IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Privacy and Security Concerns in Generative AI: A Comprehensive Survey

ABENEZER GOLDA[1], KIDUS MEKONEN[1], AMIT PANDEY[2], ANUSHKA SINGH[1], VIKAS HASSIJA[3], VINAY CHAMOLA[4] (Senior Member, IEEE), BIPLAB SIKDAR[5] (Senior Member, IEEE)

[1]Student at the School of Computer Science Engineering KIIT, Bhubaneshwar, India - 751024,(email: motolomygolda@gmail.com, kidusabebe1921@gmail.com, 21052568@kiit.ac.in)
[2]Ph.D. Student at the School of computer science engineering and technology, Bennett University greater Noida 201310,(email:e21soep0035@bennett.edu.in)
[3]Associate Professor at the School of Computer Science Engineering KIIT, Bhubaneshwar, India - 751024,(email: vikas.hassijafcs@kiit.ac.in)
[4]Associate Professor at the EEE Department, BITS Pilani, Pilani Campus, India,(email: vinay.chamola@pilani.bits-pilani.ac.in)
[5]Associate Professor in the Department of Electrical and Computer Engineering, National University of Singapore, Singapore (e-mail: bsikdar@nus.edu.sg).

ABSTRACT

Generative Artificial Intelligence (GAI) has sparked a transformative wave across various domains, including machine learning, healthcare, business, and entertainment, owing to its remarkable ability to generate lifelike data. This comprehensive survey offers a meticulous examination of the privacy and security challenges inherent to GAI. It provides five pivotal perspectives essential for a comprehensive understanding of these intricacies. The paper encompasses discussions on GAI architectures, diverse generative model types, practical applications, and recent advancements within the field. In addition, it highlights current security strategies and proposes sustainable solutions, emphasizing user, developer, institutional, and policymaker involvement.

INDEX TERMS Generative Artificial Intelligence, Privacy Concerns, Security Concerns, Deep Learning, Adversarial Attacks, Synthetic Data, Deepfake, Ethical Implications, Cybersecurity, Machine Learning, Privacy Protection, Ethical Responsibility, Misinformation, Social Engineering, Regulatory Compliance, Artificial Intelligence, Privacy Preservation, Data Security, Threat Analysis.

I. INTRODUCTION

Due to the recent advancements in Deep Learning methods and the gradual increase in computational power, there has been a proliferation of publicly available Generative AI [1] products. These include ChatGPT [2] and DALL E from openAI, Github Copilot, AlphaCode from Deepmind, and many more. They are all applications of Generative AI in some way or another. Although these tools have greatly contributed to the common good, it is equally important not to overlook the negative implications they have. We have conducted a comprehensive analysis of these concerns from five distinct perspectives, considering the rising occurrences of Deepfake incidents [3], breaches of privacy in synthetic data [4], and adversarial attacks on generative models [5]. These perspectives are user, ethical, regulatory and legal, technological, and institutional perspectives. As shown in Figure 2, we have developed a novel classification for the above-stated viewpoints and addressed possible ways to mitigate the concerns associated with the privacy and security of the models.

Several approaches are being employed to address the privacy and security concerns in Generative AI, such as Privacy-Preserving Techniques (PPTs), Adversarial Defense Mechanisms, and Regulatory Measures and Policies. PPTs such as differential privacy [6], federated learning [7], [8], [9], and secure multi-party computation [10] are used to generate synthetic data or perform computations while preserving privacy during the data training and inference phase. Privacy-preserving generative models based on Generative Adversarial Networks (GANs) or Variational Autoencoders (VAEs) can provide privacy guarantees during the data generation process. These techniques aim to limit the exposure of sensitive information during the generative phase. However, there are challenges in balancing privacy and utility since they involve adding noise to preserve privacy, which may impact the quality of the generated data.

In contrast, adversarial defense mechanisms, including techniques such as adversarial training [11], input validation,

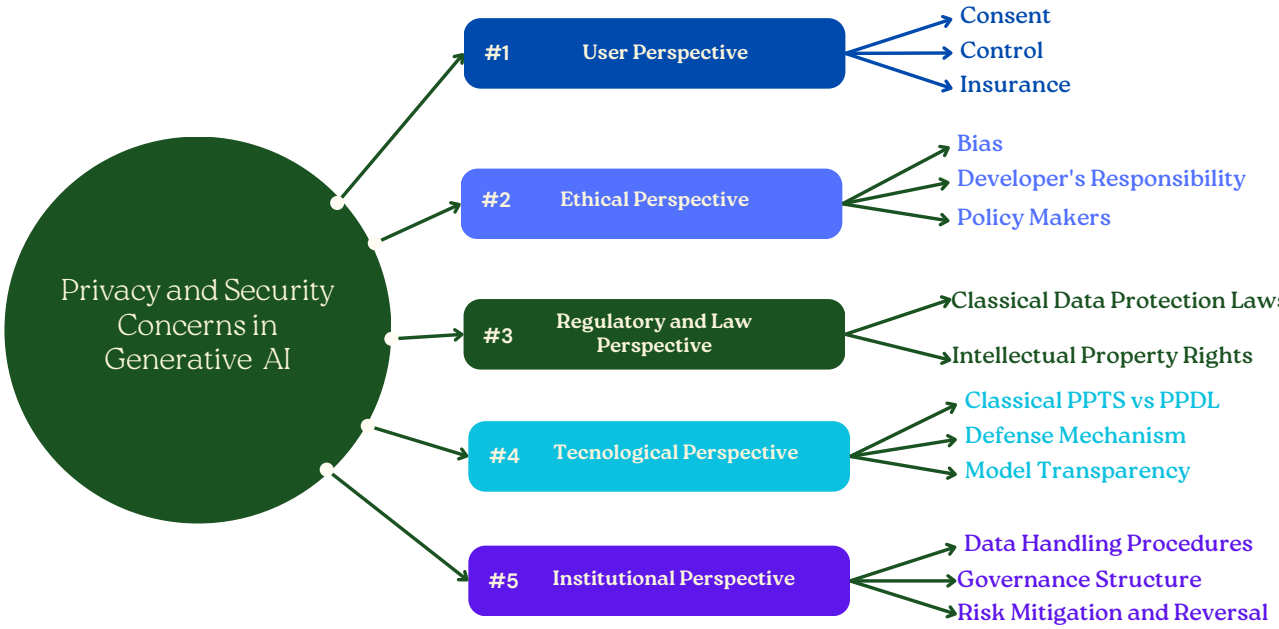

FIGURE 1: Privacy and Security Concerns in Generative AI in 5 perspectives

and the development of robust model architectures [12], are employed to counteract adversarial attacks on generative models. Although these defense mechanisms can enhance the security of generative models, adversarial attacks continue to advance in their strategies and models. Adversarial training is an approach where generative models are trained with adversarial examples to enhance their robustness against attacks. Validating and sanitizing inputs can help detect and filter out potentially malicious inputs before they reach the generative model. Moreover, continuous monitoring of generative models is crucial to detect adversarial attacks and system vulnerabilities.

PPTs have huge privacy-utility trade-offs. An illustrative instance can be found in the realm of medical research [13], where synthetic data generated using PPTs may fall short of fully replicating the statistical properties of the original data, thereby constraining its suitability for precise analysis and informed decision-making. Adversarial attacks [14] can compromise the security and trustworthiness of generative models used for image generation. Consider the scenario where an adversary may tamper with input data or introduce imperceptible noise to deceive the model into generating deceptive images, such as Deepfakes. The consequences of such manipulations extend to diverse domains, including but not limited to forensic analysis [15], content authentication [16], and autonomous vehicles [17].

Taking into consideration all the dangers associated with generative AI, this paper is aimed at providing a comprehensive and extensive survey of all the privacy and security concerns under the shadows of Generative AI. Certain countries such as Italy have feared the worst and have taken measures against ChatGPT. The Cybersecurity Hub states

TABLE 1: Abbreviations used in the text.

Abbreviation	Stands for
AI	Artificial Intelligence
ANN	Artificial Neural Networks
AR	Augmented Reality
CV	Computer Vision
DL	Deep Learning
DGMs	Deep Generative Models
DT	Deepfake Technology
GANs	Generative Adversarial Networks
ML	Machine Learning
NLP	Natural Language Processing
PPDL	Privacy-Preserving Deep Learning
RNN	Recurrent Neural Networks
VAEs	Variational Autoencoders

that The Garante, the Italian data protection agency, pointed out that the extensive gathering and retention of personal data to "train" ChatGPT lacks a legitimate legal foundation [18]. Furthermore, there exist lingering questions regarding the novelty and ownership of artworks created by generative AI models [19]. Deepfakes, a product of this technology, have found utility in diverse domains, spanning from mere entertainment to nefarious activities targeting individuals and organizations [20]. An illustrative incident involved a false tweet about an explosion at the Pentagon, which propagated widely and triggered a market selloff [21]. Notably, a search for 'Johannes Vermeer' yields an AI-generated image of 'Girl With a Pearl Earring' as the top result, supplanting the

original artwork, as illustrated in Figure 3. Despite efforts by concerned parties to mitigate the vulnerabilities of generative AI through the utilization of both classical and contemporary techniques like Privacy-Preserving Deep Learning (PPDL) [22], a compelling need persists for the development of more robust and advanced security measures for these systems. With these considerations in mind, a considerable gap remains in addressing the full spectrum of potential negative consequences associated with generative AI. Hence, a comprehensive and in-depth examination is imperative to make substantial progress in mitigating these challenges.

On that account, it is boldly imperative to bring a substantial survey in the areas of generative AI, specifically on the privacy and security of the users and the model. Through this, we get a step closer to a safe and secure AI ecosystem with resilient models and responsible developers. Our research distinguishes itself from other reviews or surveys regarding the privacy and security of Generative AI due to the following noteworthy contributions:

- **A holistic and extensive coverage:** Our survey offers a comprehensive and in-depth examination of the privacy and security concerns associated with Generative AI, encompassing various perspectives. Through a comprehensive investigation, our goal is to illuminate the multifaceted challenges associated with the privacy and security dimensions of Generative AI. We approach this subject from various perspectives, delving into its far-reaching implications across diverse domains, encompassing themes such as data protection, ethical considerations, and potential vulnerabilities. Through our extensive study, we provide an understanding of the challenges and risks inherent in Generative AI systems, helping researchers, practitioners, and policymakers to navigate this evolving landscape. Our work addresses the intricacies of privacy and security in Generative AI, contributing to the existing knowledge and providing insights for enhancing safeguards in this rapidly evolving field.
- **Developed a novel classification:** Our paper introduces a novel classification framework that examines privacy and security in Generative AI from five detailed perspectives. This comprehensive approach allows for a thorough understanding of the challenges and implications involved. By exploring legal, technical, user trust, transparency, and ethical aspects, our work provides valuable insights for addressing privacy and security concerns in Generative AI systems.
- **In-depth analysis:** Through meticulous research and thorough investigation, our work goes beyond the surface-level examination. Starting by noting previous incidents related to the systems and considering public complaints against the efficiency of the products, we offer a comprehensive resource for researchers who aim to tackle one or more of the privacy and security issues.

Our survey is primarily intended for researchers seeking a comprehensive view of security and privacy in generative AI model products while also offering valuable insights for practitioners, students, and anyone interested in AI and ML, facilitating secure and effective product utilization and understanding of their operation. The organization of the paper is given in Figure 1.

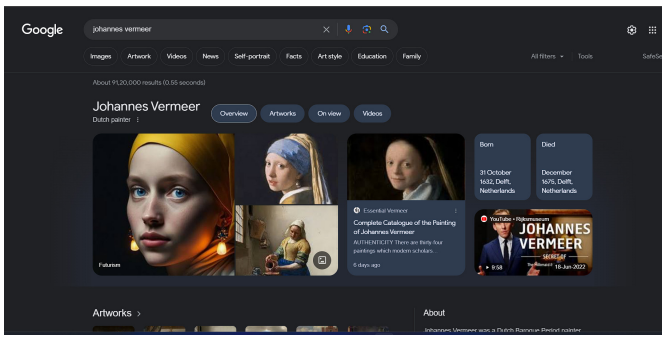

FIGURE 2: Google's top result for "Johannes Vermeer"

II. RELATED WORKS

As the discipline of generative AI has been getting traction in recent years, there have been various developments and surveys. To highlight the successes and gaps, we have done extensive research by reading and analyzing peer-reviewed surveys. The prior studies have focused on the applications of the models and how to further evolve them. However sharp double-edged sword they are, we can utilize these models in the field of cybersecurity [28] as well. Having said all that, we can not help but notice there is not a whole rounded survey encompassing the different perspectives on the privacy and security of these machines.

Harry Tanuwidjaja *et al.* [23] have discussed the challenges of data privacy and security in Machine Learning as a Service (MLaaS) platforms. PPDL is a set of techniques that allow data owners to keep their data private while still allowing MLaaS platforms to train models on the data. They discussed different PPTs and provided a detailed comparison of the surveyed PPDL works based on their own defined metrics. The paper provides a comprehensive survey of both classical and well-known PPDL techniques. Moreover, they discussed security goals and attack models with possible countermeasures for each scenario.

Qiao Zhang *et al.* [24] have identified the need for PPDL due to concerns about exposing and collecting large amounts of data, the high cost of locally deploying computation resources, and the threat of releasing well-trained model parameters. They categorized the techniques based on linear and nonlinear computations. Moreover, they provided a detailed overview of each primitive and introduced the way each primitive is used. In conclusion, they discussed several promising directions such as developing more efficient and scalable techniques and highlighting the primary technical hurdles that need to be addressed.

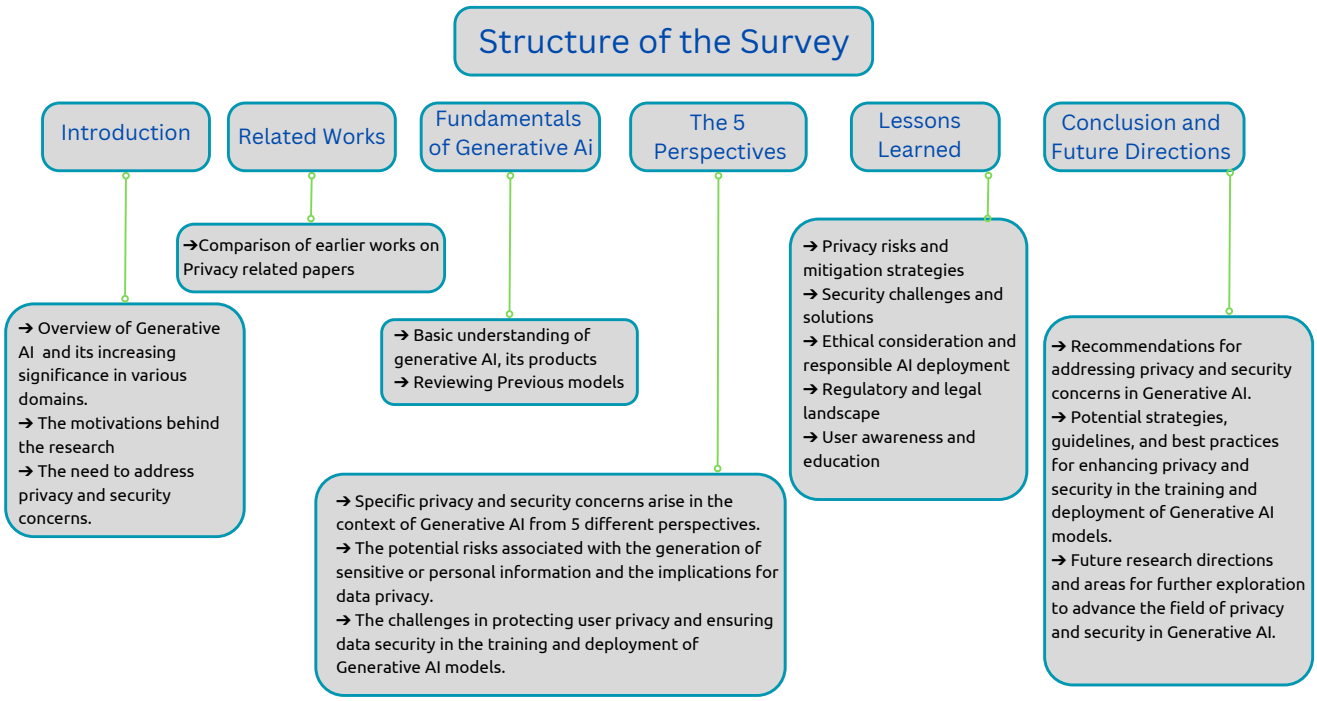

FIGURE 3: Organization of the paper

Hui Sun *et al.* [25] provides a comprehensive overview of adversarial attacks targeting deep generative models (DGMs), which are machine learning models used for generating data like images, text, and audio. The survey covers various types of attacks on DGMs, including those targeting training data, latent codes, generators, discriminators, and the generated data itself. The paper discusses the attack methods, their impact, and proposed defense mechanisms. It highlights the challenges associated with defending against these attacks, such as identifying adversarial inputs and finding defenses that are effective against diverse attack strategies while preserving model utility. The paper concludes by outlining future research directions, emphasizing the need for more robust DGMs, improved defense mechanisms, and methods for detecting and classifying adversarial inputs.

Wajiha Shahid *et al.* [26] did a survey on detecting fake news spreaders which sheds light on the current state of detecting fake news spreaders. The authors categorize features into four main types: content-based, user-based, network-based, and hybrid features, which combine elements from the previous categories. They evaluate the performance of various machine learning algorithms on different datasets and find that hybrid features yield the best results. The paper highlights challenges in fake news detection, such as the ever-changing nature of fake news, the difficulty in obtaining reliable training data, and the lack of a shared benchmark dataset for evaluation. In conclusion, the authors emphasize the importance of developing effective detection algorithms to combat the spread of misinformation.

Katina Michael *et al.* [27] explores the utilization of AI in cybersecurity. It discusses the potential benefits of AI in automating tasks, identifying threats, and responding to incidents more efficiently. Therefore, AI has been used to respond to incidents more quickly and effectively than humans can. However, the paper also highlights challenges such as the need for extensive data, ensuring the security of AI systems, and the possibility of AI being utilized by attackers. The authors conclude that while AI has the potential to greatly enhance cybersecurity, it is crucial to acknowledge and address the associated risks.

The studies reviewed in this section have significantly contributed to the field of Privacy and Security concerns in Generative AI. Some of the papers have provided a detailed overview of PPTs that could potentially lower the concerns on their own metrics. On the other hand, other papers provided a comprehensive analysis and survey on why these concerns should also be taken care of by ethical governance. Although different privacy-preserving techniques play a great role in securing data privacy during both the data training and inference phase, they come with quality and utility issues, which in turn leads to some other bias. In conclusion, all the papers reviewed above have provided well-put research on how to tackle privacy and security concerns. However, there is still a need for further study and research since the concerns are getting more sophisticated along with the rapid Generative AI growth.

TABLE 2: Related Works

Author(s) and Reference	Summary	Successes	Challenges
Harry Tanuwidjaja *et al.* [23] [2020]	Discussed different PPTs and provided a comprehensive survey of PPDL techniques, starting from classical privacy-preserving techniques to well-known deep learning techniques.	- Covered the most recent and groundbreaking methods in PPDL - Proposed a multi-scheme PPDL taxonomy that classifies adversarial models	- Limits its applicability to other domains. - Does not provide a detailed evaluation of the surveyed PPDL works
Qiao Zhang *et al.* [24] [2021]	Identified the need for PPDL due to concerns about exposing and collecting large amounts of data, the high cost of locally deploying computation resources, and the threat of releasing well-trained model parameters.	- Comprehensive overview of various techniques for enabling cloud servers - Categorizes the techniques with respect to linear and non-linear computation - Provides promising directions for future research.	- Limited comparison of the various techniques discussed - Limited discussion of the accuracy-efficiency
Hui Sun *et al.* [25] [2021]	Discussed the attack methods, their impact, and proposed defense mechanisms. It highlighted the challenges associated with defending against these attacks, such as identifying adversarial inputs.	- Develop new methods for detecting and preventing cyberattacks - Improve the security of critical infrastructure systems	- AI systems can be expensive to develop and deploy - AI systems can be vulnerable to adversarial attacks
Wajiha Shahid *et al.* [26] [2022]	Evaluated the performance of various machine learning algorithms on different datasets and found that hybrid features yield the best results.	- The use of machine learning algorithms has shown promise in detecting fake news spreaders - There is a growing body of research on fake news spreading detection	- The lack of ground-truth data makes it difficult to evaluate the performance of detection algorithms - There is no shared benchmark dataset for evaluating detection algorithms
Katina Michael *et al.* [27] [2023]	Discussed the potential benefits of AI in automating tasks, identifying threats, and responding to incidents more efficiently.	- Time and cost efficacy - Best autonomous and secure decisions	- Society can benefit and demerit from the same model - Too much dependence might cause de-learning of inherent humane skills
Our work [2023]	Gives a comprehensive overview of the various ways products of Generative AI could be used against their sole purpose and proposed paths to mitigate these shortcomings.	- Classified the privacy and security concerns into 5 - Complete and comprehensive survey for researchers seeking and developing solutions - Identification of key challenges	- Limited focus on specific and detailed implementation - Not enough experimentation with the suggested recommendations

III. FUNDAMENTALS OF GENERATIVE AI

A. RECURRENT NEURAL NETWORKS

Recurrent Neural Networks (RNNs) represent a pivotal advancement in the field of artificial intelligence, specifically designed to tackle tasks involving sequential data. Unlike traditional feedforward neural networks, they introduce a dynamic element by incorporating recurrent connections that enable the network to retain information about previous inputs [29]. This architectural innovation endows them with the capability to comprehend and generate sequences, making them particularly well-suited for applications in natural language processing, speech recognition, and time-series anal-

ysis. The inherent ability to capture temporal dependencies within data has positioned them as a cornerstone in the realm of generative AI, where the generation of coherent and contextually relevant sequences is a fundamental objective [30].

Despite their remarkable potential, RNNs are not without challenges. The vanishing gradient problem as shown in Figure 4, where the influence of distant inputs diminishes during training, and the exploding gradient problem [31], where gradients become excessively large, can impede the effective learning of long-range dependencies.

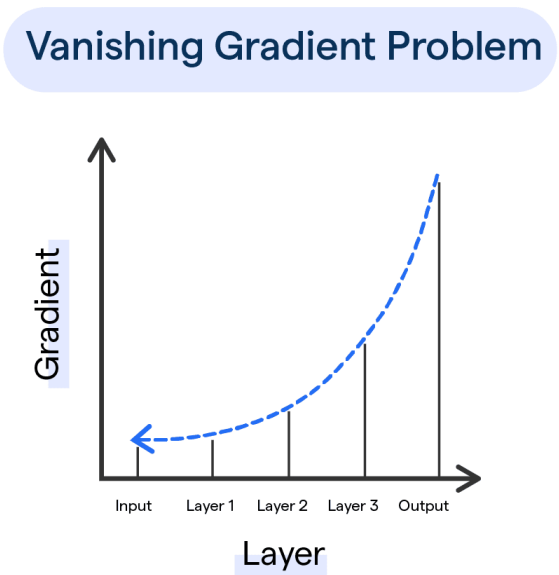

FIGURE 4: Vanishing Gradient Problem

B. RNN WITH ATTENTION

The fusion of attention mechanisms with RNNs marks a substantial advancement in sequence modeling, overcoming limitations inherent in conventional RNN structures. By incorporating attention mechanisms, they gain the ability to selectively focus on different parts of the input sequence, offering a dynamic and context-aware approach to processing sequential data [32], [33].

The augmented representation, formed by combining the context vector with the current RNN state, enhances the model's understanding of both local and broader contextual information. This improvement has far-reaching applications, particularly in NLP tasks. In machine translation, attention-equipped RNNs excel in capturing dependencies between source and target language words. Text summarization [34] benefits from the attention mechanism's ability to distill essential information for concise summaries. Similarly, question answering [35] applications achieve greater accuracy by considering pertinent details in the input.

C. TRANSFORMER

The Transformer architecture, introduced in the seminal paper "Attention is All You Need" [36] marks a transformative shift in the landscape of sequence modeling, particularly in NLP. Its departure from the sequential processing paradigm of traditional models, such as RNNs, has paved the way for more efficient and effective approaches to capturing complex patterns in sequential data [32]. Unlike RNNs, which process sequences sequentially, the Transformer's self-attention mechanism allows for the simultaneous consideration of all elements in the sequence. This parallelization not only accelerates training but also facilitates more efficient hardware utilization.

Architecturally, the Transformer is composed of modular and identical layers, each containing self-attention mechanisms and feedforward neural networks [37]. This modular design not only contributes to the model's scalability but also simplifies the addition of more layers, thereby facilitating the development of larger and more powerful models. The parallelization [38], combined with layer normalization and residual connections, contributes to more stable and faster training, especially for large models. In contrast, training deep RNNs can be a delicate task, requiring careful initialization and regularization to mitigate challenges related to the sequential nature of computation.

The Transformer architecture, as in Figure 5, represents a significant leap forward in sequence modeling. Its ability to parallelize computation, capture long-range dependencies, and handle input sequences with positional encoding has revolutionized the field. The Transformer's versatility and effectiveness have undoubtedly reshaped our approach to complex sequence modeling challenges.

D. ROLE OF GENERATIVE AI

Generative AI is a specialized branch of AI that focuses on developing models capable of generating new data resembling a given dataset. Unlike traditional AI models that are designed for specific tasks, generative AI models aim to understand the underlying distribution of the training data [39], allowing them to produce novel content that shares characteristics with the original data. Techniques such as GANs and VAEs [40] are commonly used in generative AI, enabling the creation of realistic images, videos, audio, text, and more. This field has gained significant attention due to its potential applications in creative tasks like art generation and practical uses such as data augmentation [41], drug discovery [42], and image-to-image translation [43], among others. The role of generative AI in artificial intelligence extends beyond just data generation, machine translation, and text summarizing; it plays a pivotal role in addressing several key challenges and contributing to advancements across different domains.

– **Creative Content Generation:** One of the most exciting aspects of generative AI is its ability to foster creativity. It empowers AI systems to produce art, music, literature, and other forms of creative content. Artists,

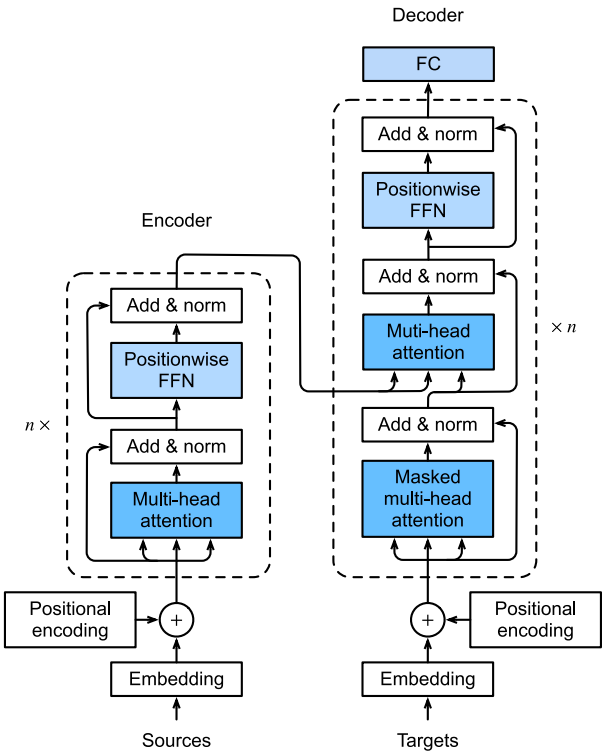

FIGURE 5: Transformer Architecture

designers, and musicians have embraced generative AI as a tool for inspiration and collaboration, leading to groundbreaking new forms of expression and artistry [44].

– **Drug Discovery and Molecule Design:** Can be employed in the pharmaceutical industry to find new drug candidates and design molecules with desired properties. By exploring extensive chemical spaces [45], these models propose potential compounds that interact with biological targets, accelerating drug discovery and potentially leading to faster development of life-saving medications.
– **Anomaly Detection and Data Security:** It can also be employed for anomaly detection in various domains, such as identifying fraudulent transactions in financial data or detecting abnormal behavior in cybersecurity [46]. By learning the normal distribution of data, generative models can identify deviations that may signify potential threats or anomalies.
– **Robotics and Autonomous Systems:** - Generative AI enables sim-to-real transfer in robotics, pre-training models in simulations and fine-tuning them with real-world data [47] for more effective control of physical robots. This approach enhances autonomous systems' capabilities by bridging the gap between simulation and reality, facilitating efficient development and deployment of reliable robotics.

The application of generative AI is continually expanding. With its ability to imagine and create, it promises to reshape various industries, foster innovation, and pave the way for more sophisticated AI systems with increasingly human-like capabilities. However, it also raises ethical considerations and challenges related to the potential misuse of generative AI for deceptive or harmful purposes, necessitating responsible development and deployment in AI applications.

E. KEY CONCEPTS OF GENERATIVE MODELS

Generative models learn the distribution of training data to produce new samples with similar patterns. Key techniques include autoencoders for efficient data representations [48], GANs for realistic content creation through adversarial training, and VAEs for meaningful latent representations using probabilistic modeling.

Autoencoders are neural networks designed for unsupervised representation learning [49], aiming to learn a compact and meaningful latent space of input data. The architecture consists of an encoder and a decoder, with the encoder mapping input data to a lower-dimensional bottleneck layer and the decoder reconstructing the original data. Training minimizes reconstruction error, enabling efficient and informative data representation. Autoencoders are vital in dimensionality reduction [50], data compression, denoising, and feature extraction, playing a significant role in various downstream tasks.

GANs are a potent category of generative models capable of producing new data similar to a given training dataset. GANs involve two neural networks: the generator, which creates synthetic data samples, and the discriminator, which serves as a binary classifier to differentiate between real training data and generated data. Adversarial training drives these models, as the generator and discriminator [51] compete to outperform each other, representing the key idea behind GANs.

– **The Generator and Discriminator:** The generator network in a GAN takes random noise as input and maps it to the data space to produce synthetic data samples. The goal of the generator is to generate data that is so realistic that the discriminator cannot distinguish it from real data [52] using the generator loss given by:

$$\mathcal{L}_{\text{gen}} = -\frac{1}{m}\sum_{i=1}^{m}\log(D(G(z_i))) \tag{1}$$

The variables in the generator loss formula are integral to understanding the dynamics of Generative Adversarial Networks (GANs). The generator loss ($\mathcal{L}_{\text{gen}}$) serves as a metric, gauging the efficacy of the generator in producing realistic data. The parameter m signifies the number of samples within the training batch, with i denoting the index for an individual sample in the batch. The random noise vector z_i serves as input to the generator for the i-th sample, contributing to the diversity

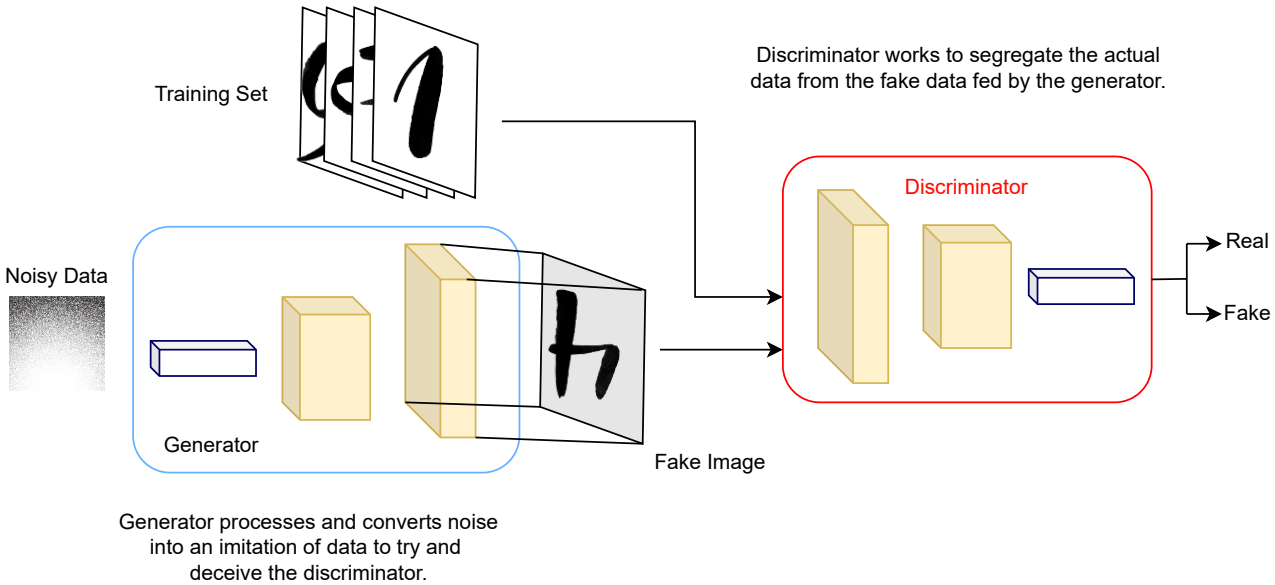

FIGURE 6: Generator and Discriminator

of generated outputs. The function $G(z_i)$ represents the generator's output when provided with the specific noise z_i, showcasing the model's ability to transform random input into meaningful data. The discriminator function $D(\cdot)$ is pivotal; it evaluates the likelihood that its input is a real sample, a crucial aspect of the adversarial interplay. Lastly, the natural logarithm function $\log(\cdot)$ is employed in the loss computation, emphasizing the importance of the logarithmic scale in assessing the divergence between generated and real samples. Together, these variables and functions underpin the intricate dance between generator and discriminator in the GAN framework [53].

On the other hand, the discriminator is trained to correctly classify data as either real or generated using a discriminator loss given by [53]:

$$\mathcal{L}_{\text{disc}} = -\frac{1}{m}\sum_{i=1}^{m}[\log(D(x_i)) + \log(1 - D(G(z_i)))] \tag{2}$$

As the training progresses, the generator improves its ability to produce more realistic data, while the discriminator enhances its capability to differentiate real from fake data as depicted in Figure 6.

– **Min-max Game in GANs:** The training process of GANs can be seen as a min-max game between the generator and the discriminator. The objective is to minimize the discriminator's ability to correctly classify generated data (minimize its loss) while maximizing the discriminator's ability to correctly classify real data (maximize its loss) [54]. This adversarial training process leads to a dynamic equilibrium where the generator produces increasingly realistic data that becomes indistinguishable from real data. The GAN framework has shown remarkable success in generating high-quality, diverse data in various domains, such as images, videos, and even text.

VAEs belong to the category of generative models and adopt a probabilistic technique for encoding and decoding data. Unlike traditional autoencoders, VAEs use probability distributions to model the encoder and decoder. The encoder maps input data to a distribution in the latent space, while the decoder generates data by sampling from this distribution [55]. VAEs aim to maximize the evidence lower bound (ELBO), a balancing objective that considers the reconstruction error and a regularization term to shape the learned latent space to adhere to a predetermined prior distribution, commonly a simple Gaussian.

F. TYPES OF GENERATIVE MODELS

Generative models stand out as a class of algorithms that hold the unique ability to learn and understand the underlying probability distribution of the data they are trained on. Unlike their discriminative counterparts, generative models venture beyond classification tasks, aiming to create new data points that closely resemble the original dataset. This subsection delves into the intriguing realm of various generative models, exploring their advantages and shortcomings. Apart from autoregressive models, VAEs, and GANs, We have explored the various types of generative models in Table 3.

G. RECENT ADVANCEMENTS AND TRENDS

AI-powered content and data generation have advanced significantly since its inception. From creating photorealistic images and videos to composing music and writing stories, AI is now capable of producing creative content that is indistinguishable from human-made work.

TABLE 3: Generative Models Comparison

Name	Description	Pros	Cons
Flow-based Models	Flow-based models use invertible transformations to map data space to latent space, enabling efficient sampling and tractable likelihood computation. [56]	- Suitable for image generation and density estimation tasks. Efficient sampling and computation of likelihood make them attractive for high-quality sample generation.	- Limited scalability in handling complex data distributions. Computationally expensive for high-dimensional data.
Boltzmann Machines	These stochastic neural networks learn joint probability distributions over binary data. They find applications in modeling and generating binary data.	- Useful in tasks like collaborative filtering and unsupervised learning, where modeling binary data is crucial. [57] Applications in various domains requiring binary data modeling.	- Computationally expensive for large datasets. Deep structures can lead to training challenges.
Deep Belief Networks (DBNs)	DBNs are generative models with multiple layers of stochastic, latent variables. They excel in unsupervised learning tasks and feature learning.	- Valuable for unsupervised learning and feature learning tasks. Effective in representing complex data distributions. [58]	- Training process can be slow with deep architectures. Interpretability of latent variables is limited.
Helmholtz Machines	Helmholtz Machines combine recognition models (e.g., autoencoders) with generative models, often using Boltzmann Machines.	- Effectively models complex data distributions. Enables data generation in diverse domains. [59]	- Two-stage learning process can result in challenging optimization. Scalability to large datasets may be limited.
Implicit Generative Models	Implicit generative models do not model explicit probability distributions, but generate samples through optimization methods like variational inference. [60]	- Offers flexibility for scenarios where explicitly modeling probability distributions is challenging or computationally expensive. Useful in tasks like data generation, completion, and synthesis.	- Less interpretable compared to explicit generative models. More training iterations and tuning may be required.

– **The Emergence of Explainable Generative AI:** Explainable Generative AI [61], [62] has emerged to address the complexity of understanding how Generative AI generates output, focusing on transparency and interpretability in decision-making. The lack of transparency has hindered generative AI's widespread adoption, particularly in industries like finance and healthcare that require accountability. The goal is to enhance transparency by developing specialized visualization and interpretation techniques tailored to generative models.

– **Integration with the Internet of Things (IoT):** IoT [63], [64], [65] encompasses interconnected objects capable of collecting and exchanging data; When combined with generative AI, it enables machines to generate content using real-time data from the physical world. This integration has the potential to create smart devices capable of generating personalized and valuable content based on user behavior and the surrounding environment. The applications span from smart home environments to industrial settings.

– **Integration with Blockchain Technology:** Blockchain [66], [67], as a secure and decentralized digital ledger, has the potential to revolutionize generative AI applications by enhancing security and transparency. The integration of generative AI with blockchain technology enables the development of decentralized AI networks that are more robust against cyber threats and offer greater privacy in their operations. The combination presents a transformative opportunity for secure and transparent applications.

– **Integration with the Augmented Reality (AR):** The integration creates a captivating synergy that elevates user experiences, facilitates interactive content, and enables realistic simulations. AR [68] overlays digital elements onto the user's real-world environment, often via smartphones or AR glasses. Combining Generative AI with AR unlocks dynamic, personalized, and immersive experiences, seamlessly integrating digital content into the user's natural surroundings.

IV. PRIVACY AND SECURITY CONCERNS IN GENERATIVE AI FROM 5 PERSPECTIVES

Generative AI has made a significant contribution to different sectors ranging from Healthcare [69], [70], [71] to Education [72], [73], from Finance to Arts [74], [75], from Autonomous Vehicles [76], [77] to Drug Discovery [78], [79], and more. In the above sections, we discussed the overall positive impacts of this advanced technology. However, this technology comes up with potential threats and dangers when it's used in a negative manner. In this section, we will discuss the negative implications of Generative AI from 5 different perspectives, namely:

1) User Perspective
2) Ethical Perspective
3) Regulatory and Law Perspective
4) Technological Perspective
5) Institutional Perspective

The rapid advancement of Generative AI has given rise to Deepfake Technology (DT) [80], [81], presenting a significant challenge to individuals and global institutions such as financial organizations and entertainment industry at large. DT is an AI-driven innovation that manipulates visual, auditory, and video content to fabricate events that never occurred. It is powered by GANs, which employs two Artificial Neural Networks (ANNs) such as detector and synthesizer [82].

A. USER PERSPECTIVE

The impact of deepfake incidents on individuals from the perspectives of consent [83], control, and insurance, is multifaceted and can have wide-ranging implications for personal, professional, and societal aspects. Deepfakes can violate individuals' consent by using their likeness, voice, or identity without their permission, leading to a loss of control over how their image is portrayed. The act of cyberbullying using deepfake technology has affected a lot of people's life without their consent. For example, an attacker might wish to publish a video of the target doing some unlawful activity just by using their headshots and training them in the deep learning model. At this time, even though the target didn't commit the crime, these high realistic videos might get them jailed or whatever the attackers plan to do with the deepfake.

DT has significant implications for privacy and security from a user's perspective. Here's an overview of how deepfakes can affect individuals in these two areas: Privacy and Security.

1. **Manipulation of Personal Content:** DT allows for the manipulation of personal photos and videos with remarkable realism [84]. This can infringe upon an individual's privacy when their images are used without consent. For example, imagine a deepfake video is created that shows a famous celebrity seemingly endorsing a controversial product, such as a dietary supplement. This video is convincingly crafted, making it appear as though the celebrity genuinely supports the product. In this case, the privacy of the celebrity is infringed upon as their likeness and reputation are manipulated without their consent. This intrusion can damage the celebrity's brand and trustworthiness. The broader issue is the erosion of trust in video content, as the public becomes increasingly skeptical of the authenticity of what they see.

2. **Identity Theft:** Deepfakes can be used to create convincing audio recordings, mimicking a person's voice and speech patterns. These fake audio clips can be used to deceive individuals or automated voice recognition systems, potentially leading to identity theft or unauthorized access to personal information. Users may unknowingly share sensitive information with attackers who convincingly mimic trusted voices, undermining their trust in communication channels.

3. **Authentication Challenges:** Deepfake technology raises concerns about the reliability of authentication methods that rely on biometric data [85] like facial recognition [86], [87]. An attacker could use a deepfake to gain unauthorized access to secure systems or accounts by mimicking the targeted individual's biometric data. For example, a criminal obtains a highly realistic deepfake mask of an individual's face and uses it to gain unauthorized access to the person's smartphone through facial recognition, subsequently accessing sensitive data. This scenario underscores the vulnerability of facial recognition systems to deepfake attacks. It suggests that additional security layers and biometric verification techniques should be implemented to safeguard personal devices and data.

Deepfake technology is continually evolving, and users must adapt their behaviors and security practices to mitigate the risks associated with its misuse. In response to these privacy and security concerns, there are ongoing efforts to develop detection and mitigation techniques for deepfakes. Additionally, legal and regulatory frameworks are being established in some jurisdictions to address the misuse of deepfake technology.

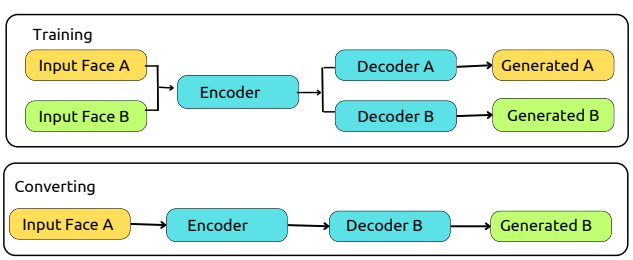

FIGURE 7: Deepfake generation

B. ETHICAL PERSPECTIVE

Deepfake technology presents several ethical challenges and considerations in terms of privacy and security. From an ethical perspective [88], here are some of the key impacts and issues associated with deepfake technology: bias, developer responsibility, and policymakers.

TABLE 4: Comparison of Data Protection Laws

Data Protection Law	Jurisdiction	Principles	Rights of Data Subjects	Data Transfers	Enforcement
GDPR	European Union	Lawfulness, Purpose Limitation, Data Minimization, Accuracy, Storage Limitation, Integrity, Confidentiality	Right to Access, Right to Erasure, Right to Rectification, Right to Object, Right to Data Portability	Subject to Adequacy Decisions, Standard Contractual Clauses, Binding Corporate Rules	Data Protection Authorities, Fines
CCPA	California, USA	Notice, Opt-out, Non-Discrimination, Data Minimization, Purpose Limitation, Security	Right to Know, Right to Deletion, Right to Opt-out of Sale	Limited	California Attorney General, Private Right of Action for Breaches
LGPD	Brazil	Purpose Limitation, Data Minimization, Transparency, Security, Non-Discrimination	Right to Access, Right to Erasure, Right to Rectification, Right to Data Portability	Subject to Adequacy Decisions	Brazilian Data Protection Authority (ANPD), Fines
PDPA	Singapore	Consent, Purpose Limitation, Notification, Accuracy, Protection, Retention	Right to Access, Right to Correction, Right to Withdraw Consent	Subject to Adequacy Decisions, Model Contractual Clauses	Personal Data Protection Commission (PDPC), Fines
POPIA	South Africa	Accountability, Lawfulness, Purpose Limitation, Minimality, Transparency	Right to Access, Right to Correction, Right to Erasure	Limited	Information Regulator, Fines

1) Bias

Deepfake technology can introduce or exacerbate biases, both overt and subtle, in various ways. Biases may be present in the training data used to create deepfake algorithms [89], or they may be introduced intentionally by creators. These biases can manifest in terms of race, gender, age, and other characteristics. Ethical principles demand that technology, including deepfake technology, should not perpetuate or amplify biases. A fair and just society should strive for equal treatment and respect for all individuals, regardless of their background.

2) Developer Responsibility

Developers and researchers are at the forefront of creating and advancing deepfake technology. Their ethical responsibility extends to how they develop, use, and regulate this technology. Developers have an ethical duty to prioritize the responsible use of deepfake technology. Ethical principles [88] such as transparency, consent, and accountability should guide their actions.

3) Policymakers

Policymakers play a pivotal role in crafting laws and regulations that govern the responsible development, distribution, and use of deepfake technology. Policymakers must navigate a complex ethical landscape [90] when regulating deepfake technology. They need to balance the protection of individual rights and public welfare with the preservation of freedom of expression and innovation.

Balancing these ethical considerations within deepfake technology requires a holistic approach involving not only developers and policy makers but also researchers, civil society organizations, and the broader public. Collaboration and dialogue among these stakeholders are crucial to ensure that

deepfake technology aligns with ethical principles and serves the best interests of society.

C. REGULATORY AND LAW PERSPECTIVE

The rise of generative AI has left regulators grappling with the challenge of keeping pace with this transformative technology. As we explore the regulatory and legal perspective, it becomes evident that a delicate balance is needed to harness the potential benefits while safeguarding against risks.

1) Classical Data Protection Laws

Data privacy laws worldwide are the first line of defense against the potential misuse of personal information by generative AI platforms [91]. Governments have introduced regulations like the General Data Protection Regulation (GDPR) [92] in the European Union, California Consumer Privacy Act (CCPA) [93] in the US, and other similar laws globally. These laws mandate that companies must handle personal data with transparency and obtain explicit consent from individuals for its use.

However, generative AI introduces new challenges as AI models may inadvertently generate content that contains personal information. This raises questions about how AI-generated data should be treated under data privacy laws and who bears responsibility for its protection. When GDPR arrived, it had some tricky terms like "undue delay" and "disproportionate effort." Though the rules seemed clear, companies sometimes played with their meanings. For example, Facebook took almost two months to report a problem and said it followed the "undue delay" rule [94].

Similarly, the global nature of AI development introduces complexities regarding cross-border data flows [95]. Data used to train generative AI models is sourced from multiple jurisdictions as shown in Table 4, each with its data protection laws. Ensuring compliance with varying regulations while maintaining seamless data access for AI development becomes a challenge.

2) Intellectual Property Rights

AI often learns from existing data. When generative AI incorporates copyrighted elements into its creations [96], understanding where fair use ends and copyright infringement begins can be challenging. This raises concerns about derivative works and their legal implications.

The determination of whether credit should be attributed to the AI itself or the human creator of the AI becomes intricate, potentially impacting copyright assertions and the acknowledgment of inventive contributions [97], [98]. Moreover, the concept of ownership is enigmatic when it comes to AI-generated content. Unlike human creators, AI lacks legal personhood [99], raising questions about who possesses the rights and how safeguarding these rights can be ensured. These uncertainties influence the potential for economic gains from such creations.

Furthermore, collaborative endeavors between humans and AI to craft content introduce the concept of joint authorship [100]. This adds a layer of complexity to copyright law as it becomes necessary to navigate the distribution of rights and contributions in these partnerships. The matter of primary ownership and the equitable sharing of recognition are puzzles that require careful consideration [101]. The evolution of copyright laws to encompass AI-generated content is another challenge, given that existing regulations were not designed with AI in mind.

D. TECHNOLOGICAL PERSPECTIVE

This section of the paper highlights critical dimensions that encompass classical Privacy-Preserving Techniques (PPTs) versus Privacy-Preserving Deep Learning (PPDLs), defense mechanisms against adversarial attacks, and strategies to enhance model transparency. These sub-sections collectively offer insights into the evolution of privacy and security safeguards within the dynamic domain of Generative AI.

1) Classical PPTs Vs PPDLs

Traditional privacy-preserving technologies (PPTs) use well-established methods to protect privacy, but they may not be enough to counter advanced privacy threats [102]. On the other hand, privacy-preserving differential learning (PPDLs) use cutting-edge techniques to address the weaknesses of traditional methods [103]. In Table 5, we have shown the comparison of computational and communication cost of each of the schemes desribed below.

- **Federated Learning [104]:** updates models locally on user devices, keeps raw data on devices, reducing the risks associated with centralized data storage.
- **Homomorphic Encryption [105]:** enables secure computations on encrypted data, allowing privacy without sacrificing data utility. This is a type of encryption that allows computations to be performed on encrypted data. This means that data can be kept encrypted throughout the computation process.
- **Differential Privacy [106]:** adds noise to aggregated results, as shown in Figure 8, protecting individual privacy while still allowing meaningful insights.

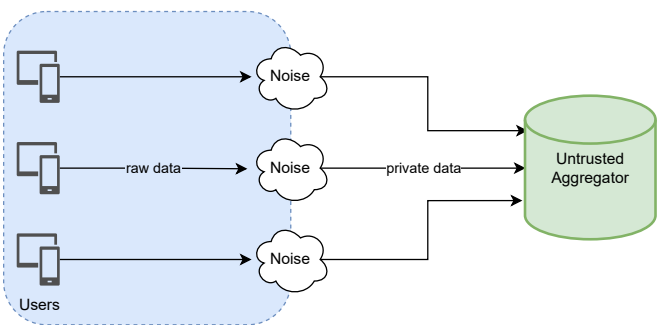

FIGURE 8: Differential Privacy

2) Defence Mechanism

Two distinct defence strategies stand out: adversarial training and detection techniques. Adversarial training enhances model resilience by incorporating adversarial examples into training data, boosting generalization [107]. Yet, it can raise computational load and vulnerability to transferability attacks [108]. Detection techniques proactively identify adversarial inputs using methods like anomaly detection, though they can grapple with false positives and evasion by adaptive adversaries.

TABLE 5: Comparison of Computation and Communication Costs

Privacy & Security Feature	Computation Cost	Communication Cost
Differential Privacy	High	Moderate
Federated Learning	Moderate	High
Homomorphic Encryption	High	High
Secure Multi-Party Computation	Moderate to High	Moderate to High

Another comparison lies between secure aggregation and model verification. Secure aggregation preserves data privacy in collaborative training by allowing updates without sharing raw data [109]. However, it introduces communication overhead and privacy challenges [110]. In contrast, model verification certifies model integrity before deployment, ensuring adherence to intended behavior and standards, but might demand extra resources and specialized tools [111].

3) Model Transparency

Transparent models help us see how these models make decisions, which is important when their results affect real life. Techniques like showing model internals visually and focusing on important parts of input help us understand these models [112]. However, very complex models might not be fully explainable. Transparent models are also important ethically, as they can prevent biases and unfair behavior. But there are challenges, like the trade-off between transparency and performance [113]. Some models are so advanced that they're hard to explain simply. Certain industries might not want to share their secret methods. Trust in these models grows when we know how they make decisions, especially in important areas like healthcare. Balancing how much we understand how complex the models are, is key and there are methods to help with this [114].

E. INSTITUTIONAL PERSPECTIVE

The risks associated with generative AI do not only arise from the models' flow or the underlying algorithm's inadequacy but also from the way the data is collected, processed and generally how an institution takes care of the data that flows through the system. Giving necessary attention to these details will be an effective risk avoidance procedure.

1) Data Handling Procedures

Organizations need to establish clear protocols for data collection, storage, sharing, and disposal [115], [116]. Key considerations include:

- **Data Storage:** secure storage solutions, including encryption and access controls, are necessary to prevent unauthorized access and data breaches [117].
- **Data Sharing:** if data is shared with third parties, ensuring that proper agreements and safeguards are in place to maintain privacy and security [118].

2) Governance Structure

The governance structure involves roles, responsibilities, and decision-making processes that oversee Generative AI's use regarding privacy and security [119]. A Data Privacy Officer (DPO) ensures regulatory compliance, manages privacy assessments, and addresses related concerns [120]. An Ethics Review Board provides guidance on ethical considerations, especially with sensitive content. Internal Policies offer clear guidelines for data handling, security protocols, and ethical Generative AI use. Together, these elements shape an organization's commitment to privacy and security.

3) Risk Mitigation and Reversal

To ensure comprehensive protection, it is vital to establish effective strategies for managing risks. This entails the continuous identification, assessment, and reduction of potential threats associated with the applications [121]. This process involves conducting routine evaluations of the underlying system and updating them to the state-of-the-art techniques, considering factors like data types, potential misuse, and unintended repercussions [122].

Additionally, having well-defined response plans is crucial in the event of privacy breaches or security incidents [123]. These plans encompass notifying affected parties and regulatory bodies, alongside implementing measures to mitigate the extent of damage. Moreover, it is essential to account for the possibility of malicious entities attempting to reverse engineer Generative AI models to gain access to sensitive data. By implementing safeguards against such endeavors, the overall security of these applications can be significantly enhanced [124].

V. LESSONS LEARNED AND OPEN ISSUES

Considering that we are currently in the initial phases of generative AI, and given the continuous escalation in computational capabilities and the widespread availability of information, it's foreseeable that more potent models will emerge. Simultaneously, this progression will likely give rise to more robust attack methods, capitalizing on vulnerabilities related to data collection, generation, processing, and

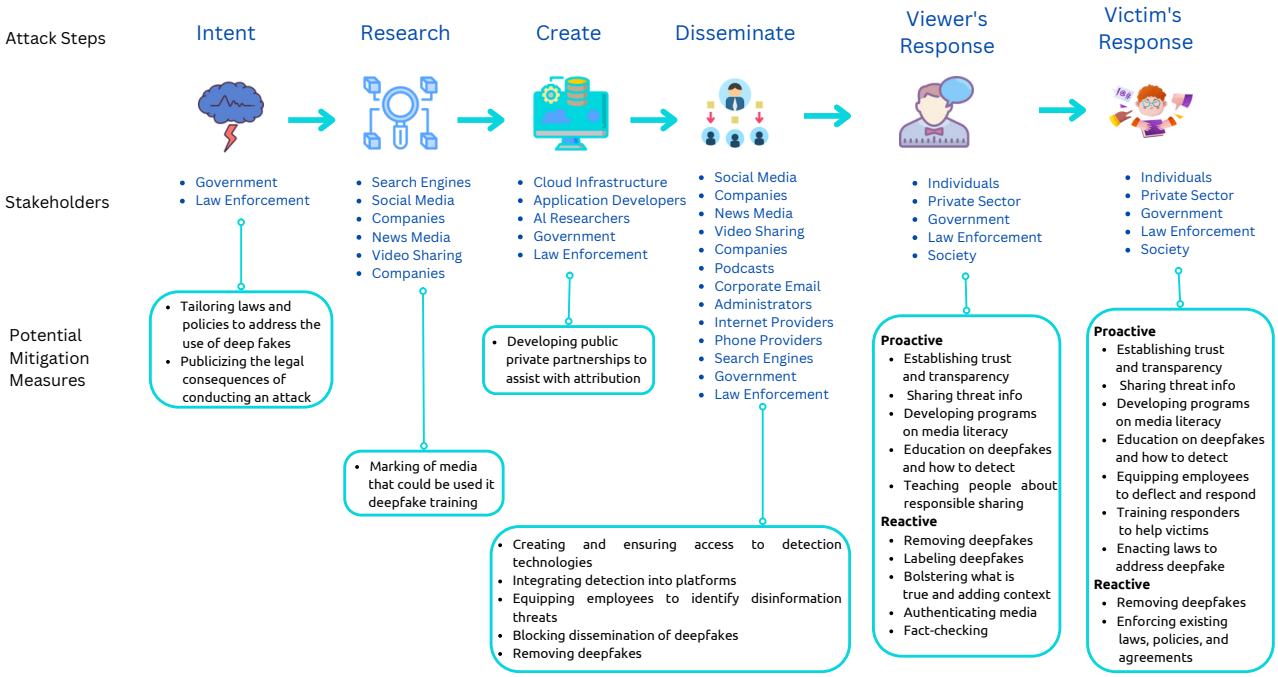

FIGURE 9: Mitigation and Reversal Strategies

model development techniques. Before outlining the lessons learned, we have taken a case study of Zao mobile application which accurately shows the concerns stated in this paper.

In September 2019, Momo Inc. unveiled the Zao mobile application, unleashing a viral sensation that swept across the globe. Zao's appeal lay in its user-friendly interface and remarkable deepfake technology, allowing users to insert their faces into scenes from famous movies and TV shows with a single uploaded selfie. What started as an entertainment app, however, swiftly unearthed a web of ethical, privacy, and security concerns.

Zao's rise to stardom was meteoric, with millions of downloads within days of its launch. Users were captivated by the prospect of starring alongside renowned actors or taking on the roles of beloved movie characters. Yet, the very allure that propelled Zao into the limelight soon led to unsettling questions about its potential for misuse.

Privacy and Security Concerns began to surface early in its journey. Initially, the app's terms of service granted its developers expansive rights to user-generated content, raising concerns about data ownership and control. Users fretted about the possibility of their images being utilized without consent. Users discovered that their deepfake video created with Zao was being used in advertisements without their knowledge or consent, emphasizing the privacy risks associated with unregulated deepfake applications. In another instance, a user generated explicit content by superimposing their face onto adult film scenes, potentially leading to the unauthorized distribution of explicit material featuring their likeness as shown in Figure 10.

The ease with which Zao facilitated deepfake creation raised grave concerns about misuse. Critics feared the app could be used to produce misleading or defamatory content, impersonate individuals, or tarnish the reputations of public figures. In a distressing scenario, a malicious user created a deepfake video of a political leader making incendiary remarks, aiming to sow disinformation and erode public trust. This illustrated the potential weaponization of deepfake technology for the spread of false information and social disruption.

Response and Repercussions followed the mounting backlash and privacy concerns. Zao pledged to revise its user agreement to address data ownership and privacy issues. Simultaneously, Chinese authorities initiated investigations into the app's data collection and privacy practices, signaling the need for regulatory intervention. The Zao case engendered discussions about the role of regulatory bodies in mitigating the ethical and privacy implications of deepfake technology. It underscored the necessity of comprehensive regulations to ensure the responsible development and use of deepfake applications.

Zao developers responded by refining their algorithms, making it more challenging to misuse the app for explicit or harmful content. They also sought to clarify their data usage policies in an effort to rebuild trust with users. This iterative development of Zao's algorithms emphasized the

importance of technological safeguards in curbing the misuse of deepfake technology. However, it also raised questions about the ongoing battle between creators of deepfake apps and those working to detect and counteract deepfakes.

In conclusion, Zao's rapid rise to prominence and subsequent ethical, privacy, and security concerns shed light on the intricate interplay between entertainment and the responsible use of emerging AI technologies like deepfakes. While the app took steps to address some of these issues, its brief yet impactful presence in the online world serves as a potent reminder of the need for vigilance, comprehensive regulations, and ethical considerations in the swiftly evolving landscape of AI and deepfake applications.

1) Emerging Privacy and Security Challenges

- **Hallucinations and Fabrications:** Generative AI crafts realistic fake content—images, videos, text—for spreading misinformation, forging identities, or creating deepfakes [125]. A fake news article resembling genuine ones could deceive individuals or manipulate opinions.
- **Cybersecurity Threats:** might spawn new malware and cyber threats harder to detect and counter than traditional attacks. For instance, authentic-looking AI-generated files could trick users into launching malware, leading to infections [126].
- **Data Privacy:** models use data for training, potentially including personal info. Inadequate anonymization could lead to identification and surveillance risks [127].

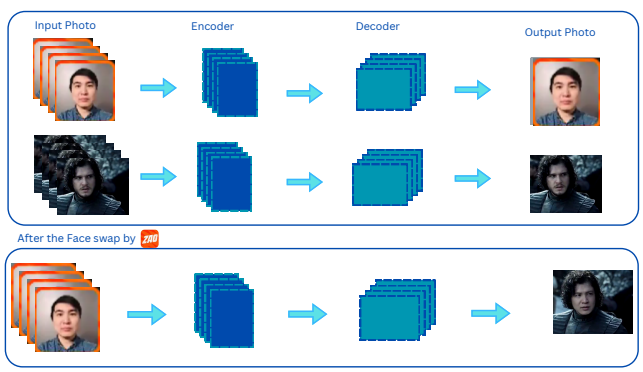

FIGURE 10: Zao's Face swap architecture

2) Interdisciplinary Collaboration

By working together, experts from a variety of disciplines can help to ensure that generative AI is developed and used in a safe, responsible, and ethical way. Undoubtedly, the synergistic efforts between industries and developers have demonstrated a cooperative approach [128]. Past instances of IT developers effectively collaborating with businesses have yielded success and engagement for both the developers and business representatives [129]. This approach serves the purpose of enhancing the safety of generative AI products. One of the organizations taking affirmative steps to promote comprehensive collaboration for enhancing the safety of generative AI products is the Partnership on AI (PAI) [130]. This initiative engages experts in the responsible AI domain, including a specialized group focusing on generative AI, to establish optimal practices. Carnegie Mellon University's Center for Applied Ethics and Technology (CEAT) is actively involved in researching ethical aspects of generative AI and providing education about its potential implications. This approach is crucial for effectively addressing all five dimensions of product safety and ensuring a positive user experience.

3) Mitigation and Reversal Strategies

Effective risk management strategies are vital for comprehensive protection. This involves continuous identification, assessment, and reduction of potential threats associated with applications. Regular evaluations of the system, considering factors like data types and unintended consequences, are crucial [131].

We have taken a case of deep fake mitigation strategies as shown in Figure 9.

VI. CONCLUSION AND FUTURE DIRECTION

In this comprehensive article, we give a holistic overview of generative AI. We curate a taxonomy that considers the privacy and security concerns of generative AI from user, ethical, regulatory, technological, and institutional perspectives. In Table 6, we show the weakness and strength of each of the perspectives. We delve into the multifaceted dimensions of this evolving landscape, examining the implications across diverse domains, ranging from data protection to ethical considerations and potential vulnerabilities. By offering an extensive analysis, we have shed light on the intricate challenges and risks that accompany Generative AI systems.

Several potential pathways hold promise in shaping the future of data privacy and security. One such avenue involves the establishment of an interdisciplinary institution tasked with overseeing global data protection efforts [132]. By uniting experts from diverse fields, this organization would foster international cooperation and the development of standardized strategies to counter evolving threats. Simultaneously, the advancement of defensive techniques through cutting-edge neural networks, such as multimodal neural networks [133], emerges as a proactive measure against sophisticated adversarial activities. These networks offer multifaceted protection, enhancing system resilience across various data types. Additionally, enhancing the clarity and precision of regulations and policies represents a critical step [134]. By collaboratively crafting interpretable and unambiguous frameworks, legal experts, policymakers, and technologists can lay the groundwork for ethical data practices and transparent compliance. As these pathways intersect, they offer a collective approach to fortify the foundations of data privacy and security in an increasingly complex digital landscape.

TABLE 6: Weakness and strength of the Perspectives

Perspective	Weaknesses	Strengths
User	Lack of awareness of privacy risks and settings	Ability to control personal data through informed consent
Ethical	Potential for biased or harmful outcomes in AI-generated content	Promotion of fairness, accountability, and transparency
Regulatory and Legal	Challenges in keeping pace with rapidly evolving technology	Establishment of guidelines and regulations to protect users
Technological	Vulnerabilities to adversarial attacks and data breaches	Advancements in encryption,privacy-preserving techniques
Institutional	Dependence on centralized data repositories and governance	Adoption of decentralized approaches and accountability

Generative AI faces several challenges and open research questions that impact its effectiveness, robustness, and ethical implications.

- **Ownership and Control of User-Generated Content:** The lack of clear guidelines for ownership and control of data generated by users using generative AI, which can potentially lead to misuse and unauthorized distribution of created content.
- **Training Instability:** Training instability is a challenge often encountered when training generative models, particularly VAEs and GANs. It refers to the difficulty in optimizing the model's parameters during the training process, leading to slow or even failed convergence [135].
- **Lack of Interpretability:** This raises concerns about the potential biases or undesirable characteristics learned by the models.
- **Adversarial Attacks:** Adversarial attacks make it challenging to validate the authenticity of data, especially in scenarios where synthetic data is used for training other models. This can undermine trust in AI systems that use generative models.
- **Regulatory Frameworks and Compliance Challenges**
- **Data Privacy Risks and User Anonymity**

REFERENCES

[1] P. Eigenschink, T. Reutterer, S. Vamosi, R. Vamosi, C. Sun, and K. Kalcher, "Deep generative models for synthetic data: A survey," IEEE Access, vol. 11, pp. 47 304–47 320, 2023.

[2] A. Shoufan, "Exploring students' perceptions of chatgpt: Thematic analysis and follow-up survey," IEEE Access, vol. 11, pp. 38 805–38 818, 2023.

[3] J. Pitt, "Deepfake videos and ddos attacks (deliberate denial of satire) [editorial]," IEEE Technology and Society Magazine, vol. 38, no. 4, pp. 5–8, 2019.

[4] J. Yoon, L. N. Drumright, and M. van der Schaar, "Anonymization through data synthesis using generative adversarial networks (ads-gan)," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 8, pp. 2378–2388, 2020.

[5] X. Tong and W. Qi, "False data injection attack on power system data-driven methods based on generative adversarial networks," in 2021 IEEE Sustainable Power and Energy Conference (iSPEC), 2021, pp. 4250–4254.

[6] N. Wu, C. Peng, and K. Niu, "A privacy-preserving game model for local differential privacy by using information-theoretic approach," IEEE Access, vol. 8, pp. 216 741–216 751, 2020.

[7] F. Sattler, S. Wiedemann, K.-R. Müller, and W. Samek, "Robust and communication-efficient federated learning from non-i.i.d. data," IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 9, pp. 3400–3413, 2020.

[8] A.-H. Rasha, T. Li, W. Huang, J. Gu, and C. Li, "Federated learning in smart cities: Privacy and security survey," Information Sciences, 2023.

[9] H. Yang, J. Zhao, Z. Xiong, K.-Y. Lam, S. Sun, and L. Xiao, "Privacy-preserving federated learning for uav-enabled networks: Learning-based joint scheduling and resource management," IEEE Journal on Selected Areas in Communications, vol. 39, no. 10, pp. 3144–3159, 2021.

[10] Q. Zhang, C. Xin, and H. Wu, "Privacy-preserving deep learning based on multiparty secure computation: A survey," IEEE Internet of Things Journal, vol. 8, no. 13, pp. 10 412–10 429, 2021.

[11] S. Chaudhury, H. Roy, S. Mishra, and T. Yamasaki, "Adversarial training time attack against discriminative and generative convolutional models," IEEE Access, vol. 9, pp. 109 241–109 259, 2021.

[12] D. Vasan, M. Alazab, S. Venkatraman, J. Akram, and Z. Qin, "Mthael: Cross-architecture iot malware detection based on neural network advanced ensemble learning," IEEE Transactions on Computers, vol. 69, no. 11, pp. 1654–1667, 2020.

[13] A. J. Rodriguez-Almeida, H. Fabelo, S. Ortega, A. Deniz, F. J. Balea-Fernandez, E. Quevedo, C. Soguero-Ruiz, A. M. Wägner, and G. M. Callico, "Synthetic patient data generation and evaluation in disease prediction using small and imbalanced datasets," IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 6, pp. 2670–2680, 2023.

[14] T. Bai, J. Zhao, J. Zhu, S. Han, J. Chen, B. Li, and A. Kot, "Ai-gan: Attack-inspired generation of adversarial examples," in 2021 IEEE International Conference on Image Processing (ICIP), 2021, pp. 2543–2547.

[15] L. Verdoliva, "Media forensics and deepfakes: An overview," IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910–932, 2020.

[16] A. Yazdinejad, R. M. Parizi, G. Srivastava, and A. Dehghantanha, "Making sense of blockchain for ai deepfakes technology," in 2020 IEEE Globecom Workshops (GC Wkshps, 2020, pp. 1–6.

[17] J. Xiong, R. Bi, Y. Tian, X. Liu, and D. Wu, "Toward lightweight, privacy-preserving cooperative object classification for connected autonomous vehicles," IEEE Internet of Things Journal, vol. 9, no. 4, pp. 2787–2801, 2022.

[18] O. Powell, "OpenAI confirms ChatGPT data breach — cshub.com," https://www.cshub.com/data/news/openai-confirms-chatgpt-data-breach, 2023, [Accessed 07-Jul-2023].

[19] G. Franceschelli and M. Musolesi, "Copyright in generative deep learning," Data & Policy, vol. 4, p. e17, 2022.

[20] K. A. Pantserev, "The malicious use of ai-based deepfake technology as the new threat to psychological security and political stability," Cyber defence in the age of AI, smart societies and augmented humanity, pp. 37–55, 2020.

[21] C. HETZNER, "Pentagon attack hoax illustrates investor pitfalls of A.I.-driven fake news — fortune.com," https://fortune.com/2023/05/23/twitter-elon-musk-pentagon-attack-deepfake-capital-markets-investors-deutsche-bank/, 2023, [Accessed 07-Jul-2023].

[22] Y. Aono, T. Hayashi, L. Wang, S. Moriai et al., "Privacy-preserving deep learning via additively homomorphic encryption," IEEE transactions on information forensics and security, vol. 13, no. 5, pp. 1333–1345, 2017.
[23] H. C. Tanuwidjaja, R. Choi, S. Baek, and K. Kim, "Privacy-preserving deep learning on machine learning as a service—a comprehensive survey," IEEE Access, vol. 8, pp. 167 425–167 447, 2020.
[24] Q. Zhang, C. Xin, and H. Wu, "Privacy-preserving deep learning based on multiparty secure computation: A survey," IEEE Internet of Things Journal, vol. 8, no. 13, pp. 10 412–10 429, 2021.
[25] H. Sun, T. Zhu, Z. Zhang, D. Jin, P. Xiong, and W. Zhou, "Adversarial attacks against deep generative models on data: A survey," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 4, pp. 3367–3388, 2023.
[26] W. Shahid, Y. Li, D. Staples, G. Amin, S. Hakak, and A. Ghorbani, "Are you a cyborg, bot or human?—a survey on detecting fake news spreaders," IEEE Access, vol. 10, pp. 27 069–27 083, 2022.
[27] K. Michael, R. Abbas, and G. Roussos, "Ai in cybersecurity: The paradox," IEEE Transactions on Technology and Society, vol. 4, no. 2, pp. 104–109, 2023.
[28] M. Wigan, "Cyber security and securing subjective patient quality engagements in medical applications: Ai and vulnerabilities," IEEE Transactions on Technology and Society, vol. 3, no. 3, pp. 185–188, 2022.
[29] S. P. Yadav, S. Zaidi, A. Mishra, and V. Yadav, "Survey on machine learning in speech emotion recognition and vision systems using a recurrent neural network (rnn)," Archives of Computational Methods in Engineering, vol. 29, no. 3, pp. 1753–1770, 2022.
[30] Y. K. Saheed and M. O. Arowolo, "Efficient cyber attack detection on the internet of medical things-smart environment based on deep recurrent neural network and machine learning algorithms," IEEE Access, vol. 9, pp. 161 546–161 554, 2021.
[31] A. H. Ribeiro, K. Tiels, L. A. Aguirre, and T. Schön, "Beyond exploding and vanishing gradients: analysing rnn training using attractors and smoothness," in International conference on artificial intelligence and statistics. PMLR, 2020, pp. 2370–2380.
[32] X. Yang, T. Wang, X. Ren, and W. Yu, "Survey on improving data utility in differentially private sequential data publishing," IEEE Transactions on Big Data, vol. 7, no. 4, pp. 729–749, Oct 2021.
[33] Z. Chen, M. Ma, T. Li, H. Wang, and C. Li, "Long sequence time-series forecasting with deep learning: A survey," Information Fusion, vol. 97, p. 101819, 2023.
[34] A. Jadhav, R. Jain, S. Fernandes, and S. Shaikh, "Text summarization using neural networks," in 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), 2019, pp. 1–6.
[35] C. Zheng, Z. Wang, and J. He, "Bert-based mixed question answering matching model," in 2022 11th International Conference of Information and Communication Technology (ICTech)), 2022, pp. 355–358.
[36] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023.
[37] M. M. Lau, J. T. S. Phang, and K. H. Lim, "Convolutional deep feedforward network for image classification," in 2019 7th International Conference on Smart Computing Communications (ICSCC), 2019, pp. 1–4.
[38] L. Diop and C. Ba, "Parallelization of sequential pattern sampling," in 2021 IEEE International Conference on Big Data (Big Data), 2021, pp. 5882–5884.
[39] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F.-Y. Wang, "Generative adversarial networks: introduction and outlook," IEEE/CAA Journal of Automatica Sinica, vol. 4, no. 4, pp. 588–598, 2017.
[40] R. Wei and A. Mahmood, "Recent advances in variational autoencoders with representation learning for biomedical informatics: A survey," Ieee Access, vol. 9, pp. 4939–4956, 2020.
[41] N.-T. Tran, V.-H. Tran, N.-B. Nguyen, T.-K. Nguyen, and N.-M. Cheung, "On data augmentation for gan training," IEEE Transactions on Image Processing, vol. 30, pp. 1882–1897, 2021.
[42] A. Ramesh, A. S. Rao, S. Moudgalya, and K. Srinivas, "Gan based approach for drug design," in 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2021, pp. 825–828.
[43] H. Emami, M. M. Aliabadi, M. Dong, and R. B. Chinnam, "Spa-gan: Spatial attention gan for image-to-image translation," IEEE Transactions on Multimedia, vol. 23, pp. 391–401, 2020.
[44] P. Chen, Y. Zhang, M. Tan, H. Xiao, D. Huang, and C. Gan, "Generating visually aligned sound from videos," IEEE Transactions on Image Processing, vol. 29, pp. 8292–8302, 2020.
[45] M. Hua, C. Liu, S. Yang, S. Liu, K. Fu, Z. Dong, Y. Cai, B. Zhang, and K. J. Chen, "Characterization of leakage and reliability of sin x gate dielectric by low-pressure chemical vapor deposition for gan-based mis-hemts," IEEE Transactions on Electron Devices, vol. 62, no. 10, pp. 3215–3222, 2015.
[46] X. Zhou, J. Xiong, X. Zhang, X. Liu, and J. Wei, "A radio anomaly detection algorithm based on modified generative adversarial network," IEEE Wireless Communications Letters, vol. 10, no. 7, pp. 1552–1556, 2021.
[47] S. Eiffert, K. Li, M. Shan, S. Worrall, S. Sukkarieh, and E. Nebot, "Probabilistic crowd gan: Multimodal pedestrian trajectory prediction using a graph vehicle-pedestrian attention network," IEEE Robotics and Automation Letters, vol. 5, no. 4, pp. 5026–5033, 2020.
[48] E. Hosseini-Asl, J. M. Zurada, and O. Nasraoui, "Deep learning of part-based representation of data using sparse autoencoders with nonnegativity constraints," IEEE transactions on neural networks and learning systems, vol. 27, no. 12, pp. 2486–2498, 2015.
[49] A. Xiao, J. Huang, D. Guan, X. Zhang, S. Lu, and L. Shao, "Unsupervised point cloud representation learning with deep neural networks: A survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.
[50] Y. Yang, Q. J. Wu, and Y. Wang, "Autoencoder with invertible functions for dimension reduction and image reconstruction," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 48, no. 7, pp. 1065–1079, 2016.
[51] L. Gonog and Y. Zhou, "A review: generative adversarial networks," in 2019 14th IEEE conference on industrial electronics and applications (ICIEA). IEEE, 2019, pp. 505–510.
[52] C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (medical) time series generation with recurrent conditional gans," arXiv preprint arXiv:1706.02633, 2017.
[53] Z. Pan, W. Yu, B. Wang, H. Xie, V. S. Sheng, J. Lei, and S. Kwong, "Loss functions of generative adversarial networks (gans): Opportunities and challenges," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 4, no. 4, pp. 500–522, 2020.
[54] S. Bernard, P. Bas, J. Klein, and T. Pevny, "Explicit optimization of min max steganographic game," IEEE Transactions on Information Forensics and Security, vol. 16, pp. 812–823, 2020.
[55] F. Ye and A. G. Bors, "Lifelong mixture of variational autoencoders," IEEE Transactions on Neural Networks and Learning Systems, 2021.
[56] Q. Zhou, C. Du, D. Li, H. Wang, J. K. Liu, and H. He, "Neural encoding and decoding with a flow-based invertible generative model," IEEE Transactions on Cognitive and Developmental Systems, 2022.
[57] C. P. Chen and S. Feng, "Generative and discriminative fuzzy restricted boltzmann machine learning for text and image classification," IEEE transactions on cybernetics, vol. 50, no. 5, pp. 2237–2248, 2018.
[58] P. Zhong, Z. Gong, S. Li, and C.-B. Schönlieb, "Learning to diversify deep belief networks for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 6, pp. 3516–3530, 2017.
[59] P. Radanliev and D. De Roure, "Review of algorithms for artificial intelligence on low memory devices," IEEE Access, vol. 9, pp. 109 986–109 993, 2021.
[60] Y. C. Subakan and P. Smaragdis, "Generative adversarial source separation," in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018, pp. 26–30.
[61] X.-H. Li, C. C. Cao, Y. Shi, W. Bai, H. Gao, L. Qiu, C. Wang, Y. Gao, S. Zhang, X. Xue, and L. Chen, "A survey of data-driven and knowledge-aware explainable ai," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 1, pp. 29–49, 2022.
[62] A. Rawal, J. McCoy, D. B. Rawat, B. M. Sadler, and R. S. Amant, "Recent advances in trustworthy explainable artificial intelligence: Status, challenges, and perspectives," IEEE Transactions on Artificial Intelligence, vol. 3, no. 6, pp. 852–866, 2022.
[63] S. De, M. Bermudez-Edo, H. Xu, and Z. Cai, "Deep generative models in the industrial internet of things: A survey," IEEE Transactions on Industrial Informatics, vol. 18, no. 9, pp. 5728–5737, 2022.
[64] M. Casillo, F. Colace, B. B. Gupta, A. Lorusso, F. Marongiu, and D. Santaniello, "A deep learning approach to protecting cultural heritage buildings through iot-based systems," in 2022 IEEE International Conference on Smart Computing (SMARTCOMP), 2022, pp. 252–256.
[65] H. Benaddi, M. Jouhari, K. Ibrahimi, A. Benslimane, and E. M. Amhoud, "Adversarial attacks against iot networks using conditional gan based learning," in GLOBECOM 2022 - 2022 IEEE Global Communications Conference, 2022, pp. 2788–2793.

[66] L. Ouyang, Y. Yuan, and F.-Y. Wang, "Learning markets: An ai collaboration framework based on blockchain and smart contracts," IEEE Internet of Things Journal, vol. 9, no. 16, pp. 14 273–14 286, 2022.

[67] T. N. Dinh and M. T. Thai, "Ai and blockchain: A disruptive integration," Computer, vol. 51, no. 9, pp. 48–53, 2018.

[68] R. R, A. S, A. N, and V. Vivek, "A survey on advanced text recognition and projection in augmented reality," in 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 2022, pp. 1085–1089.

[69] N. Taimoor and S. Rehman, "Reliable and resilient ai and iot-based personalised healthcare services: A survey," IEEE Access, vol. 10, pp. 535–563, 2022.

[70] S. Baker and W. Xiang, "Artificial intelligence of things for smarter healthcare: A survey of advancements, challenges, and opportunities," IEEE Communications Surveys Tutorials, vol. 25, no. 2, pp. 1261–1293, 2023.

[71] M. Kuzlu, Z. Xiao, S. Sarp, F. O. Catak, N. Gurler, and O. Guler, "The rise of generative artificial intelligence in healthcare," in 2023 12th Mediterranean Conference on Embedded Computing (MECO), 2023, pp. 1–4.

[72] J. Qadir, "Engineering education in the era of chatgpt: Promise and pitfalls of generative ai for education," in 2023 IEEE Global Engineering Education Conference (EDUCON), 2023, pp. 1–9.

[73] V. Volz, N. Justesen, S. Snodgrass, S. Asadi, S. Purmonen, C. Holmgård, J. Togelius, and S. Risi, "Capturing local and global patterns in procedural content generation via machine learning," in 2020 IEEE Conference on Games (CoG), 2020, pp. 399–406.

[74] Z. Chen, L. Chen, Z. Zhao, and Y. Wang, "Ai illustrator: Art illustration generation based on generative adversarial network," in 2020 IEEE 5th International Conference on Image, Vision and Computing (ICIVC), 2020, pp. 155–159.

[75] S. S. Nasrin and R. I. Rasel, "Hennagan: Henna art design generation using deep convolutional generative adversarial network (dcgan)," in 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), 2020, pp. 196–199.

[76] J. Betz, H. Zheng, A. Liniger, U. Rosolia, P. Karle, M. Behl, V. Krovi, and R. Mangharam, "Autonomous vehicles on the edge: A survey on autonomous vehicle racing," IEEE Open Journal of Intelligent Transportation Systems, vol. 3, pp. 458–488, 2022.

[77] Y. Fang, H. Min, W. Wang, Z. Xu, and X. Zhao, "A fault detection and diagnosis system for autonomous vehicles based on hybrid approaches," IEEE Sensors Journal, vol. 20, no. 16, pp. 9359–9371, Aug 2020.

[78] N. R. C. Monteiro, B. Ribeiro, and J. P. Arrais, "Drug-target interaction prediction: End-to-end deep learning approach," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, no. 6, pp. 2364–2374, Nov 2021.

[79] B. Ru, D. Li, Y. Hu, and L. Yao, "Serendipity—a machine-learning application for mining serendipitous drug usage from social media," IEEE Transactions on NanoBioscience, vol. 18, no. 3, pp. 324–334, July 2019.

[80] N. Waqas, S. I. Safie, K. A. Kadir, S. Khan, and M. H. Kaka Khel, "Deepfake image synthesis for data augmentation," IEEE Access, vol. 10, pp. 80 847–80 857, 2022.

[81] P. Korshunov and S. Marcel, "Improving generalization of deepfake detection with data farming and few-shot learning," IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 4, no. 3, pp. 386–397, July 2022.

[82] H. Khalid and S. S. Woo, "Oc-fakedect: Classifying deepfakes using one-class variational autoencoder," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 2020, pp. 2794–2803.

[83] M. L. Jones, E. Kaufman, and E. Edenberg, "Ai and the ethics of automating consent," IEEE Security Privacy, vol. 16, no. 3, pp. 64–72, May 2018.

[84] S. Java, F. L. Basheer, S. Riaz, M. J. Kaur, and A. Mushtaq, "Detection of online manipulation to prevent users victimization," in 2019 Amity International Conference on Artificial Intelligence (AICAI), Feb 2019, pp. 593–599.

[85] Y. Wu, J. Weng, Z. Wang, K. Wei, J. Wen, J. Lai, and X. Li, "Attacks and countermeasures on privacy-preserving biometric authentication schemes," IEEE Transactions on Dependable and Secure Computing, vol. 20, no. 2, pp. 1744–1755, March 2023.

[86] S. Li and W. Deng, "Deep facial expression recognition: A survey," IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1195–1215, July 2022.

[87] L. Li, X. Mu, S. Li, and H. Peng, "A review of face recognition technology," IEEE Access, vol. 8, pp. 139 110–139 120, 2020.

[88] C. Huang, Z. Zhang, B. Mao, and X. Yao, "An overview of artificial intelligence ethics," IEEE Transactions on Artificial Intelligence, vol. 4, no. 4, pp. 799–819, Aug 2023.

[89] C.-Y. Yeh, H.-W. Chen, S.-L. Tsai, and S.-D. Wang, "Disrupting image-translation-based deepfake algorithms with adversarial attacks," in 2020 IEEE Winter Applications of Computer Vision Workshops (WACVW), 2020, pp. 53–62.

[90] D. Peters, K. Vold, D. Robinson, and R. A. Calvo, "Responsible ai—two frameworks for ethical design practice," IEEE Transactions on Technology and Society, vol. 1, no. 1, pp. 34–47, March 2020.

[91] J. Yoon, L. N. Drumright, and M. Van Der Schaar, "Anonymization through data synthesis using generative adversarial networks (ads-gan)," IEEE journal of biomedical and health informatics, vol. 24, no. 8, pp. 2378–2388, 2020.

[92] W.-S. Lee, A. John, H.-C. Hsu, and P.-A. Hsiung, "Spchain: A smart and private blockchain-enabled framework for combining gdpr-compliant digital assets management with ai models," IEEE Access, vol. 10, pp. 130 424–130 443, 2022.

[93] W. Stallings, "Handling of personal information and deidentified, aggregated, and pseudonymized information under the california consumer privacy act," IEEE Security & Privacy, vol. 18, no. 1, pp. 61–64, 2020.

[94] C. Li, "A repeated call for omnibus federal cybersecurity law," Notre Dame L. Rev., vol. 94, p. 2211, 2018.

[95] D. S. Guamán, J. M. Del Alamo, and J. C. Caiza, "Gdpr compliance assessment for cross-border personal data transfers in android apps," IEEE Access, vol. 9, pp. 15 961–15 982, 2021.

[96] D. S. Ong, C. S. Chan, K. W. Ng, L. Fan, and Q. Yang, "Protecting intellectual property of generative adversarial networks from ambiguity attacks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3630–3639.

[97] J. Fei, Z. Xia, B. Tondi, and M. Barni, "Supervised gan watermarking for intellectual property protection," in 2022 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE, 2022, pp. 1–6.

[98] W. Xia, T. Li, and C. Li, "A review of scientific impact prediction: tasks, features and methods," Scientometrics, vol. 128, no. 1, pp. 543–585, 2023.

[99] S. Chakraborty, "Intellectual property in the era of generative ai," Jun 2023. [Online]. Available: https://www.analyticsinsight.net/intellectual-property-in-the-era-of-generative/#:~:text=In%20response%20to%20the%20challenges,them%20certain%20rights%20and%20protections.

[100] Z. Cai, "Usage of deep learning and blockchain in compilation and copyright protection of digital music," Ieee Access, vol. 8, pp. 164 144–164 154, 2020.

[101] Y.-C. F. Wang, "Generative ai: How it changes our lives? take vision & language as an example," in 2023 International VLSI Symposium on Technology, Systems and Applications (VLSI-TSA/VLSI-DAT). IEEE, 2023, pp. 1–1.

[102] I. K. Dutta, B. Ghosh, A. Carlson, M. Totaro, and M. Bayoumi, "Generative adversarial networks in security: a survey," in 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE, 2020, pp. 0399–0405.

[103] H. C. Tanuwidjaja, R. Choi, S. Baek, and K. Kim, "Privacy-preserving deep learning on machine learning as a service—a comprehensive survey," IEEE Access, vol. 8, pp. 167 425–167 447, 2020.

[104] Y. Liu, X. Yuan, Z. Xiong, J. Kang, X. Wang, and D. Niyato, "Federated learning for 6g communications: Challenges, methods, and future directions," China Communications, vol. 17, no. 9, pp. 105–118, 2020.

[105] S. Li, S. Zhao, G. Min, L. Qi, and G. Liu, "Lightweight privacy-preserving scheme using homomorphic encryption in industrial internet of things," IEEE Internet of Things Journal, vol. 9, no. 16, pp. 14 542–14 550, 2021.

[106] B. Jiang, J. Li, G. Yue, and H. Song, "Differential privacy for industrial internet of things: Opportunities, applications, and challenges," IEEE Internet of Things Journal, vol. 8, no. 13, pp. 10 430–10 451, 2021.

[107] F. Ding, K. Yu, Z. Gu, X. Li, and Y. Shi, "Perceptual enhancement for autonomous vehicles: Restoring visually degraded images for context prediction via adversarial training," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 7, pp. 9430–9441, 2021.

[108] Y. Zhu, Y. Chen, X. Li, K. Chen, Y. He, X. Tian, B. Zheng, Y. Chen, and Q. Huang, "Toward understanding and boosting adversarial transferability from a distribution perspective," IEEE Transactions on Image Processing, vol. 31, pp. 6487–6501, 2022.

[109] S. Dietzel, E. Schoch, B. Konings, M. Weber, and F. Kargl, "Resilient secure aggregation for vehicular networks," IEEE network, vol. 24, no. 1, pp. 26–31, 2010.

[110] H. Fereidooni, S. Marchal, M. Miettinen, A. Mirhoseini, H. Möllering, T. D. Nguyen, P. Rieger, A.-R. Sadeghi, T. Schneider, H. Yalame et al., "Safelearn: Secure aggregation for private federated learning," in 2021 IEEE Security and Privacy Workshops (SPW). IEEE, 2021, pp. 56–62.

[111] D. C. Nguyen, M. Ding, Q.-V. Pham, P. N. Pathirana, L. B. Le, A. Seneviratne, J. Li, D. Niyato, and H. V. Poor, "Federated learning meets blockchain in edge computing: Opportunities and challenges," IEEE Internet of Things Journal, vol. 8, no. 16, pp. 12 806–12 825, 2021.

[112] A. F. Winfield, S. Booth, L. A. Dennis, T. Egawa, H. Hastie, N. Jacobs, R. I. Muttram, J. I. Olszewska, F. Rajabiyazdi, A. Theodorou et al., "Ieee p7001: A proposed standard on transparency," Frontiers in Robotics and AI, vol. 8, p. 665729, 2021.

[113] A. Bhaskara, M. Skinner, and S. Loft, "Agent transparency: A review of current theory and evidence," IEEE Transactions on Human-Machine Systems, vol. 50, no. 3, pp. 215–224, 2020.

[114] F. Rajabiyazdi and G. A. Jamieson, "A review of transparency (seeing-into) models," in 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2020, pp. 302–308.

[115] T. Hafeez, L. Xu, and G. Mcardle, "Edge intelligence for data handling and predictive maintenance in iiot," IEEE Access, vol. 9, pp. 49 355–49 371, 2021.

[116] S. Wang, C. Li, and A. Lim, "Rophs: Determine real-time status of a multi-carriage logistics train at airport," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 7, pp. 6347–6356, 2021.

[117] H. Cai, B. Xu, L. Jiang, and A. V. Vasilakos, "Iot-based big data storage systems in cloud computing: perspectives and challenges," IEEE Internet of Things Journal, vol. 4, no. 1, pp. 75–87, 2016.

[118] J. Yoon, L. N. Drumright, and M. Van Der Schaar, "Anonymization through data synthesis using generative adversarial networks (ads-gan)," IEEE journal of biomedical and health informatics, vol. 24, no. 8, pp. 2378–2388, 2020.

[119] A. Majeed and S. O. Hwang, "When ai meets information privacy: The adversarial role of ai in data sharing scenario," IEEE Access, 2023.

[120] P. S. Chauhan and N. Kshetri, "2021 state of the practice in data privacy and security," Computer, vol. 54, no. 8, pp. 125–132, 2021.

[121] C. Benzaid and T. Taleb, "Zsm security: Threat surface and best practices," IEEE Network, vol. 34, no. 3, pp. 124–133, 2020.

[122] M. Gupta, C. Akiri, K. Aryal, E. Parker, and L. Praharaj, "From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy," IEEE Access, 2023.

[123] C. Benzaïd and T. Taleb, "Ai for beyond 5g networks: a cyber-security defense or offense enabler?" IEEE network, vol. 34, no. 6, pp. 140–147, 2020.

[124]

[125] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, "Deepfake detection: A systematic literature review," IEEE access, vol. 10, pp. 25 494–25 513, 2022.

[126] A. Dunmore, J. Jang-Jaccard, F. Sabrina, and J. Kwak, "A comprehensive survey of generative adversarial networks (gans) in cybersecurity intrusion detection," IEEE Access, 2023.

[127] H. Lee, M. U. Kim, Y. Kim, H. Lyu, and H. J. Yang, "Development of a privacy-preserving uav system with deep learning-based face anonymization," IEEE Access, vol. 9, pp. 132 652–132 662, 2021.

[128] G. Hurlburt, "What if ethics got in the way of generative ai?" IT Professional, vol. 25, no. 2, pp. 4–6, 2023.

[129] C. Li, B. Cheang, Z. Luo, and A. Lim, "An exponential factorization machine with percentage error minimization to retail sales forecasting," ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 15, no. 2, pp. 1–32, 2021.

[130] G. Leenders, B. Human, I. Apr, F. Source, and I. Policy, "The regulation of artificial intelligence—a case study of the partnership on ai," Becoming Human: Artificial Intelligence Magazine, vol. 13, 2019.

[131] M. Jovanovic and M. Campbell, "Generative artificial intelligence: Trends and prospects," Computer, vol. 55, no. 10, pp. 107–112, 2022.

[132] H. Suryotrisongko and Y. Musashi, "Review of cybersecurity research topics, taxonomy and challenges: Interdisciplinary perspective," in 2019 IEEE 12th conference on service-oriented computing and applications (SOCA). IEEE, 2019, pp. 162–167.

[133] G. Joshi, R. Walambe, and K. Kotecha, "A review on explainability in multimodal deep neural nets," IEEE Access, vol. 9, pp. 59 800–59 821, 2021.

[134] J. Salgado-Criado and C. Fernández-Aller, "A wide human-rights approach to artificial intelligence regulation in europe," IEEE Technology and Society Magazine, vol. 40, no. 2, pp. 55–65, 2021.

[135] Z. Wu, C. Li, M. Li, and A. Lim, "Inertial proximal gradient methods with bregman regularization for a class of nonconvex optimization problems," Journal of Global Optimization, vol. 79, pp. 617–644, 2021.

• • •

ABENEZER GOLDA is currently pursuing B.Tech. degree from Kalinga Institute of Industrial Technology (KIIT), Bhubaneshwar. He has completed a few projects in Machine Learning, Deep Learning, and Data Science. He is currently (the summer of 2023) pursuing his research internship at the Birla Institute of Technology and Science (BITS), Pilani under Dr. Vikas Hassija. His research interests include machine learning, reinforcement learning, deep learning, Privacy Preservation, and Data Science.

KIDUS MEKONNEN is currently pursuing B.Tech. degree from Kalinga Institute of Industrial Technology (KIIT), Bhubaneshwar. He is currently (the summer of 2023) pursuing his research internship at the Birla Institute of Technology and Science (BITS), Pilani under Dr. Vikas Hassija. His research interests include machine learning, reinforcement learning, quantum computing, and deep learning.

AMIT PANDEY is a research scholar in the computer science department in Bennett University Greater Noida India. His research areas are semantic segmentation for biomedical image segmentation and explainable artificial intelligence. He completed MTech (Computer Science) from Deenbandhu ChotuRam University of Science and Technology, Sonipat, Haryana, India.

ANUSHKA SINGH is currently pursuing B.Tech. degree from Kalinga Institute of Industrial Technology (KIIT), Bhubaneshwar. She is currently (the summer of 2023) pursuing her research internship at the Birla Institute of Technology and Science (BITS), Pilani under Dr. Vikas Hassija. Her research interests include machine learning, reinforcement learning, quantum computing, and deep learning.

VIKAS HASSIJA is currently working is a post-doc researcher at the National University of Singapore, Singapore. He received the B.Tech. degree from M.D.U University, Rohtak, India, in 2010, and the M.S. degree in telecommunications and software engineering from the Birla Institute of Technology and Science (BITS), Pilani, India, in 2014. He received his Ph.D. degree in IoT security and blockchain from the Jaypee Institute of Information and technology (JIIT), Noida. He has also worked as an Assistant Professor with JIIT for 4 years. He has eight years of industry experience and has worked with various telecommunication companies like Tech Mahindra and Accenture. His research interests include the IoT security, network security, blockchain,and distributed computing.

VINAY CHAMOLA (Fellow, IET) received the B.E. degree in electrical and electronics engineering and master's degree in communication engineering from the Birla Institute of Technology and Science, Pilani, India, in 2010 and 2013, respectively. He received his Ph.D. degree in electrical and computer engineering from the National University of Singapore, Singapore, in 2016. In 2015, he was a Visiting Researcher with the Autonomous Networks Research Group (ANRG), University of Southern California, Los Angeles, CA, USA. He also worked as a post-doctoral research fellow at the National University of Singapore, Singapore. He is currently an Associate Professor with the Department of Electrical and Electronics Engineering, BITS-Pilani, Pilani, where he heads the Internet of Things Research Group / Lab. His research interests include IoT Security, Blockchain, UAVs, VANETs, 5G, and Healthcare. He serves as an Area Editor for the Ad Hoc Networks Journal, Elsevier and the IEEE Internet of Things Magazine. He also serves as an Associate Editor in the IEEE Transactions on Intelligent Transportation Systems, IEEE Networking Letters, IEEE Consumer electronics magazine, IET Quantum Communications, IET Networks, and several other journals. He serves as co-chair of various reputed workshops like IEEE Globecom Workshop 2021, IEEE INFOCOM 2022 workshop, IEEE ANTS 2021, and IEEE ICIAfS 2021, to name a few. He is listed in the World's Top 2% Scientists identified by Stanford University. He is co-founder and President of a healthcare startup Medsupervision Pvt. Ltd. He is a senior member of the IEEE.

BIPLAB SIKDAR (Senior Member, IEEE) received the B.Tech. degree in electronics and communication engineering from North Eastern Hill University, Shillong, India, in 1996, the M.Tech. degree in electrical engineering from the Indian Institute of Technology Kanpur, Kanpur, India, in 1998, and the Ph.D. degree in electrical engineering from the Rensselaer Polytechnic Institute, Troy, NY, USA, in 2001. He was a Faculty with the Rensselaer Polytechnic Institute, from 2001 to 2013, an Assistant Professor and an Associate Professor. He is currently a Professor and Head of Department of the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. His current research interests include wireless networks, and security for Internet of Things and cyber physical systems. He has served as an Associate Editor for the IEEE Transactions on Communications, IEEE Transactions on Mobile Computing, IEEE Internet of Things Journal and IEEE Open Journal of Vehicular Technology.