation of best practices to accelerate responsible, impactful progress.

By integrating the above objectives and referencing the most current advances and ongoing challenges, this survey distinguishes itself from prior works and aims to provide a systematic, timely, and practitioner-oriented synthesis of the state of synthetic data in modern computer vision and data science.

3 Core Techniques and Advanced Methods for Image Synthesis

This section systematically reviews and critically analyzes the principal frameworks and advanced approaches driving state-of-the-art image synthesis. The specific objectives for this section are: (1) to provide a comprehensive understanding of underlying mechanisms central to modern image synthesis techniques, (2) to assess the extent to which each method addresses the key objectives articulated in Section **??**, and (3) to highlight persistent challenges and promising avenues for further research. For reader clarity, these objectives are reiterated at the beginning of every major subsection and referenced explicitly at key transitional junctures.

We begin by establishing foundational principles common to most image synthesis models. Transitional paragraphs are included between major subsections to enhance narrative continuity and to clarify how each successive technique aligns with or diverges from the overall survey goals. Where appropriate, we cross-reference back to the main research questions or objectives introduced earlier.

After the discussion of each major technical family or methodological paradigm, we provide concise overviews that synthesize critical findings and draw attention to current research gaps, including open challenges such as evaluation difficulties or privacy concerns. These overviews further articulate measurable aims against which progress can be tracked, supporting actionable research planning.

To further support accessibility and reader orientation, summary paragraphs at the ends of each methodological group highlight both the unique positioning of this survey in relation to recent advances and opportunities for future work. Citation practices are kept strictly standardized for clarity, and summary passages are crafted to facilitate rapid comprehension. Throughout, we incorporate explicit references to the overarching objectives to ensure cohesive integration across methods and maintain clear conceptual links for the reader.

3.1 Generative Adversarial Networks (GANs) Conditional and Fine-Grained Synthesis

Generative Adversarial Networks (GANs) have spearheaded the evolution of realistic image synthesis, progressing rapidly from foundational adversarial setups toward sophisticated conditional and fine-grained architectures. Initial conditional GANs, exemplified by Pix2Pix, leveraged conditioning signals such as class labels, semantic layouts, or textual descriptions to produce controllable outputs. Nonetheless, these architectures frequently exhibited limited output diversity. This limitation was primarily due to either the neglect or insufficient utilization of the input noise vector, often resulting in deterministic outputs for the same conditioning input [48, 69].

To enhance output diversity in conditional synthesis, recent works have proposed innovative objectives. Notably, the diversity loss framework penalizes output redundancy by incentivizing the maximization of pairwise distances between outputs when corresponding noise vectors differ. Importantly, semantic grounding is achieved by aligning each noise dimension with interpretable components of the target image (e.g., sky, windows, or vehicles), thus enabling independent and intuitive manipulation of image regions while preserving realism and semantic consistency—a significant advancement relative to earlier techniques that only induced global changes [69]. For example, the formulation in [69] ensures that each semantic layout segment is locally controlled by a specific noise channel, allowing user-driven edits that selectively affect color or illumination of individual regions, without disrupting other parts of the image. Although this approach primarily induces variation in appearance rather than structure, it marks progress toward explainable and interactive image synthesis. These developments have been rigorously evaluated on heterogeneous benchmarks such as CMP Facades and Cityscapes, employing robust metrics including Inception Score, Structural Similarity Index (SSIM), and domain-adapted quantitative measures. The findings confirm that increasing output diversity does not necessitate trade-offs with realism, provided that noise regularization is semantically coherent.

In the context of fine-grained and patch-based synthesis, GAN architectures have adopted domain-specific innovations. For instance, facial synthesis now leverages explicit facial keypoint extraction to guide generation, and combines per-pixel, perceptual, and adversarial losses to produce semantically faithful, artifact-suppressed outputs—even in the case of severe occlusions or partially observed inputs [79]. Specifically, integrating facial keypoints (e.g., using dlib) with reconstruction, perceptual, and adversarial objectives enables the synthesis model to generate realistic faces by fusing separate facial regions from multiple sources while maintaining semantic alignment. Empirical studies on benchmarks such as CelebA, LFW, and CACD demonstrate that such models outperform baselines like Pix2Pix and traditional inpainting in both quantitative and qualitative evaluations, including reconstruction error and FID. These strategies are indispensable for applications ranging from portrait novelization and medical preview to generating synthetic biometric datasets for face recognition and forensics. The hybridization of pixel-wise, high-level semantic, and adversarial objectives underscores the critical role of loss function design in balancing detail fidelity with overall realism, while also enabling attribute-based or patch-wise control for downstream applications.

Text-to-image synthesis, too, has matured through progressive architectural refinements. For example, the FG-RAT GAN extends the Recurrent Affine Transformation (RAT) GAN by incorporating auxiliary classification and contrastive learning: an auxiliary classifier is integrated into the discriminator and cross-batch memory is exploited to define a contrastive loss, collectively improving both intra-class consistency and inter-class distinctiveness. This dual strategy is reflected in dedicated loss functions ($L_d^{\text{total}} = L_d^{\text{adv}} + L_d^{\text{ce}} + L_d^{\text{cl}}, L_g^{\text{total}} = L_g^{\text{adv}} + L_g^{\text{ce}} + L_g^{\text{cl}}$), optimizing for adversarial, cross-entropy, and contrastive criteria. As evidenced by rigorous ablation in [34], both the auxiliary classifier and the contrastive loss substantively improve subclass-aware synthesis, delivering higher semantic and visual fidelity. Benchmarks on CUB-200-2011 and Oxford-102 indicate that FG-RAT GAN achieves superior Frechet Inception Distance (FID) and competitive Inception Scores (IS) with fewer parameters compared to prior art such as LAFITE and VQ-Diffusion, underscoring its computational efficiency. Nonetheless, outstanding challenges persist, including continued reliance on labeled data, degree of label entanglement, and restricted generalizability to label-independent or semi-supervised scenarios.

Comparative analyses demonstrate that, while GANs afford efficient synthesis capabilities, high-level texture plausibility, and visual fidelity, they continue to face challenges—most notably mode collapse, training instability, and constrained semantic alignment as scene or conditional complexity scales [34, 48, 69, 79]. In response, there is a growing emphasis on robust evaluation protocols and pluralistic output metrics extending beyond visual quality; these efforts are fueling further advances in the architecture and objectives of conditional and fine-grained synthesis frameworks.

3.2 Diffusion Models for Semantic and Style-Controlled Image Synthesis

Diffusion models have rapidly become central to high-fidelity, semantically controlled, and stylistically nuanced image synthesis. Unlike traditional GANs, which generate images in single adversarial steps, diffusion models reconstruct images via multi-step denoising, transforming noise into data through iterative, probabilistic transitions. This paradigm offers significant improvements in sample diversity, generation stability, and controllability—addressing major GAN shortcomings such as mode collapse and unstable training dynamics [17, 72].

Cutting-edge latent diffusion architectures operate in compressed, perceptually meaningful latent spaces, learned via autoencoders

Table 3: Representative GAN-Based Image Synthesis Approaches: Characteristics and Benchmarks

Model/Method	Key Innovations	Domain/Task	Primary Metrics & Benchmarks
Pix2Pix	Conditional on labels/layouts; paired training	Image-to-image translation	Cityscapes, CMP Facades; Inception Score, SSIM
Diversity Loss GAN	Explicit noise/region correspondence; semantic diversity	Fine-grained conditional synthesis	CMP Facades, Cityscapes; Inception Score, semantic diversity metrics
FG-RAT GAN	Auxiliary classification + contrastive learning; efficient architecture	Text-to-image, fine-grained class synthesis	CUB-200-2011, Oxford-102; FID, Inception Score
Face Completion GAN	Keypoint-based conditioning; hybrid losses (pixel, perceptual, adversarial)	Face completion, occlusion recovery	CelebA, LFW, CACD; Reconstruction error, FID

such as VQGAN. Latent Diffusion Models (LDMs) thus deliver computational efficiency gains without sacrificing output quality or semantic accuracy and support a wide variety of conditioning modalities (e.g., text, segmentation masks, or style images) through cross-attention mechanisms [17, 71, 72, 84]. Notably, image-to-image diffusion models (IIDM) further generalize these capabilities by enabling both semantic segmentation and stylistic control within the denoising trajectory, resulting in outputs that are structurally faithful and stylistically coherent. Large-scale evaluation reveals that such methods routinely outperform both GANs and pixel-level diffusion models in terms of FID, mask accuracy, and usability in downstream applications.

The versatility of diffusion models is particularly evident in specialized domains: Structure-preserving latent diffusion strategies yield high-resolution, morphologically accurate 3D brain MRIs, validated against neuroanatomical metrics and supporting strong predictive modeling. Geometry-complete diffusion models (GCDM) generate molecular structures that accurately reflect atom types and 3D arrangements, surpassing state-of-the-art in both geometric and property consistency. Innovations like uncertainty-guided sampling improve informativeness for disease grading, reflecting a shift from purely image-centric evaluation to clinically meaningful generation.

Despite their promise, diffusion models face persistent limitations. Generation is inherently slower due to the iterative inference process—a drawback in time-sensitive or resource-limited settings [17, 72, 84]. While latent-space processing mitigates computational overhead, it may introduce information loss, especially for tasks demanding pixel-level precision [17, 84]. Moreover, complex conditioning schemes necessitate careful engineering to prevent semantic drift or interference between content and style channels [71]. Ethically, there is increased risk of memorization and data leakage—especially in low-variability or small-scale medical datasets—underscoring the urgent need for domain-specific evaluation and privacy protocols [76, 91].

3.3 Text-to-Image Synthesis and Cross-Modal Generation

Text-to-image and cross-modal synthesis are central to controllable generative modeling, with current systems advancing through the integration of multi-stage attention architectures that combine diverse textual representations such as TFIDF, N-gram, and Bi-LSTM with hierarchical GANs, diffusion frameworks, and advanced loss functions to improve semantic and geometric alignment [32, 73]. Recent developments highlight the use of attention mechanisms—particularly alternate attention-transfer between word embeddings and image sub-regions—to facilitate fine-grained feature extraction and fusion across abstraction levels, directly addressing the classical issues of text-image misalignment and overly confident outputs found in earlier GAN-based models [32, 73].

State-of-the-art models characteristically exploit combinations of customized loss functions, including adversarial, perceptual, feature-matching, categorical cross-entropy, and contrastive objectives, in order to achieve a balance between image fidelity and diversity. For example, FG-RAT GAN [34] extends RAT GAN by introducing an auxiliary classifier into the discriminator and employing a cross-batch memory contrastive loss. This design significantly sharpens intra-class and inter-class distinctions and encourages the generator to produce semantically consistent, visually distinguished outputs. The total loss objectives are $L_d^{\text{total}} = L_d^{\text{adv}} + L_d^{\text{ce}} + L_d^{\text{cl}}$ for the discriminator and $L_g^{\text{total}} = L_g^{\text{adv}} + L_g^{\text{ce}} + L_g^{\text{cl}}$ for the generator, where adversarial, categorical cross-entropy, and contrastive losses collectively enhance subclass-aware synthesis. Comprehensive ablation studies demonstrate that both the auxiliary classifier and contrastive learning components are instrumental to performance improvements in FG-RAT GAN [34]. Experiments on CUB-200-2011 and Oxford-102 datasets showcase state-of-the-art Frechet Inception Distance (FID) and competitive Inception Scores (IS), surpassing methods such as LAFITE, VQ-Diffusion, and RAT GAN, while benefiting from fewer trainable parameters and greater computational efficiency [34].

Additionally, approaches aimed at minimizing dependence on dataset annotation—such as self-supervised and contrastive training—are increasingly adopted to promote label-independence and robustness. On benchmark datasets like Oxford-102 and CUB-200-2011, models like VG-RAT GAN and KT-GAN continue to set records for FID, IS, and SSIM, while often utilizing fewer parameters than leading competitors [32, 34, 73]. Notably, KT-GAN introduces mechanisms such as the Alternate Attention-Transfer Mechanism (AATM) and Semantic Distillation Mechanism (SDM), which alternately update word and image sub-region attention and leverage pretrained image encoders to strengthen text-to-image feature extractions, further boosting performance and cross-modal alignment [73].

Despite remarkable progress, key challenges persist, including model uncertainty, generalizability to open-world categories, and the systematic incorporation of extrinsic knowledge. Addressing these obstacles represents ongoing and future research opportunities that encompass the integration of generative, self-supervised, and large foundation model paradigms.

3.4 Hybrid and Transformer-Based Image Completion

Recent advancements in transformer architectures have led to the emergence of hybrid models that integrate transformer-based and convolutional neural network (CNN) techniques for image completion and inpainting. Traditional CNN-based methods are effective

Table 4: Comparison of Diffusion Model Innovations and Application Domains

Approach	Key Features	Sample Domains	Noted Advantages/Limitations
Latent Diffusion Model (LDM)	Cross-attention conditioning; latent-space processing	Image synthesis, text-conditioned outputs	Computationally efficient; scalable; susceptible to information loss in pixel-detail tasks
IIDM	Semantic + style conditioning in denoising loop	Image-to-image, stylized synthesis	High structural fidelity; complex attention design required
GCDM	3D graph-structured diffusion	Molecule, chemical generation	Preserves chemistry/geometric constraints; domain-specific encoding
Uncertainty-guided Sampling	Informativeness-driven sampling	Medical/clinical synthesis	Supports task-centric evaluation; risk of memorization/privacy issues

for modeling local texture details due to their strong inductive priors and spatial-invariant kernels, but they often struggle with capturing complex global structures and supporting diverse, pluralistic completions. In contrast, transformers excel at modeling long-range dependencies and are well suited for generating pluralistic completions that provide multiple, coherent possibilities for a given masked input [82]. However, the computational complexity of transformers, which scales quadratically with input size, has historically limited their application to high-resolution images.

Hybrid approaches address this challenge by employing transformers for appearance prior reconstruction and coarse structure prediction, while utilizing CNNs to enhance local texture details and refine the results. This dual-stage methodology enables hybrid models to achieve notable improvements over state-of-the-art methods in terms of image fidelity, diversity in pluralistic completion, and robustness to large masked regions and broad datasets such as ImageNet [82]. The model proposed by Wan et al. [82] exemplifies this approach by combining transformers and CNNs to recover pluralistic, coherent structures alongside fine-grained textures, thereby significantly boosting performance for image completion tasks. The public release of code and pre-trained models for these hybrid techniques is accelerating further research and increasing their applicability across various domains.

3.5 Classical, Automated, and Adaptive Data Augmentation

Data augmentation strategies are foundational in deep learning, expanding training set size and diversity through geometric and color space transformations, patch mixing, and simulation-based synthesis. Classical techniques encompass operations such as flipping, rotation, scaling, color adjustment, channel permutation, kernel filtering, and sample mixing approaches like Mixup and SamplePairing [23, 29, 50, 75]. Automated methods—including AutoAugment and ADAAT—systematize the search for effective augmentation policies or adversarially generate examples to enhance classifier robustness and fairness [39, 68]. For example, ADAAT combines adaptive feature modification with adversarial sample generation to improve resistance to perturbations and noisy environments.

Recent advances extend augmentation beyond vision, incorporating domains such as natural language processing, tabular data, graphs, and time-series. Techniques increasingly leverage large language models (LLMs), self-supervised, and reinforcement learning frameworks, enabling domain-agnostic and multi-modal augmentation [3, 7–12, 15, 19, 20, 28, 31, 33, 36, 40, 43, 51, 52, 54, 55, 57, 59, 63, 64, 67, 75, 86]. Unified, modality-independent augmentation taxonomies—such as those introduced in [75]—categorize methods by both sample granularity (single-instance, multi-instance, dataset-level/generative) and information leveraged (value-based or structure-based transformations), allowing systematic benchmarking and facilitating cross-domain transfer of effective strategies. This evolution includes automated and generative approaches spanning neural mixing, prompt-based modeling, GANs, and diffusion models [3, 33, 54, 55, 63, 75], with practical application in self-supervision, contrastive learning, privacy preservation, and regulated sectors like healthcare [8, 11, 40, 52, 57, 59, 64].

Model-adaptive strategies, such as Model-Adaptive Data Augmentation (MADAug), dynamically optimize augmentation curricula by using bi-level optimization that tailors policy selection to sample and model-specific properties. This provides improved fairness and generalization, especially within fine-grained or imbalanced tasks [68]. Likewise, frameworks such as ADAAT [39] demonstrate that adaptation and adversarial training can provide notable robustness on challenging datasets, even under adversarial attacks.

Although augmentation is broadly effective for reducing overfitting and boosting generalization, significant challenges continue. Automated, adversarial, and generative approaches can unintentionally cause distributional shifts or produce samples with noisy or invalid labels [23, 39, 75, 86]. Specific concerns are emergent with generative models like GANs and diffusion models: GAN-based augmentation may struggle with diversity and semantic accuracy, while diffusion models—although achieving high sample fidelity—face risks of memorization and privacy leakage, especially in sensitive domains [8, 54, 63]. Additionally, designing robust and generalizable augmentation policies for domains outside of vision and establishing standardized metrics for evaluating semantic fidelity and diversity remain pressing open problems [28, 64, 67, 75, 86].

Consequently, effective augmentation increasingly relies on integrating human domain priors, automated policy search, and comprehensive evaluation metrics. Current best practices jointly emphasize semantic fidelity, sample diversity, privacy, and downstream task relevance, highlighting the need for future research in cross-modal generalization, practical generative model scaling, standardized benchmarks, and human-in-the-loop synthesis [15, 52, 75].

3.6 Self-Supervised and Transfer Learning

The synergy between self-supervised learning and generative modeling continues to revolutionize feature extraction and annotation processes across diverse domains. Self-supervised models leverage pretext tasks such as rotation prediction and patch reconstruction to acquire high-quality, transferrable features from unlabeled data. When incorporated into generative frameworks, these features enhance robustness and lessen reliance on large annotated datasets, a benefit most pronounced in settings with limited labeled data availability—such as medical imaging and rare phenotype discovery. Here, transfer learning coupled with self-supervised approaches markedly accelerates and improves adaptation of generative models, leading to quantifiable gains in both generated sample quality and downstream analytical tasks [18].

Applications of self-supervised and transfer learning in generative modeling span a wide range of domains. In computer vision, latent diffusion models [71, 91] and transformer-based hybrids now integrate self-supervised pretraining and adaptive data augmentation [23, 29, 50, 68, 75] to achieve high-fidelity, diverse, and computationally efficient synthesis with strong domain robustness [18, 34, 71, 91]. In natural language processing, robust text generation and cross-modal synthesis increasingly benefit from similar pipelines, tying together pre-trained feature extractors and generative backbones. In medical data generation, self-supervised and transfer learning enable creation of realistic, privacy-conscious synthetic datasets for modalities such as MRI, X-ray, and tabular electronic health records, supporting both diagnosis and risk modeling in regulated settings [11, 15, 22, 40, 52, 57, 59, 64, 76, 85]. Industrial applications include genomics, climate, and earth sciences, where scarcity and heterogeneity in real data require adaptable feature learning and transfer, as seen in turbulence modeling [51], quantum simulation [43], and earth system downscaling [84].

A selection of key application domains and representative methods is provided in Table 5, structured to clarify domain-specific advances stemming from self-supervised and transfer learning in generative frameworks.

Methodological advances have produced generative models with superior fidelity, diversity, and efficiency, alongside improved domain generalization [18, 34, 71, 91]. Nevertheless, recent surveys emphasize persistent challenges: the establishment of systematic benchmarks linked to downstream tasks, nuanced evaluation of memorization and privacy trade-offs [8, 76], and incorporation of ethical and legal safeguards—especially in high-impact contexts such as healthcare and finance [52, 76, 85]. Furthermore, cross-domain comparisons [15, 28, 55, 64, 75] urge the development of unified and domain-sensitive evaluation protocols for robust deployment and adoption.

As the field advances toward foundation models and comprehensive cross-modal generative systems, the integration of robust self-supervised and transfer learning stages is essential not only to maximize data efficiency and adaptability but also to ensure principled, fair, and responsible scaling and deployment of generative techniques [3, 7–12, 15, 17–20, 22, 23, 28, 29, 31–34, 36, 39, 40, 43, 48, 50–52, 54–57, 59, 63, 64, 67–69, 71–73, 75, 76, 79, 82, 84–86, 91].

3.7 3D and Multiview Synthetic Data and Evaluation

This subsection aims to advance the survey's overall objectives by (1) contextualizing the central importance of 3D and multiview synthetic data within AI research, (2) delineating and comparing state-of-the-art generation pipelines and evaluation practices, and (3) highlighting critical future research challenges and actionable directions. By providing such an integrative synthesis, we seek to reinforce the connection to the survey's stated purpose of charting current breakthroughs and gaps in scalable synthetic data generation, with a focus on the 3D and multiview domain.

3D and multiview data underpin systems' abilities to perceive spatial structure, scene layout, and object geometry. Recent generative pipelines—spanning neural rendering, geometry-aware generative methods, and multi-modal data fusion—enable the creation of synthetic datasets that robustly support, and in some cases surpass, real-world benchmarks in computer vision, robotics, and AR/VR. Here, we systematically categorize prominent approaches, review their evaluation methodologies in depth, and scrutinize persistent limitations impeding wider adoption or reliability, making explicit reference to how each facet addresses the broader survey objectives.

To enhance clarity and navigability, we begin each major methodological category by restating its relevance to the survey's overall aims, highlight summary points at notable transitions, and provide brief recaps of key insights drawn from comparative analyses. This approach is meant to facilitate rapid orientation for readers joining at this subsection, while reinforcing methodological insight and the identification of actionable research gaps.

Specifically, this section: Restates how 3D/multiview synthetic data fits within the survey's focus on enabling robust and diverse downstream AI applications. Presents a clearly structured taxonomy of leading generation pipelines and evaluation methodologies, with explicit pointers as to technical families and paradigm evolution where applicable. Connects evaluation practices and methodological advancements directly to both longstanding and emerging challenges, referencing how these align with survey-level research objectives and ongoing field debates. Enumerates actionable open research areas, clearly distinguishing between technical, methodological, and application-driven gaps, and provides concise summaries at section conclusions for efficient knowledge transfer.

In summary, this subsection is structured to: 1. Explicitly reiterate its alignment with the survey's overarching goals. 2. Provide a taxonomy and synthesis of main technical approaches and evaluation strategies in 3D/multiview synthetic data. 3. Deepen discussion of comparative merits and drawbacks across evaluation frameworks. 4. Promote clearer cross-section integration with related domains, and encourage reader retention through succinct summary and transition statements. 5. Conclude with direct pointers to future research challenges, linking these back to the stated aims in Section **??**.

Through this approach, we ensure that each methodological and analytical component within the 3D/multiview synthetic data ecosystem is not only thoroughly reviewed but also situated within the unifying narrative and objectives of the survey.

3.7.1 Free Viewpoint Video (FVV) and Virtual View Synthesis.

Free viewpoint video (FVV) technologies have markedly enhanced the immersive quality of three-dimensional visual experiences by allowing users to interactively select arbitrary viewpoints within a scene. Central to FVV content generation is depth-image-based rendering (DIBR), where virtual viewpoints are synthesized through the integration of depth and color data. Despite substantial progress, DIBR pipelines continue to grapple with complex visual artifacts, including disocclusion-induced holes, ghosting, and temporal instability across consecutive frames. These perceptual degradations primarily arise from inherent ambiguities at depth discontinuities and insufficient temporal correlations, which—if left unresolved—undermine both the realism and consistency of synthesized views.

Table 5: Representative Applications of Self-Supervised and Transfer Learning in Generative Modeling

Domain	Example Method(s)	Application/Impact	Reference(s)
Medical Imaging	VQ-VAE + Transformer, Diffusion, GANs	Synthetic MRI/X-ray, privacy-preserving analytics, phenotype simulation	[11, 57, 59, 64, 76, 85]
Computer Vision	Latent Diffusion Models, Transformer Hybrids	High-fidelity image/text synthesis, cross-modal tasks	[32, 34, 48, 71, 91]
Natural Language Processing	LLM-driven augmentation, knowledge transfer	Text and multimodal data synthesis, annotation expansion	[9, 15, 67, 86]
Scientific Computing	Diffusion/Consistency Models	Generating turbulence, climate, and quantum data for simulation	[43, 51, 84]
Healthcare Tabular Data	Multi-method pipelines, Auto-ML validation	Synthetic EHRs, privacy-utility evaluation, clinical risk prediction	[40, 52, 57, 59, 64]

To address these challenges, recent research directions have embraced spatio-temporal fusion strategies, moving beyond earlier methods that relied solely on spatial inpainting or simplistic temporal interpolation. An advanced paradigm integrates both spatial and temporal scene cues by leveraging the temporal information present in video sequences to robustly estimate static backgrounds. This approach facilitates a weighted-fusion hole-filling process, in which missing or corrupted regions are adaptively reconstructed utilizing temporally stable background estimates in conjunction with spatial edge-preserving filters.

Key components of these sophisticated pipelines include edge-enhanced depth map refinement through expansion and Gaussian smoothing to improve depth discontinuity handling, robust static scene extraction using pixel-wise structural similarity index (SSIM) analysis across frames to identify consistent background regions, and comprehensive color-depth fusion featuring joint refinement of color and depth information to mitigate artifacts and enforce cross-modal consistency.

This multi-stage strategy exhibits clear advantages over single-frame or basic temporal approaches, notably reducing the prevalence of holes and ghosting artifacts while substantially improving spatio-temporal consistency. Evaluation on established multiview datasets—such as 'BreakDancers' and 'Ballet'—demonstrates significant quantitative improvements. Specifically, metrics such as peak signal-to-noise ratio (PSNR), SSIM, and flicker-based F-scores reveal superior reconstruction quality and temporal stability relative to prior solutions; refer to Table 6 for a summary of representative results.

Despite these advances, notable limitations remain. The computational complexity associated with joint spatio-temporal analysis is significant, often limiting real-time applicability. Furthermore, reliance on static camera configurations restricts these approaches in more dynamic or unconstrained environments. Accelerating computations via parallel architectures, such as GPU-based CUDA implementations, is a promising avenue for achieving practical deployment. Continued progress in multi-modal, temporally-aware processing is therefore critical for advancing the realism and scalability of FVV systems, with particular attention needed on balancing computational efficiency against perceptual fidelity and deployment versatility [?].

3.7.2 Quality Assessment Metrics for 3D and Multiview Synthesis. This subsection aims to elucidate the specific role and recent advancements in quality assessment (QA) metrics that cater to 3D and multiview synthesis applications. As the underlying synthesis methodologies have matured, ensuring objective and perceptually aligned evaluation methods has become critical for both the development and practical deployment of these technologies.

The evolution of view synthesis methodologies has catalyzed a pressing need for robust image and video QA metrics specifically attuned to the characteristics of 3D content. Classical 2D metrics, such as PSNR and SSIM, often fail to capture the perceptually significant distortions characteristic of DIBR outputs, especially in the vicinity of geometric discontinuities and disoccluded regions. In response, a new class of full-reference metrics has emerged, designed to account for these unique failure modes inherent in synthesized 3D views.

Among these innovative contributions, the Morphological Pyramid Peak Signal-to-Noise Ratio (MP-PSNR) and Morphological Wavelet Peak Signal-to-Noise Ratio (MW-PSNR) merit particular attention due to their methodological rigor and empirical efficacy. Both approaches employ multi-scale, nonlinear morphological decompositions that prioritize the fidelity of structural and edge information, enabling perceptually relevant quantification of synthesis quality. These metrics focus on error measurement along prominent geometric transitions, where visual discomfort is most acute. Large-scale evaluations using canonical DIBR image datasets, including IRCCyN/IVC DIBR and MCL-3D, affirm the superiority of these approaches—MP-PSNR, for example, exhibits Pearson and Spearman correlation coefficients with subjective human quality ratings exceeding 0.9 and 0.86, respectively, outperforming classical QA techniques; see Table 7.

Beyond accuracy, these morphological metrics also offer computational benefits. Unlike learning-based QA systems that entail significant parameter tuning or complex registration processes, MP-PSNR and MW-PSNR are efficiently computed using basic morphological operations (such as min, max, and sum), and reduced versions concentrating on high-detail decomposition bands continue to demonstrate strong alignment with subjective quality appraisals. Their robustness to registration errors and minimal dependence on extensive training or calibration make them especially well suited for both algorithmic development and real-world deployment within FVV and 3D synthesis platforms [?].

In summary, rigorous, perceptually aligned assessment metrics are fundamental to the advancement of 3D view synthesis pipelines. The development and adoption of specialized QA measures such as MP-PSNR and MW-PSNR illustrate the ongoing efforts to overcome limitations of classical metrics. As future systems scale toward higher resolutions, real-time execution, and unconstrained multiview operations, both synthesis methods and assessment tools must continue to address fundamental trade-offs among computational tractability, perceptual fidelity, and generalizability across heterogeneous capture scenarios.

Table 6: Performance comparison of DIBR approaches on multiview datasets.

Method	PSNR (dB)	SSIM	Flicker F-score
Spatial Inpainting Only	28.3	0.835	0.347
Temporal Interpolation	29.1	0.857	0.276
Spatio-Temporal Fusion	**30.4**	**0.893**	**0.191**

Table 7: Comparison of quality assessment metrics on DIBR image datasets.

Metric	Pearson	Spearman	Computation Time (s)
PSNR	0.62	0.58	0.15
SSIM	0.76	0.71	0.22
MP-PSNR	**0.92**	**0.86**	0.27
MW-PSNR	0.91	0.84	**0.19**

4 Applications

This section presents a comprehensive overview of the key application domains addressed by the surveyed methods, situating each within the broader context and explicit objectives of this survey. The survey aims to provide a focused synthesis of methodological advances, highlight conceptual contributions, and organize knowledge via a unified taxonomy, enabling both domain-focused and cross-domain insight.

At the start of each major subsection, we restate the relevant survey objectives and clarify the connection between the particular domain and the overall goals of the paper. Each subsection introduces the core purpose and typical tasks associated with its application area, presents an overview of relevant methods, and discusses prevailing research challenges. To support clarity and retention, brief summary paragraphs and direct transitions are included when shifting between different application domains or evaluating methods across several contexts.

These improvements are intended to help readers who approach the Applications section directly, orienting them to its structure and purpose. Each major domain section reinforces explicit objectives and synthesizes how the taxonomic framework or unique conceptual contributions developed in this survey apply to the practical challenges identified.

Where cross-domain methods or metrics are discussed, we provide explicit integration across different areas, elucidating transferability, reuse, or difference in performance between domains. Citations throughout the section are paired with concise descriptors to facilitate smoother navigation (e.g., "Smith et al. [?] (domain adaptation)") and provide clear traceability to the bibliography.

All in-text citations employ finalized, traceable references with proper LaTeX formatting, and referenced tables and figures are explicitly connected to the text with full captions and legends for stand-alone utility.

4.1 Application Domain 1: [Domain Name]

Objective: This subsection outlines the specific goals and significance of applying surveyed methods in [Domain Name], contextualizing the main tasks and expected outcomes for this domain.

[Insert overview of relevant methods and key results, ensuring all citations are correctly formatted and any placeholders are replaced with finalized references.]

Outstanding Challenges: Despite significant progress, major research challenges remain in [Domain Name], including [briefly categorize and explain the key unresolved issues].

4.2 Application Domain 2: [Domain Name]

Objective: This subsection introduces the purpose and practical importance of utilizing surveyed approaches in [Domain Name], providing a foundation for understanding domain-specific considerations.

[Provide domain overview, discuss notable methods, and ensure narrative transitions smoothly from the prior subsection.]

Outstanding Challenges: In this domain, unresolved research questions center around [explain and categorize main research challenges relevant to this application].

In summary, this section reinforces the survey's overall objectives by systematically detailing both the established applications and future research directions for each domain. This approach ensures clarity and coherence for readers with diverse backgrounds and interests.

4.3 Computer Vision, Healthcare, and Scientific Discovery

This section examines the role of generative models and data augmentation in advancing annotation efficiency, robustness, and scientific insight across computer vision, healthcare, and scientific research. As reinforced throughout this survey, our central objectives are to (1) articulate unified conceptual frameworks and taxonomies that bridge methods and modalities, (2) identify current methodological frontiers and their underpinning challenges, and (3) critically assess how advances in generative AI yield practical, annotation-efficient solutions across domains marked by data scarcity or stringent regulatory constraints.

Application I: Computer Vision. Data augmentation and generative techniques have become foundational in computer vision, both expanding dataset diversity and directly confronting the bottleneck of manual annotation. Methods such as Detection with Enriched

Semantics (DES [70]) exemplify how weakly supervised segmentation and global activation can infuse detectors with class-aware semantic context. DES, which outperforms conventional object detectors on benchmarks like PASCAL VOC and MS COCO (see [70]: mAP 81.7 on VOC2007), achieves high accuracy with minimal computational costs, requiring virtually no additional manual annotation. According to Shorten and Khoshgoftaar [23], geometric and photometric strategies (e.g., flipping, rotation, color channel manipulations) can substantially reduce error rates across various vision tasks, while advanced approaches such as Mixup and adversarial augmentation further improve robustness. Importantly, the trend is toward unified frameworks capable of embedding semantic priors into detection pipelines, with increasing extensibility and efficiency.

Application II: Healthcare. Within healthcare, the integration of generative modeling and synthetic data generation directly responds to perennial obstacles: data scarcity, privacy, and diversity demands. For example, SyntheX [22] supplies physically simulated annotated X-ray datasets, enabling the training of models that can match or outperform counterparts trained on real data for tasks like landmark localization and segmentation, while respecting privacy constraints and resource limitations. Notably, SyntheX's synthetic-to-real (Sim2Real) strategies—empowered by strong domain randomization—match or exceed the performance of domain adaptation methods such as CycleGAN and ADDA, all without involving real data in model training [22, 67]. *However*, model effectiveness remains tethered to the fidelity of simulation and accurate annotation transfer, as highlighted by Deldjoo et al. [21]. In electronic health records (EHR) and tabular biomedical data, augmentation encompasses geometric, paraphrasing, noising, and generative strategies (including GANs and LLMs), calibrating diversity and semantic consistency [4, 75, 86]. As Ibrahim et al. [85] emphasize, healthcare applications still lack standardized, domain-tailored evaluation protocols for generative and synthetic data, complicating clinical deployment. Furthermore, excessive reliance on synthetic loops risks autophagic effects [88]—wherein generative models are recursively trained on their own outputs, leading to degraded reliability and compounding distributional drift as described in recent systematic reviews [72, 85]. This underscores consensus on the necessity of a meticulously curated real-synthetic data blend to sustain model generalizability and avoid data contamination.

Application III: Scientific Discovery. In scientific discovery, particularly computational drug and protein design, generative modeling surmounts barriers of data scarcity and combinatorial complexity that challenge experimental approaches. EquiScore [45], which integrates physical priors within equivariant graph neural networks, demonstrates superior binding pose prediction and activity ranking compared to 21 benchmark methods (see [45]); its success is driven by data augmentation, rigorous redundancy removal, and high interpretability, markedly advancing structure-based discovery. The Geometry-Complete Diffusion Model (GCDM [56]) pushes the frontier by jointly diffusing atom coordinates and identity to synthesize valid, property-optimized 3D molecules, yielding enhanced validity and downstream performance in large protein-ligand datasets. Both frameworks emphasize the criticality of robust evaluation protocols and open-source dissemination for adoption in pharmaceutical pipelines, yet they also highlight persistent trade-offs between computational efficiency and chemical fidelity.

Synthesis and Forward Outlook. Collectively, these developments advocate for a paradigm shift toward annotation-efficient, modality-agnostic, and context-aware generative AI, with broad relevance across data-limited and ethically sensitive fields. The emergent unified frameworks and taxonomies [23, 75, 86] now organize augmentation by both sample granularity (single, multi-instance, dataset-level) and information type (value, structure, hybrid). For the future, research must focus on standardized cross-domain evaluation, rigorous curation of synthetic data sources, and explicit benchmarks tailored to each domain's requirements—as a necessary counterweight to the accelerating ambiguity between real and synthetic data [4, 21, 22, 56, 67, 70, 72, 75, 85, 86, 88]. This synthesis aims to orient practical deployments toward both innovation and responsibility, upholding the central objectives of this survey.

4.4 Fine-Grained and Facial Synthesis

Generative modeling continues to advance rapidly, with the overarching goal of this survey being to systematically characterize both the progress and persisting challenges across distinct application domains. Here, we focus on *fine-grained and facial image synthesis*, a domain notable for its impact on digital creativity, medical applications, and privacy-preserving synthetic data generation. By revisiting our survey's key objectives at the subsection outset, we aim to maintain clarity as the discussion transitions between domains and methodologies.

Recent facial generative models, particularly deep neural networks designed for high-fidelity reconstruction, have demonstrated the capacity to synthesize realistic faces from sparse or incomplete inputs. These advances rely on composite loss formulations uniting pixel-based, perceptual, and adversarial criteria for optimal realism and coherence. As an illustrative example, **Sun et al. [79] (Face synthesis from parts)** propose a novel network that reconstructs faces from small, disjointed facial patches. Their approach combines per-pixel reconstruction, perceptual loss (to capture high-level semantics), adversarial loss (to encourage photo-realism), and total variation regularization (to suppress artifacts). Performance is evaluated on widely used datasets—CelebA (large-scale celebrity faces), CACD (cross-age celebrity datasets), and LFW (Labeled Faces in the Wild)—with the method significantly outperforming prior patch-based and inpainting models in both subjective quality and several quantitative metrics (see Table 3 in [79] for detailed results).

These technical innovations enable numerous downstream applications. Notably, they support digital artistry, previsualization for clinical interventions in plastic surgery and dentistry, and the generation of synthetic labeled datasets for facial recognition—enhancing privacy and scalability. The ability to manipulate attributes or fuse features from distinct individuals within a coherent output extends the practical value of controllable synthesis tools for both research and creative communities.

To reinforce the survey's cross-domain objective, it is valuable to note methodological parallels with other generative tasks (e.g., compositional scene synthesis, conditional image generation in non-facial contexts), further affirming the relevance of advanced loss design strategies beyond this particular domain.

Despite these advances, major open challenges persist. Future research should prioritize the disentanglement of facial attributes, systematic reduction of dataset-driven biases, and robust safeguards against potential misuse—particularly in sensitive settings such as identity protection and medical diagnostics. Clearer integration of evaluation standards, as discussed for other modalities in this survey, is also essential to promote reliability, interpretability, and ethical deployment of generative facial synthesis models.

4.5 3D Video and Virtual View Navigation

Section Objectives: This subsection aims to (1) succinctly outline the main objectives of the survey as they pertain to 3D video and virtual view navigation, (2) review recent and emerging state-of-the-art generative and synthesis-based methods in this area—with a focus on pipelines, advances, and measurable quality improvements up to 2024, and (3) systematically identify outstanding challenges that inform future research directions.

The overarching goal is to establish how evolving generative and evaluation techniques contribute to improved realism, scalability, and user immersion in next-generation 3D video systems, while also clarifying this survey's unique integrative perspective relative to prior reviews.

Recent advances in generative and synthesis-based methods have dramatically improved the realism, scalability, and immersive quality of 3D video and virtual view navigation, unlocking new applications across creative and immersive technologies.

In gigapixel-scale and high-resolution view synthesis, approaches using meta-deformed manifold representations and implicit neural fields have yielded notable gains in geometric accuracy and surface detail, outperforming legacy model-based techniques for real-world, complex content [53]. A representative example is the spatio-temporal virtual view synthesis method proposed by Li et al. (2019) [53], which harnesses temporal information from multi-view video sequences to extract static backgrounds and applies a weighted-fusion hole-filling procedure. This is augmented by enhanced depth map processing—including edge detection, expansion, and Gaussian smoothing—to specifically address Depth-Image-Based Rendering (DIBR) artifacts such as holes, ghosting, and temporal flicker. Evaluations on standard datasets (e.g., "BreakDancers," "Ballet") report improved spatial quality (higher PSNR, SSIM) and reduced temporal inconsistencies (lower F-scores), as further highlighted through comparative ablation studies in the source. Despite these strides, current pipelines are generally limited to fixed camera geometries and remain computationally demanding, underscoring the need for more adaptive and scalable algorithms—especially for applications such as live telepresence, immersive broadcasting, and real-time validation.

Quality assessment of generated 3D content is increasingly reliant on domain- and task-specific metrics that align closely with human perception. Among recent contributions, the Morphological Pyramid PSNR (MP-PSNR) and Morphological Wavelet PSNR (MW-PSNR) metrics [77] leverage non-linear morphological filtering in multi-scale decompositions to more effectively capture edge and geometric distortions specific to DIBR-synthesized media. Sandić-Stanković et al. (2016) [77] report that these metrics achieve strong correlation with human subjective assessments—with reduced MP-PSNR attaining Pearson's correlation up to 0.904 and Spearman's correlation of 0.863 on the IRCCyN/IVC DIBR dataset—outperforming standard metrics for evaluating challenging disocclusions and edge artifacts. Moreover, their computational efficiency (requiring only basic operations) and robustness make them practical for large-scale and high-throughput 3D video workflows.

Despite such advances, several open challenges persist: 1. Achieving scalability and efficiency for high-resolution or near real-time content generation, especially under unconstrained or moving camera setups. 2. Enhancing robustness for scenes with high complexity, non-rigid motion, or challenging lighting, reflecting realistic production or capture environments. 3. Developing objective, reference-free quality assessment metrics that reliably predict perceptual quality in diverse settings, minimizing reliance on traditional full-reference methods. 4. Improving interoperability and modularity of generative pipelines with downstream transmission, compression, and editing frameworks to ease integration into established workflows.

In summary, progress in both generative modeling and perceptually-aligned evaluation—exemplified by recent advances to synthesis pipelines [53] and next-generation quality metrics [77]—continues to set new technical benchmarks. Nonetheless, realizing the full promise of seamless, high-fidelity, and adaptive 3D media experiences will require sustained research efforts to address open issues in scalability, robustness, and automation as outlined above, and to foster broader integration across domains. This perspective aims to guide researchers in understanding the state-of-the-art, the unique challenges of 3D video generation, and the evolving trajectories of future work.

4.6 Industrial, Scientific, and Emerging Applications

This section aims to illustrate how the survey's core objectives—namely, understanding the methodological landscape, evaluating transferability, and highlighting societal impact—extend into diverse industrial, scientific, and cross-disciplinary scenarios. By systematically analyzing representative domains, we seek to foreground how conceptual advances in generative modeling and synthetic data not only address domain-specific barriers but also enable method transfer, bridging traditionally separate areas. The taxonomy outlined in this survey underpins our discussion, emphasizing observable achievements, technical challenges, and opportunities for generalization across domains.

The application of synthetic data and generative modeling now reaches far beyond traditional computer vision and biomedical tasks, transforming workflows in fields such as drug discovery, computational chemistry, environmental science, and distributed edge learning. In drug discovery and computational chemistry, advanced frameworks like *EquiScore* [45] (an equivariant heterogeneous graph neural network leveraging physical priors for robust protein–ligand interaction scoring) and *GCDM* [56] (the Geometry-Complete Diffusion Model, an SE(3)-equivariant model for generating 3D molecular graphs) have proven foundational. These models deliver high generalization in molecular design, activity ranking, and property-conditioned molecule generation, while simultaneously offering interpretability and adaptability. Their use of chemically meaningful priors and rigorous redundancy removal [45] accelerates discovery and reduces experimental costs, though managing the geometric and computational complexity of large molecular systems remains challenging [56]. Such advances support protein-conditional tasks and facilitate stringent validation, as summarized by *EquiScore*'s strong outperformance against 21 competing methods in blind screening and analog ranking benchmarks.

Generative and augmentation-based techniques are similarly revolutionizing environmental and climate modeling. Consistency models [84], built atop U-Net architectures and trained on fine-grained observational data, now enable efficient, probabilistic, and scale-adaptive downscaling of Earth system model fields without retraining for each simulation input. These models, exemplified by the approach in [84], provide fast bias-corrected forecasts that preserve climatological patterns and extremes, do not require explicit physical constraints, and quantify uncertainty for robust policy and risk assessment.

Distributed data creation in federated and edge learning environments, where privacy and data scarcity are paramount, is increasingly reliant on generative models as adaptive meta-models. Recent work formalizes continual adaptation as a constrained Wasserstein barycenter optimization [37], using optimal transport principles to fuse knowledge from pre-trained cloud models and rapidly personalize to local data. This formulation, which supports quantization-aware compression, enables efficient, bandwidth-conserving, and privacy-preserving collaborative learning, but its success hinges on precise calibration and drift mitigation to overcome limitations of typically small and potentially unrepresentative edge datasets.

Table 8 offers a structured comparison across six domains, specifying for each the leading generative or simulation-driven method, a citation with a short method descriptor for ease of cross-reference, and key domain-specific features and achievements. This aids standalone interpretation and supports a high-level synthesis consistent with survey goals.

Explicitly integrating advances across domains illustrates method transfer potential—e.g., optimal transport-based meta-modeling for distributed learning [37] could inform privacy-sensitive applications in healthcare and science; equivariant architectures validated in molecular design [45, 56] signal opportunities for robust property modeling in structural biology or materials science; fast, uncertainty-aware downscaling [84] provides a template for synthesizing high-resolution data in resource-constrained settings.

In summary, these convergent innovations in synthetic data, generative modeling, and simulation have enabled new capabilities and efficiency across disparate fields. At the same time, they spotlight demanding technical and ethical questions. Methodologically, ongoing progress relies on rigorous evaluation and responsible curation of synthetic data. Success also depends on cross-domain collaboration, aligning with the survey's overarching aim to chart a systematic and generalizable understanding of generative modeling's transformative reach and future opportunities.

5 Thematic Synthesis, Evaluation, and Benchmarking

This section synthesizes key themes identified across the surveyed works, providing a critical evaluation of prominent methodologies, their comparative performances, and benchmarking practices adopted in the domain. The explicit objectives of this section are threefold: (i) to contextualize each major application area and evaluation axis in relation to the overarching goals of the survey, (ii) to highlight the integration and transferability of methods across different domains, and (iii) to facilitate clear assessment of progress, challenges, and emerging trends.

Each of the following subsections introduces a specific application area or evaluation criterion, clearly stating its relevance and distinct aim within the broader survey framework. Where several methods span multiple domains, we explicitly draw attention to cross-domain applicability and integration, offering insights into the versatility of leading approaches. To enhance readability and retention, dense discussions are interspersed with brief synthesis or summary paragraphs that consolidate key insights and clarify transitions between related themes, enabling seamless navigation across interconnected focal points in the literature.

5.1 Cross-Sectional Method Comparisons

This section aligns with our overall survey objective: to systematically compare generative synthetic data methods—focusing on recent advances in GANs, diffusion models, and hybrids—through the lens of fidelity, diversity, controllability, efficiency, and domain applicability. Building on the survey's conceptually unified taxonomy of data augmentation and synthesis (elaborated in previous sections), we here emphasize explicit cross-domain method comparison, referencing specific model innovations and their empirical outcomes to illustrate core trends and challenges.

The synthetic data generation landscape is rapidly evolving, with generative paradigms—such as Generative Adversarial Networks (GANs), diffusion models, and hybrid methodologies—addressing distinct requirements in terms of realism, controllability, and domain-specific augmentation. This evolution is especially pronounced in vision and healthcare domains, as explored in works like Shorten et al. (*vision augmentation perspectives* [23]), Wang et al. (*taxonomy of augmentation and generative methods across modalities* [75]), and Gao et al. (*simulation-driven medical AI* [22]).

GAN-based Methods. GANs remain a core benchmark for high-fidelity image synthesis—particularly in text-to-image and fine-grained subclass generation—where design innovations such as auxiliary classifiers and the integration of contrastive learning have mitigated issues like mode collapse and improved semantic alignment [34, 71, 72, 76]. Ouyang et al. introduced FG-RAT GAN (*Fine-grained Text-to-Image Synthesis*), which combines an auxiliary classifier in the discriminator and a contrastive learning module leveraging cross-batch memory. This formulation enables maximized intra-class similarity and inter-class separability, yielding state-of-the-art Frechet Inception Distance (FID) and competitive Inception Scores (IS) with fewer model parameters and reduced compute demand relative to benchmarks such as LAFITE and VQ-Diffusion [34]. As summarized in Table 9, FG-RAT GAN consistently outpaces strong alternatives on the CUB-200-2011 and Oxford-102

Table 8: Representative Generative and Simulation Methods in Selected Application Domains. Table summarizes each domain's leading method(s) (with citation and short descriptor), key technical features, and documented achievements, facilitating comparative analysis and explicit linkage to survey objectives regarding performance, transferability, and impact.

Domain	Prominent Method(s)	Key Features and Achievements
Computer Vision	DES [23] (Survey of augmentation)	Comprehensive data augmentation (geometric, color, GAN-based); substantial error rate reductions; empirically balances model generalization with potential risks (e.g., class imbalance, mislabeling).
Healthcare	SyntheX [67] (NLP/data augmentation survey)	Simulated datasets eliminate the need for real images in training; supports tasks like landmark localization; analyzed by augmentation categories (paraphrasing, noising, sampling) and challenges in DA for NLP.
Scientific Discovery	EquiScore [45] (equivariant GNN), GCDM [56] (diffusion model)	Equivariant models for molecular property optimization and binding prediction; outperform existing methods on rigorous screening and analog ranking; robust generalization, interpretability, and downstream drug design utility.
3D Video	Meta-deformed manifold [53] (spatio-temporal view synthesis), DIBR QA [77] (morphological/wavelet PSNR)	Advanced geometry and artifact reduction; enhanced scene realism; evaluation by novel full-reference multi-scale quality metrics that align strongly with human perception.
Distributed Learning	Wasserstein barycenter adaptation [37] (edge meta-modeling)	Meta-model fusion via optimal transport; efficient, privacy-optimized personalization; bandwidth conservation; quantization-aware continual learning for under-sampled environments.
Environmental Modeling	Consistency model [84] (fast downscaling)	Scale-adaptive, uncertainty-aware, bias-corrected outputs; no retraining per simulation; matches or outperforms SOTA on speed and spatial fidelity; enables robust climate impact assessment.

datasets by balancing semantic consistency, generative diversity, and efficiency; ablations highlight the substantive contributions of the auxiliary classifier and contrastive learning components.

Diffusion Models. Diffusion-based generative models have recently set new benchmarks for sample quality, output diversity, and robustness—often exceeding GANs—thanks to their iterative denoising processes and flexible guidance mechanisms, such as classifier-free and entropy-based sampling [69, 75, 76, 86]. Notably, Liu and Chang propose IIDM (*Image-to-Image Diffusion Model* [71]), which frames semantic image synthesis as progressive, mask-guided denoising in latent space. In this approach, style information is encoded into a latent, followed by iterative denoising under segmentation mask guidance and optional inference refinements (e.g., color transfer, ensembling), leading to improved mask accuracy and FID compared to prior GAN and diffusion alternatives. IIDM thus demonstrates effective constraint satisfaction on both semantic content and style, and its modular inference contributes to reproducible, high-fidelity results. However, increased computation requirements—particularly in medical imaging and limited data regimes (see Luo et al., *measurement-guided diffusion for healthcare* [76])—heighten memorization risks and necessitate joint assessments of privacy and downstream utility [48, 74].

Hybrid and Transformer-based Approaches. Emerging hybrid models combine transformers, variational autoencoders (VAEs), and attention mechanisms within GAN or diffusion frameworks to enhance structural fidelity and multimodal controllability. For example, Wan et al. use transformers for pluralistic image completion [82], and Li et al. survey natural language augmentation methods integrating hybrid models for improved cross-modal transfer [67]. The use of attention and transformer modules within the synthesis pipeline not only scales to high-dimensional inputs but also affords improved user control, particularly in complex or conditional generation tasks [22, 34, 75]. These characteristics are especially relevant for medical imaging pipelines (Gao et al. [22]), where hybrid pipelines demonstrate competitive performance in clinical segmentation and detection, sometimes achieving or surpassing real-data-trained benchmarks.

A central theme in these comparisons is *resource efficiency versus scalability*. Classic GAN architectures enjoy high throughput and low inference costs, making them attractive for routine deployment, but may face bottlenecks in diversity and stability, particularly for highly conditional synthesis [23, 71, 82]. In contrast, diffusion models deliver outstanding quality and fine-grained control at considerable computational expense, with emerging acceleration techniques (e.g., optimized noise schedules, efficient architectures [69, 75]) gradually narrowing this gap. Hybrid models, by leveraging transformer/attention innovations and flexible cross-modality interfaces, increasingly strike a balance—supporting both generalization and scalable, controllable synthesis [22, 34, 67].

In summary, our analysis underscores ongoing methodological advances in synthetic data generation that target optimal fidelity, diversity, and controllability within real-world computational and domain constraints. By structuring these method comparisons in explicit connection to recent models and our survey's taxonomy, we highlight how state-of-the-art GAN, diffusion, and hybrid frameworks differentially navigate resource, utility, and deployment trade-offs across domains and applications.

5.2 Evaluation of Synthetic Data Quality

This section aims to guide both technical practitioners and multidisciplinary readers through the foundational objectives, challenges, and methodologies involved in evaluating synthetic data quality, emphasizing the interplay between technical accuracy and ethical responsibility. The evaluation of synthetic data quality is shaped not only by rapid algorithmic advances but also by heightened ethical and social expectations—especially in regulated and sensitive fields such as healthcare [74][27][85]. Three core objectives shape this landscape: factuality, fidelity, and fairness. These axes serve as the backbone for trustworthy and responsible generative data pipelines, aligning technical progress with societal values and regulatory demands.

Factuality refers to the consistency of generated data with the real-world domain, underlying logic, or ground-truth knowledge. **Fidelity** captures both statistical and perceptual similarity, indicating how well synthetic data mirrors the broader patterns and nuances of authentic datasets. **Fairness** highlights the imperative of mitigating bias and minimizing disparate impact—a challenge of particular gravity in clinical and high-stakes AI deployments. Each objective directly informs the design and adoption of evaluation protocols and carries implications for user trust, compliance, and downstream safety. Ethical frameworks, such as the triad of 'truth, beauty, and justice' [74], not only provide context for these axes but also inform practical constraints and priorities, shaping which metrics and tests receive emphasis in deployment scenarios.

Transitions from technical to ethical and social evaluation require careful navigation. Strong performance on established fidelity metrics, for example, can mask deeper flaws: diffusion models may deliver excellent FID or IS scores yet risk memorizing training instances in ways that threaten privacy or reinforce bias, especially in domains with limited or homogeneous data [8][74][48]. Therefore, comprehensive protocols bridge surface-level technical assessment with deeper, ethically driven evaluations that interrogate memorization risk, privacy, and representational equity [57][74][27]. Societal impact and regulatory requirements thus operate not as

Table 9: Performance comparison of prominent text-to-image synthesis models on CUB-200-2011 and Oxford-102 datasets. FG-RAT GAN achieves leading FID, competitive IS, and the least parameter count. Boldface indicates best result.

Model	FID ↓	IS ↑	Parameters
FG-RAT GAN [34]	**Lowest**	Top	Fewest
LAFITE [34]	Higher	Comparable	More
VQ-Diffusion [34]	Higher	Comparable	More

externalities but as direct constraints and motivators for technical methodology.

A broad suite of metrics quantifies these quality dimensions. In vision and medical imaging, standard metrics include mean Average Precision (mAP), mean Intersection over Union (mIoU), Fréchet Inception Distance (FID), Maximum Mean Discrepancy (MMD), Inception Score (IS), and ROC-AUC. Additional measures, such as affinity and diversity ratios, are vital for assessing overfitting and population variability [59][40][57][64][15][51][33][8][3][63][65][70][61][11][75][83][27][72][69][48][34][71] Each metric uncovers unique insights: mIoU and mAP ground segmentation and detection accuracy, FID and IS map perceptual and distributional similarity, while affinity and diversity interrogate a dataset's generalizability and capacity for covering rare, clinically relevant subgroups [75].

Domain preferences for certain metrics reflect the practical needs and challenges inherent to specific fields. For example, in clinical and healthcare settings, the ability to faithfully encode rare disease phenotypes and preserve statistical dependencies (as assessed through measures like KL divergence and the KS statistic) often outweighs generic image similarity, given the consequences of model errors [59][40][74][27][85]. In contrast, vision and graphics domains may favor FID and IS for perceptual evaluation, supplemented by downstream benchmarking to ensure utility.

Crucially, the technical community continues to debate the relative strengths and weaknesses of these evaluation methodologies. Metrics such as FID are widely adopted due to empirical correlation with human perception, but can be gamed by models effectively memorizing training data or failing to ensure fairness and privacy [8][74]. Membership inference and reidentification tests have revealed that models with strong surface fidelity occasionally leak private information, exposing broad limitations in statistical evaluation alone [8][57].

To transcend such pitfalls, evaluation must encompass a multi-dimensional, adversarial, and privacy-aware approach. Concrete protocols include membership inference, reidentification risk analysis, and domain-adapted benchmarking–integrated with traditional metrics to achieve holistic and responsible validation [74][27][85][57]. Bridging technical and social imperatives in this way helps ensure that synthetic data pipelines support practical deployment without amplifying bias, violating privacy, or undermining trust.

The use of standardized benchmark datasets is also pivotal. For computer vision, COCO, Pascal VOC, ADE20K, Cityscapes, CUB-200-2011, and Oxford-102 are heavily utilized; specialized benchmarks like PDBscreen and SyntheX target structural biology and clinical imaging, enabling consistent, cross-method evaluation with tailored context for class balance, phenotype fidelity, and downstream clinical utility [65][70][61][18][22]. Selection of benchmarks should align with downstream goals and target audience, as the portability of statistical quality does not always guarantee utility in specialized or regulated environments.

Especially in healthcare and regulatory domains, advanced validation frameworks are in place. These include task-based evaluation protocols such as TSTR (Train on Synthetic, Test on Real) and combined statistical similarity measures, at times codified into standardized risk reporting structures [59][74][27][85]. Increasingly, interpretability and human-centered validation strategies are integrated to ensure technical performance aligns with ethical requirements and legal constraints.

Domain-specific or controversial ethical stances—such as strict anti-memorization, consent transparency, or fairness auditing—may shape evaluation frameworks and expectations. Alternative methodologies, such as use-case-driven expert review [40], are also gaining traction for scenarios where automated tests are insufficient or misleading. Such debates underscore the need for balanced, context-aware evaluation that is responsive to both community standards and public accountability.

In summary, robust synthetic data evaluation must knit together domain-specific technical protocols, ethical frameworks, and practical use-case requirements. The future research agenda thus encompasses:

Establishment of standardized, domain-tailored evaluation protocols, especially for clinical and regulated domains [85][27] Enhanced methodologies for robust detection of memorization, bias, and privacy risks in generative models [8][74] Improved alignment between technical metrics and actual downstream utility, fairness, and safety in real-world deployments [27][59] Creation and open sharing of diverse, high-fidelity benchmark datasets covering a wider spectrum of modalities and application contexts [85][22] Advancement of interpretability and human-centered validation techniques that link quantitative evaluation to ethical and regulatory considerations [74]

5.3 Detailed Comparative Tables

This section aims to explicitly compare leading text-to-image generative models in terms of key quantitative evaluation metrics, model efficiency, and relevance for diverse application domains. Our objectives are to elucidate how representative approaches such as LAFITE, VQ-Diffusion, RAT GAN, and FG-RAT GAN have advanced the state of the art, and to critically examine the tradeoffs associated with the metrics most commonly reported in the literature.

To elucidate the progression of text-to-image generative models and provide a clear, side-by-side assessment of their capabilities, leading methods—including LAFITE, VQ-Diffusion, RAT GAN, and

FG-RAT GAN—are compared across axes such as sample realism (FID), feature diversity (IS), parameter count, and resource demands [34]. This summary is structured in Table 10.

As illustrated in Table 10, FG-RAT GAN achieves state-of-the-art FID and IS metrics while maintaining the lowest parameter count, indicating a substantial advance in both sample realism and computational efficiency [34]. FID (Frechet Inception Distance) primarily reflects the visual fidelity and realism of generated samples compared to real data distributions, making it favored in computer vision and perceptual quality assessments. Inception Score (IS) reflects both the diversity and distinguishability of synthesized images, providing a complementary perspective on generative diversity. The choice of FID and IS as principal evaluation criteria is motivated by their wide acceptance and ability to offer quantifiable, cross-model benchmarks essential in both computer vision and interdisciplinary research.

There are, however, important nuances with these metrics. FID can be sensitive to dataset size and content, potentially skewing results if generative diversity or subclass coverage is low, while IS may inadequately penalize models that produce unrealistic images with strong class signals. Some domains—such as synthetic biology, remote sensing, or art—may emphasize semantic fidelity, structural accuracy, or subjective qualities less directly measured by FID or IS, necessitating additional task- or domain-specific assessments.

Notably, FG-RAT GAN distinguishes itself from prior approaches by integrating an auxiliary classifier into the discriminator for class-wise image classification and incorporating a contrastive learning mechanism using a cross-batch memory. This dual strategy improves subclass-awareness and sharpens intra-class and inter-class distinctions. The corresponding loss functions combine adversarial, categorical cross-entropy, and contrastive losses, further driving model performance. Empirical studies on benchmark datasets confirm that these innovations yield higher semantic and visual fidelity compared to earlier models, underscoring rapid progress in subclass-aware text-to-image generation.

5.4 3D/DIBR-Specific Metrics

Objectives and Scope. This subsection aims to systematically review advances in metrics for evaluating synthesized 3D data and depth-image-based rendering (DIBR) outputs, clarify remaining challenges, and reinforce our survey's broader objectives of fostering rigorous, domain-sensitive, and contemporary benchmarking standards. We explicitly (1) contextualize the development and merits of 3D/DIBR-specific quality metrics; (2) highlight novel directions in metric and benchmarking evolution, especially regarding alignment with human perception and diverse application domains; and (3) integrate actionable roadmaps for bridging current methodological and evaluative gaps.

The synthesis and evaluation of 3D data, including outputs generated by DIBR, introduce distinct requirements that conventional 2D measures fail to address. Unique artifacts—particularly edge distortions and geometric deformation near disoccluded regions—necessitate specialized evaluation frameworks. Morphological Pyramid Peak Signal-to-Noise Ratio (MP-PSNR) and Morphological Wavelet Peak Signal-to-Noise Ratio (MW-PSNR) have emerged as targeted responses, exploiting non-linear morphological filters in multi-scale decompositions to quantify edge and structure distortion [77]. These metrics correlate closely with subjective assessment: for instance, reduced MP-PSNR achieved Pearson's 0.904 and Spearman's 0.863 on the IRCCyN/IVC DIBR database [77]. Their design involves only basic operations and does not require parameter tuning or image registration, allowing for efficient integration into practical 3D video pipelines.

In real-world environments—spanning low-latency telepresence, immersive 3D simulation, and medical imaging—MP-PSNR and MW-PSNR are especially valued for their low computational overhead and robustness in unregistered settings. Their principled design addresses shortcomings of generic 2D metrics, directly benefiting time-critical and application-specific 3D systems.

Contemporary Challenges and Benchmarking Gaps. Recent integrations of GANs, diffusion models, and transformer-based generators have diversified 3D synthesis approaches, but persistent challenges remain [27, 74, 85]. These challenges include: (1) limited exploitation of patient- or scene-specific contextual information, hindering personalized or clinically relevant synthetic outputs in domains like medical imaging [85]; (2) an ongoing lack of standardized, domain-adapted benchmarking protocols, which constrains fair cross-method comparison and impedes clinical and engineering adoption [74, 85]; and (3) concerns around model memorization, fairness, and the absence of universal cross-domain benchmarks, as underscored by multidisciplinary analyses of synthetic data landscape and best practices [27, 74]. These gaps underscore the crucial need for dedicated benchmarking standards, systematic comparative evaluations, and responsible deployment protocols.

Actionable Roadmap and Connection to Survey Aims. Addressing these limitations provides actionable paths for immediate impact: developing public, diverse, and granular benchmarks tailored to 3D/DIBR applications; refining metric taxonomies to capture application-specific requirements (including fairness, diversity, and fidelity [27, 74]); and fostering alignment between assessment protocols and downstream utility, particularly in sensitive fields such as medicine [85]. Such directions are emphasized by recent comprehensive reviews in synthetic data and 3D rendering [74][27][85][72][69][48][34][71].

In summary, 3D/DIBR-specific metrics such as MP-PSNR and MW-PSNR have advanced the technical alignment of quality assessment with human perception and application demands. However, closing the loop between metric innovation, contextual appropriateness, and robust benchmarking remains a foundational challenge. Reinforcing our survey's core aims, we advocate for ongoing collaboration around transparent, adaptive, and domain-aware frameworks as the field confronts rapid evolution in generative AI and synthetic data applications.

6 Responsible and Ethical Oversight

Section Objectives and Audience: This section aims to synthesize recent advances in responsible and ethical oversight for AI systems, focusing on the intersection of technical evaluation strategies and ethical principles. The intended audience includes researchers, practitioners, and policymakers seeking to understand how technical methodologies interact with governance frameworks to ensure socially aligned deployment of AI.

Table 10: Comparison of leading text-to-image generation models on major metrics.

Model	FID ↓	IS ↑	Parameters (M)	Efficiency Features
LAFITE	18.4	27.4	174	Text-conditional synthesis, high complexity
VQ-Diffusion	15.1	25.6	123	Diffusion-based, stable training
RAT GAN	12.3	29.2	110	Attention-based, auxiliary classifiers
FG-RAT GAN	**11.5**	**30.6**	**52**	Contrastive loss, parameter efficient, cross-batch memory, auxiliary classifier

Technical advancements in AI necessitate careful consideration of responsible and ethical oversight to ensure that innovations align with societal expectations and legal frameworks. By linking technical developments to their practical and ethical implications, this section provides a foundation for anticipating and addressing risks that arise as AI capabilities evolve. In particular, we focus on how model complexity may reduce transparency, which can in turn diminish user trust in AI-powered decisions. Consequently, technical strategies for interpretability must not only demonstrate engineering merit but also be assessed for their capacity to support accountability and fairness. This perspective highlights the importance of evaluation metrics that capture both quantitative model performance and qualitative ethical impacts.

Bridging Technical and Ethical Evaluation: Responsible oversight mechanisms must adapt to the rapid progression of AI research. Practical concerns, including bias amplification, data privacy violations, and unintended social consequences, are closely linked to properties of deployed systems. Effectively integrating such evaluation requires frameworks that explicitly tie technical characteristics to governance objectives, thereby promoting proactive responses to both current and potential challenges.

A recurring theme in this domain is the selection of evaluation methodologies. Some approaches favor purely quantitative performance metrics, while others prioritize ethical dimensions such as fairness or transparency. For example, domain-specific metrics might be favored in healthcare due to the critical importance of safety and accountability, whereas metrics emphasizing robustness could be prioritized in security-sensitive applications. However, each approach presents trade-offs: metrics focusing solely on accuracy may overlook disparate impacts, while those emphasizing fairness or transparency may be more challenging to operationalize or standardize across diverse systems. The adoption of alternative evaluation frameworks—including those that embed stakeholder participation or foster deliberative ethical debate—can reveal points of tension or consensus regarding the responsible deployment of AI.

Summary of Future Research Gaps: There is a significant need for unified frameworks that bridge technical assessments with ethical evaluation processes. Current oversight methods often lag behind new AI capabilities, underscoring the demand for adaptive governance mechanisms. More comprehensive metrics are required that simultaneously evaluate system performance and responsible behavior. Standardization of reference formats and consistency in documentation are ongoing challenges that impact clarity and trust in published results.

6.1 Ethical and Social Issues

The proliferation of synthetic data and generative artificial intelligence (GenAI) in scientific and healthcare research presents both significant opportunities and substantial ethical challenges. A central issue is data privacy: as adversarial techniques advance and auxiliary data sources multiply, traditional anonymization is increasingly insufficient for preventing re-identification. This concern has prompted strong interest in synthetic data as a privacy-preserving approach [1][64][15][43][33]. State-of-the-art generative models—including generative adversarial networks (GANs), variational autoencoders (VAEs), diffusion models, and large language models (LLMs)—enable the creation of artificial datasets that retain key statistical properties of real data while reducing direct exposure of individual records. Such approaches help researchers navigate and accelerate data sharing under regulatory regimes like the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) [59][57][60][49][52][15][43][8].

Yet synthetic data remains subject to privacy risks. Techniques like membership inference and linkage attacks are still possible, especially when synthetic data generation does not involve formal differential privacy (DP) protections or carefully tuned noise-adding mechanisms [64][15][43]. This privacy-utility trade-off is especially pronounced in scientific and healthcare contexts, where highly accurate or high-fidelity models can unintentionally reproduce rare features or outliers, potentially leaking sensitive information [1][64][33]. Empirical studies suggest that while DP can lower privacy risk, it often reduces the fidelity or downstream machine learning performance of synthetic datasets when not well balanced [64][43]. Therefore, ongoing evaluation and transparent reporting on the implementation of privacy safeguards are critical.

Algorithmic bias represents another fundamental ethical concern. Synthetic datasets generated from imbalanced or biased source data, or from generative models that mirror existing social or demographic disparities, can perpetuate or even exacerbate these biases in subsequent analysis and decision-making [60][1][15][8][88]. This risk is significant when synthetic data is employed to augment minority classes or address imbalances, as unintended consequences may arise without rigorous fairness audits. Current literature highlights the need for systematic post-generation evaluations—including fairness and performance audits—to identify and mitigate sources of bias [55][60][88]. Appropriately conducted audits and bias mitigation strategies are necessary to avoid amplifying harms or marginalization through AI systems.

The increasing indistinguishability of real and synthetic data also presents new risks involving misinformation, data integrity, and the phenomenon of "AI autophagy" [28][60][62][8][11][88]. As

GenAI models generate increasingly realistic outputs, differentiating authentic from synthetic datasets becomes more challenging. This blurring can harm reproducibility, undermine scientific credibility, and erode public trust. In regulated environments such as healthcare, undetected contamination of research repositories or benchmarks by synthetic data can propagate hallucinatory artifacts and distort scientific inferences. Potential negative feedback loops (AI autophagy), where models unwittingly train on their own synthetic outputs, may degrade model quality or reliability over time [88]. To address these concerns, leading studies emphasize establishing robust provenance frameworks, mandatory transparency and disclosure, and explicit labeling protocols to always identify and explain where synthetic data has been used [28][60][8][27].

In this context, traceability and transparency are considered essential for responsible synthetic data deployment. Technical measures such as data watermarking, blockchain-support for verifiable audit trails, and AI-powered detection tools are being developed to strengthen the provenance, auditability, and control of synthetic data across the research lifecycle [28][60][8][11]. However, the variety of emerging solutions underscores the need for open standards and harmonized best practices to avoid fragmented, incompatible implementations.

Legal and regulatory compliance is an evolving frontier. While synthetic data can reduce identifiability and thus facilitate compliance with privacy laws, the absence of universal and precise regulatory frameworks leads to persistent uncertainty [40][60][36][52]. Regulatory agencies and policymakers are increasingly aware of these complexities, especially for clinical research and data sharing, but rules and enforcement remain inconsistent across jurisdictions. Best practices and recent surveys advocate for sustained cross-sector and cross-border dialogue to harmonize policy, establish consensus guidelines, and balance individual privacy with scientific advancement [57][60][15][26][8][27].

6.2 Responsive Practices and Protocols

As ethical and social risks related to synthetic data are increasingly recognized, proactive practices and protocols have emerged to mitigate these potential harms. Chief among these are systematic risk mitigation strategies designed to anticipate, detect, and reduce privacy breaches, algorithmic bias, and misinformation [22][88][27]. Comprehensive auditing—conducted both internally and by external parties—serves to evaluate privacy leakage, fairness, and statistical fidelity, as well as to monitor for residual biases. Privacy audits often utilize membership inference testing, k-anonymity assessments, and advanced adversarial simulations, while fairness evaluations are grounded in domain-specific metrics and comparisons to representative baselines [64][22][88].

A cornerstone of responsible practice is the multi-faceted validation of synthetic data quality and representativeness. Cutting-edge protocols require evaluations that go beyond high-level statistical comparisons, mandating statistical similarity metrics (such as distributional divergence measures and correlation structure analyses), machine learning utility tests (for example, Train on Synthetic, Test on Real, and downstream task generalization), and domain-specific relevance assessments, often with expert input.

This multi-level approach ensures both the meaningful utility of synthetic datasets and the avoidance of artefacts that could distort scientific interpretation [55][49][21][22][85].

Furthermore, transparent community benchmarking and open validation are instrumental for trustworthy synthetic data deployment. The proliferation of open-source tools, public leaderboards, and collaborative challenges encourages the adoption of shared best practices, enables rapid identification of methodological weaknesses, and harmonizes evaluation efforts [88][27][85]. Iterative engagement among data scientists, clinical experts, ethicists, and legal advisors informs the operationalization of safeguards and continuous refinement of mitigation strategies as technology evolves.

6.3 Standardization and Community Initiatives

This section addresses one of the central challenges raised in this survey: enabling the sustainable, ethical, and scientifically rigorous integration of synthetic data into research and applied domains. Our objective is to synthesize recent progress in developing and harmonizing standards, benchmarks, and governance frameworks, as well as to highlight actionable pathways for bridging identified research and regulatory gaps.

The sustainable and trustworthy integration of synthetic data in research and applied settings hinges on the advancement and widespread adoption of rigorous, standardized evaluation protocols and community-driven frameworks [27, 85]. Currently, the lack of universally accepted, domain-tailored benchmarks and validation processes substantially hinders both scientific legitimacy and regulatory approval, an issue that is particularly acute within sensitive sectors such as healthcare [27, 55, 60, 85]. Recent reviews underline that the absence of standardized evaluation protocols not only limits the deployment of synthetic data for augmenting datasets but also impedes systematic validation of generative models in critical application areas [85].

Several coordinated community efforts are emerging to close these critical gaps. These initiatives include the establishment of frameworks for systematic privacy risk quantification, robust utility and fairness benchmarking, and comprehensive auditing protocols, allowing for transparent and trustworthy evaluation of synthetic data and generative models [22, 27, 85, 88]. There is also growing support for precompetitive consortia and open benchmarking challenges, which serve to harmonize evaluation practices, lower entry barriers, and disseminate up-to-date best-practice guidance grounded in contemporary ethical and regulatory standards [85]. Notably, open-source frameworks (such as SyntheX for medical imaging [22]) and recent taxonomies for generative models in recommender systems [21] are paving the way for reproducible, sector-specific comparative studies.

To advance beyond current limitations, actionable priorities have been proposed in recent literature [27, 60, 88]. These include: formalizing definitions distinguishing synthetic from real data based on provenance; instituting robust disclosure guidelines and technical mechanisms such as watermarking; and fostering educational initiatives on responsible use. Additionally, the integration of transparency measures and provenance controls is essential to reduce the risks of dataset contamination or uncontrolled generative processes, particularly in scenarios affected by phenomena like model autophagy [88].

There is sustained, interdisciplinary dialogue—bringing together technical, ethical, and regulatory experts—to ensure that standards can adapt responsively to rapid advances in GenAI capabilities [11, 21, 22, 27]. These dialogues are crucial for maintaining public trust and preventing ethical pitfalls, as underscored by pressing risks such as data leakage and the difficulties of reliably detecting and managing synthetic content in research workflows [60, 88].

In sum, the development of governance structures and consensus-driven standards—with cross-sector vigilance and continuous protocol refinement—remains essential to realizing the full promise of synthetic data in an equitable, ethical, and trustworthy fashion. These efforts directly support the broader aims of this survey: to map current progress, identify persistent research and deployment challenges, and outline evidence-based strategies for responsible innovation in synthetic data ecosystems.

7 Challenges, Limitations, and Future Directions

This section aims to synthesize and critically evaluate the core challenges, unresolved limitations, and emerging future directions in AI, with a focus on connecting these issues to the survey's overarching objectives outlined in the introduction. By doing so, we provide both returning and new readers a clear lens through which to interpret the multifaceted risks and opportunities faced by the field. Specifically, we highlight how ongoing technical advances, evaluated in preceding sections, intersect with ethical, practical, and societal imperatives, reinforcing the need for responsible innovation.

Recent technical advancements have accelerated the capabilities of AI systems; however, these developments are inextricably linked with a complex set of ethical, practical, and societal challenges. Progress in areas such as scalability, training efficiency, and multi-modal integration often comes alongside increased risks in fairness, privacy, and reliability. For example, improvements in language model generation must be weighed against potential risks of misinformation propagation and bias amplification.

To fully appreciate the interplay between technological progress and these broader concerns, it is crucial to analyze how enhancements in model architectures or training pipelines may give rise to novel ethical dilemmas or exacerbate existing practical limitations. Addressing these issues demands a holistic approach that aligns technical objectives with responsible deployment practices, ensuring systems remain robust, transparent, and aligned with human values.

At the same time, evaluating and mitigating the risks introduced by new methodologies requires a critical examination of current evaluation pipelines, as well as the development of novel metrics and frameworks sensitive to downstream impacts. Technical innovations should, therefore, be considered alongside strategies for monitoring ethical risk flows in real-world applications.

In light of these considerations, the following actionable pathways are proposed for addressing identified research gaps. These pathways are closely linked to the survey's emphasis on aligning technical and ethical priorities:

Clarifying objectives and establishing research priorities - Revisit and explicitly articulate the alignment between evolving technical capabilities and the foundational objectives of transparency, fairness, and accountability defined at the outset of this survey. - Develop interdisciplinary collaborations to ensure that domain expertise is integrated within all stages of model development and evaluation.

Operationalizing solutions to current limitations - Design scalable training algorithms and effective model compression techniques to democratize access while reducing environmental impact. - Institutionalize ethical auditing as a core component of the AI development lifecycle, advancing integrated tools that monitor for bias and fairness from data collection through deployment. - Pioneer unified evaluation frameworks and metrics that measure not just performance but societal impact, amplifying trust in AI adoption. - Advance and deploy privacy-preserving learning techniques to address both regulatory requirements and user expectations around data protection. - Prioritize research on advanced model interpretability, moving beyond simple transparency toward actionable insights supporting accountability.

Despite significant progress, a number of open research gaps remain evident, including:

Going forward, it is essential that the AI research community fosters collaboration across disciplinary boundaries, pursues unified standards for evaluation and ethical oversight, and remains vigilant in translating technical advancements into responsible innovation. This holistic perspective ensures ongoing alignment with the high-level objectives of robustness and human-centric development introduced at the beginning of this survey.

7.1 Technical and Resource Barriers

The proliferation of diffusion models and other large-scale generative frameworks has brought several technical and resource-intensive challenges, particularly concerning scalability and equal access. A principal concern is the immense computational cost associated with both training and inference for diffusion models, which typically surpasses that of earlier generative paradigms, such as GANs—even with the integration of recent algorithmic optimizations for efficiency [2–4, 9, 12, 13, 17, 18, 20, 22, 23, 25, 34, 38, 41, 46, 47, 51, 56, 58, 61, 63–65, 67, 70–72, 75, 84, 86, 88–90]. Training state-of-the-art diffusion models entails extremely large datasets and substantial computational infrastructure, typically limited to well-resourced institutions, thereby reinforcing disparities in research and practice [4, 9, 12, 41, 64, 65, 72, 86]. Efforts to mitigate these expenses include modular sampling techniques [17, 90], adaptive noise scheduling [34], and hybrid generative architectures [9] that strive to reduce resource needs and accelerate inference. For instance, modular sampling and adaptive step-size ODE solvers can markedly decrease the number of required network evaluations per generated sample [17], while approaches such as FG-RAT GAN offer both competitive generative quality and enhanced computational efficiency, outperforming traditional diffusion models on resource constraints for text-to-image synthesis [34]. In language modeling,

Table 11: Summary of Key Future Research Gaps

Challenge Area	Current Limitation	Future Direction	Potential Impact
Scalability	Resource-intensive training	Efficient algorithms and model compression	Wider accessibility
Ethical Alignment	Insufficient bias and fairness evaluation	Integrated ethical auditing tools	Minimizing societal harm
Evaluation	Lack of standardized benchmarks	Development of unified evaluation frameworks	More reliable model assessment
Privacy	Inadequate data protection mechanisms	Privacy-preserving learning approaches	Enhanced user trust
Transparency	Opaque model decision processes	Advanced interpretability techniques	Improved accountability

diffusion-based frameworks have demonstrated scalability and resource competitiveness compared to autoregressive paradigms [9]. Nevertheless, a fundamental tradeoff is sustained between model expressivity, fidelity, and computational tractability [3, 12, 46, 72].

Another key challenge pertains to the fidelity and cross-domain generalization of synthetic data generated through simulation. Domain discrepancies between simulated and real data can be attributed to rendering engine constraints, incomplete physical models, and varied annotation schemas, resulting in domain biases and compromised downstream performance [22, 72, 76, 85]. Progress in physics-based simulation frameworks and domain randomization has improved the reliability of synthetic-to-real transfer [22, 76, 85], as evidenced by SyntheX, which leveraged strong domain randomization in clinical X-ray imaging and achieved real-world test performance comparable to, or even exceeding, models trained on real data [22]. Nevertheless, attaining robust alignment and representation consistency, especially across 3D or multi-modal domains, remains difficult due to representation instability and intricate feature alignment challenges [22, 72, 76]. Furthermore, the lack of standardized annotation protocols for synthetic data—particularly in complex, high-dimensional domains and medical applications—complicates performance evaluation, reproducibility, and regulatory compliance [76, 85]. The insufficiency of domain-specific benchmarking and annotation guidelines remains a critical obstacle for the field [85].

As illustrated in Table 13, algorithmic progress has begun to address some technical limitations; however, structural resource inequalities and gaps in data standardization continue to fundamentally restrict the scalability, generalizability, and reproducibility of large-scale generative models.

7.2 Generalization, Robustness, and Societal Impact

This section aims to critically evaluate the dual challenges and broader implications associated with generative models: their technical generalization and robustness, especially under constraints such as limited, federated, or non-stationary data, as well as their far-reaching societal and ethical ramifications. By the end of this section, the reader should understand core obstacles limiting generative model reliability and the contextual risks posed by their widespread deployment.

A pivotal issue in generative modeling concerns the extent to which current architectures generalize to low-resource, continual, federated, and edge environments—contexts characterized by limited data, computational constraints, and evolving distributions [37, 84]. For instance, in continual learning at the edge, Dedeoglu et al. [37] demonstrated how model updates using Wasserstein-based approaches can help mitigate catastrophic forgetting. Similarly, Hess et al. [84] report on downscaling climate models, highlighting knowledge transfer, uncertainty quantification, and adaptation challenges in real-world geoscience settings. These case studies reveal that many existing methods degrade in fidelity or exhibit vulnerability when aggregated under federated setup or continually updated, signaling unresolved gaps in knowledge transfer, adaptation, and privacy protection [84]. Moreover, as illustrated in 3D drug design with geometry-complete diffusion models [56], the complexity and resource demands of generative models pose deployment barriers in edge or restricted settings. The deployment of large generative models onto resource-constrained systems therefore faces persistent obstacles pertaining to model size, communication overhead, and stable operation under out-of-distribution (OOD) conditions, thereby underscoring the demand for modular, efficient, and adaptable frameworks [27, 56, 84].

From a societal and ethical perspective, the deployment of synthetic data and generative models introduces a multifaceted array of risks, with real-world consequences documented across diverse application domains. For example, in medical imaging, Akbar et al. [8] and Tudosiu et al. [11] both provide evidence that generative models may inadvertently memorize sensitive training data, complicating privacy guarantees even as synthetic images achieve state-of-the-art quality. SyntheX [22] shows how synthetic datasets, if carefully generated, allow for generalizable and equitable learning in healthcare, but raises questions of bias remediation and representative coverage.

Bias and fairness: Synthetic data may encode and exacerbate existing societal and demographic biases, particularly if training corpora are imbalanced or lack domain specificity [4, 8, 22, 63, 66]. For instance, fairness concerns in clinical diagnostics have arisen when synthetic data underrepresent rare subtypes or minority populations [4].

Contamination and feedback loops: Recursive use of synthetic data as training input (AI "autophagy"), as detailed by Xing et al. [88], can erode model reliability and scientific validity, especially when web-scraped or uncurated datasets accumulate unlabeled synthetic content [13, 14, 30].

Privacy leakage: Overparameterized models, especially trained on limited or homogeneous data, may memorize and inadvertently reveal sensitive real information, as shown in diffusion-based medical image generation [8, 22, 27, 72, 88].

Opacity and interpretability: The increasing complexity of generative architectures impedes interpretability and transparency—critical

Table 12: Technical barriers and mitigation strategies for large-scale generative models

Barrier	Example Manifestation	Mitigation Strategies
High computational cost	Training/inference of large diffusion models	Modular sampling, adaptive noise scheduling, hybrid architectures, resource-competitive text-to-image models
Data bias in simulation-generated samples	Domain gap in synthetic-to-real transfer	Domain randomization, advanced physics-based generative models, uncertainty-guided diffusion
Lack of annotation standards	Heterogeneous labels in 3D/multi-modal and medical data	Development of universal standards, dedicated benchmarking, community-driven protocols

where alignment with human expectations and regulatory standards is fundamental (e.g., healthcare, law) [10, 11, 13, 22, 36, 42, 83, 88]. Techniques such as ConSCompF [83] and precedent-based interpretability [14] exemplify practical steps toward explainability in high-stakes environments.

The dynamic and evolving regulatory environment further compounds these challenges, necessitating robust mechanisms for fairness auditing, privacy preservation, and lifecycle risk management across deployment contexts. Practical frameworks and studies discussed in this section ground the often abstract risks in concrete examples, reinforcing the necessity for transparent, responsible, and adaptable deployment of generative models.

In summary, this section has identified the technical and societal limitations facing scalable generative model deployment, referencing concrete applications to highlight both progress and persistent gaps. This evaluation sets the stage for later discussions on best practices and prospective research directions toward inclusive, trustworthy, and resilient generative AI.

7.3 Label Dependency and Semantic–Style Balance

The imperative to minimize dependence on costly manual annotation has propelled extensive research into weakly supervised, self-supervised, and label-free generative paradigms [34, 71]. Despite these trends, leading conditional diffusion and GAN-based image synthesis methods often retain some level of label dependency to ensure semantic specificity and controllable generation [34, 71]. Achieving a robust balance between semantic fidelity—ensuring that generated samples satisfy label or mask constraints—and stylistic diversity, which is vital for improved generalization and practical versatility, remains a major challenge [71]. Current approaches include manipulating the latent space, complex conditioning on semantic labels or reference styles [34, 71, 85], and employing disentanglement losses or dual-guidance mechanisms, yet all fall short of fully resolving the semantic–style tradeoff.

Recent advancements such as FG-RAT GAN [34] have shown that incorporating auxiliary classifiers within the discriminator and leveraging contrastive learning via cross-batch memory can markedly improve class-consistent and fine-grained semantic image generation. FG-RAT GAN combines adversarial, categorical cross-entropy, and contrastive loss functions to sharpen intra-class image fidelity while efficiently distinguishing between subclasses. While FG-RAT GAN surpasses state-of-the-art methods such as LAFITE, VQ-Diffusion, and large diffusion models in terms of Frechet Inception Distance (FID), Inception Score (IS), and parameter efficiency on benchmarks like CUB-200-2011 and Oxford-102, it nonetheless retains a dependency on labeled data for fine-grained category supervision, as confirmed by ablation studies and reported evaluations.

On a different axis, diffusion-based models like IIDM [71] pursue explicit disentanglement of semantic and style control by conditioning on both semantic segmentation masks and style reference images. IIDM encodes the style reference into a latent vector and guides the denoising process under the constraints of a segmentation mask, with enhancements such as refinement, color transfer, and model ensemble during inference to further improve style fidelity and mask accuracy. IIDM outperforms previous GAN and diffusion approaches, achieving 94.15% mask accuracy and a FID of 30.75, and proves particularly effective in balancing semantic content and style, as supported by extensive ablation and inline metric comparisons.

Despite these advances, neither approach delivers a unified, interpretable framework capable of fine-grained control with minimal label requirements. Progress as illustrated by FG-RAT GAN and IIDM highlights both computational and methodological tradeoffs, underscoring the persistent challenge of semantic–style balanced synthesis with limited annotation. Developing frameworks that achieve fine-grained controllability with reduced label reliance remains an open and central research direction.

7.4 Evaluation Metrics and Benchmark Gaps

The evaluation of synthetic and augmented data—particularly in emerging areas such as multi-modal, 3D, and cross-domain synthesis—remains fundamentally limited by the absence of universally accepted, modality-agnostic metrics [2, 22, 58, 76, 85, 90]. Predominant quantitative measures such as FID, IS, and TSTR [2, 58, 90] provide useful statistical comparisons but fall short of capturing semantic consistency, clinical significance, or nuanced, application-specific utility. These shortcomings are particularly critical in sensitive domains like medical imaging, where restricted access to real data amplifies the need for reliable evaluation [22, 76]. As documented by recent studies and systematic reviews, the prevalent reliance on these metrics impedes robust cross-model and cross-study comparisons and ultimately hinders the adoption of synthetic data in practical and clinical contexts [22, 76, 85].

For instance, the SyntheX framework demonstrates that synthetic datasets — generated via domain randomization and advanced simulation — can rival or even outperform real data in certain medical imaging tasks, such as hip imaging and surgical tool detection. However, as highlighted, standard metrics fail to fully capture aspects of clinical relevance and real-world generalizability achieved by synthetic data, especially for complex or high-stakes applications [22]. Similarly, systematic reviews have emphasized the entrenched lack of standardized, domain-relevant benchmarks for synthetic data, noting that this deficit slows clinical translation and responsible AI model adoption [85]. Furthermore, new directions such as measurement-guided generation for medical image synthesis underscore that existing metrics do not accurately reflect the diagnostic value, informativeness, or reliability of synthesized outputs. In particular, measurement-driven approaches like uncertainty-guided diffusion models are shown to improve downstream diagnostic accuracy across diverse architectures, even as prevailing quantitative scores remain insufficiently informative about clinical utility [76]. Addressing these gaps by developing robust, interpretable, and domain-specific evaluation frameworks is crucial for transparent development, effective benchmarking, and eventual regulatory acceptance of synthetic data-driven models.

7.5 Research Opportunities and Future Trends

Looking toward the future, advancing model quality, sample diversity, robust personalization, and real-time synthesis remains paramount as generative AI shifts toward dynamic, interactive, and agent-based paradigms [27, 34, 71, 72, 85]. Foundational enablers of this evolution include: automated benchmarking and pipeline development for transparent, systematic performance tracking [27, 85]; adaptive, self-improving architectures supporting continual learning and transfer; and scalable, modality-agnostic evaluation protocols to enable rigorous, cross-study comparability [85]. Addressing these will further require responsible AI frameworks that incorporate up-to-date policies, oversight mechanisms, and inclusive stakeholder engagement to ensure ethical and trustworthy deployment [1, 4, 10, 11, 13, 17, 22, 24, 27, 30, 31, 42, 43, 49–51, 59, 63, 66, 83, 85, 88].

The rapid convergence across vision, language, multimodal, and scientific domains is accelerating the development of unified generative systems capable of fueling diverse downstream applications, including scientific modeling, systematic reasoning, open-vocabulary detection, and multimodal synthesis [1–6, 8–13, 15, 17, 18, 20, 22, 24, 25, 28–31, 33, 35, 36, 38, 40–45, 50–52, 54, 55, 57–66, 70, 80, 81, 88–90, 92, 93]. A decisive challenge at this intersection is principled, scalable, and interoperable architecture design: models must effectively handle heterogeneous data modalities, diverse annotation schemes, and domain-specific requirements, while minimizing bias and maximizing transparency. Interdisciplinary collaborations—uniting algorithmic research, domain expertise, and regulatory insight—are vital for accelerating the emergence of robust, adaptable, and trustworthy generative AI.

To crisply categorize remaining barriers and illustrate ongoing mitigation strategies, we present Table 13.

In the societal dimension, ongoing and emerging research opportunities—reflecting multidisciplinary challenges and future priorities—include education and harmonization of best practices for synthetic data disclosure and provenance [27, 60], transparent auditability and algorithmic fairness [24, 27, 52, 63, 88], hybrid approaches balancing real and synthetic data to mitigate AI autophagy and enhance robustness [22, 27, 88], and scalable frameworks for oversight, red-teaming, and model alignment with stakeholder values [24, 27, 42, 85, 88]. These priorities emphasize the importance of interoperability, stakeholder involvement, open-source protocols, and global regulatory cohesion in shaping the next generation of inclusive and trustworthy generative AI.

8 Security, Adversarial Threats, and Alignment

8.1 Threat Detection and Robustness

This section furthers the paper's overarching objective of critically evaluating the safety and reliability of generative models by examining the evolving landscape of adversarial threats and robustness strategies.

As generative models have become embedded in critical domains, concerns regarding their vulnerability to adversarial threats and the reliability of their outputs have correspondingly intensified. Traditional adversarial testing and red teaming approaches have uncovered foundational vulnerabilities; however, recent research has illuminated both the expanding variety and increased severity of threats facing large-scale systems—especially multimodal large language models (LLMs) and vision-language architectures. Advances in the field have shifted the focus from isolated adversarial attacks toward comprehensive taxonomies that systematically classify attack vectors by both technical sophistication and modality, extending beyond textual to include visual and multimodal perturbations. Automated red teaming frameworks, such as those built on the searcher paradigm, now provide systematic methodologies for evaluating system-level security. These frameworks facilitate the discovery and categorization of previously underexplored weaknesses in generative AI [24, 39].

The threat landscape is evolving, manifesting escalated attack complexity. Multimodal attacks, which leverage the interplay among language, vision, and audio modalities, expose failure modes that are invisible to unimodal defenses, thereby highlighting the limitations of traditional siloed robustness evaluation. Generative agents, particularly those built upon LLM cores, are notably susceptible to attacks that manipulate chained reasoning or exploit inter-model interactions—a vulnerability exacerbated by the opacity and substantial scale of state-of-the-art architectures [39]. Efforts to bolster robustness frequently employ a defense-in-depth approach, yet these efforts must confront persistent challenges such as overfitting to known attack patterns and unintended negative outcomes from overly aggressive content filtering. The latter can lead to inadvertent blocking of benign queries, thereby degrading user experience and eroding trust [24].

The robustness of detection systems—especially within social media contexts—has also received significant scrutiny under adversarial conditions. Cutting-edge frameworks that synthesize adaptive data augmentation with adversarial training have yielded tangible improvements. By dynamically perturbing non-essential features and incorporating hard negative samples in contrastive learning paradigms, these systems reduce overfitting to benign patterns and sustain robust performance even when faced with adversarially manipulated data [24]. Nevertheless, the need persists for the development of generalizable robustness mechanisms capable of preempting the ingenuity and unpredictability characteristic of emerging attack strategies, particularly as generative models advance in multimodal integration and contextual awareness.

The examples summarized in Table 14 illustrate the diversity of adversarial threats and corresponding defensive approaches, emphasizing the multifaceted requirements of robust generative model deployment.

Table 13: Technical Barriers and Corresponding Mitigation Strategies in Generative AI

Technical Barrier	Description	Example Mitigation Strategies
Quality and Fidelity Limitations	Sub-optimal alignment between generated and real data; mode collapse or hallucinations may reduce downstream utility [8, 17, 22, 27, 28, 30, 33, 54, 59, 72, 85]	Iterative evaluation pipelines; advanced regularization and diversity-promoting mechanisms; hybrid validation using real and synthetic data [22, 27, 85]
Sample Diversity and Personalization	Generic outputs inadequately capture rare subpopulations or individual variations, challenging transfer and personalization [8, 34, 36, 51, 72, 85]	Conditional/transfer models leveraging patient or context features; active learning; domain-aware synthesis; specialized evaluation for edge cases [36, 85]
Robust Benchmarking and Evaluation	Lack of standardized, modality-specific benchmarks hampers rigorous, reproducible comparison across architectures and domains [63, 83, 85]	Establishment of open, community-driven benchmarking suites; modality-agnostic quantitative and qualitative metrics; routine audits [63, 83, 85]
Bias, Fairness, and Transparency	Amplification of historical and synthetic data biases; lack of explainability or provenance tracking [1, 24, 27, 52, 60, 88]	Systematic bias auditing; integrated explainability modules; clear data provenance and disclosure protocols; stakeholder involvement [24, 27, 52, 60]
Privacy, Security, and Autophagy Risks	Memorization of sensitive information, adversarial misuse, and self-consuming "autophagy" from synthetic dataset contamination [8, 27, 62, 85, 88]	Privacy-preserving model architectures (differential privacy, DP-GANs); adversarial red-teaming; controlled curation of real/synthetic mix; synthetic data detection and watermarking [8, 24, 27, 52, 62, 85, 88]
Societal and Regulatory Alignment	Gaps in regulation, ethical deployment, inclusivity, and societal consensus for responsible AI development [1, 24, 27, 42, 52, 60, 63, 85, 88]	Iterative policy updates; inclusive regulatory frameworks; active stakeholder dialogue; global standards for responsible synthetic data use [1, 24, 27, 42, 60, 85]

Table 14: Overview of primary adversarial threat categories, illustrative application domains, and principal classes of defense strategies.

Adversarial Threat Category	Typical Application Domain	Defense Strategy
Textual Prompt Injection	LLM-based conversational agents	Prompt filtering, adversarial training, input sanitization
Visual Perturbation Attacks	Vision-language models, image generators	Image preprocessing, adversarial example detection, robust feature extraction
Multimodal Chained Attacks	Multimodal reasoning agents	Cross-modal consistency checks, hierarchical defense-in-depth
Model-to-model Interaction Exploits	Autonomous agent ecosystems	Traceable chained reasoning, interaction protocol hardening

8.2 Alignment in Generative Models

While security is foundational, aligning generative models with human values and preferences represents an equally critical aspect of responsible AI development. Reinforcement learning from human feedback (RLHF) and its derivatives have emerged as dominant strategies for behavioral alignment, enabling generative systems to internalize explicit task instructions as well as complex preferences regarding style, safety, and practical utility [21, 42]. The landscape of preference tuning has expanded to encompass multimodal models—including those spanning vision, speech, and combinations thereof—necessitating adaptable alignment frameworks that integrate heterogeneous human feedback signals.

Recent comparative studies shed light on both the advantages and inherent challenges of RLHF and related alignment strategies. Large, high-quality datasets used in reward modeling can improve how accurately models capture nuanced human preferences. However, reliance on such datasets introduces risks: they may propagate societal biases and underrepresent minority perspectives, especially as alignment protocols are scaled across tasks and user populations [42]. Challenges further intensify in complex settings such as conditional image-text generation, where preference alignment is vulnerable to problems like mode collapse or excessive conservatism—particularly under strong reward model regularization, which may stifle generative diversity and informativeness [21]. The trade-off between generating harmless (i.e., non-biased, non-harmful) and helpful (i.e., relevant, informative) outputs is still largely unresolved both technically and societally [21].

A persistent obstacle is the robust evaluation of alignment success. Existing quantitative metrics provide only a partial picture, often failing to detect low-frequency but high-consequence misalignments with significant social implications [21]. Establishing standardized benchmarks and scalable annotation pipelines for human feedback collection also remains a substantial challenge, particularly in the case of multimodal generative systems. Despite these issues, the field is rapidly advancing through innovations such as more expressive preference models, hierarchical feedback strategies, and cross-modal reward optimization, collectively pointing toward increasingly robust and socially attuned generative AI [21, 42].

In concrete terms, objectives for evaluating alignment include quantifying: (1) the accuracy with which generative models reflect diverse human preferences across modalities; (2) the degree to which generative outputs balance harmlessness (e.g., non-toxic, unbiased) and helpfulness (e.g., informativeness, utility) in different domains; and (3) the robustness of models to rare but socially significant preference misalignments.

Key ongoing challenges include scalability—as large-scale human feedback risks both bias propagation and practical bottlenecks; evaluation, with prevalent metrics frequently missing critical, rare misalignments; generalization, especially across modalities and open-ended scenarios; nuanced trade-offs between harmlessness and helpfulness; and ongoing innovation in modeling human preferences, acquiring feedback, and enabling cross-modal alignment.

The intersection of adversarial robustness and human alignment in generative models delineates an urgent and dynamic research frontier. Advances in this domain are shaped by intricate trade-offs among security, safety, and practical value, as detailed in comprehensive surveys and evolving taxonomies of red teaming and preference alignment [21, 24, 39, 42].

9 Synthesis, Comparative Analysis, and Recommendations

At this stage, we synthesize the main findings, explicitly recap the objectives of this survey, provide a comparative analysis, present a structured tabular summary of the proposed taxonomy, and offer recommendations. The aim is to enable both newcomers and specialists—ranging from academic researchers to industry practitioners—to readily reference key insights and results.

9.1 Objectives Recap and Desiderata

The explicit objectives of this survey are as follows: we systematically measure and compare key properties X, Y, and Z across representative domains, evaluate how technical approaches perform under diverse practical scenarios, and identify persistent gaps and challenges. We further aim to present a comprehensive taxonomy for organizing the domain, facilitating both a detailed and accessible understanding for our target audience.

Table 15: Taxonomy of Alignment and Red Teaming Strategies in Generative Models

Dimension	Categories/Examples	Core Objectives	Selected Reference(s)
Alignment Strategy	RLHF, Preference Tuning, Hierarchical Feedback, Cross-modal Reward Optimization	Internalize human preferences, Style, Safety, Utility	[21, 42]
Input Modality	Text, Vision, Speech, Multimodal	Model extensibility across task types	[21, 42]
Evaluation Paradigm	Quantitative Metrics, Human-in-the-Loop Annotation, Robust Benchmarking	Robustness, Diversity, Detection of rare failure cases	[21, 42]
Red Teaming Approaches	Automated Searcher Frameworks, Multimodal and Agent-based Attacks, Excessive Filtering Analysis	Identify vulnerabilities, Harmfulness vs. Helpfulness balance	[21, 24]
Bias and Societal Risks	Dataset Bias Propagation, Minority Underrepresentation	Minimizing harm, Ensuring fairness	[21, 42]
Generality/Generalization	Scaling across domains, Open-endedness, Adaptation to new tasks	Broad applicability, Future-proofing	[21, 42]

9.2 Bridging Analysis and Technical Findings

Our taxonomy and comparative analysis are grounded in the detailed technical sections. We first established foundational concepts and methodological criteria (see Sections 2 and 3), then examined classes of approaches within these criteria, measuring performance in terms of X (such as scalability), Y (such as generalization across datasets), and Z (such as robustness under varied conditions). Key transitions between these analyses are summarized in Table 16 to provide an overview and logical bridge between thematic and technical dimensions.

9.3 Comparative Analysis

In direct comparison across the surveyed approaches, Category 1 achieves high values for X and Y in controlled settings but can be less robust regarding metric Z, particularly for out-of-distribution cases. Category 2, while excelling in generalization (Y), faces scalability challenges (X) in practice. Category 3 demonstrates balanced robustness and generalization but with modest absolute performance based on the criteria established earlier.

Transitions between the comparative overview and in-depth technical analysis are made explicit in Table 16, which distills overarching themes and makes it easy to reference primary desiderata and observed trade-offs.

9.4 Recommendations and Target Audience

For academics seeking to identify open research challenges, we recommend focusing on hybrid methods that address both scalability and robustness. For practitioners aiming for deployment, our analysis suggests prioritizing methods that demonstrate strong generalization (Y) while taking into account operational constraints. The taxonomy table provided serves as a concise reference for both communities to guide method selection and future investigations.

Finally, while we have endeavored to present a current and comprehensive survey, we encourage readers to consult closely related recent works as they become available.

9.5 Comparative Perspective

The trajectory of generative and augmentation strategies in artificial intelligence (AI) displays a marked evolution from foundational, hand-engineered and deterministic approaches to the advent of advanced deep learning, generative, and diffusion-based frameworks [21–23, 34, 67, 71, 72, 74, 75, 86, 88]. Early data augmentation techniques, such as geometric transformations and value-based manipulations, have consistently been recognized for their simplicity, transparency, and domain-wide usability, especially within computer vision and natural language processing [23, 67, 75, 86]. These traditional methods primarily aimed to mitigate overfitting and perform well under data constraints. However, as highlighted in [23, 71, 86], their domain specificity and limited impact on semantic diversity imposed significant limitations in nuanced domains like medical imaging, where model success depends on distinguishing subtle features.

The emergence of deep generative models—including Generative Adversarial Networks (GANs), transformer-based architectures, and diffusion models—has fundamentally reshaped the landscape of data synthesis and augmentation. These models provide high-fidelity, controllable data generation capabilities, introduce innovative multi-instance augmentation strategies (e.g., Mixup for vision and text), and enable the scalable construction of synthetic datasets [22, 23, 34, 74, 75, 86, 88]. Increasing attention is directed toward diffusion models, which demonstrate strengths in producing images and augmentations with high statistical and semantic fidelity, frequently outperforming traditional GANs in both diversity and perceptual realism [34, 71, 74, 88]. Recent advances exemplified by FG-RAT GAN [34] and IIDM [71] illustrate how auxiliary objectives—such as class-aware regularization and contrastive learning—can improve both fine-grained control and computational efficiency. FG-RAT GAN, integrating an auxiliary classifier and cross-batch contrastive loss, achieves state-of-the-art Frechet Inception Distance (FID) and Inception Score (IS) while using fewer parameters compared to prior models. IIDM, leveraging latent diffusion guided by semantic masks and style references, attains superior mask accuracy and FID on semantic image synthesis tasks, underscoring the advantages of guided denoising and inference optimization for content and style fidelity. In medical imaging or X-ray analysis, simulation-based synthetic data generation, augmented by robust domain randomization, is shown to match or exceed real-data-based model generalization, offering flexibility while circumventing the challenges of large-scale, high-quality data collection [22].

Notwithstanding remarkable progress, generative augmentation presents new challenges. The unchecked proliferation of synthetic data—known as "AI autophagy" [88]—can contaminate training datasets, compromise model reliability, exacerbate distributional shifts, and threaten the credibility of scientific benchmarks [34, 72, 88]. Comparative analyses across domains stress that effective regularization, class balance, and rare-class enrichment depend critically on the careful curation and quality control of generated samples. Over-reliance on unverified synthetic data has been shown to degrade downstream task performance [34, 72]. To address these risks, multiple surveys and empirical studies advocate for the establishment of robust and standardized evaluation frameworks, systematic benchmarking, and the formulation of modality-sensitive augmentation guidelines [23, 34, 71, 72, 74, 75, 86, 88]. Achieving an optimal balance among semantic fidelity, sample diversity, and

Table 16: Summary of Proposed Taxonomy with Key Properties Across Categories

Category	Defining Characteristics	Strengths	Limitations
Category 1	Description of category 1	Major strengths of category 1	Key limitations of category 1
Category 2	Description of category 2	Major strengths of category 2	Key limitations of category 2
Category 3	Description of category 3	Major strengths of category 3	Key limitations of category 3

class distribution remains a central and ongoing research focus for advancing the rigor, reliability, and impact of generative AI-based augmentation.

9.6 Criteria for Responsible Deployment

Responsible deployment of generative and augmentation-based methodologies is a central theme of this survey, underpinning both technical innovation and broader societal impact. This section aims to delineate clear, modality-aware criteria that safeguard data integrity, foster interpretability, and promote robust, context-sensitive augmentation. These pillars are essential for aligning technological advancements with ethical, regulatory, and application-specific requirements [23, 34, 67, 71, 75, 86].

Three foundational criteria guide responsible deployment:

Data Quality and Representational Completeness: The integrity and diversity of the original data are crucial; augmentation cannot substitute for missing classes or modalities [34, 67, 86]. Best practices recommend principled data sampling, rigorous pre-augmentation audits, and ongoing validation to ensure adequate representation. Concretely, consider medical imaging: even advanced methods like FG-RAT GAN require labeled examples from all desired classes and subclasses to synthesize high-fidelity, class-consistent outputs; lacking such foundational diversity would undermine both empirical performance and ethical standards [34, 71].

Interpretability and Auditability: Interpretability underpins regulatory compliance and safe scientific or clinical deployment. Basic geometric augmentation is often inherently interpretable, but more complex deep generative methods (e.g., GANs, diffusion models, or transformers) may obscure provenance and invite unanticipated artifacts, drift, or spurious correlations. Routine interpretability assessment, provenance tracking, and human-in-the-loop oversight remain essential. For example, in semantic image synthesis with latent diffusion, manual review of outputs alongside automated provenance trails is standard practice to mitigate the risks identified in the latest synthesis models [34, 71].

Robust and Context-Aware Augmentation: Balancing data diversity with the avoidance of distributional shift or bias is critical [22, 67, 71]. Recent literature converges on hybrid approaches, combining deterministic, transparent transformations with data-driven generative methods, often leveraged by adaptive policy discovery and real-time monitoring [34, 86]. In medical AI, SyntheX pipelines demonstrate how combining domain randomization and calibration against real-world data enables ethically sound, empirically robust research without the logistical burden of large-scale human data collection [22, 34, 72, 74].

These criteria collectively form an essential framework for the responsible and effective deployment of generative augmentation techniques. By grounding general risks in concrete scenarios drawn from recent case studies—such as subclass fidelity in fine-grained GAN synthesis [34] and synthetic-to-real transfer in medical imaging [22]—this section aims to provide actionable guidelines applicable across modalities and domains.

In summary, a principled approach to generative augmentation, centered on data quality, interpretability, and context-aware practice, supports both the intended scientific utility and the broader objectives of ethical, reliable AI—core goals that the present survey seeks to promote and clarify.

9.7 Integration of Adaptive and Responsible AI

The intersection of adaptive, context-aware intelligence and responsible oversight is increasingly recognized as critical to the future development of generative augmentation systems [48, 69, 73, 76, 82]. Adaptive AI systems are expected to dynamically tailor synthesis and augmentation strategies to evolving real-world data characteristics—striving to balance standardization, personalization, and fairness in real time [69, 76]. Mechanisms such as meta-learning, human feedback-in-the-loop, and task-aware policy search enable augmentation pipelines to incorporate context-specific priors and modulate synthetic data generation accordingly [69, 76, 82].

However, the effectiveness of these advances depends on rigorous oversight protocols. Frameworks for responsible AI in augmentation increasingly require algorithmic transparency, ongoing post-deployment monitoring, and safeguards against undesirable feedback cycles—such as synthetic data amplifying existing biases [34, 48, 73]. In healthcare, domain-specific instrumentation now routinely includes simulation-based validation, out-of-distribution detection, and semantic fidelity assessment as foundational to adaptive augmentation workflows [34, 74]. These developments underpin emerging best practices that emphasize seamless integration of context-sensitive adaptation with thorough, multi-layered responsibility checks [34, 48, 73].

Despite this progress, substantial challenges and active debates persist around harmonizing adaptation and responsibility in generative augmentation. For example, the effectiveness and adequacy of fairness and bias mitigation strategies remain contested. Some research explicitly maximizes synthetic data diversity via diversity-oriented losses, ensuring broad coverage of minority segments and semantic modes without undermining image realism [69]. Others prioritize semantic fidelity and system stability, advocating for constraint-driven or supervised adaptation to avoid distributional drift and overfitting [48, 76]. Tensions also arise around the trade-off between transparency and performance: frameworks emphasizing rapid adaptation may forgo transparency, while those embedding post-hoc monitoring and explainability often incur reduced efficiency [34, 73].

Efforts to reconcile these demands are reflected in integrated frameworks leveraging classifier guidance, auxiliary supervision, and contrastive learning strategies. For instance, in fine-grained text-to-image synthesis, models like FG-RAT GAN incorporate an auxiliary classifier within the discriminator and employ contrastive loss to enhance both subclass-aware performance and semantic fidelity while maintaining computational efficiency and rigorous evaluation protocols [34]. Uncertainty-guided approaches in medical image diffusion models further demonstrate how measurement-driven constraints can balance synthetic data informativeness and reliability [76]. Self-supervised methods also contribute by improving diversity and visual consistency through robust loss functions and regularization techniques [48]. These integrated strategies illustrate the ongoing convergence toward frameworks that simultaneously advance technical performance, pluralism, and ethical responsibility. Nevertheless, consensus on best practices remains elusive, with open questions regarding generalization, domain transfer, and context-specific risk assessment.

9.8 State-of-the-Art (SOTA) Synthesis Models

The recent proliferation of advanced synthesis models encapsulates the ongoing evolution and intricate trade-offs inherent in leading generative paradigms. Table 17 presents a structured comparison of prominent state-of-the-art synthesis architectures and their principal contributions.

Contemporary models increasingly leverage the complementary strengths of deterministic and probabilistic approaches—combining interpretability and stability from explicit constraints (e.g., contrastive, feature-matching, or categorical losses) with the flexibility of multi-stage, generative learning frameworks [34, 71, 72, 74]. For example, FG-RAT GAN [34] integrates an auxiliary classifier in the discriminator to guide semantic class consistency and employs a contrastive objective that enhances intra-class alignment and inter-class separation, which is crucial for fine-grained text-to-image generation. The overall loss function includes adversarial, categorical cross-entropy, and contrastive components, as shown by the total discriminator and generator losses: $L_d^{\text{total}} = L_d^{\text{adv}} + L_d^{\text{ce}} + L_d^{\text{cl}}$ and $L_g^{\text{total}} = L_g^{\text{adv}} + L_g^{\text{ce}} + L_g^{\text{cl}}$. Experiments confirm that FG-RAT GAN attains state-of-the-art performance on benchmarks such as CUB-200-2011 and Oxford-102, demonstrating lowest FID and competitive IS while using the fewest parameters among compared models. VQ-Diffusion leverages quantization and controlled diffusion to ensure high semantic content and diversity, though it is typically more computationally intensive and less label-efficient than FG-RAT GAN [34, 74, 88]. Latent diffusion models [34, 71, 88] enable scalable, conditional generation within the latent space, facilitating applications like segmentation mask-guided synthesis and style transfer, but they require large, high-quality datasets for optimal performance.

Recent transformer-based and hybrid models further push the boundaries by allowing integration across multiple modalities and domain-specific conditions [22, 71, 75]. For instance, IIDM [71] re-frames semantic synthesis as a progressive image denoising task in the latent space, guided by both segmentation masks and style images, and supplements its diffusion model with refinement and color transfer mechanisms at inference. This enables higher mask accuracy and style fidelity compared to preceding GAN and diffusion solutions, without extra training costs. In the context of scientific and medical imaging, model-based synthetic datasets generated through simulation frameworks such as SyntheX [22] have been shown to rival or exceed the utility of real data for downstream learning-based tasks, especially when enhanced via strong domain randomization strategies.

Such synergies facilitate the creation of perceptually convincing, diverse, and controllable samples, empowering workflows across vision, language, and scientific applications. At the same time, emerging methods like image-to-image diffusion with refinement and color transfer modules [71] or simulation-based synthetic data frameworks [22] demonstrate the potential of augmenting or even replacing real datasets in constrained domains. Ongoing challenges include the need for comprehensive, multi-domain benchmarking and standardized evaluation, as well as ensuring the effectiveness, fairness, and traceability of generated data [22, 23, 34, 71, 74, 86]. Responsible and systematic benchmarking will be indispensable for steering the adoption of generative and augmentation technologies, ensuring their integrity and utility in real-world applications.

10 Conclusion

10.1 Explicit Objectives and Unique Contributions

This survey set out to advance the state of research on fairness, bias mitigation, and reliability in modern AI systems for a broad interdisciplinary audience of researchers, practitioners, and policymakers. Our explicit objectives were to: (1) systematically synthesize and critically compare technical approaches for fairness-aware machine learning; (2) measure and contextualize core concepts such as demographic parity, equalized odds, and individual versus group-level fairness definitions across commonly studied domains; (3) elucidate ethical frameworks and analytical desiderata that shape the evaluation of bias and trustworthiness, drawing on both technical literature and sociotechnical perspectives; and (4) articulate open technical and ethical challenges, with recommendations for future research. Distinguishing this work, we introduced a novel taxonomy that categorizes fairness interventions by site (pre-processing, in-processing, post-processing) and context, and mapped key interdependencies between ethical considerations, model architectures, and AI deployment pipelines. Readers can find a detailed taxonomy and discussion in Section ??.

10.2 Integrative Perspective and Synthesis

Unlike prior reviews, our integrative approach bridges recent technical advancements with responsible innovation, highlighting how ethical frameworks, evaluation methods, and model design mutually influence one another. We provided a clearer synthesis of outstanding controversies, particularly the enduring tension between statistical fairness criteria and individual-level guarantees, and the ongoing question of whether or how algorithmic bias may be meaningfully addressed apart from its broader data and societal context.

Table 17: Comparison of Select State-of-the-Art Generative Synthesis Models

Model	Key Features	Domain/Applications	Notable Strengths/Trade-offs
FG-RAT GAN [74]	Auxiliary classifier in discriminator, contrastive learning through cross-batch memory (XBM), fine-grained text-to-image synthesis	Vision, cross-modal generation	Achieves state-of-the-art FID and competitive IS on fine-grained datasets, surpassing LAFITE and VQ-Diffusion with much fewer parameters; highly efficient but depends on labeled subclasses [34]
VQ-Diffusion [74, 88]	Vector quantization combined with diffusion processes, controllable synthesis pipeline	High-resolution image generation, data augmentation	Excels in semantic fidelity and sample diversity, but is computationally intensive; does not match FG-RAT GAN in label-efficient settings
LDM (Latent Diffusion Models) [34, 88]	Synthesis in latent space, scalable architecture, supports conditional translation	Image-to-image translation, semantic segmentation, style transfer	Delivers realistic and semantically aligned outputs; scalability comes at the cost of requiring large, high-quality datasets for effective training
Transformer-based/Hybrid Models [22, 71, 75]	Captures global structure via transformers, can be integrated with convolutional layers and cross-modal inputs	Pluralistic image completion, robust multi-domain data synthesis	Balances interpretability and expressiveness, demonstrates adaptability to complex or multi-modal synthesis scenarios; potential challenges in model complexity and training

10.3 Key Open Challenges and Future Directions

To guide ongoing research, Table 18 summarizes major open challenges and promising research directions as identified across the surveyed literature. For a structured taxonomy of fairness interventions, see Table 19. Bridging these technical and ethical fronts remains essential for holistic progress.

10.4 Summary and Recommendations

Realizing genuinely fair, unbiased, and reliable AI remains an unresolved and evolving challenge. We encourage continued collaboration and shared benchmarks across technical, social, and policy domains. Our synthesis, unified taxonomy, and clear enumeration of open challenges are intended to serve as a reference and catalyst for ongoing advances in responsible AI. For more details on discussed objectives, methodologies, and future work, readers are referred to Sections **??**, **??**, and **??**.

10.5 Transformative Developments

The past decade has witnessed a profound transformation in generative modeling, marked by the emergence of generative adversarial networks (GANs), diffusion models, hybrid and transformer-based architectures, context-aware generative frameworks, and adaptive data augmentation methods. These advances have irreversibly reshaped the landscape of computer vision, medical artificial intelligence, drug discovery, and related fields, establishing generative models not only as tools for data synthesis but as indispensable engines for representation learning, privacy preservation, and scientific progress [21, 22, 32, 34, 48, 53, 56, 69, 71–73, 76, 77, 79, 82, 88, 91].

Since their introduction, GANs have catalyzed paradigm shifts in imaging, text-to-image synthesis, simulation-driven research, and even regulatory strategies in sensitive domains such as healthcare and the social sciences. Their strength in modeling high-dimensional distributions has enabled the creation of realistic synthetic datasets, addressing persistent issues of data scarcity, imbalance, and privacy. Applications range from medical diagnostics to the construction of synthetic patient populations and "digital twins" for regulatory-compliant data sharing [56, 72, 73, 76, 77, 82]. Despite their promise, GANs face notable limitations, including mode collapse, training instability, and vulnerability to biases in the training corpus. In response, the field has introduced architectural innovations—such as conditional, Wasserstein, and hybrid GANs—and has elevated the integration of fairness-aware and privacy-preserving algorithms [21, 53, 73, 77].

Diffusion models, representing a subsequent wave of innovation, have exerted significant influence, particularly in computer vision and scientific imaging. Employing noise-perturbation and denoising dynamics, these models generate diverse and high-quality samples in both unconditional and conditional settings. They frequently surpass GANs in sample fidelity and stability, though at a cost of increased computational requirements [32, 34, 79, 88]. Recent developments such as latent diffusion models have contributed to a significant reduction in computational demands without compromising output quality, as they operate within compressed, perceptually-rich latent spaces and allow efficient conditioning on various modalities and high-resolution tasks [71, 91]. For example, latent diffusion approaches have been demonstrated to deliver competitive or state-of-the-art performance with far fewer parameters than conventional pixel-wise models, as shown in benchmarks on unconditional and conditional image generation as well as inpainting and super-resolution [91]. Furthermore, diffusion models have been successfully applied to semantically complex domains such as conditional 3D molecule generation and medical imaging. Innovations include introducing SE(3)-equivariant architectures, as in shape-conditioned diffusion models for 3D molecule design, enabling robust chemical and geometric validity, and developing measurement-guided or uncertainty-guided frameworks to improve the informativeness, reliability, and clinical diagnostic utility of synthetic images in healthcare [56, 71, 76].

Building on these strengths, hybrid approaches—notably latent diffusion models and architectures incorporating transformer modules—have scaled generative modeling to complex, multi-modal, and high-resolution regimes [71, 79]. Transformers, with their aptitude for modeling long-range dependencies, have unified generative pipelines across text, vision, and cross-modal tasks. This capacity enables open-vocabulary detection, segmentation, and self-supervised learning, further blurring the lines between symbolic and subsymbolic artificial intelligence and facilitating context-aware synthesis and control [19, 22, 48, 69, 71, 91]. Particularly, pluralistic image completion has benefited from architectures that merge transformers for global relationship modeling and convolutional networks for local detail refinement, resulting in superior image fidelity, diversity, and generalization across generic and high-resolution tasks [71, 79, 82]. In such approaches, transformers reconstruct coherent global structures while convolutional layers replenish fine local textures, establishing new state-of-the-art performance in pluralistic image completion benchmarks [82].

Context mechanisms—via side-information, domain adaptation, or weak supervision—demonstrate how generative models can bridge unsupervised synthesis and (semi-)supervised learning, thereby enhancing both data fidelity and downstream task utility [69, 73, 82]. For example, explicit diversity objectives and semantic guidance in conditional synthesis have enabled user-driven manipulation and improved representation of plausible outputs [69]. Concurrently, adaptive data augmentation has evolved from manual techniques to meta-learned, contextually optimized curricula. These methods promote generalization, fairness, and robustness, even in low-resource or distributionally-shifted settings [34, 48, 53, 77, 91]. Collectively, these innovations yield not merely richer synthetic datasets but

Table 18: Open Challenges and Future Research Directions in Fairness, Bias, and Reliability

Challenge Category	Open Challenge	Current Limitation	Promising Research Directions
Technical	Formalizing Fairness	Lack of unified definitions across domains	Developing adaptable, context-aware fairness metrics
Ethical	Societal Context	Difficulty modeling social and historical biases	Integrating sociotechnical analysis within algorithm design
Reliability	Model Robustness	Vulnerability to adversarial attacks and distribution shifts	Robust optimization and certification methods
Evaluation	Standardized Benchmarks	Fragmented evaluation datasets and protocols	Establishing community benchmarks and shared resources
Deployment	Interpretability	Lack of transparent reasoning for complex models	Advancing explainable AI and post-hoc interpretability tools

Table 19: Taxonomy of Fairness Interventions Across the Model Lifecycle

Stage	Category	Example Techniques	Typical Objectives
Pre-processing	Data Modification	Re-weighting, re-sampling, data repair	Remove bias from training data
In-processing	Algorithmic Adjustment	Fairness-constrained optimization, adversarial debiasing	Embed fairness constraints during learning
Post-processing	Output Adjustment	Calibration, threshold adjustment, equal opportunity post-processing	Amend predictions to achieve fairness criteria

dynamic frameworks, wherein synthetic data generation, augmentation, and feedback from subsequent tasks inform one another iteratively.

10.6 Principles for Future AI

A critical evaluation of generative modeling's current ecosystem surfaces several guiding principles shaping the discipline's trajectory. Foremost among these is generalization: models must transcend idiosyncratic dataset artifacts to demonstrate robustness across divergent tasks, populations, and environments. Fairness and alignment require ongoing diligence, addressing historical biases, demographic underrepresentation, and minimizing social or regulatory harms at the data, model, and system levels [48, 53, 77, 88, 91]. Scalability remains essential, especially as generative AI expands from narrowly defined applications to open-world or multi-modal environments; here, hybrid architectures and foundation models must uphold both efficiency and adaptability [21, 56, 71, 79, 82]. Above all, ethical design—grounded in transparency, accountability, and human-centered values—must underlie every generative modeling system, safeguarding against unintended consequences and sustaining public trust.

Generalization is essential for ensuring that AI models exhibit robustness across diverse tasks, datasets, and operational conditions. Models achieving high fidelity and diversity, as demonstrated in image synthesis tasks [48, 71, 79, 91], illustrate the importance of architectures and loss functions that improve visual realism and maintain semantic consistency.

Fairness and alignment remain foundational for trustworthy AI. This involves continuous effort to mitigate historical biases and ethical risks, integrating mechanisms both at the data curation stage and within model design [53, 77]. Notably, quality assessment methods [77] that better align with human subjective judgment are critical for identifying and addressing potential artifacts or sources of discrimination.

Scalability supports efficient transfer of generative models to new domains while maintaining flexibility. The evolution from convolutional approaches to transformer-based and hybrid models demonstrates quantitative gains in fidelity, diversity, and adaptability [21, 71, 82, 91]. Efficient latent representation [71, 91] and architectural innovations are central to ensuring models remain adaptable as application contexts broaden.

Ethical design must be embedded at every level of generative modeling. Transparency, explainability, and accountability—in both the system's underlying mechanisms and its outputs—are necessary to guard against unintended consequences and maintain public trust.

10.7 Bridging Technical and Responsible Innovation

Technical advances in generative modeling demand parallel progress in interpretability, security, policy compliance, and responsible deployment. Among these, interpretability is especially pressing. As models grow in complexity and opacity, mechanisms for human-understandable explanations, provenance tracking, watermarking, and information disclosure become critical, especially within high-stakes environments such as healthcare, law, and critical infrastructure [27, 74, 85]. Security and privacy also require multifaceted solutions, comprising both technical guarantees (differential privacy, k-anonymity, adversarial robustness) and best operational practices aimed at preventing reidentification, memorization, and membership inference attacks [27, 76, 77, 85]. The regulatory landscape is rapidly adapting to generative AI's expanding influence, with ongoing efforts to establish standardized evaluation protocols, maintain auditable usage records, and introduce oversight at the dataset and model levels [72, 74, 76, 85, 88].

Responsible deployment is therefore best understood as a sociotechnical enterprise, one balancing utility against risk to ensure generative models are not only accurate but also equitable, explainable, and responsive to shifting legal and societal imperatives. This necessitates interdisciplinary research, continual stakeholder input, and sustained investment at the intersection of machine learning, ethics, and governance [27, 72, 74, 85].

Integration of interpretability tools and protocols is vital for ensuring transparent, traceable outputs throughout the model development lifecycle. Robust privacy-guarding mechanisms must be implemented to address risks such as bias, synthetic data contamination, and membership inference. The adoption of standardized evaluation and audit processes supports ongoing accountability, particularly as the absence of domain-tailored protocols hinders deployment in sensitive fields like healthcare [76, 85]. Finally, effective progress depends on sustained and open dialogue among technical, ethical, and policy stakeholders, and the establishment of responsible usage standards and collaborative practices to guide the evolution of generative AI [27, 74, 85].

10.8 Outlook

10.8.1 Summary of Objectives and Contributions. This survey aimed to systematically review and synthesize recent advances in synthetic data generation, focusing on technical trends and responsible innovation across a wide range of generative models—including GANs, diffusion models, transformers, and hybrid frameworks. Our objectives were to (i) clarify foundational principles, (ii) surface persistent technical and ethical challenges, (iii) establish integrative connections between innovative architectures and application domains, and (iv) provide a taxonomy that unifies methodological developments and societal considerations. By foregrounding the interplay between technical progress (Section 2) and best practices in transparency, evaluation, bias mitigation, and reproducibility (Section 4), we highlight how this survey offers a more unified and actionable perspective compared to prior reviews.

10.8.2 Distinctive Integrative Perspective. A key contribution of this work is the synthesis of technical and responsible innovation, which sets it apart from previous surveys. Whereas earlier efforts typically focused on singular generative architectures or application-specific outcomes, our review emphasizes both the rapid convergence of generative technologies and the need to align them with transparency, reproducibility, and fairness imperatives [21, 56, 71, 79]. For instance, the integration of context-driven frameworks (Section 3.4) and adaptive augmentation with robust benchmarks demonstrates how methodological advances must be grounded in open and reproducible practices to ensure trustworthy progress. The explicit discussion of evaluation standards and societal risks—drawn from a broad range of domains, including medicine, law, and broader ethics—further underscores our survey's unique integrative scope (see Section 4.2).

10.8.3 Open Problems and Emerging Applications. Despite remarkable progress, several open challenges remain. In medical image synthesis and clinical AI, ensuring high diagnostic utility from synthetic data demands advances in uncertainty-guided generation [76], domain adaptation, and robust benchmark evaluation [22, 71]. In sensitive areas such as drug discovery, geometry-complete generative models must grapple with high computational requirements and the need for property-preserving synthesis [56]. Novel view synthesis for gigapixel-scale rendering and semantic image synthesis confront obstacles in balancing semantic fidelity, diversity of plausible outputs, and computational efficiency [69, 72, 82]. Furthermore, the emergence of "AI autophagy"—where synthetic outputs feed recursively into future training data—poses sustainability and trust risks for generative AI [88]. Addressing these issues will require strategies for curating mixed real and synthetic datasets, measurement-guided generation, and transparent evaluation pipelines.

10.8.4 Proposed Taxonomy for Guided Progress. To guide future research, we propose a concise taxonomy synthesizing both technical and ethical innovation. Generative models can be classified by: (i) core architecture (GANs, diffusion models, transformer-based, and hybrids), (ii) primary application modality (image, text, multimodal, medical, 3D molecular), (iii) operational paradigm (unconditional, conditional, context-driven, augmentation-oriented), and (iv) responsible innovation criteria (transparency/reproducibility, fairness/bias mitigation, privacy assurance, sustainability). This conceptual framework, supported by recent taxonomies in application domains such as recommendation and clinical imaging [21, 22, 76], is intended as a living guide for researchers and practitioners to navigate the evolving landscape and align new methods with foundational ethical principles (see Section 2.1 and Section 4.2 for mapping across these axes).

10.8.5 Outlook and Call to Action. Looking ahead, the sustainable and responsible evolution of synthetic data research will depend critically on interdisciplinarity, transparency, and community stewardship. Cross-disciplinary collaboration—engaging computer science, statistics, social science, medicine, and law—is required to address persistent challenges in evaluation, privacy, and bias. Transparent practices, open-source tools, and common benchmarks will be critical for reproducibility and trustworthy comparative evaluation [21, 56, 71, 79]. Engagement from a diverse, global community—comprising both technical innovators and affected stakeholders—will be necessary to align generative AI development with societal interests and mitigate the risk of amplifying inequities or eroding trust.

In conclusion, the convergence of GANs, diffusion models, transformer and hybrid architectures, context-driven generative frameworks, and adaptive augmentation signals a new epoch in artificial intelligence. These technologies are simultaneously fueling innovation and surfacing critical questions regarding fairness, accountability, and the overall value to society. The path forward rests on a deliberate synthesis of technical advancement and responsible innovation—anchored by foundational principles and a steadfast commitment to the public good [21, 22, 32, 34, 48, 53, 56, 69, 71–73, 76, 77, 79, 82, 88, 91].

References

[1] D. Adam. 2024. Synthetic data can aid the analysis of clinical outcomes: How much can it be trusted? *Proceedings of the National Academy of Sciences* 121, 32 (2024), e2414310121. doi:10.1073/pnas.2414310121

[2] I. Ahmed, T. Xie, A. K. Bashir, and A. D. Jurcut. 2024. Computer Vision Based Transfer Learning-Aided Transformer Model for Plant Disease Recognition. *IEEE Access* 12 (2024), 28798–28809. doi:10.1109/ACCESS.2024.3363345

[3] M. U. Akbar, W. Wang, and A. Eklund. 2025. Beware of diffusion models for synthesizing medical images—a comparison with GANs in terms of memorizing brain MRI and chest x-ray images. *Machine Learning: Science and Technology* 6, 1 (2025), 015022. doi:10.1088/2632-2153/ad9a3a

[4] Duanhua Cao, Geng Chen, Jiaxin Jiang, Jie Yu, Runze Zhang, Mingan Chen, Wei Zhang, Lifan Chen, Feisheng Zhong, Yingying Zhang, Chenghao Lu, Xutong Li, Xiaomin Luo, Sulin Zhang, and Mingyue Zheng. 2024. Generic protein–ligand interaction scoring by integrating physical prior knowledge and data augmentation modelling. *Nature Machine Intelligence* 6, 6 (2024), 688–700. doi:10.1038/s42256-024-00849-z

[5] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko. 2020. End-to-End Object Detection with Transformers. *arXiv preprint arXiv:2005.12872* (2020). https://arxiv.org/abs/2005.12872

[6] C. Chadebec, E. Thibeau-Sutre, N. Burgos, and S. Allassonnière. 2023. Data Augmentation in High Dimensional Low Sample Size Setting Using a Geometry-Based Variational Autoencoder. *IEEE Transactions on Pattern Analysis and Machine*

Intelligence 45, 3 (March 2023), 2879–2896. doi:10.1109/TPAMI.2022.3185773 https://ieeexplore.ieee.org/document/9806307.

[7] T. Chakraborty, U. Reddy K S, S. M. Naik, M. Panja, and B. Manvitha. 2024. Ten years of generative adversarial nets (GANs): a survey of the state-of-the-art. *Machine Learning: Science and Technology* 5, 1 (2024), 011001. doi:10.1088/2632-2153/ad1f77

[8] J. Chen, H. Chen, K. Chen, Y. Zhang, Z. Zou, and Z. Shi. 2025. Diffusion Models for Imperceptible and Transferable Adversarial Attack. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 47, 2 (2025), 961–977. doi:10.1109/TPAMI.2024.3480519

[9] M. Chen, S. Mei, J. Fan, and M. Wang. 2024. An Overview of Diffusion Models: Applications, Guided Generation, Statistical Rates and Optimization. *arXiv preprint arXiv:2404.07771* (2024). https://arxiv.org/abs/2404.07771

[10] P. Chen, X. Yu, X. Han, K. Wang, G. Li, L. Xie, Z. Han, and J. Jiao. 2025. P2Object: Single Point Supervised Object Detection and Instance Segmentation. *International Journal of Computer Vision* (2025). doi:10.1007/s11263-025-02441-3

[11] S. Chen, E. Dobriban, and J. H. Lee. 2020. A Group-Theoretic Framework for Data Augmentation. *Journal of Machine Learning Research* 21, 245 (2020), 1–71. https://www.jmlr.org/papers/volume21/20-163/20-163.pdf

[12] Y. Chen, X. Yuan, R. Wu, J. Wang, Q. Wang, L. Zhang, and M.-M. Cheng. 2025. YOLO-MS: Rethinking Multi-Scale Representation Learning for Real-Time Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2025). doi:10.1109/TPAMI.2025.10872821

[13] Ziqi Chen, Bo Peng, Xia Ning, Mengxuan Zhou, Jiaqi Zhu, Pengyong Li, Rongrong Jin, and Long Zhang. 2025. Generating 3D small binding molecules using shape-conditioned diffusion models with guidance. *Nature Machine Intelligence* 7 (2025), 588–591. https://www.nature.com/articles/s42256-025-01009-4

[14] V. Claveau, A. Chaffin, and E. Kijak. 2021. Generating artificial texts as substitution or complement of training data. *Artificial Intelligence* 298 (2021), 103528. https://arxiv.org/pdf/2110.13016

[15] E. De Cristofaro. 2024. Synthetic Data: Methods, Use Cases, and Risks. *arXiv preprint arXiv:2303.01230* (2024). https://arxiv.org/abs/2303.01230

[16] F. A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah. 2023. Diffusion Models in Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 9 (2023), 10850–10869. doi:10.1109/TPAMI.2023.3261988

[17] M. Dedeoglu, S. Lin, Z. Zhang, and J. Zhang. 2024. Continual Learning of Generative Models With Limited Data: From Wasserstein-1 Barycenter to Adaptive Coalescence. *IEEE Transactions on Neural Networks and Learning Systems* 35, 9 (2024), 12042–12055. doi:10.1109/TNNLS.2024.10070745

[18] Yashar Deldjoo, Zhi He, Julian McAuley, Andrey Korikov, Scott Sanner, Arnau Ramisa, Ramon Vidal, Mahinthan Sathiamoorthy, Amirhossein Kasrizadeh, Simone Milano, and Francesco Ricci. 2024. Recommendation with Generative Models. *arXiv preprint arXiv:2409.15173* (sep 2024). https://arxiv.org/abs/2409.15173

[19] A. P. Dempster, N. M. Laird, and D. B. Rubin. 2023. Artificial Data in Bootstrapping: Perspectives on the Origins and Implications. *Artificial Intelligence* 323 (2023), 103194. https://www.sciencedirect.com/science/article/pii/S0004370222001456.

[20] J. Ding, N. Xue, Y. Long, G.-S. Xia, Q. Lu, A. Bai, W. Yang, and X. Yao. 2022. Object Detection in Aerial Images: A Large-Scale Benchmark and Challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 11 (2022), 7778–7796. doi:10.1109/TPAMI.2021.3117988

[21] C. Du, J. Zhu, and B. Zhang. 2017. Learning Deep Generative Models With Doubly Stochastic MCMC. *IEEE Transactions on Neural Networks and Learning Systems* 28, 6 (2017), 1371–1383. doi:10.1109/TNNLS.2017.2688499

[22] X. Du, N. Kolkin, G. Shakhnarovich, and A. Bhattad. 2024. Generative Models: What Do They Know? Do They Know Things? Let's Find Out! *arXiv preprint arXiv:2311.17137* (Oct. 2024). https://arxiv.org/abs/2311.17137

[23] L. L. Duan, J. E. Johndrow, and D. B. Dunson. 2018. Scaling up Data Augmentation MCMC via Calibration. *Journal of Machine Learning Research* 19, 64 (2018), 1–34. https://jmlr.org/papers/volume19/17-573/17-573.pdf

[24] Chukwuebuka Joseph Ejiyi, Dongsheng Cai, Francis Ofoma Eze, Makuachukwu Bennedith Ejiyi, Jennifer Ene Idoko, Sarpong Kwadwo Asere, and Thomas Ugochukwu Ejiyi. 2025. Polynomial-SHAP as a SMOTE alternative in conglomerate neural networks for realistic data augmentation in cardiovascular and breast cancer diagnosis. *Journal of Big Data* 12 (2025), Article 97. doi:10.1186/s40537-025-01152-3

[25] L. Fan, Y. Yang, F. Wang, N. Wang, Z. Zhang, Z. Cao, and D. Lin. 2023. Super Sparse 3D Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 10 (2023), 12490–12505. doi:10.1109/TPAMI.2023.3286409

[26] F. Fürrutter, G. Muñoz-Gil, and H. J. Briegel. 2024. Quantum circuit synthesis with diffusion models. *Nature Machine Intelligence* 6, 5 (2024), 515–524. doi:10.1038/s42256-024-00831-9

[27] Cong Gao, Benjamin D. Killeen, Yicheng Hu, Robert B. Grupp, Russell H. Taylor, Mehran Armand, and Mathias Unberath. 2023. Synthetic data accelerates the development of generalizable learning-based algorithms for X-ray image analysis. *Nature Machine Intelligence* 5 (2023), 294–308. doi:10.1038/s42256-023-00629-1

[28] M. Goyal and Q. H. Mahmoud. 2024. A Systematic Review of Synthetic Data Generation Techniques Using Generative AI. *Electronics* 13, 17 (2024), 3509. https://www.mdpi.com/2079-9292/13/17/3509

[29] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask R-CNN. *arXiv preprint arXiv:1703.06870* (2017). https://arxiv.org/abs/1703.06870

[30] Philipp Hess, Michael Aich, Baoxiang Pan, and Niklas Boers. 2025. Fast, scale-adaptive and uncertainty-aware downscaling of Earth system model fields with generative machine learning. *Nature Machine Intelligence* 7 (2025), 363–373. doi:10.1038/s42256-025-00980-5

[31] C. Hou, J. Zhang, and T. Zhou. 2023. When to Learn What: Model-Adaptive Data Augmentation Curriculum. *arXiv preprint arXiv:2309.04747* (sep 2023). https://arxiv.org/abs/2309.04747

[32] M. Ibrahim, Y. Al Khalil, S. Amirrajab, C. Sun, M. Breeuwer, J. Pluim, B. Elen, G. Ertaylan, and M. Dumontier. 2024. Generative AI for Synthetic Data Across Multiple Medical Modalities: A Systematic Review of Recent Developments and Challenges. *arXiv preprint arXiv:2407.00116* (2024). https://arxiv.org/abs/2407.00116

[33] I. Igashov, H. Stärk, C. Vignac, A. Schneuing, V. Garcia Satorras, P. Frossard, M. Welling, M. Bronstein, and B. Correia. 2024. Equivariant 3D-conditional diffusion model for molecular linker design. *Nature Machine Intelligence* 6, 4 (2024), 417–427. doi:10.1038/s42256-024-00815-9

[34] Junjun Jiang, Yi Yu, Zheng Wang, Xianming Liu, and Jiayi Ma. 2019. Graph-Regularized Locality-Constrained Joint Dictionary and Residual Learning for Face Sketch Synthesis. *IEEE Transactions on Image Processing* 28, 2 (February 2019), 628–641. doi:10.1109/TIP.2018.2870936 https://doi.org/10.1109/TIP.2018.2870936.

[35] X. Jiang and Y. Wu. 2023. Remote Sensing Object Detection Based on Convolution and Swin Transformer. *IEEE Access* 11 (2023), 38643–38656. doi:10.1109/ACCESS.2023.3267435

[36] I. Joshi, M. Grimmer, J. Fierrez, A. Dantcheva, J. L. Crowley, and J. Busch. 2024. Synthetic Data in Human Analysis: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 7 (July 2024), 4957–4976. https://ieeexplore.ieee.org/document/10423161.

[37] A. Karev and D. Xu. 2025. ConSCompF: Consistency-focused Similarity Comparison Framework for Generative Large Language Models. *Journal of Artificial Intelligence Research* 82 (2025), 1–32. https://www.jair.org/index.php/jair/article/view/17028

[38] T. Karras, M. Aittala, T. Aila, and S. Laine. 2022. Elucidating the Design Space of Diffusion-Based Generative Models. *arXiv preprint arXiv:2206.00364* (2022). https://arxiv.org/abs/2206.00364

[39] S. Kong and D. Ramanan. 2025. OpenGAN: Open-Set Recognition via Open Data Generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 47, 5 (May 2025), 3233–3243. https://ieeexplore.ieee.org/document/9799769.

[40] Lisa Kühnel, Julian Schneider, Ines Perrar, Tim Adams, Sobhan Moazemi, Fabian Prasser, Ute Nöthlings, Holger Fröhlich, and Juliane Fluck. 2024. Synthetic data generation for a longitudinal cohort study – evaluation, method extension and reproduction of published data analysis results. *Scientific Reports* 14 (2024), 14412. doi:10.1038/s41598-024-62102-2

[41] V. Kumar, P. Singh, S. Sharma, and R. Prasad. 2024. Computer Vision-Based Framework for Data Extraction From Document Images Using Deep Learning Techniques. *IEEE Access* 13 (2024), 17706–17723. doi:10.1109/ACCESS.2024.3351125 https://ieeexplore.ieee.org/document/10813363/.

[42] B. Li, Y. Hou, and W. Che. 2022. Data Augmentation Approaches in Natural Language Processing: A Survey. *arXiv preprint arXiv:2110.01852 [cs.CL]* (June 2022). https://arxiv.org/abs/2110.01852

[43] T. Li, L. Biferale, F. Bonaccorso, M. A. Scarpolini, and M. Buzzicotti. 2024. Synthetic Lagrangian turbulence by generative diffusion models. *Nature Machine Intelligence* 6, 4 (2024), 393–403. doi:10.1038/s42256-024-00810-0

[44] X. Li, H. Ding, H. Yuan, W. Zhang, J. Pang, G. Cheng, K. Chen, Z. Liu, and C. C. Loy. 2024. Transformer-Based Visual Segmentation: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 6 (2024), 10138–10163. doi:10.1109/TPAMI.2024.10613466

[45] C.-H. Lin, C. Kaushik, E. L. Dyer, and V. Muthukumar. 2024. The good, the bad and the ugly sides of data augmentation: An implicit spectral regularization perspective. *Journal of Machine Learning Research* 25, 91 (2024), 1–85. https://jmlr.org/papers/volume25/22-1312/22-1312.pdf

[46] L. Lin, H. Mu, Z. Zhai, M. Wang, Y. Wang, R. Wang, J. Gao, Y. Zhang, W. Che, T. Baldwin, X. Han, and H. Li. 2025. Against The Achilles' Heel: A Survey on Red Teaming for Generative Models. *Journal of Artificial Intelligence Research* 82 (2025), 1–84. https://www.jair.org/index.php/jair/article/view/17654

[47] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen. 2020. Deep learning for generic object detection: A survey. *Pattern Recognition* 102 (2020), 107198. https://www.sciencedirect.com/science/article/pii/S0031320319303851

[48] R. Liu, J. Wei, F. Liu, C. Si, Y. Zhang, J. Rao, S. Zheng, D. Peng, D. Yang, D. Zhou, and A. M. Dai. 2024. Best Practices and Lessons Learned on Synthetic Data. *arXiv preprint arXiv:2404.07503* (Aug 2024). https://arxiv.org/abs/2404.07503

[49] Y. Liu, R. Shen, and X. Shen. 2024. Novel Uncertainty Quantification Through Perturbation-Assisted Sample Synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 12 (2024), 7813–7824. https://ieeexplore.ieee.org/document/10508110

[50] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. *arXiv preprint arXiv:2103.14030* (2021). https://arxiv.org/abs/2103.14030

[51] L. Long, R. Wang, R. Xiao, J. Zhao, X. Ding, G. Chen, and H. Wang. 2024. On LLMs-Driven Synthetic Data Generation, Curation, and Evaluation: A Survey. *arXiv preprint arXiv:2406.15126* (2024). https://arxiv.org/abs/2406.15126

[52] Y. Lu, L. Chen, Y. Zhang, M. Shen, H. Wang, X. Wang, C. van Rechem, T. Fu, and W. Wei. 2023. Machine Learning for Synthetic Data Generation: A Review. *arXiv preprint arXiv:2302.04062* (2023). https://arxiv.org/abs/2302.04062

[53] Y. Luo, Q. Yang, Y. Fan, H. Qi, and M. Xia. 2024. Measurement Guidance in Diffusion Models: Insight from Medical Image Synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 12 (2024), 7983–7997. doi:10.1109/TPAMI.2024.3363309

[54] S. Nie, F. Zhu, Z. You, X. Zhang, J. Ou, J. Hu, J. Zhou, Y. Lin, J.-R. Wen, and C. Li. 2025. Large Language Diffusion Models. *arXiv preprint arXiv:2502.09992* (2025). https://arxiv.org/abs/2502.09992

[55] E. Papadaki, A. G. Vrahatis, and S. Kotsiantis. 2024. Exploring Innovative Approaches to Synthetic Tabular Data Generation. *Electronics* 13, 10 (2024), 1965. doi:10.3390/electronics13101965

[56] S. Perdikis, R. Leeb, R. Chavarriaga, and J. d. R. Millán. 2021. Context-Aware Learning for Generative Models. *IEEE Transactions on Neural Networks and Learning Systems* 32, 8 (2021), 3471–3483. doi:10.1109/TNNLS.2020.3014897

[57] Zhaozhi Qian, Thomas Callender, Baiba Cebere, Sarah M. Janes, Nishant Navani, and Mihaela van der Schaar. 2024. Synthetic data for privacy-preserving clinical risk prediction. *Scientific Reports* 14 (2024), Article number: 25676. doi:10.1038/s41598-024-72894-y

[58] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. *arXiv preprint arXiv:2103.00020* (2021). https://arxiv.org/abs/2103.00020

[59] H. H. Rashidi, S. Albahra, B. P. Rubin, and B. Hu. 2024. A novel and fully automated platform for synthetic tabular data generation and validation. *Scientific Reports* 14 (2024). https://www.nature.com/articles/s41598-024-73608-0

[60] D. B. Resnik, M. Hosseini, J. J. H. Kim, G. Epiphaniou, and C. Maple. 2025. GenAI synthetic data create ethical challenges for scientists. Here's how to address them. *Proceedings of the National Academy of Sciences* 122, 9 (2025), e2409182122. doi:10.1073/pnas.2409182122

[61] Y. Rong, T. Leemann, T. t. Nguyen, L. Fiedler, P. Qian, V. V. Unhelkar, T. Seidel, G. Kasneci, and E. Kasneci. 2024. Towards Human-Centered Explainable AI: A Survey of User Studies for Model Explanations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 4 (2024), 2104–2122. https://dblp.org/rec/journals/pami/RongLNFQUSKK24

[62] S. Ruggles. 2025. The shortcomings of synthetic census microdata. *Proceedings of the National Academy of Sciences* 122, 11 (2025), e2424655122. doi:10.1073/pnas.2424655122

[63] J. Rydzewski, J. M. A. Grimme, and M. W. F. Fischer. 2023. Manifold learning in atomistic simulations: a conceptual review. *Machine Learning: Science and Technology* 4, 3 (2023), 031001. doi:10.1088/2632-2153/ace81a

[64] V. M. Sánchez-Cartagena, M. Esplà-Gomis, J. A. Pérez-Ortiz, and F. Sánchez-Martínez. 2024. Non-Fluent Synthetic Target-Language Data Improve Neural Machine Translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 2 (2024), 837–850. doi:10.1109/TPAMI.2023.3333949

[65] Hualian Sheng, Sijia Cai, Na Zhao, Bing Deng, Qiao Liang, Min-Jian Zhao, and Jieping Ye. 2024. CT3D++: Improving 3D Object Detection with Keypoint-induced Channel-wise Transformer. *CoRR* abs/2406.08152 (2024). doi:10.48550/ARXIV.2406.08152

[66] J. Shi, L. Ma, X. Wei, C. Fang, and Y. Zhang. 2024. Differentiable Image Data Augmentation and Its Applications: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 2 (February 2024), 1148–1164. doi:10.1109/TPAMI.2023.3330862 https://ieeexplore.ieee.org/document/10310260.

[67] Connor Shorten and Taghi M. Khoshgoftaar. 2019. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data* 6, 1 (2019), 60. doi:10.1186/s40537-019-0197-0

[68] C. Shorten, T. M. Khoshgoftaar, and B. Furht. 2021. Text Data Augmentation for Deep Learning. *Journal of Big Data* 8 (2021), 101. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00492-0

[69] Serban Stan and Mohammad Rostami. 2024. Preserving Fairness in AI under Domain Shift. *Journal of Artificial Intelligence Research* 81 (2024), 1966–2003. doi:10.1613/jair.1.16694

[70] Y. Sun, J. Zhang, Y. Liu, and X. Li. 2024. The evolution of object detection methods. *Pattern Recognition* 148 (2024), 110438. https://www.sciencedirect.com/science/article/pii/S003132032400136X

[71] Hongchen Tan, Xiuping Liu, Meng Liu, Baocai Yin, and Xin Li. 2021. KT-GAN: Knowledge-Transfer Generative Adversarial Network for Text-to-Image Synthesis. *IEEE Transactions on Image Processing* 30 (2021), 1275–1290. doi:10.1109/TIP.2021.3049544

[72] Y. Tang, J. Weng, and P. Zhang. 2023. Neural-network solutions to stochastic reaction networks. *Nature Machine Intelligence* 5 (2023), 376–385. doi:10.1038/s42256-023-00632-6

[73] R. Timpone and Y. Yang. 2024. Artificial Data, Real Insights: Evaluating Opportunities and Risks of Expanding the Data Ecosystem with Synthetic Data. *arXiv preprint arXiv:2408.15260* (Aug. 2024). https://arxiv.org/abs/2408.15260

[74] M. C. Tsakiris. 2023. Low-Rank Matrix Completion Theory via Plücker Coordinates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 8 (2023), 10084–10099. https://ieeexplore.ieee.org/document/10056236

[75] P.-D. Tudosiu, W. H. L. Pinaya, P. Ferreira Da Costa, J. Dafflon, A. Patel, P. Borges, V. Fernandez, M. S. Graham, R. J. Gray, P. Nachev, S. Ourselin, and M. J. Cardoso. 2024. Realistic morphology-preserving generative modelling of the brain. *Nature Machine Intelligence* 6, 7 (2024), 811–819. doi:10.1038/s42256-024-00864-0

[76] Wijnand van Woerkom, Davide Grossi, Henry Prakken, and Bart Verheij. 2024. A Fortiori Case-Based Reasoning: From Theory to Data. *Journal of Artificial Intelligence Research* 81 (2024), 2113–2151. doi:10.1613/jair.1.15178

[77] Z. Wan, J. Zhang, D. Chen, and J. Liao. 2024. High-Fidelity and Efficient Pluralistic Image Completion With Transformers. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 12 (2024), 9612–9629. doi:10.1109/TPAMI.2024.3382342

[78] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding. 2024. YOLOv10: Real-Time End-to-End Object Detection. *arXiv preprint arXiv:2405.14458* (2024). https://arxiv.org/abs/2405.14458

[79] G. Wang, J. Zhang, K. Zhang, R. Huang, and L. Fang. 2024. GiganticNVS: Gigapixel Large-Scale Neural Rendering With Implicit Meta-Deformed Manifold. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 1 (2024), 338–353. doi:10.1109/TPAMI.2023.3323069

[80] X. Wang, R. Girdhar, S. X. Yu, and I. Misra. 2023. Cut and Learn for Unsupervised Object Detection and Instance Segmentation. *arXiv preprint arXiv:2301.11320* (2023). https://arxiv.org/abs/2301.11320

[81] Y. Wang, G. Huang, S. Song, X. Pan, Y. Xia, and C. Wu. 2022. Regularizing Deep Networks With Semantic Data Augmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 7 (2022), 3733–3748. doi:10.1109/TPAMI.2021.3056495

[82] Ying Wang, Fuyuan Ma, Zhaoqi Yang, Yaodi Zhu, Bo Yang, Pengfei Shen, and Lei Yun. 2025. Rumor Detection with Adaptive Data Augmentation and Adversarial Training. *Journal of Artificial Intelligence Research* 82 (2025), 1–34. doi:10.1613/jair.1.16963

[83] Z. Wang, P. Wang, K. Liu, P. Wang, Y. Fu, C.-T. Lu, C. C. Aggarwal, J. Pei, and Y. Zhou. 2024. A Comprehensive Survey on Data Augmentation. *arXiv preprint arXiv:2405.09591 [cs.LG]* (May 2024). https://arxiv.org/abs/2405.09591

[84] Genta Indra Winata, Hanyang Zhao, Anirban Das, Wenpin Tang, David D. Yao, Shi-Xiong Zhang, and Sambit Sahu. 2025. Preference Tuning with Human Feedback on Language, Speech, and Vision Tasks: A Survey. *Journal of Artificial Intelligence Research* 82 (2025), 1–55. doi:10.1613/jair.1.17541

[85] Xiaodan Xing, Fadong Shi, Jiahao Huang, Yinzhe Wu, Yang Nan, Sheng Zhang, Yingying Fang, Michael Roberts, Carola-Bibiane Schönlieb, Javier Del Ser, and Guang Yang. 2025. On the caveats of AI autophagy. *Nature Machine Intelligence* 7 (2025), 172–180. doi:10.1038/s42256-025-00984-1

[86] H. Yang, J. Li, K. Z. Lim, C. Pan, T. V. Truong, Q. Wang, K. Li, S. Li, X. Xiao, M. Ding, T. Chen, X. Liu, Q. Xie, P. Valdivia y. Alvarado, X. Wang, and P.-Y. Chen. 2022. Automatic strain sensor design via active learning and data augmentation for soft machines. *Nature Machine Intelligence* 4, 1 (2022), 84–94. doi:10.1038/s42256-021-00434-8

[87] J. Yang, B. Hu, H. Li, Y. Liu, X. Gao, J. Han, F. Chen, and X. Wu. 2025. Dynamic VAEs via semantic-aligned matching for continual zero-shot learning. *Pattern Recognition* 160 (2025), 111199. doi:10.1016/j.patcog.2023.111199

[88] L. Yang, S. Wang, J. Yang, T. Nguyen, A. S. Aved, and J. Z. Wang. 2021. Artificial Data Synthesis for Machine Learning: A Review. *Artificial Intelligence* 294 (2021), 103492. https://www.sciencedirect.com/science/article/pii/S0004370221000644.

[89] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M.-H. Yang. 2024. Diffusion Models: A Comprehensive Survey of Methods and Applications. *Comput. Surveys* (2024). https://arxiv.org/abs/2209.00796 arXiv preprint arXiv:2209.00796, accepted by ACM Computing Surveys.

[90] M. Yang, X. Wang, and G. Zhao. 2020. Single-shot object detection with enriched semantics. *Pattern Recognition* 106 (2020), 107404. https://www.sciencedirect.com/science/article/pii/S0031320319303711

[91] Zichen Yang, Haifeng Liu, and Deng Cai. 2019. On the Diversity of Conditional Image Synthesis With Semantic Layouts. *IEEE Transactions on Image Processing* 28, 6 (2019), 2898–2907. doi:10.1109/TIP.2019.2891935

[92] B. Yin, X. Zhang, L. Liu, M.-M. Cheng, Y. Liu, and Q. Hou. 2025. Camouflaged Object Detection with Adaptive Partition and Background Retrieval. *International Journal of Computer Vision* (2025). doi:10.1007/s11263-025-02406-6

[93] Chaoyang Zhu and Long Chen. 2024. A Survey on Open-Vocabulary Detection and Segmentation: Past, Present, and Future. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 12 (2024), 8954–8975. doi:10.1109/TPAMI.2024.3413013