

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.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, Washington, DC, USA

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

In . ACM, New York, NY, USA, 20 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 Introduction

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 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, photo-realistic 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, 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].

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]. Despite these technological advances, several ethical and practical challenges persist:

- Generative models may inadvertently leak sensitive information through memorization or attribute disclosure.
- Comprehensive data sanitization remains technically elusive.
- There is potential for bias propagation, challenges in interpreting synthetic samples, and unforeseen outcomes, including erosion of trust in synthetic data and the emergence of novel forms of algorithmic discrimination [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 are progressively becoming indispensable for ensuring responsible advancement in this domain [15, 74].

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:** Theoretical and technical foundations, including data-centric motivations, core generative models, and foundational taxonomies.
- **Section 3:** Principal generative methods, tracing progress from classical simulation to state-of-the-art deep generative models, with critical evaluation of merits and limitations.
- **Section 4:** Survey of key application domains, encompassing object detection, segmentation, domain adaptation, healthcare, and privacy-preserving analytics.
- **Section 5:** Assessment methodologies covering quality, utility, and ethical dimensions of synthetic data, alongside current metrics and outstanding research challenges.
- **Section 6:** Discussion of open challenges, including adversarial risks, fairness, scalability, and resolving the indistinguishability between synthetic and real data.
- **Section 7:** Perspectives on future directions, emphasizing opportunities for the development of 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.

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

- **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 person-abots [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

Historical methods of synthetic data generation predate modern deep generative modeling and are rooted primarily in classical statistical modeling and explicit simulation. Key statistical models—including Gaussian mixtures, copulas, and Markov models—constituted the earliest systematic approaches to sampling from learned or assumed data distributions, typically under assumptions of independence or stationarity [52, 74, 75]. Explicit simulation, utilizing agent-based or rule-driven systems, has been especially valuable in

generating complex synthetic environments with explicit control over latent variables. Notable applications include medical imaging phantoms, physical simulations, and synthetic population modeling [22, 74].

Early synthetic data pipelines often employed value- and structure-based transformations, such as geometric manipulations (rotations or translations), noise injection, or parametric sampling [6, 52, 75, 81]. While such traditional approaches offered high mathematical tractability and interpretability, their expressive power and diversity were limited. Consequently, they often failed to capture the rich multimodal dependencies, high-dimensional relationships, or rare-event features that characterize real-world datasets [22, 46, 74, 81].

Despite these limitations, the interpretability and theoretical rigor of traditional methods provided a crucial foundation for advanced techniques. However, the increasing complexity and scale of data-centric tasks—particularly those requiring realistic domain shifts or significant annotation diversity—began to expose their boundaries. This motivated the transition towards more flexible, data-driven frameworks for synthetic data generation [22, 52, 75].

2.3 Emergence of GANs and Diffusion Models

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 min-imax 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].

Nevertheless, GANs exhibit inherent challenges, such as mode collapse, training instability, and an inability to directly estimate sample likelihoods or uncertainty [21, 28, 52, 72]. 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, 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, 63, 85]. Their compatibility with structured or multimodal conditioning (e.g., class labels, text prompts, and

domain attributes) further enhances their suitability for annotation-efficient or cross-modal synthesis tasks [7, 8, 63, 72, 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 large language models (LLMs) as both generators and evaluators in multimodal workflows [3, 7, 33, 40, 43, 52, 55, 72, 85, 88]. These platforms provide unprecedented 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 7, these paradigms differ substantially in generative capabilities, stability, and suitability for complex, multimodal data.

2.4 Importance in Computer Vision and Data Science

Synthetic data has attained a central role in advancing the frontiers of computer vision and broader data science. Its impact transcends the mere expansion of training sets, addressing challenges such as aggressive data augmentation, annotation scarcity, and robustness to distributional shifts. In computer vision, synthetic datasets:

- Enable large-scale augmentation and balance data imbalances;
- Foster robustness to new domains and unseen conditions;
- Support effective model training where labeled data is sparse or highly biased [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].

These advantages extend to other fields, including natural language processing, electronic health records, and time-series analysis, where synthetic data provides a unified toolkit for ameliorating data scarcity [4, 22, 52, 75, 86].

In annotation-scarce regimes, synthetic data is transformative for weakly supervised, unsupervised, and few-shot learning paradigms. Its use is particularly compelling in domains where human labeling costs are prohibitive—such as dense segmentation, anomaly detection, or multimodal medical tasks [7, 10, 22, 29, 41, 45, 50, 54, 56, 58, 63–65]. Furthermore, the capacity to interpolate and bridge across disparate modalities has catalyzed advances in vision-language integration, graph representation augmentation, and cross-domain generalization [15, 22, 33, 72, 75, 85].

Significantly, synthetic data generation also addresses pressing challenges in privacy and algorithmic fairness. By decoupling model development from real, sensitive datasets, synthetic data facilitates collaborative research and accelerates industry adoption. It also provides a valuable resource for probing and mitigating biases, reducing data leakage risks, and analyzing representational disparities [2, 27, 41, 61, 74, 86].

Nevertheless, the benefits of synthetic data must be weighed against persistent open challenges. Ensuring high-fidelity representation of ground-truth distributions, establishing robust evaluation protocols, and maintaining transparency in synthetic data usage remain essential for responsible and effective application across the AI ecosystem [4, 27, 52, 74, 86].

3 Core Techniques and Advanced Methods for Image Synthesis

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]. 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 [48, 69]. 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. Such methods are indispensable for applications in face completion, medical preview, and synthetic biometric data generation [79]. The hybridization of reconstruction-based, high-level semantic, and adversarial objectives underscores the critical role of loss function design in balancing detail fidelity with overall realism.

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 ensures semantic and visual fidelity, particularly in fine-grained domains such as CUB-200-2011 and Oxford-102 datasets [34]. Relative to robust baselines such as LAFITE and VQ-Diffusion, FG-RAT GANs yield superior Frechet Inception Distance (FID) and

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
Key Strengths	Easy to interpret, tractable, controlled latent factors	High-dimensional, realistic sample generation; creative tasks; conditional synthesis	Stable training, high-fidelity outputs, controllable sample 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	High computational demand; large sample generation times; overfitting risk with limited data
Effective Domains	Simple tabular data, simulations, controlled environments	Images, audio, tabular data, domain adaptation	Images, video, 3D data, molecular data, multimodal synthesis

competitive Inception Scores, using fewer parameters, thereby highlighting the increased efficiency of contemporary designs. Persistent challenges remain, however, including reliance on annotated data, label entanglement, and generalizability limitations in label-independent or semi-supervised contexts.

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]. 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 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:

- **Medical Imaging:** Structure-preserving latent diffusion strategies yield high-resolution, morphologically accurate 3D brain MRIs, validated against neuroanatomical metrics and supporting strong predictive modeling.
- **Chemistry:** 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.

- **Clinical Tasks:** 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 stand at the forefront of controllable generative modeling. Contemporary systems increasingly integrate multi-stage attention architectures, pairing textual representations (e.g., TFIDF, N-gram, Bi-LSTM) with hierarchical GANs or diffusion frameworks and advanced loss functions to achieve faithful semantic and geometric alignment [32, 73]. Multi-level attention mechanisms allow for greater flexibility in extracting and fusing cross-modal features at various levels of abstraction, directly addressing historical weaknesses such as text-image misalignment or excessive visual confidence observed in early GAN-based approaches [32].

Customized loss strategies—incorporating adversarial, perceptual, feature-matching, and label-smoothing objectives—consistently drive improvements in both image diversity and fidelity. Recent architectures favor label-independence, using self-supervised and contrastive methods to reduce annotation dependence and bolster robustness across datasets such as Oxford-102 and CUB-200-2011. These cross-modal models recurrently register improved Inception Scores and SSIM values, reflecting gains in realism, semantic faithfulness, and output variability [32, 34, 73].

- Despite these gains, unresolved issues include model uncertainty, generalizability to open-world categories, and the systematic integration of extrinsic knowledge—each representing avenues for further research combining generative, self-supervised, and foundational model paradigms.

3.4 Hybrid and Transformer-Based Image Completion

Recent developments in transformer architectures have facilitated the rise of hybrid models that combine transformer-based and convolutional neural network (CNN) techniques for image completion

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	Face completion, occlusion recovery	CelebA, LFW; Reconstruction error, FID

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	Informative-driven sampling	Medical/clinical synthesis	Supports task-centric evaluation; risk of memorization/privacy issues

and inpainting. While traditional CNN-based methods excel at modeling local textures, they lack the global context awareness required to reconstruct complex spatial relationships. Transformers, with their inherent capacity for long-range dependency modeling, enable pluralistic image completion—producing multiple, coherent completion candidates for the same masked input [82]. Integrating transformer-driven structural prediction with CNN-based fine texture synthesis yields state-of-the-art results for image fidelity, completion diversity, and robustness to large occlusions, while also optimizing computational demands. The availability of open-source hybrid models is catalyzing rapid experimentation and expansion of applicability across domains.

3.5 Classical, Automated, and Adaptive Data Augmentation

Data augmentation strategies remain foundational in deep learning, expanding effective training set size through geometric and color space transformations, patch mixing, and simulation-based synthesis. Classical techniques include operations such as flipping, rotation, scaling, color adjustment, channel permutation, kernel-based filtering, and approaches like Mixup and SamplePairing [23, 29, 50, 75]. Advanced methods such as AutoAugment and ADAAT automate the search for optimal augmentation policies or adversarially generate examples to improve classifier robustness and fairness [39, 68]. Notably, ADAAT integrates adaptive feature modification with adversarial sample generation, enhancing resistance even to adversarial perturbations and noisy environments.

Modern augmentation extends these paradigms beyond visual data to domains including natural language processing, tabular datasets, graphs, and time-series. Techniques increasingly employ large language models (LLMs), self-supervised, and reinforcement learning frameworks to enable multi-modal and domain-agnostic augmentation [3, 7–12, 15, 19, 20, 28, 31, 33, 36, 40, 43, 51, 52, 54, 55, 57, 59, 63, 64, 67, 75, 86]. The advent of unified and modality-agnostic augmentation taxonomies has enabled systematic benchmarking and cross-domain transfer of augmentation strategies [11, 86]. Progressive approaches, exemplified by Model-Adaptive Data Augmentation (MADAUG), dynamically optimize augmentation curricula via bi-level optimization—adapting policies in response to both sample and model-specific properties, and yielding gains in fairness and generalization, especially in fine-grained or imbalanced contexts [68].

While augmentation strategies reduce overfitting and enhance generalization, several challenges persist:

- Automated and adversarial methods may introduce undesirable distributional shifts or erroneous labels;
- GAN-produced augmentations occasionally lack sufficient diversity or semantic accuracy;
- Discovering broadly applicable augmentation policies remains an open problem, particularly outside the vision domain [39, 68, 86].

Consequently, effective augmentation now requires the integration of human priors, automated search, and evaluation metrics that jointly prioritize semantic fidelity, output diversity, and task relevance.

3.6 Self-Supervised and Transfer Learning

The increasing synthesis between self-supervised learning and generative modeling is reshaping both feature extraction and annotation efficiency. Through pretext tasks that predict contextual or structural transformations (e.g., rotation prediction, patch reconstruction), self-supervised models yield robust, generalizable features. Incorporating these self-supervised features into generative frameworks reinforces model robustness and diminishes the need for extensive labeled data [18]. When combined with transfer learning, such approaches accelerate the adaptation of generative models to new, data-scarce domains, such as medical imaging and rare phenotype identification, thereby improving both generative quality and performance on downstream tasks.

Synthesizing advancements from classic GANs to latent diffusion models, transformer-based hybrids, adaptive augmentation, and self-supervision, this section delineates the expanding repertoire and nuanced trade-offs of contemporary image synthesis. Across the spectrum of fidelity, diversity, computational efficiency, and domain specificity, recent literature underscores the necessity of integrating methodological rigor, robust benchmarking, and ethical safeguards to guide further progress in generative image modeling [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

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.

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:** Utilization of expansion and Gaussian smoothing to improve depth discontinuity handling.
- **Robust static scene extraction:** Application of pixel-wise structural similarity index (SSIM) analysis across frames to identify consistent background regions.
- **Comprehensive color-depth fusion:** 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 5 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. The evolution of view synthesis methodologies has catalyzed a pressing need for robust image and video quality assessment

(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 6.

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 sum, the intertwined evolution of synthesis pipelines and specialized QA metrics signals increasing maturity within the 3D vision field—highlighting the necessity of rigorous, perceptually aligned evaluation in tandem with methodological innovation. 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.

4 Applications

4.1 Computer Vision, Healthcare, and Scientific Discovery

The integration of advanced generative models and data augmentation strategies has significantly transformed domains such as computer vision, healthcare, and scientific discovery. These innovations not only yield improved model performance but also reshape conventional workflows.

In computer vision, semantic enrichment and generative techniques directly address persistent challenges in annotation efficiency and multimodal learning. Architectures exemplified by Detection with Enriched Semantics (DES) employ weakly supervised segmentation branches and global activation mechanisms to infuse

Table 5: 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 6: 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

high-level, class-aware context into object detectors. DES achieves high accuracy with minimal computational overhead and eliminates the need for additional manual annotations, directly alleviating the limitation of scarce ground-truth data prevalent in datasets such as PASCAL VOC and MS COCO [23]. This trend marks a departure from narrowly specialized methods, toward unified frameworks capable of embedding semantic priors into advanced detection pipelines with extensibility and minimal annotation burden.

In healthcare, synthetic data generation and generative modeling have become indispensable for circumventing hurdles posed by data scarcity, privacy constraints, and the demand for diverse labeled datasets. Simulation frameworks, notably SyntheX, employ physically accurate domain randomization and simulation methodologies to generate extensive annotated datasets, obviating the reliance on ethically fraught and costly real-world data acquisition. Evidence demonstrates that models trained solely on synthetic X-ray datasets, as produced by SyntheX, can match or exceed the performance of real-data-trained models in tasks such as landmark localization, segmentation, and detection within hip imaging and robot-assisted surgery [67]. Strategic approaches like synthetic-to-real (Sim2Real) training—particularly when bolstered by robust domain randomization—equal or surpass classical domain adaptation techniques, such as CycleGAN and ADDA, while entirely bypassing real data in the training loop. Nonetheless, these advancements are constrained by simulator fidelity and the risk of annotation mismatch, emphasizing the need for ongoing research in simulation realism and transferability [21, 67].

Beyond imaging, synthetic and augmented data are increasingly deployed in electronic health records (EHR) and tabular biomedical datasets to enhance the robustness and generalizability of deep learning models. Here, augmentation techniques span diverse modalities: geometric and pixel-level methods in vision, paraphrasing and noising in natural language processing, and sophisticated generative schemes in tabular or time-series contexts. Recent frameworks have unified augmentation taxonomies across image, text, graph, and time-series data, organizing techniques by granularity and information type [4, 75, 86]. Blending single-instance, multi-instance, and generative augmentation methods (including GAN- and LLM-based data synthesis) consistently yields measurable gains in downstream model performance, affinity, and diversity. Despite

these benefits, unresolved challenges endure. For instance, GAN- and diffusion-generated data risk diminished diversity, label inconsistencies, and problematic autophagic training loops—scenarios in which models are repeatedly exposed to their own synthetic outputs, leading to compounded distributional drift and undermining reliability [72, 85]. Consequently, there is a growing consensus that a balanced, meticulously curated combination of real and synthetic data is essential to preserve and enhance model utility [85].

Generative modeling also catalyzes breakthroughs in scientific discovery, especially in drug design and protein-ligand interaction modeling, where combinatorial complexity and limited empirical data challenge traditional experimental approaches. Cutting-edge models such as EquiScore, which employs equivariant graph neural networks, integrate physical priors with structural interaction data to outperform more than twenty established methods in binding pose prediction and analog activity ranking. These successes are underpinned by strategic augmentation, redundancy reduction, and interpretability mechanisms [45]. Similarly, the Geometry-Complete Diffusion Model (GCDM) facilitates the joint diffusion of atom coordinates and types, permitting the synthesis of valid, large molecules optimized for desired drug-like properties [56]. These innovations are further supported by advanced evaluation protocols and committed open-source dissemination, promoting adoption in pharmaceutical applications while illuminating an ongoing trade-off between computational requirements and chemical validity.

Collectively, these advancements signal a paradigm shift toward annotation-efficient, modality-agnostic, and context-aware generative AI applications, especially in environments characterized by data scarcity or ethical considerations. Nonetheless, responsible deployment calls for rigorous methodologies, standardized evaluation practices, and vigilant oversight to prevent data contamination, as the demarcation between real and synthetic data grows increasingly ambiguous [4, 21, 22, 56, 67, 70, 72, 75, 85, 86, 88].

4.2 Fine-Grained and Facial Synthesis

Generative modeling has markedly advanced fine-grained and facial image synthesis, driving progress in both creative and clinical spheres. State-of-the-art facial generative networks are now

capable of synthesizing high-fidelity face images from sparse, incomplete, or disjointed patches by employing composite loss functions—harmonizing pixel-based, perceptual, and adversarial criteria. The resulting outputs display superior visual plausibility and semantic integrity over traditional inpainting or patch-based benchmarks. These models facilitate diverse applications, including digital artistry, virtual previewing for plastic surgery and dentistry, and scalable production of synthetic, labeled datasets to bolster face recognition and classification—an especially vital advance for privacy preservation and demographic diversity [79].

Importantly, the capacity for attribute and style manipulation within these models enables controllable synthesis, accommodating a spectrum of research and creative demands. Nevertheless, technical and ethical challenges persist, notably in disentangling attributes, mitigating dataset biases, and safeguarding against misuse in identity-sensitive or medical diagnostic contexts.

4.3 3D Video and Virtual View Navigation

Advances in generative and synthesis-based methods have greatly enhanced the realism, scalability, and user experience in 3D video and virtual view navigation, unlocking new possibilities for both creative media and immersive technologies.

In gigapixel-scale novel view synthesis, approaches utilizing meta-deformed manifold representations and implicit neural fields have demonstrated significant improvements in geometric correspondence and high-fidelity surface detail, outperforming traditional methods for complex, real-world and high-resolution scenarios [53]. Simultaneously, depth-image-based rendering (DIBR) has benefited from innovative spatio-temporal synthesis and hole-filling techniques, employing strategies such as scene background extraction, depth edge refinement, and weighted-fusion mechanisms to alleviate well-known rendering artifacts—including holes, ghosts, and flicker. These enhancements yield measurable gains in both spatial and temporal reconstruction quality, as supported by quantitative metrics (PSNR, SSIM) and subjective user studies [77]. However, current pipelines frequently rely on fixed camera geometries and suffer from substantial computational demands, indicating the ongoing need for scalable, adaptive processing—particularly for live telepresence, immersive broadcasting, and real-time validation applications.

Rigorous quality assurance for generated 3D media relies on domain-specific, perceptually aligned metrics. Recent innovations such as Morphological Pyramid PSNR and Wavelet PSNR utilize multi-scale, edge-preserving decompositions to more accurately assess geometric and boundary distortions in DIBR outputs. These metrics exhibit improved alignment with human perception, supporting more reliable deployment in dynamic or artifact-sensitive usage scenarios [77]. Collectively, such advances in generative modeling and quality assessment set elevated standards for the creative and technical validation of next-generation 3D video systems.

4.4 Industrial, Scientific, and Emerging Applications

The application of synthetic data and generative modeling extends well beyond conventional vision and biomedical domains, enabling

a diverse array of industrial and scientific advancements with profound societal ramifications. In drug discovery and computational chemistry, architectures such as diffusion-based and equivariant generative models (e.g., EquiScore and GCDM) have established themselves as foundational tools for molecular design, property optimization, and virtual screening. Their integration with advanced chemical priors facilitates accelerated discovery cycles and substantial reductions in experimental expense, while also affording interpretability and adaptability for specialized tasks ranging from analog ranking to protein-conditional generation. Yet, maintaining the balance between geometric complexity and computational efficiency remains a persistent technical challenge [45, 56].

Generative and augmentation-based approaches have similarly transformed environmental and climate modeling. Consistency models trained on fine-grained observational data not only reduce computational downscaling costs, but also provide uncertainty-calibrated, bias-corrected forecasts that generalize across diverse climate scenarios [84]. These methodologies circumvent the need for retraining per simulation, delivering rapid probabilistic predictions that preserve key climatological patterns and extremes—integral for policy-making and risk assessment under evolving global conditions.

Distributed data creation—fundamental in federated and edge learning environments constrained by privacy and infrastructure—leverages generative models as adaptive meta-models that can be rapidly customized to local data. By formalizing continual adaptation as a Wasserstein barycenter problem and integrating quantization-aware compression, recent frameworks enable bandwidth-efficient, privacy-preserving model personalization and collaborative learning without direct transmission of sensitive real-world data [37]. However, such systems demand precise calibration to prevent data drift and overfitting, as edge devices typically possess limited and potentially unrepresentative datasets.

In summary, the convergence of synthetic data, generative modeling, and simulation-driven innovations has triggered breakthroughs across traditionally segregated disciplines, while simultaneously raising complex technical and ethical challenges that necessitate enhanced methodological rigor and cross-domain collaboration as these approaches transition toward wide-scale, high-impact deployments [37, 45, 56, 84].

Table 7 provides a structured comparison of representative generative and simulation-driven methods, highlighting their principal features and domain-specific achievements.

Together, these ongoing advances reinforce the need for methodical evaluation, careful curation of synthetic data, and sustained interdisciplinary cooperation to maximize the positive impact of generative modeling across a burgeoning array of applications.

5 Thematic Synthesis, Evaluation, and Benchmarking

5.1 Cross-Sectional Method Comparisons

The field of synthetic data generation is undergoing a rapid evolution, with generative paradigms such as Generative Adversarial Networks (GANs), diffusion models, and hybrid methodologies addressing distinct requirements regarding realism, controllability, and domain-specific augmentation. This progression is particularly

Table 7: Representative Generative and Simulation Methods in Selected Application Domains

Domain	Prominent Method(s)	Key Features and Achievements
Computer Vision	DES [23]	Weakly-supervised segmentation; semantic enrichment; reduces annotation cost with minimal computational overhead.
Healthcare	SyntheX [67]	Physically realistic simulated X-ray datasets; supports landmark localization and detection; eliminates need for real images in training.
Scientific Discovery	EquiScore [45], GCDM [56]	Equivariant graph neural networks and diffusion models; excel in molecular property optimization and binding prediction.
3D Video	Meta-deformed manifold, DIBR [53, 77]	Improved geometric correspondence, scene realism; novel QA metrics (Wavelet/Morphological PSNR) for artifact-sensitive deployment.
Distributed Learning	Wasserstein barycenter adaptation [37]	Model personalization without real data transmission; bandwidth and privacy optimized; supports federated/edge scenarios.
Environmental Modeling	Consistency models [84]	Bias-corrected, high-resolution outputs; efficiency without retraining for each simulation input.

pronounced within vision and healthcare applications, as documented in recent literature [21–23, 34, 48, 67, 69, 71–76, 82, 86, 88]. Traditionally, GANs have been regarded as the benchmark for high-fidelity image synthesis. Innovations in their architecture—such as the introduction of auxiliary classifiers and the utilization of contrastive learning—have significantly mitigated challenges like mode collapse and improved semantic alignment. These advancements enable task-specific performance enhancements, particularly in the realms of subclass differentiation and fine-grained, text-conditioned synthesis [34, 71, 72, 76]. A notable advancement is the FG-RAT GAN, which excels at generating class-consistent and diverse images, while reducing both model complexity and resource consumption compared to models such as LAFITE and VQ-Diffusion [34].

Diffusion models, in turn, have established new state-of-the-art benchmarks, regularly surpassing GANs in terms of sample quality, diversity, and robustness. They are especially effective in conditional and multi-modal settings, owing to their iterative denoising processes and advanced guidance mechanisms, such as classifier-free and entropy-based sampling [69, 75, 76, 86]. However, these improvements entail increased computational demands and, in some cases—such as in medical imaging—heightened memorization risks in limited data regimes [48, 74]. These risks necessitate balanced assessments encompassing privacy and utility. Hybrid approaches, which incorporate components such as transformers, variational autoencoders (VAEs), or attention mechanisms within GAN or diffusion frameworks, have demonstrated quantitative gains in structural fidelity and multimodal controllability, particularly within both medical and general vision domains [22, 34, 67, 75].

A central consideration when comparing these approaches is their respective resource efficiency and scalability:

- **GANs:** Offer high throughput and low inference cost, facilitating deployment. However, they sometimes encounter scalability challenges and difficulties in complex conditional synthesis due to stability and diversity bottlenecks [23, 71, 82].
- **Diffusion Models:** Deliver superior synthesis quality and controllability, albeit with greater training and inference costs. Recent research on acceleration—such as optimized noise schedules—is narrowing the performance gap [69, 75].
- **Hybrid Models:** Transformer-augmented and attention-enhanced pipelines strike a balance, enabling cross-modality generalization, effective high-dimensional input handling, and refined user control [22, 34, 67].

In summary, ongoing methodological innovation seeks to optimize fidelity, diversity, explicit control, and computational feasibility, with careful prioritization of domain constraints and downstream application requirements.

5.2 Evaluation of Synthetic Data Quality

Evaluating the quality of synthetic data demands rigorous attention to three core objectives: factuality, fidelity, and fairness. These criteria are essential to ensure authenticity, practical utility, and responsible deployment—especially in sensitive domains such as healthcare [74][27][85].

- **Factuality** is defined as the consistency between generated instances and ground-truth domain logic or factual knowledge.
- **Fidelity** pertains to both statistical and perceptual similarity between synthetic and real data.
- **Fairness** insists on minimizing bias and disparate impact, a pressing concern in both healthcare and high-stakes vision tasks.

The quantitative evaluation of these dimensions relies on a suite of metrics, frequently adapted to both the target domain and the underlying synthesis modality. For vision and 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. In addition, affinity and diversity ratios are used to assess overfitting and population variability [59][40][57][64][15][51][33][8][3][63][65][70][61][11][75][8]. Each metric addresses specific aspects of quality: for instance, mIoU/mAP reflect segmentation and detection accuracy, FID and IS measure perceptual and feature diversity, and affinity/diversity ratios illuminate synthetic data generalizability.

Relying solely on these fidelity metrics, however, risks masking underlying issues such as model memorization or hidden biases. Diffusion models, for example, may achieve superior FID and IS performance while inadvertently reproducing memorized training instances, particularly on homogeneous datasets [74][48]. Therefore, comprehensive evaluation protocols—encompassing adversarial, privacy-oriented, and domain-adaptive assessments—are strongly advocated [74][27][85]. These include membership inference, re-identification risk analyses, and performance benchmarking using domain-adapted metrics.

The availability of robust benchmark datasets is pivotal for standardized method comparison:

- **Vision Benchmarks:** COCO, Pascal VOC, ADE20K, Cityscapes, CUB-200-2011, Oxford-102
- **Domain-Specific Resources:** PDBscreen (structural biology), SyntheX (clinical X-rays), various medical imaging corpora

Such diversity facilitates not only the assessment of generic perceptual quality but also ensures relevance regarding phenotype fidelity, class balance, and clinical applicability [65][70][61][18][22]. In healthcare settings, task-based validation protocols (e.g., TSTR—Train

on Synthetic, Test on Real) and advanced statistical similarity measures (e.g., KL divergence, KS statistic) are employed to further contextualize synthetic data performance [74][27][85].

5.3 Detailed Comparative Tables

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 8.

As illustrated in Table 8, 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 efficiency. Distinctions in architectural features—such as the use of auxiliary classifiers, contrastive learning, and transformer-based conditioning—further highlight the rapid developments in this domain.

5.4 3D/DIBR-Specific Metrics

The synthesis and evaluation of 3D data, including outputs generated by depth-image-based rendering (DIBR), introduce distinct requirements not addressed by conventional 2D measures. Artifacts such as edge distortions and geometric deformation—particularly near disoccluded regions—necessitate specialized evaluation. Morphological Pyramid Peak Signal-to-Noise Ratio (MP-PSNR) and Morphological Wavelet Peak Signal-to-Noise Ratio (MW-PSNR) have been introduced to directly target edge preservation and structural coherence across scales [77]. These metrics offer enhanced alignment with human visual assessment and are optimized for use in demanding environments, such as real-time streaming, immersive 3D simulation, and medical imaging systems, where computational overhead is a constraint.

Collectively, the integration of GANs, diffusion models, hybrid strategies, and transformer-based architectures—coupled with rigorous, context-sensitive benchmarking protocols—has resulted in a highly robust methodological ecosystem. Persistent challenges remain, including model memorization, insufficiently evaluated fairness, and the lack of universal cross-domain benchmarks. Continued innovation in both generative methodology and quality assessment remains essential for the responsible advancement of the field [74][27][85][72][69][48][34][71].

6 Responsible and Ethical Oversight

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 considerable ethical challenges. Of foremost concern is data privacy: traditional anonymization measures are increasingly vulnerable to re-identification, particularly as adversarial capabilities grow and auxiliary data sources become ubiquitous. This escalating risk has catalyzed the adoption of synthetic data as a privacy-enhancing 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)—have advanced the ability to create artificial datasets that preserve the statistical properties of the original data while reducing direct exposure of individual records. Such innovations mitigate regulatory constraints imposed by regimes like the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA), and accelerate data sharing for machine learning (ML) research [59][57][60][49][52][15][43][8].

Nevertheless, these developments do not render synthetic data immune to privacy risks. Techniques such as membership inference and linkage attacks remain viable, especially when synthetic data generation is not supplemented with formal differential privacy (DP) guarantees or comparable noise infusion mechanisms [64][15][43]. The privacy-utility trade-off is particularly acute in scientific and healthcare domains, as high-fidelity models may inadvertently reproduce features or outliers that enable sensitive information leakage [1][64][33]. Empirical studies indicate that while DP can reduce privacy risks, it often does so at the expense of downstream ML task performance if not implemented judiciously [64][43].

Algorithmic bias represents a further axis of ethical concern. Synthetic datasets generated from skewed or unrepresentative source data—or from generative models that inadvertently encode social or demographic disparities—may perpetuate or amplify these biases in downstream analytics and decision-making systems [60][1][15][8][88]. This challenge is especially salient when synthetic data is used to augment minority classes, correct imbalances, or fill missing modalities. As such, diligent post-generation audits and systematic fairness assessments are imperative to mitigate unintended harm [55][60][88].

The increasing indistinguishability of real and synthetic data introduces additional risks related to misinformation and data integrity. With GenAI models achieving greater fidelity, the line between authentic and synthetic datasets becomes blurred, threatening reproducibility, scientific credibility, and public trust [28][62][8][11]. Synthetic data, if undocumented, can corrupt research repositories and benchmarks, or propagate hallucinated artifacts that compromise scientific inference. To address these risks, robust provenance frameworks, mandatory transparency and disclosure, and explicit labeling protocols have been advocated to ensure that any use of synthetic data in research outputs is clearly identified [28][8].

In this context, transparency and traceability are fundamental requirements for responsible synthetic data deployment. Technical measures—including data watermarking, cryptographic blockchain certification, and AI-driven detection tools—are emerging to ensure the provenance and audibility of synthetic datasets throughout the research lifecycle [28][8][11]. Such technologies, however, demand standardization to prevent fragmentation of practices and to achieve broad adoption.

Legal and regulatory compliance represents an evolving frontier. Synthetic data facilitates compliance with privacy legislation by reducing identifiability, but the lack of universal, precise regulatory guidelines continues to create uncertainty [40][60][36][52]. Regulatory agencies are increasingly alert to such nuances, particularly in clinical research and data-sharing contexts; nonetheless, their approaches diverge across jurisdictions. Sustained, cross-sector

Table 8: 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

dialogue is necessary to establish consensus guidelines that appropriately balance individual protection with scientific progress [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

The sustainable and trustworthy integration of synthetic data in research and applied settings is contingent on the development and adoption of robust, standardized evaluation protocols and community-driven frameworks [27][85]. At present, the absence of

universally accepted, domain-adapted benchmarking and validation procedures limits scientific acceptance and regulatory clearance—an impediment that is especially pronounced in sensitive sectors such as healthcare [55][60][27][85].

Emergent community efforts address these gaps through several coordinated strategies:

- Establishing standardized frameworks for privacy risk quantification, utility benchmarking, and fairness auditing, thereby enabling rigorous comparison and trustworthy integration of synthetic data and generative models [22][88][27][85].
- Supporting precompetitive consortia, open benchmarking challenges, and collaborative infrastructure to promote harmonization, reduce adoption barriers, and disseminate best-practice tools and guidance aligned with contemporary ethical and regulatory expectations [85].
- Enhancing transparency and provenance controls to minimize the risk of contamination by synthetic data or untracked generative processes in downstream applications.
- Fostering sustained, interdisciplinary dialogue spanning technical, ethical, and regulatory domains to adapt standards dynamically as GenAI capabilities advance [11][21][22][27].

Collectively, governance structures and consensus-driven standards—supported by cross-sector vigilance and continual protocol development—are pivotal in ensuring that the transformative benefits of synthetic data are realized in an ethical, equitable, and trustworthy manner.

7 Challenges, Limitations, and Future Directions

7.1 Technical and Resource Barriers

The proliferation of diffusion models and other large-scale generative frameworks has underscored numerous technical and resource-intensive obstacles that constrain scalability and impede equitable access. A chief concern is the substantial computational burden required for both the training and inference phases of diffusion models, which significantly exceeds that of previous generative paradigms, such as GANs—even in light of recent algorithmic advances aimed at improving 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 such models necessitates vast datasets and computational infrastructure, resources predominantly accessible to well-funded organizations, thereby accentuating disparities in research and deployment [4, 9, 12, 41, 64, 65, 72, 86]. Strategies including modular sampling [90], adaptive noise scheduling [34], and hybrid generative architectures are being pursued to accelerate inference and reduce resource consumption. Nevertheless, the fundamental

tradeoff between model expressivity, fidelity, and computational feasibility persists [3, 12, 46, 72].

Equally pressing are the challenges related to the fidelity and cross-domain transferability of simulation-generated or synthetic data. The disparity between simulated and real-world data—stemming from rendering engine limitations, incomplete physical modeling, and non-standardized annotation—introduces domain biases that can impair downstream performance [22, 72, 76, 85]. Recent advances in highly realistic simulation frameworks, domain randomization, and physics-based generative models have improved transferability [76, 85]; however, achieving strong cross-modality consistency, particularly in 3D or multi-modal settings, remains restricted by representation instability and complex alignment issues [22, 72, 76]. Furthermore, the absence of standardized annotation protocols for synthetic data—especially in intricate, high-dimensional domains—complicates model evaluation, reproducibility, and regulatory scrutiny.

As shown in Table 9, targeted algorithmic advances partially address these technical constraints; however, resource inequalities and deficiencies in standardization continue to challenge both scalability and reproducibility.

7.2 Generalization, Robustness, and Societal Impact

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]. Many existing methods are vulnerable to catastrophic forgetting in continual learning scenarios and degrade in fidelity under federated aggregation, signaling unresolved gaps in knowledge transfer, adaptation, and privacy protection [84]. Furthermore, the deployment of large generative models onto resource-constrained systems 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 perspective, the deployment of synthetic data and generative models introduces a multifaceted array of risks:

- **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].
- **Contamination and feedback loops:** Recursive use of synthetic data as training input (AI “autophagy”) can erode model reliability and scientific validity [13, 14, 30].
- **Privacy leakage:** Overparameterized models, especially trained on limited or homogeneous data, may memorize and inadvertently reveal sensitive real information [22, 27, 72, 88].
- **Opacity and interpretability:** The increasing complexity of generative architectures impedes interpretability and transparency—critical where alignment with human expectations and regulatory standards is fundamental (e.g., healthcare, law) [10, 11, 13, 22, 36, 42, 83, 88].

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.

7.3 Label Dependency and Semantic–Style Balance

The necessity to reduce reliance on expensive manual annotation has catalyzed significant research into weakly supervised, self-supervised, and label-free generative paradigms [34, 71]. Nonetheless, state-of-the-art conditional diffusion and GAN-based synthesis methods frequently incorporate some degree of label dependency to guarantee semantic richness and enable controlled generation [34, 71]. A critical unresolved issue is the balance between semantic fidelity—ensuring generated samples adhere strictly to designated label or mask constraints—and stylistic richness, which is vital for both generalization and downstream utility [71]. Existing approaches, such as manipulation of the latent space, sophisticated input conditioning [34, 85], and the utilization of disentanglement losses or dual-guidance mechanisms, offer only partial solutions; a comprehensive, robust framework for interpretable, finely controllable, and minimally label-dependent synthesis remains an open challenge.

7.4 Evaluation Metrics and Benchmark Gaps

The evaluation of synthetic and augmented data—especially in nascent frontiers like multi-modal, 3D, and cross-domain synthesis—remains severely hindered by the scarcity of universally accepted, modality-spanning metrics [2, 22, 58, 76, 85, 90]. Conventional statistics (e.g., FID, IS, TSTR) [2, 58, 90] are popular yet often inadequately sensitive to semantic consistency, clinical relevance, or application-specific utility. This deficiency is particularly pronounced in domains where genuine data are tightly restricted, curtailing reliable cross-model and cross-study comparisons and retarding the translation of synthetic data into real-world contexts [22, 76, 85]. The lack of comprehensive, interpretable, and context-aware benchmarks further complicates both model development and regulatory approval, mandating focused research into new evaluation paradigms.

7.5 Research Opportunities and Future Trends

Looking ahead, future research efforts should prioritize advancements in model quality, sample diversity, robust personalization, and real-time synthesis—imperatives as generative AI evolves toward dynamic, interactive, and agent-based applications [27, 34, 71, 72, 85]. Key enablers of this evolution include:

- Automated benchmarking and pipeline development for systematic performance tracking [27, 85]
- Adaptive, self-improving model architectures capable of continual learning and transfer
- Scalable, modality-agnostic evaluation frameworks that facilitate rigorous cross-study comparability
- Responsible AI practices encompassing updated policies, oversight mechanisms, and inclusive stakeholder engagement to secure 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]

Table 9: 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
Data bias in simulation-generated samples	Domain gap in synthetic-to-real transfer	Domain randomization, physics-based generative models
Lack of annotation standards	Heterogeneous labels in 3D/multi-modal data	Development of universal standards, community benchmarks

The convergence of vision, language, multimodal, and scientific domains is fostering the development of unified generative systems that support a wide array of applications—from scientific modeling and systematic reasoning to open-vocabulary detection and 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 the design of principled, scalable, and interoperable architectures that can effectively manage heterogeneous data modalities, diverse annotation schemes, and domain-specific requirements while minimizing bias and maximizing transparency. Progress in these areas will further hinge on interdisciplinary collaboration, uniting foundational research, applied expertise, and regulatory acumen to accelerate the emergence of the next generation of robust, adaptable, and trustworthy generative AI.

This enhanced section streamlines the presentation of technical, methodological, and societal challenges, introduces Table 9 for crisp visualization of technical barriers and ongoing mitigation strategies, and employs itemized bullet points for multifaceted societal risks and research opportunities, ensuring academic rigor, clarity, and optimized readability in full compliance with LaTeX academic formatting.

8 Security, Adversarial Threats, and Alignment

8.1 Threat Detection and Robustness

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 10 illustrate the diversity of adversarial threats and corresponding defensive approaches, emphasizing the multifaceted requirements of robust generative model deployment.

8.2 Alignment in Generative Models

While security is fundamental, the alignment of generative models with human values and preferences constitutes an equally critical pillar of responsible AI development. Reinforcement learning from human feedback (RLHF), along with its extensions, has become the prevailing methodology for behavioral alignment, enabling generative systems to internalize not only explicit task instructions but also intricate preferences involving style, safety, and utility [21, 42]. The scope of preference tuning has widened to address multimodal architectures—including those spanning vision, speech, and mixed-modality generation—necessitating flexible alignment frameworks capable of integrating heterogeneous human feedback signals.

Recent comparative studies have offered critical insights into the strengths and constraints of RLHF and related alignment strategies. Large, high-quality datasets employed for reward modeling enhance the fidelity with which models capture complex human preferences. However, such datasets also pose significant risks: they can inadvertently encode prevailing societal biases and underrepresent minority viewpoints, especially when alignment protocols are scaled across diverse tasks and populations [42]. Furthermore, preference alignment applied in highly complex environments, such as paired or conditional image-text generation, is susceptible to issues including mode collapse or excessive conservatism—especially

Table 10: 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

when reward models are over-regularized, thereby curtailing generative diversity and informativeness [21]. Managing the trade-off between producing harmless (i.e., non-biased or non-harmful) and helpful (i.e., contextually relevant and informative) outputs remains an unresolved research and societal challenge [21].

Evaluation of alignment success constitutes another persistent challenge. While existing quantitative metrics provide partial insight, they frequently fail to reliably detect value misalignment in low-frequency, high-impact failure cases—those with substantial social consequences despite their rarity [21]. For multimodal generative systems, the creation of standardized benchmarks and scalable annotation pipelines for collecting human feedback remains a formidable obstacle. Nevertheless, ongoing innovation—such as more expressive preference modeling, hierarchical feedback mechanisms, and cross-modal reward optimization—highlights a dynamic trajectory toward more robust and socially aligned generative AI [21, 42].

- **Scalability:** Aligning models via human feedback at scale risks bias propagation and capacity constraints.
- **Evaluation:** Quantitative metrics often miss edge-case misalignments with profound social impact.
- **Generalization:** Ensuring aligned behavior generalizes beyond seen scenarios, especially in multimodal or open-world settings.
- **Trade-offs:** Balancing harmlessness and helpfulness remains a nuanced, largely unsolved problem.
- **Innovation:** Advances in preference modeling, feedback acquisition, and cross-modal alignment signify ongoing progress.

The interface of adversarial robustness and human alignment in generative models thus defines a critical, fast-evolving frontier. Progress in this domain is shaped by rapid methodological advances and by the complex trade-offs involved in ensuring security, safety, and utility [21, 24, 39, 42].

9 Synthesis, Comparative Analysis, and Recommendations

9.1 Comparative Perspective

The trajectory of generative and augmentation strategies in artificial intelligence (AI) exhibits a marked evolution, signifying a shift from hand-crafted, deterministic pipelines to increasingly sophisticated deep learning, generative, and diffusion-based frameworks [21–23, 34, 67, 71, 72, 74, 75, 86, 88]. Initial data augmentation approaches—such as geometric distortions and value-based manipulations—were celebrated for their simplicity, interpretability, and applicability across diverse fields, most notably computer vision and natural language processing. However, these methods demonstrated notable limitations: their efficacy was tightly bound

to specific domains, and they often lacked the capacity for meaningful semantic diversification, constraining their utility in specialized contexts such as medical imaging or fine-grained visual tasks [23, 71, 86].

The advent of deep generative models, specifically Generative Adversarial Networks (GANs), transformer-based architectures, and diffusion models, has dramatically redefined the capabilities of data synthesis and augmentation. These advanced models offer controllable, high-fidelity sample generation, routinely incorporating multi-instance strategies such as Mixup or generating extensive synthetic datasets via GANs and diffusion methodologies [22, 23, 34, 74, 86, 88]. Diffusion models, in particular, now lead the field for generating both images and augmentations with robust statistical and semantic integrity, surpassing earlier GAN-based approaches in terms of diversity and the realism of generated samples [34, 74, 88].

Nevertheless, this progression is accompanied by new challenges, including risks of semantic drift, contamination by synthetic data (often identified as “AI autophagy”), and delicate calibration needed to balance diversity with task relevance or class equilibrium [34, 72]. Comparative studies indicate that while generative synthetic data can significantly mitigate overfitting and bolster the representation of rare classes, the absence of rigorous curation or excessive reliance on model-generated data may ultimately undermine downstream performance and reliability [34, 72]. These insights underscore the necessity for robust evaluation frameworks, comprehensive benchmarking, and modality-specific guidelines to maximize the advantages of generative augmentation while safeguarding scientific rigor [34, 71, 72].

9.2 Criteria for Responsible Deployment

The increasing adoption of generative and augmentation-based methodologies underscores an urgent need for comprehensive, modality-aware guidelines for responsible implementation—a concern increasingly echoed in recent literature [23, 34, 67, 71, 75, 86]. Three foundational pillars define 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]. Consequently, best practices advocate for principled data sampling, rigorous pre-augmentation audits, and ongoing validation to ensure representational adequacy [34, 71].
- **Interpretability and Auditability:** Interpretability is vital for both regulatory compliance and safe deployment in

scientific and clinical applications. While conventional augmentation (e.g., geometric transforms) is generally transparent, the opacity of deep generative approaches (such as diffusion or transformer-based models) raises concerns regarding unanticipated artifacts, data drift, or spurious correlations. Routine interpretability assessments, provenance tracking, and human-in-the-loop oversight are therefore essential [23, 34, 86].

- **Robust and Context-Aware Augmentation:** Effective augmentation must balance the pursuit of large-scale data diversity against the potential for introducing distributional shifts or bias [22, 67, 71]. Recent taxonomies recommend hybrid approaches that integrate deterministic, transparent transformations with data-driven, generative methods—ideally augmented by automated policy discovery and adaptive monitoring [34, 86]. In medical AI, platforms such as SyntheX illustrate the benefits of synthetic pipelines that employ domain randomization and calibration against real-world data, thereby reducing ethical and logistical barriers to empirical research [34, 72, 74].

Collectively, these criteria constitute the essential framework for responsible and effective deployment of generative augmentation techniques.

9.3 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 the evolving characteristics of real-world data—striving to balance objectives such as standardization, personalization, and fairness in real time [69, 76]. Mechanisms including meta-learning, human feedback-in-the-loop, and task-aware policy search enable augmentation pipelines to internalize context-specific priors and to modulate the generation of synthetic data accordingly [69, 76, 82].

However, the efficacy of these advances is contingent upon equally rigorous oversight protocols. Frameworks for responsible AI in augmentation increasingly mandate algorithmic transparency, persistent post-deployment monitoring, and the presence of safeguards to prevent undesirable feedback cycles—such as synthetic data inadvertently amplifying existing biases [34, 48, 73]. In fields such as healthcare, domain-specific instrumentation now embraces simulation-based validation, out-of-distribution detection, and semantic fidelity assessment as standard components of adaptive augmentation workflows [34, 74]. These priorities are reflected in emerging best practices, emphasizing the seamless integration of context-sensitive adaptation and comprehensive, multi-layered responsibility checks [34, 48, 73].

9.4 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 11 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 the interpretability and structural stability of explicit constraints (such as feature-matching losses) with the flexibility and expressiveness of multi-stage, generative learning frameworks [71, 72, 74]. Such synergies facilitate not only the creation of highly realistic and perceptually convincing samples but also enable interactive, controllable synthesis workflows applicable across vision, language, and scientific disciplines. Going forward, systematic benchmarking and comprehensive, multi-domain evaluations will be indispensable for steering the responsible adoption of generative and augmentation technologies, ensuring efficacy and scientific integrity in diverse, real-world workflows [22, 23, 34, 71, 74, 86].

10 Conclusion

10.1 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]. 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,

Table 11: 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, contrastive learning, fine-grained text-to-image synthesis	Vision, cross-modal generation	Outperforms LAFITE & VQ-Diffusion; computationally efficient
VQ-Diffusion [74, 88]	Vector quantization, diffusion processes, controllable synthesis	High-resolution image generation, augmentation	Superior semantic fidelity, high diversity; computationally intensive
LDM (Latent Diffusion Models) [34, 88]	Latent-space modeling, scalable synthesis, conditional translation	Image-to-image translation, segmentation, style transfer	High realism, effective in semantic alignment; requires extensive training data
Transformer-based/Hybrid Models [22, 71, 75]	Global structure modeling, integration with convolutional layers, cross-modal capabilities	Pluralistic image completion, robust data synthesis	Balances interpretability and expressiveness; adaptable to multiple modalities

further blurring the lines between symbolic and subsymbolic artificial intelligence and facilitating context-aware synthesis and control [19, 22, 48, 69, 71, 91].

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

10.2 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:** Ensuring robustness over diverse tasks, datasets, and operational conditions.
- **Fairness and Alignment:** Vigilantly mitigating historical biases and ethical risks through model and data design.
- **Scalability:** Supporting efficiency, flexibility, and rapid transfer to new domains as models scale.
- **Ethical Design:** Embedding transparency, explainability, and accountability into all facets of generative modeling.

10.3 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 for transparent, traceable outputs.
- Implementation of robust privacy-guarding mechanisms.
- Adoption of evaluation standards and audit processes for ongoing accountability.
- Ongoing dialogue among technical, ethical, and policy stakeholders.

10.4 Outlook

Looking ahead, the sustainable and responsible evolution of synthetic data research will depend 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. Thibaud-Sutre, N. Burgos, and S. Allasoinnière. 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 Ejayi, Dongsheng Cai, Francis Ofoma Eze, Makuachukwu Bennedith Ejayi, Jennifer Ene Idoko, Sarpong Kwadwo Asere, and Thomas Ugochukwu Ejayi. 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. Furrutter, 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. Ammirajab, C. Sun, M. Breeuwer, J. Pluim, B. Elen, G. Ertaayan, 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. Perdakis, 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. Jones, 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, Yinze 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