Exploring Dual-Use Risks and Sustainable Deployment of Large Language Models: Advances in Generative AI for Cybersecurity, Human-Robot Interaction, and Environmental Management

Mrs. Poonam M Ramgirwar[1], Ms. Priya P. Borade[2], Ashwini Shinde[3]
[1,2] Assistant professor, JNEC, MGMU
[3] Teaching assistant, MGMU JNEC

Abstract—**This paper provides a comprehensive examination of the advancements, challenges, and implications associated with large language models (LLMs) and generative artificial intelligence (AI) across multiple application domains. Emphasizing both the beneficial and potentially harmful uses of these technologies, this research surveys recent literature in cybersecurity, human-robot interaction, synthetic data augmentation, environmental impact mitigation, and AI model alignment. The study highlights critical issues such as dual-use risks of LLMs, adversarial vulnerabilities, privacy concerns, computational sustainability, and the integration of new hardware architectures. Drawing from a diverse set of case studies and empirical findings, the analysis underscores the importance of responsible AI deployment practices, including transparency, explainability, and regional workload management. The paper further explores cutting-edge applications ranging from robotic artistic creation to personalized fashion recommendations and multimedia content generation. Recommendations are provided to guide future research focused on enhancing security, ethical governance, and ecological sustainability in generative AI development. Ultimately, this work serves as a pivotal reference for researchers and practitioners aiming to harness the transformative potential of LLMs while addressing their complex technical and societal challenges.**

I. INTRODUCTION

The rapid expansion of artificial intelligence (AI), especially large language models and generative AI techniques, has profoundly reshaped the digital and physical worlds in just a few short years. From chatbots delivering personalized content to cutting-edge systems capable of creating photorealistic images and complex multimedia, the boundaries of possibility have shifted dramatically. Today, the influence of these models extends far beyond scientific laboratories and corporate research centers, reaching into the everyday experiences of society. Notably, the ability of LLMs to understand, generate, and manipulate natural language and images has led to breakthroughs in fields as diverse as healthcare, cybersecurity, entertainment, education, and environmental monitoring. Against this backdrop, it is more crucial than ever to examine the complex mechanisms, social consequences, and technical risks intertwined with the rapid adoption of such technologies. Researchers, engineers, and members of the public are increasingly engaged with both the transformative potential and the nuanced limitations of AI systems. While much attention has focused on their capacity for automation and creativity, far less is widely understood about the challenges of security, data privacy, and sustainability. The prevalence of synthetic content, from benign entertainment to high-risk scenarios like deepfakes, has generated reasoned debate about the role of explainability, human oversight, and policy frameworks. Emerging evidence suggests that the power and pervasiveness of these models demand a reconsideration of best practices and the development of entirely new methodologies to ensure AI is deployed responsibly. Given the pace of innovation, society cannot afford to rely on traditional paradigms alone for evaluating risk and safeguarding integrity. Therefore, a comprehensive and multidisciplinary understanding is required to navigate this rapidly changing landscape.

LLMs and generative AI systems have sparked both awe and apprehension among stakeholders due to their vast capabilities and opaque inner workings. On one hand, they represent a leap forward in fields requiring language understanding, reasoning, text summarization, and even artistic creation. On the other, their dual-use nature allows them to serve as tools for malicious activity as well as significant social good. For example, advancements in generative models underpin new forms of medical image analysis, enhancing the early detection of diseases or skin cancer through augmentation of rare or difficult-to-obtain datasets. Meanwhile, the very strengths that make these models desirable can be exploited in adversarial settings: the creation of spear-phishing texts, automated vulnerabilities analysis, and even the synthesis of malware. Organizations have found themselves needing to defend against highly sophisticated, machine-powered threats that mirror or even outstrip human ingenuity. Forward-thinking research, such as surveys that systematically explore both threats and defenses, is necessary to develop robust frameworks for mitigating dual-use risks. Collaboration between academia, industry, and governmental agencies is growing in urgency as the costs associated with cyber incidents rise in tandem with the sophistication of adversarial AI. At the core is a pressing need for comprehensive surveys that aggregate the state of knowledge, analyze proliferation trends, and recommend effective countermeasures.

At the intersection of technical innovation and practical implementation is an ongoing dialogue about data privacy, model transparency, and ethical constraints. Generative AI platforms, such as those developed by OpenAI, Anthropic, Google, Meta, and others, rely on massive volumes of data often collected from user interactions, internet repositories, and proprietary datasets. The tension between leveraging vast data resources for improved model accuracy and safeguarding the privacy rights of individuals is a recurring theme in contemporary discourse. Regulatory attention is increasingly focused on questions of data provenance, informed consent, and usage restrictions, especially as AI systems are embedded in sensitive applications across healthcare, finance, and national security. Annotation and curation of data require meticulous oversight, lest bias and discrimination inadvertently perpetuate in algorithmic outcomes. At the same time, recent advances in explainable AI (XAI) aim to shed light on the opaque decision-making processes of deep neural architectures. These transparency initiatives not only improve trust and user acceptance but are also a crucial defense against unintended or adversarial exploitation. The development of frameworks and toolkits for adversarial robustness further illustrates the drive to build AI systems that balance innovation with responsibility. An in-depth exploration of these issues remains essential for supporting safe and effective adoption of generative AI and LLMs.

We will present graphically a clear overview of the foundational concepts introduced in the research paper's introduction. It starts with the core technologies of large language models (LLMs) and generative artificial intelligence, highlighting their growing importance. Surrounding this core are the diverse application areas such as cybersecurity, healthcare, human-robot interaction, and environmental monitoring, showcasing the breadth of impact these models have across sectors. The framework also identifies critical challenges including dual-use risks, explainability, privacy, and environmental sustainability, which emphasize the complexity and responsibility involved in deploying these technologies. The diagram further illustrates technological advancements addressing these challenges, like model watermarking, carbon-aware computation, and federated learning, which enhance performance and security. Finally, it points toward future research priorities such as responsible deployment, cross-industry collaboration, and sustainable AI, underscoring the field's commitment to ethical progress and societal benefit. By visually organizing these interconnected themes, the diagram helps readers quickly grasp the research context and the multidisciplinary scope of current generative AI development.

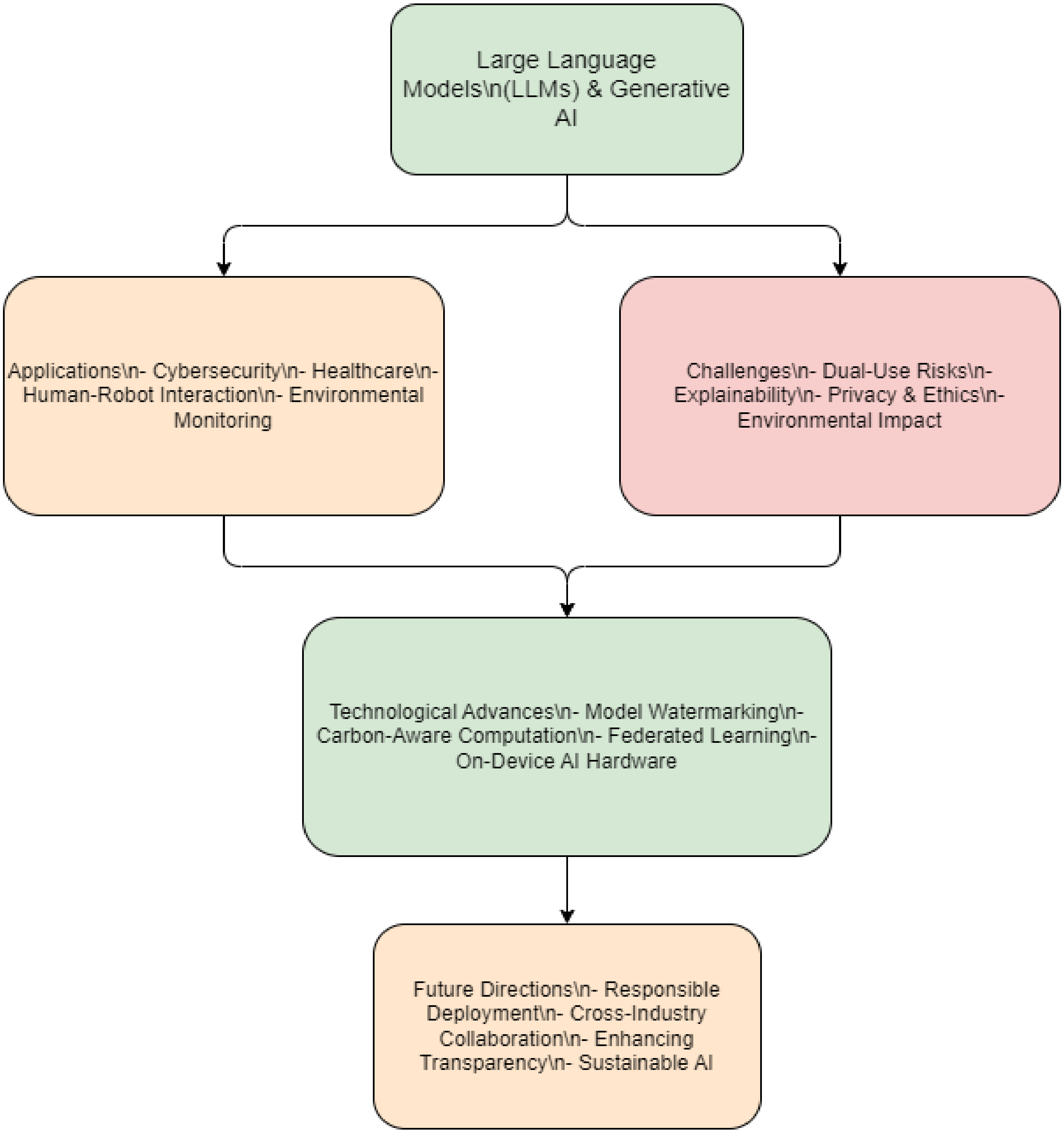
Large Language Models\n(LLMs) & Generative AI
Applications\n- Cybersecurity\n- Healthcare\n- Human-Robot Interaction\n- Environmental Monitoring
Challenges\n- Dual-Use Risks\n- Explainability\n- Privacy & Ethics\n- Environmental Impact
Technological Advances\n- Model Watermarking\n- Carbon-Aware Computation\n- Federated Learning\n- On-Device AI Hardware
Future Directions\n- Responsible Deployment\n- Cross-Industry Collaboration\n- Enhancing Transparency\n- Sustainable AI

Figure 1 Conceptual Framework of Large Language Models and Generative AI: Applications, Challenges, and Future Directions

A compelling aspect of modern generative models is their application in domains traditionally considered outside the scope of machine intelligence, such as fine art, creative writing, and human-robot interaction. For instance, recent work demonstrates how robotic arms, guided by LLMs and diffusion-based vision models, can create intricate drawings, merging algorithmic comprehension with embodied execution. These interactive systems exemplify the frontier of AI-powered creativity, providing new modes of collaboration between humans and machines. Fine-tuning language models to respond to conversational cues and channel them into visual or physical expressions opens up a wealth of opportunities, from assistive devices for artists to innovative educational tools. The challenge lies not only in training models with the appropriate supervision and prompt engineering but also in ensuring that their outputs remain under meaningful human control. As AI systems become more deeply enmeshed in co-creative processes, the dialogue around agency, authorship, and ownership becomes ever more significant. Additionally, advances in explainable generative AI are crucial to provide artists, educators, and end-users with interpretability, thereby maintaining transparency and fostering broader acceptance. These multifaceted applications highlight the need for a robust and ethically grounded research agenda.

Few topics garner as much attention in the AI community today as the environmental sustainability of large-scale models. The proliferation of LLMs and generative engines, trained and deployed using considerable computational resources, has led to mounting concern regarding energy consumption and carbon emissions. Recent analyses indicate that inference activities in serving user requests at scale often exceed the energy demands of model training, particularly as the user base for AI-powered services continues to expand globally. Regional disparities in the carbon intensity of data centers have been cited as an area of opportunity for emissions mitigation. Techniques such as carbon-aware workload shifting, whereby requests are dynamically routed to regions with greener energy grids, are emerging as practical responses to the environmental impact of AI. Empirical studies leveraging simulation data and real-world benchmarks suggest that strategic workload management could yield significant reductions in emissions. These initiatives reinforce the dual imperative to not only maximize utility and accessibility but also to proactively minimize ecological footprints. As regulatory scrutiny grows and environmental benchmarks become more stringent, novel frameworks for sustainable AI deployment will be crucial components of future infrastructure planning.

Another frontier in the deployment of LLMs and diffusion models is the augmentation of datasets for specialized machine learning applications. Data scarcity, particularly in the context of protected or sensitive domains, presents a persistent barrier to progress. In response, generative models have demonstrated effectiveness in synthesizing photorealistic data for training and evaluation. This is especially salient for applications involving human faces, rare disease detection, or underrepresented minorities, where privacy regulations and limited specimen availability constrain conventional data collection efforts. By incorporating advanced techniques such as ControlNet-guided conditioning and prompt-driven augmentation, researchers can curate datasets that capture diversity across demographic, environmental, and technical parameters. The open-sourcing of such datasets, under rigorous privacy and ethical guidelines, further expands opportunities for collaboration and benchmarking. Critically, the alignment of generative models with privacy standards is a topic of ongoing debate, underscoring an urgent need for transparent practices and robust validation. These developments signal the emergence of synthesis-driven approaches to overcoming foundational challenges in AI research and application.

In the realm of cybersecurity, the implications of generative AI and LLMs are multifaceted and profound. Defensive measures historically designed to detect signature-based threats are increasingly inadequate against AI-generated attacks, which can adapt and proliferate at machine speed. Adaptive malware, deepfake-driven social engineering, and advanced phishing techniques are now capable of evading filters through continual mutation and reinvention. Researchers have forecast that AI-generated malware will constitute a majority of new threats within the coming years, necessitating a step change in both threat assessment tools and cyber hygiene protocols. To combat these novel risks, the field is witnessing the integration of LLM-enhanced detection systems, real-time anomaly monitoring, and the development of explainable intrusion response mechanisms. Partnerships encompassing cloud infrastructure providers, endpoint security vendors, and standards organizations are beginning to coalesce around common priorities. Recommendations for responsible deployment ranging from watermarking and traceability to red-teaming and adversarial simulation are gaining momentum as best practices. The evolving threat landscape compels both industry leaders and researchers to remain proactive and adaptive in their defensive strategies.

The evolution of hardware and specialized neural processors is another critical dimension impacting the performance and efficiency of generative models. Modern mobile and edge devices are now increasingly equipped with neural processing units (NPUs) capable of supporting both transformer-based LLMs and visual generative models directly on-device. The migration from cloud-centric architectures to distributed, heterogeneous systems fundamentally changes the dynamics of model deployment, enabling faster, privacy-preserving, and context-aware computations. Innovations in wafer-level packaging, memory hierarchies, and thermal management are central to supporting the compute-intensive workloads of generative AI while maintaining acceptable power consumption profiles. Furthermore, the distinction between compute-intensive visual models and memory-bounded language models continues to drive architectural specialization. As more devices integrate these advanced processing units, the proliferation of AI becomes further democratized, opening up applications ranging from real-time image synthesis to autonomous robotics. Continued research and investment in next-generation hardware are vital for sustaining the pace of innovation and ensuring equitable access to advanced AI capabilities.

An essential challenge for both the research and practitioner communities lies in aligning generative model outputs with human preferences and real-world requirements. While reinforcement learning from human feedback (RLHF) has become a staple for language model alignment, equivalent approaches for image-based diffusion models are still nascent. Recent proposals, such as optimizing diffusion models directly on human comparison data using Direct Preference Optimization (DPO), are showing promise in bridging this gap. The capacity to generate images and content that are not only realistic but also precisely aligned with user intent marks a significant leap forward. These advances facilitate improvements in text-to-image alignment, visual appeal, and subjective satisfaction, thereby enabling more user-centric applications. The rigorous evaluation of these techniques against large, crowdsourced datasets is crucial for establishing benchmarks and validating progress. Moreover, the advent of AI-based feedback mechanisms further extends the possibilities for scalable and automated alignment. As the demand for personalized AI applications grows, the methodologies to align, fine-tune, and evaluate generative models remain a focal point of ongoing research.

A notable area of innovation is the integration of generative models into real-time, edge-enabled data processing for environmental and industrial applications. For example, in marine litter detection, the combination of diffusion models and LLMs has enabled the augmentation of datasets that improve detection accuracy for rare or underrepresented waste categories. Such improvements are of critical importance in the pursuit of sustainable and effective monitoring systems in oceans, rivers, and other vulnerable ecosystems. The use of scalable, high-performance computing resources allows for the efficient generation of training samples and real-time model updates. Object detection frameworks, like YOLOv8, have benefited from this synthetic data, demonstrating measurable improvements in recall and mean average precision. Importantly, the performance of these solutions is validated using videos and data captured in dynamic, real-world scenarios, highlighting their practical value. The implications extend beyond environmental science, offering templates for similar strategies in agriculture, urban monitoring, and other resource-constrained domains.

II. A SYSTEMATIC MODEL FOR GENERATIVE AI CYBERSECURITY: RISKS, DEFENCES, AND POLICY ROADMAP

The framework in Figure 1 (below) presents a structured workflow for analysing cybersecurity risks in generative AI systems and developing effective defence strategies. It begins with the collection of diverse data sources, including system logs, curated datasets, and platform feeds, complemented by an extensive review of existing research, reports, and case studies. These inputs support the identification of dual-use concerns, such as the misuse of large language models for phishing or the generation of malicious code. Threat modelling is then employed to explore potential attack vectors, including zero-day exploitation and automated reconnaissance, providing insight into how adversaries may leverage AI. In response, the framework outlines defensive strategies such as watermarking, adversarial training, and explainable AI to counter emerging threats. These approaches are tested through benchmarking and real-world evaluation to assess their robustness and practicality. The findings culminate in a set of recommendations and a strategic roadmap that integrates technical safeguards with policy guidance and collaborative initiatives. Case studies of real-world incidents run parallel to this workflow, grounding the analysis in practice, while a feedback loop ensures that lessons learned inform future data collection and policy adjustments. This iterative process enables the framework to remain adaptive in addressing the evolving cybersecurity challenges posed by generative AI.

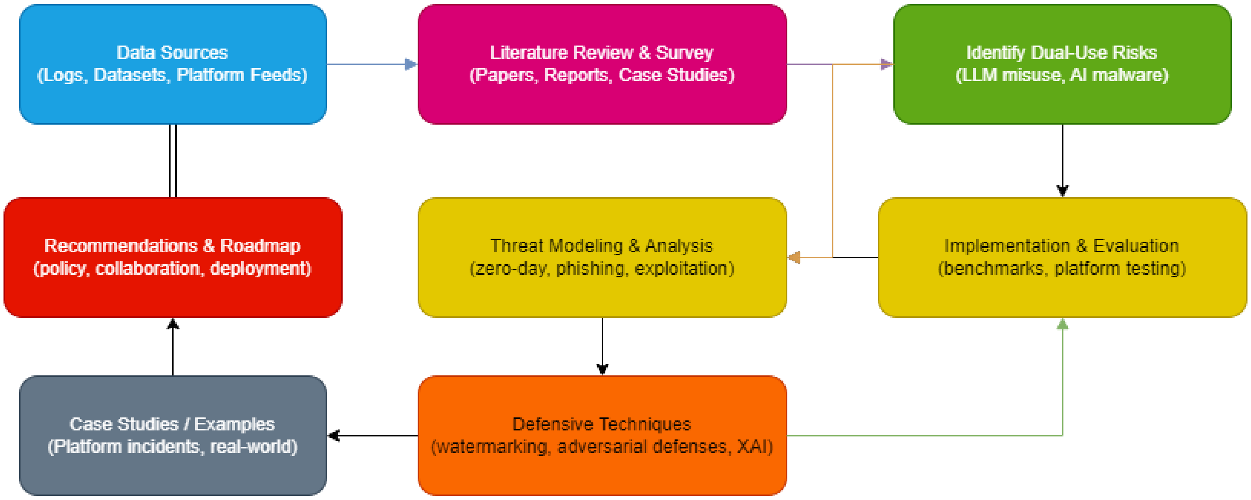
Data Sources
(Logs, Datasets, Platform Feeds)
Literature Review & Survey
(Papers, Reports, Case Studies)
Identify Dual-Use Risks
(LLM misuse, AI malware)
Recommendations & Roadmap
(policy, collaboration, deployment)
Threat Modeling & Analysis
(zero-day, phishing, exploitation)
Implementation & Evaluation
(benchmarks, platform testing)
Case Studies / Examples
(Platform incidents, real-world)
Defensive Techniques
(watermarking, adversarial defenses, XAI)

Figure 1 An End-to-End Framework for Assessing and Mitigating Cybersecurity Risks in Generative AI

As shown in Figure 1, The diagram presents a comprehensive workflow for examining cybersecurity challenges and defenses associated with generative AI systems. It highlights how data, research, and real-world insights flow through different stages to ultimately guide policy and technical safeguards.

1. Data Sources (Logs, Datasets, Platform Feeds)

This stage represents the foundation of analysis, where raw information such as system logs, curated datasets, and live feeds from online platforms is collected. These inputs are critical for identifying patterns of misuse and validating security techniques.

2. Literature Review & Survey (Research Papers, Reports, Case Studies)

Alongside raw data, existing academic publications, government reports, and case analyses are reviewed. This stage ensures that the framework builds on established knowledge, uncovers knowledge gaps, and incorporates lessons from earlier studies.

3. Identification of Dual-Use Risks (LLM Misuse, AI-Generated Malware)

At this stage, the focus shifts to the dual-use nature of generative AI technologies that offer benefits but can also be exploited. Potential risks include the creation of harmful code, AI-assisted malware, or the use of large language models (LLMs) to craft convincing phishing attempts.

4. Threat Modeling & Analysis (Zero-Day Exploits, Phishing, Exploitation Techniques)

Security experts then model how malicious actors could weaponize AI. This involves exploring how LLMs or generative tools may aid in zero-day exploitation, targeted phishing attacks, or automated vulnerability scanning, providing insight into the evolving cyber threat landscape.

5. Defensive Strategies (Watermarking, Adversarial Defenses, Explainable AI)

Countermeasures are proposed here. Examples include:

- Watermarking methods to trace AI-generated content,
- Adversarial defenses to harden models against manipulation,
- Explainable AI (XAI) techniques that improve the interpretability of AI-based defense systems.

6. Implementation & Evaluation (Benchmarking, Platform Testing)

These defensive methods are tested in practice using benchmark datasets and real-world environments. This stage provides quantitative and qualitative evaluation, ensuring that the proposed safeguards are not only theoretical but also practical.

7. Recommendations & Roadmap (Policy, Collaboration, Deployment)

The insights feed into a roadmap that outlines strategies for long-term security. This includes technical deployment guidelines, collaborative efforts between researchers and policymakers, and policies aimed at responsible use of generative AI technologies.

8. Case Studies / Real-World Incidents

Running in parallel to the main flow, this element incorporates practical examples of generative AI misuse, such as actual cyber incidents on platforms. These case studies help ground the theoretical analysis in real-world contexts and strengthen the threat modeling stage.

9. Feedback Loop

Finally, the diagram includes a feedback connection from the roadmap stage back to data collection. This emphasizes that policies and defenses influence the type of data gathered in future cycles, making the framework adaptive to new challenges.

The application of LLMs and diffusion models for personalized recommendations and user-centric interfaces represents an exciting direction for digital commerce and consumer experience. Recommendation engines that incorporate insights from user preferences, character associations, and contextual data are being developed to revolutionize domains such as fashion, entertainment, and digital marketing. For instance, recent research has shown how LLM-driven frameworks can generate personalized outfit recommendations based on users' favorite fictional characters or interests, leveraging explicit prompt engineering and visual generative techniques. This blend of text-based and visual AI opens up a spectrum of possibilities for tailoring content and product suggestions in highly granular and intuitive ways. Seamless integration of such systems into online platforms requires sophisticated approaches to model interpretability, user feedback, and privacy preservation. The potential for enhancing customer engagement through real-time, interactive AI systems is substantial, driving increased investment and innovation in consumer-facing applications.

As AI-driven content generation continues to rise, the convergence of text analysis, multimedia synthesis, and automated editing is reshaping the ways in which information is produced, consumed, and disseminated. Modern platforms are now capable of transforming textual data articles, scripts, or spoken words into cohesive multimedia experiences, complete with images, voiceovers, and synchronized audio-visual elements. The comprehensive utilization of natural language processing, optical character recognition, and deep synthesis points to a future where content creation is more efficient, accessible, and impactful than ever before. These technologies are revolutionizing not only media and journalism, but also educational resources, marketing, and entertainment delivery. Maintaining narrative coherence, ensuring quality, and respecting copyrights remain central concerns as generative systems become central to modern communication ecosystems. The integration of feedback loops, quality benchmarks, and ethical guidelines defines the evolving landscape of automated multimedia content generation.

Finally, the ongoing evolution of AI ethics, governance, and policy is shaping the strategic trajectory for LLMs and generative AI deployment. Policymakers and industry leaders are grappling with the dual imperatives of fostering innovation and mitigating harms. Key debates focus on watermarking AI-generated content, ensuring model traceability, and fostering cross-industry collaboration to address shared risks. The demand for responsible, transparent, and scalable AI solutions is reflected in the growing number of collaborative frameworks and industry standards that codify best practices. As awareness of the systemic impact of AI systems grows, the importance of holistic research bringing together technical, ethical, legal, and societal perspectives becomes ever more apparent. Sustainable advancement in the field will depend on the collective commitment of researchers, technologists, and stakeholders to align the trajectory of generative AI with the broader public interest. The future of responsible AI governance relies on both rigorous inquiry and pragmatic action, ensuring that generative models remain safe, effective, and aligned with human values.

III. LITERATURE SURVEY

Presented below is a summary table that captures major research contributions in the domains of large language models and generative AI, spanning multiple application areas and methodologies. The table organizes these works by highlighting their primary objectives, key techniques, central findings, the datasets and domains they address, and any limitations identified by the researchers. This structured overview is intended to help readers quickly compare essential literature, recognize thematic trends, and pinpoint areas that may require further investigation within the field of AI and machine learning.

Reference	Year	Main Focus	Models/Techniques	Key Findings	Dataset/Domain	Limitations
K. Ahi & S. Valizadeh	2025	Survey of LLMs & generative AI in cybersecurity and privacy	LLMs, Generative AI (ChatGPT, Claude, Gemini, LLaMA, Copilot, Stable Diffusion)	Analyzes dual-use risks, AI-generated malware, explainability, defensive strategies; presents practical recommendations for secure deployment	Cybersecurity platforms (Google Play Protect, MS Defender, AWS, Apple, OpenAI, GitHub)	Focus mainly on cybersecurity, less on other domains
S. Xie et al.	2025	Embodied generative AI art for human-robot interaction	Fine-tuned GPT-3.5 Turbo, Stable Diffusion, UFactory xArm	Presents a robotic drawing system using LLM-guided precise drawing; enhances human-robot artistic collaboration	Art/Human-Robot Interaction	Focuses only on drawing/art; real-world deployment challenges
M.A. Farooq et al.	2025	Synthetic facial data for children via diffusion models	ChildDiffusion (diffusion & LLMs), ControlNet, Vision Transformer	Generates diverse, photorealistic child faces, open-source dataset for ML; addresses scarcity & privacy	Child facial datasets (ChildRace, 2.5k samples)	Focused on faces, ethical challenges remain
E. Zhang et al.	2024	Environmental impact of generative AI, carbon-aware workload shifting	Mistral, Vicuna, Stable Diffusion; workload simulations	Shifting AI inference workloads regionally reduces carbon emissions; 1,652 lbs per 10M requests possible	Generative AI in US/EU data centers	Simulation study, real-world implementation needed
B. Wallace et al.	2024	Diffusion model	Diffusion-DPO, Stable Diffusion	Direct Preference	Image generation,	Scalability and dataset

		alignment via human preferences	XL, Pick-a-Pic dataset	Optimization aligns models to user preferences, boosts visual appeal and alignment	Pick-a-Pic (851K pairwise preferences)	quality affect results
J.-S. Park et al.	2025	On-device NPU for generative AI in mobile SoCs	Samsung Exynos 2400 NPU; FOWLP; memory hierarchy	Hardware supports transformer visual/LLMs on device, improves performance and energy efficiency	Mobile/Edge AI applications	Device constraints, limited hardware discussion
G. Mellone et al.	2025	Marine litter data augmentation with diffusion/LLMs	Stable Diffusion, Alpaca LLMs, YOLOv8	Data augmentation for marine litter detection; increased recall +7.82%, mAP50 +3.87	G-Litter marine litter dataset, real-world mission videos	Focused on marine, scalability to other domains unclear
N. Forouzandehmehr et al.	2023	Character-based outfit generation with LLMs	LVA-COG framework, Stable Diffusion, prompt engineering	Personalized outfit recommendations using character/context extraction via LLMs	Fashion/outfit recommendation datasets	Niche focus, personalization and diversity limits
D. Jadhav et al.	2024	AI-driven text-to-multimedia content generation	NLP (LLaMA2, Gemini), DALL-E, Stable Diffusion, OCR, MoviePy	Automated media creation; integrates text, images, voiceovers; enhances accessibility and engagement	Journalistic, marketing, and educational content	Quality and narrative coherence challenge automated systems
F. Cady	2025	Fundamentals and practice of deep learning and generative AI	Deep neural networks, diffusion models, transformers, LangChain	Foundation concepts, classic applications, and prompt engineering; tutorials for implementation	Classification/image analysis, educational	General overview, lacks empirical evaluation

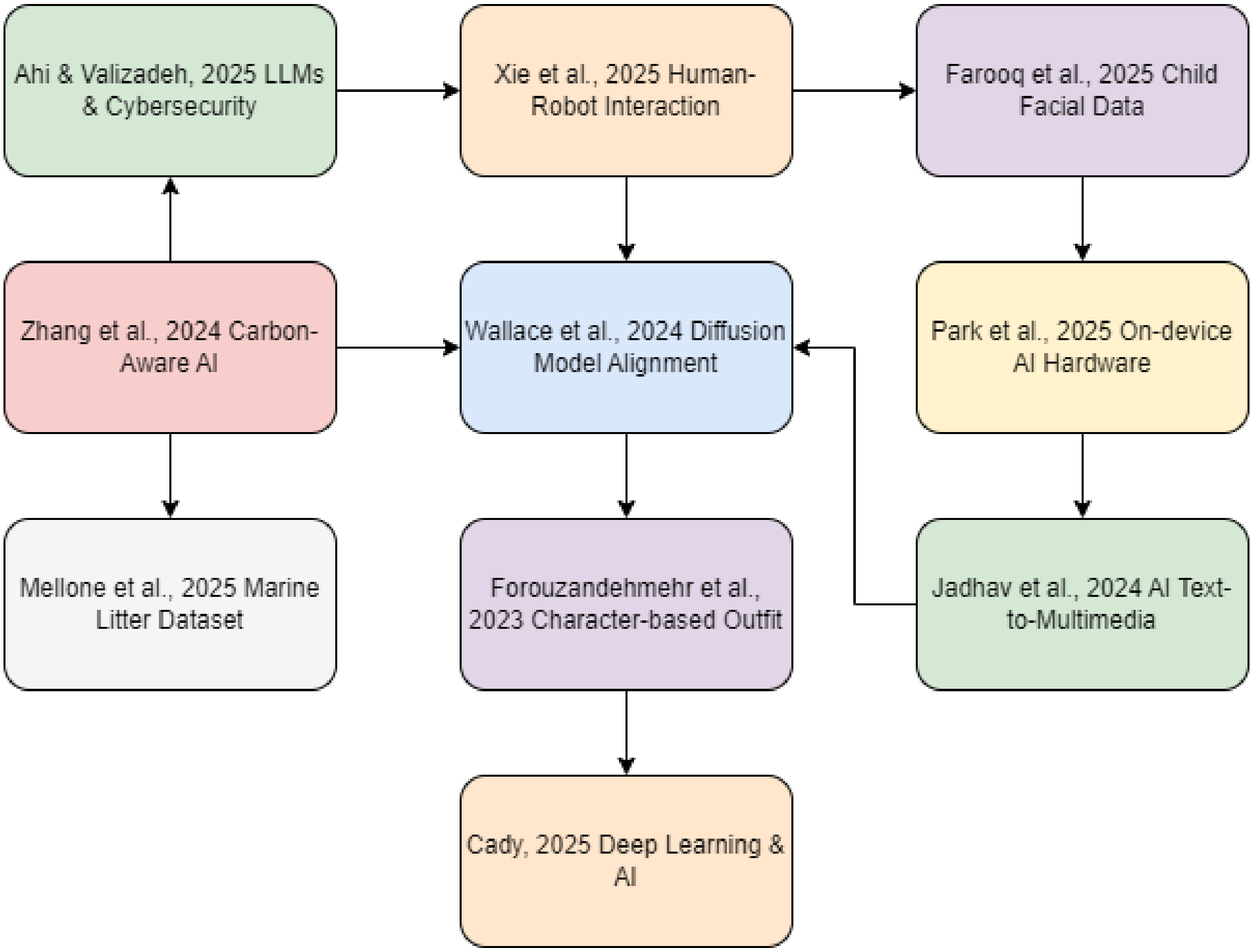
Ahi & Valizadeh, 2025 LLMs & Cybersecurity
Xie et al., 2025 Human-Robot Interaction
Farooq et al., 2025 Child Facial Data
Zhang et al., 2024 Carbon-Aware AI
Wallace et al., 2024 Diffusion Model Alignment
Park et al., 2025 On-device AI Hardware
Mellone et al., 2025 Marine Litter Dataset
Forouzandehmehr et al., 2023 Character-based Outfit
Jadhav et al., 2024 AI Text-to-Multimedia
Cady, 2025 Deep Learning & AI

Figure 2 Interconnected Themes and Progression in Recent Generative AI and Large Language Models Research

As shown in figure 2, it visually summarizes the relationships and thematic connections among key academic works on large language models (LLMs) and generative AI technologies, as reflected in the cited literature.

Starting from the top left, Ahi and Valizadeh's 2025 work on LLMs and cybersecurity leads sequentially to studies on human-robot interaction by Xie et al. (2025) and child facial data generation by Farooq et al. (2025). The research on child facial data further connects to Park et al.’s 2025 investigation into on-device AI hardware, which then flows into multimedia content generation explored by Jadhav et al. (2024).

Parallel to this, Zhang et al.’s 2024 study on carbon-aware AI workloads branches into environmental dataset augmentation by Mellone et al. (2025). Zhang et al.’s work also feeds into Wallace et al.’s 2024 research on diffusion model alignment, which connects onward to a character-based outfit generation framework by Forouzandehmehr et al. (2023). This pathway culminates in Cady’s 2025 comprehensive treatment of deep learning and AI fundamentals.

The figure 2 thus captures a network of evolving research themes, showing how foundational topics in AI sustainability, model alignment, and application-driven innovations interrelate. Arrows indicate the flow of influence or logical progression from one study to another, reflecting a collective scholarly effort advancing both theoretical and practical fronts in generative AI research. This structured representation highlights interdisciplinary connections and the continuity of research development across different application domains.

IV. CONCLUSION

This research provides a comprehensive overview of the current advancements, challenges, and opportunities presented by large language models and generative AI technologies across various domains. It highlights the dual-use nature of these models, demonstrating their transformative potential in fields such as cybersecurity, human-robot interaction, data augmentation, and environmental sustainability, while also addressing the inherent risks like AI-generated malware and privacy concerns. The survey of recent studies reveals the critical need for responsible deployment strategies, including model explainability, alignment with human preferences, and environmentally aware computational practices. Furthermore, the integration of advanced hardware and edge computing emerges as a key enabler for scalable and efficient AI applications. This work underscores the importance of interdisciplinary collaboration among researchers, practitioners, and policymakers to build secure, ethical, and sustainable AI frameworks. Future efforts should emphasize enhancing transparency, mitigating adversarial threats, and minimizing the ecological footprint of AI systems to unlock their full societal benefits. The evolving landscape of generative AI calls for ongoing research to address emerging challenges and to foster innovations that align with human values and global sustainability goals.

REFERENCES

[1] Ahi, K., and S. Valizadeh. "Large Language Models (LLMs) and Generative AI in Cybersecurity and Privacy: A Survey of Dual-Use Risks, AI-Generated Malware, Explainability, and Defensive Strategies." 2025 Silicon Valley Cybersecurity Conference (SVCC), San Francisco, CA, USA, 2025, pp. 1-8. IEEE Xplore, https://doi.org/10.1109/SVCC65277.2025.11133642.

[2] Xie, S., E. B. Sandoval, K. A. Shaik, and F. Cruz. "Embodied Generative AI Art for Enhanced Human-Robot Interaction Through a Human-Centric LLM-Guided Robotic Arm Drawing System." 2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Melbourne, Australia, 2025, pp. 1727-1730. IEEE Xplore, https://doi.org/10.1109/HRI61500.2025.10974105.

[3] Farooq, M. A., W. Yao, and P. Corcoran. "ChildDiffusion: Unlocking the Potential of Generative AI and Controllable Augmentations for Child Facial Data Using Stable Diffusion and Large Language Models." IEEE Access, vol. 13, 2025, pp. 96616-96634. IEEE Xplore, https://doi.org/10.1109/ACCESS.2025.3575964.

[4] Zhang, E., D. Wu, and J. Boman. "Carbon-Aware Workload Shifting for Mitigating Environmental Impact of Generative AI Models." 2024 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics, Copenhagen, Denmark, 2024, pp. 446-453. IEEE Xplore, https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics62450.2024.00087.

[5] Wallace, B., et al. "Diffusion Model Alignment Using Direct Preference Optimization." 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2024, pp. 8228-8238. IEEE Xplore, https://doi.org/10.1109/CVPR52733.2024.00786.

[6] Park, J.-S., et al. "16.3 An on-Device Generative AI Focused Neural Processing Unit in 4nm Flagship Mobile SoC with Fan-Out Wafer-Level Package." 2025 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, USA, 2025, pp. 286-288. IEEE Xplore, https://doi.org/10.1109/ISSCC49661.2025.10904722.

[7] Mellone, G., et al. "G-Litter: Marine Litter Dataset Augmentation with Diffusion Models and Large Language Models on GPU Acceleration." 2025 33rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP), Turin, Italy, 2025, pp. 526-535. IEEE Xplore, https://doi.org/10.1109/PDP66500.2025.00081.

[8] Forouzandehmehr, N., et al. "Character-based Outfit Generation with Vision-Augmented Style Extraction via LLMs." 2023 IEEE International Conference on Big Data (BigData), Sorrento, Italy, 2023, pp. 1-7. IEEE Xplore, https://doi.org/10.1109/BigData59044.2023.10416943.

[9] Jadhav, D., S. Agrawal, S. Jagdale, P. Salunkhe, and R. Salunkhe. "AI-Driven Text-to-Multimedia Content Generation: Enhancing Modern Content

Creation." 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Kirtipur, Nepal, 2024, pp. 1610-1615. IEEE Xplore, https://doi.org/10.1109/I-SMAC61858.2024.10714771.

[10] Cady, Field. "Deep Learning and AI." The Data Science Handbook, Wiley, 2025, pp. 309-329. doi:10.1002/9781394234523.ch24.