Knowledge-Enhanced Text-to-Image Generation for Digital Cultural Heritage Preservation on Edge Computing Platforms

Navin C. Sinha
Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
ncsinha@buffalo.edu

Krish Dutta
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
krish2003@uab.edu

Abstract

The preservation of digital cultural heritage increasingly relies on generative artificial intelligence to reconstruct, visualize, and contextualize artifacts whose physical counterparts face deterioration, destruction, or inaccessibility. Text-to-image models offer a promising pathway for generating historically grounded visual content, yet generic models struggle with the domain specificity, contextual accuracy, and cultural sensitivity required by heritage institutions. This paper presents a system-level investigation of knowledge-enhanced text-to-image generation deployed on edge computing platforms, a paradigm that reconciles the need for powerful generative capabilities with the constraints of fieldwork, museums, and heritage sites. We examine how structured domain knowledge, including ontologies, curated metadata, and verified visual references, can be integrated into the generation pipeline through retrieval-augmented conditioning, prompting strategies, and lightweight adapter modules. The architectural discussion foregrounds edge-native deployment trade-offs, encompassing model compression, distributed inference, hardware heterogeneity, and the incorporation of emerging edge-AI accelerator ecosystems. Beyond technical architecture, we analyze governance frameworks for provenance, authenticity labeling, and bias mitigation, emphasizing the risk of erasing marginalized heritage narratives when generative models amplify training data imbalances. The paper further addresses robustness under degraded communication, energy sustainability at remote sites, and long-term digital accessibility. By synthesizing technical, organizational, and ethical dimensions, we argue that knowledge-enhanced generation on edge infrastructure constitutes a sustainable and culturally responsible foundation for next-generation digital heritage preservation systems. The analysis is grounded in contemporary research on multimodal learning, edge computing, and cultural informatics, offering forward-looking perspectives for system architects, heritage professionals, and policy makers.

Keywords

text-to-image generation, edge computing, digital cultural heritage, knowledge graphs, model adaptation, provenance, sustainability.

1. Introduction

The acceleration of generative modeling has transformed the landscape of content creation, enabling algorithms to synthesize photorealistic and stylized images from natural language descriptions. Foundational contributions such as contrastive language-image pretraining [1] and latent diffusion models [2] have democratized access to high-fidelity visual synthesis, spawning a rapidly growing ecosystem of applications in design, entertainment, and education. Multimodal architectures that couple frozen image encoders with large language models, exemplified by bootstrapped vision-language frameworks [3], further tighten the loop between semantic intent and visual output, opening new possibilities for domain-specific deployment. In parallel, the digital preservation of cultural heritage has emerged as an urgent interdisciplinary challenge: artifacts, monuments, manuscripts, and intangible practices are increasingly documented in digital form, yet the production of rich, contextually accurate visual reconstructions remains labor-intensive and reliant on scarce expert knowledge.

Within this nexus, knowledge-enhanced text-to-image generation holds exceptional promise. By conditioning the generative process not only on short textual prompts but also on structured information from cultural ontologies, historical records, and iconographic databases, systems can produce images that respect provenance, material authenticity, and stylistic conventions. Early attempts to embed commonsense and scene-level constraints through spatial knowledge priors [4] demonstrate that significant gains in controllability and semantic fidelity are achievable when generation is guided by external structured knowledge rather than purely statistical correlations. Cultural heritage applications, however, impose stricter demands: an ill-informed reconstruction of a medieval manuscript illumination or an eroded temple relief can propagate misinformation, distort historical understanding, and offend communities for whom heritage objects carry deep symbolic meaning. A comprehensive survey of machine learning approaches for cultural heritage [5] underscores the domain's unique blend of high-resolution imaging, sparse annotations, and the need for interpretable outputs, all of which complicate the direct application of conventional generative models.

The deployment of such knowledge-intensive models in heritage contexts is further constrained by the computational environments in which they must operate. Many heritage sites, archaeological excavations, and community museums are situated in locations with limited network bandwidth, intermittent connectivity, or strict data sovereignty requirements that preclude reliance on cloud-centric inference. This has motivated a turn toward edge computing architectures, which distribute computation closer to data sources and end users. However, the computational demands of modern text-to-image models, often requiring hundreds of millions of parameters and substantial GPU memory, clash with the resource-constrained nature of edge devices. The resulting tension between model capability and deployability defines the central architectural problem addressed in this paper. A recent comprehensive survey of edge-deep learning convergence [6] catalogs the challenges of collaborative inference, model partitioning, and energy-aware scheduling, but the intersection with knowledge-enhanced generation for cultural heritage has not been systematically examined.

This paper provides a system-level analysis that integrates perspectives from generative AI, knowledge engineering, edge computing, and heritage governance. Rather than proposing a single algorithm or benchmark, we survey architectural patterns, structural trade-offs, and policy implications that shape the design of knowledge-enhanced text-to-image pipelines on edge platforms. We examine how machine-readable heritage knowledge, encoded as scene graphs [7] or semantic reference models such as the CIDOC Conceptual Reference Model [8], can be injected into the generation stack via retrieval-augmented conditioning, parameter-efficient fine-tuning, and hybrid prompting. We analyze how edge-native strategies including model quantization [9], knowledge distillation [10], cooperative inference with caching layers [11], and federated fine-tuning [12] respond to the need for low-latency, privacy-preserving generation in the field. Furthermore, we situate these technical mechanisms within broader governance frameworks, discussing provenance documentation through datasheet-like structures [13], UNESCO digital heritage charters [14], and the environmental sustainability of edge-AI lifecycles. Throughout, we emphasize the non-technical requirements that distinguish heritage applications from generic content generation: authenticity verification, community co-design, intergenerational accessibility, and resilience against cultural erasure. The concluding sections synthesize the trade-offs and propose directions for policy-oriented systems research.

2. Domain-Specific Requirements and Knowledge Structures

Cultural heritage image generation is fundamentally a knowledge-intensive task that cannot be reduced to text-to-pixel mapping learned from web-scale image-text pairs. A heritage institution seeking to visualize a partially destroyed temple must incorporate architectural treatises, archaeological reports, epigraphic readings, and comparative iconography. Generic foundation models, trained predominantly on contemporary photographs with loose captions, lack the factual grounding needed to reconstruct, for example, the correct sequence of cornice moldings for an Angkorian sanctuary or the precise pigment palette of an Etruscan tomb painting. The resulting images, while visually plausible, risk propagating anachronisms or culturally inappropriate details that undermine public trust and scholarly authority. The gap between statistical plausibility and historical veracity must be closed by embedding structured domain knowledge directly into the generative pipeline.

One prominent knowledge representation paradigm builds on scene graphs, which decompose a visual scene into objects, attributes, and pairwise relationships. By parsing an expert-authored description of a historical event or an archaeological context into a graph of entities and relations, a generation system can enforce spatial consistency, hierarchical composition, and stylistic constraints that are difficult to convey through a flat textual prompt alone. Early work on iterative scene graph generation [7] provides the representational backbone, and contemporary text-to-image models have shown that conditioning on graph-structured layout information significantly improves object count fidelity and inter-object spatial reasoning, a capability essential for reconstructing ritual processions, architectural elevations, or artifact ensembles. For heritage applications, scene graphs can be populated from curated databases that link iconographic motifs to specific periods and regions, thereby anchoring the generation to authoritative scholarship rather than to the model's internal distributional priors.

At a higher level of abstraction, formal ontologies developed for the cultural heritage sector offer interoperability and semantic precision that shallow keyword-based conditioning cannot achieve. The CIDOC Conceptual Reference Model [8] is the de facto standard for harmonizing heterogeneous cultural information, providing classes and properties that connect physical artifacts, digital surrogates, temporal periods, persons, places, and events. Integrating CIDOC-CRM with a text-to-image pipeline enables concept-level queries such as "generate a watercolor rendition of a Late Classic Maya vase from the Petén region depicting a royal accession scene" to be grounded in unambiguous identifiers rather than ambiguous natural language terms. A retrieval-augmented module can traverse the ontology graph to collect verified reference images, material descriptions, and stylistic annotations, which are then encoded as conditioning vectors injected into cross-attention layers via lightweight adapters. This architecture simultaneously improves factual accuracy and facilitates provenance tracing, because every generated output can be linked back to the specific knowledge graph triples and source documents that influenced it.

Beyond formal ontologies, domain-specific visual knowledge base s—including iconographic thesauri, color analysis reports, and multispectral imaging archives—serve as additional conditioning resources. Unlike general image retrieval systems that prioritize visual similarity over factual correctness, heritage-oriented retrieval must incorporate authentication filters and temporal ordering constraints to avoid feeding anachronistic references into the generator. For instance, a system tasked with depicting a Tang dynasty mural should not retrieve a Qing dynasty reproduction, even if it is visually similar, without explicit scholarly annotation of its later provenance. Implementing such temporal and attributional constraints within a retrieval index requires close collaboration between system architects and domain curators, a socio-technical design process that conventional machine learning pipelines rarely accommodate. The edge deployment context further complicates this integration, as full-scale knowledge graphs may exceed the storage capacity of a field device, necessitating hierarchical caching and relevance-based subgraph extraction strategies that we examine in the following section.

3. Edge Computing Architecture and Deployment Trade-offs

The decision to deploy knowledge-enhanced text-to-image generation at the network edge rather than in centralized cloud data centers is driven by a confluence of practical, legal, and cultural factors. Archaeological sites in remote locations often operate with satellite links that impose high latency and stringent data caps, making continuous cloud inference infeasible. Indigenous and local communities may assert data sovereignty over heritage imagery, requiring that generation and curation remain within territorial boundaries. Even in well-connected urban museums, interactive visitor experiences demand sub-second response times that rule out round-trip cloud communication. Thus, the edge becomes the natural locus for computation, but the mismatch between the immense parameter counts of state-of-the-art generative models and the modest capabilities of edge hardware presents a formidable systems challenge.

Bridging this gap involves a combination of model compression, collaborative inference, and hardware-aware optimization. Efficient model scaling techniques such as compound scaling, originally proposed for convolutional networks [9], have been extended to vision transformers and diffusion backbones, enabling their footprint to be reduced without catastrophic loss of generation quality. Post-training quantization and integer-only inference pipelines [10] further shrink memory requirements and accelerate matrix multiplications on edge processors equipped with neural processing units. Knowledge distillation, whereby a compact student generator is trained to mimic the output distribution of a large teacher model under the guidance of domain-knowledge conditioning, preserves the teacher's sensitivity to ontological constraints while operating within the power envelope of a fanless embedded system [11]. However, distillation alone cannot capture all the nuances of cultural styles, and a careful balance must be struck between model capacity and the granularity of heritage-specific conditioning signals.

When a single edge node remains insufficient, the system can be partitioned across a hierarchy of devices. Cooperative edge caching strategies [12] precompute and store reusable components of the image generation stack, such as style feature banks derived from periodspecific iconographic corpora, so that only lightweight personalization prompts need to be processed at inference time. A tiered architecture can place a distilled base model on a micro-edge device worn by a field archaeologist, while a more capable edge server at a nearby camp runs retrieval-augmented conditioning and feeds structured scene graphs to the handheld unit. Federated learning protocols [13] enable these dispersed nodes to collaboratively fine-tune adapters on newly digitized artifacts without centralizing sensitive cultural data, a crucial capability for community-owned heritage collections. The resulting architecture is a heterogeneous, intermittently connected system whose operational envelope must be characterized not only by throughput and latency but also by metrics of historical accuracy, ontological consistency, and community acceptance.

The hardware landscape for edge-native generative AI is evolving rapidly, with the emergence of custom AI accelerators designed outside traditional data-center supply chains. Recent developments have demonstrated the feasibility of training a full text-to-image foundation model entirely on a cluster of China-developed edge-oriented AI accelerators, achieving competitive fidelity while consuming substantially less power than comparable GPU-based systems [16]. Although this model was not originally designed for cultural heritage, its existence signals that the hardware asymmetry between cloud and edge is narrowing, and that domain-adapted training on heritage corpora using such edge-native compute fabrics is a plausible near-term trajectory. Integrating such accelerators into museum kiosks or portable documentation kits would dramatically reduce dependence on foreign cloud infrastructure and align with institutional mandates for technological sovereignty. Architectural decisions must therefore anticipate a multi-vendor edge ecosystem, necessitating hardware abstraction layers, portable model formats, and vendor-agnostic runtime compilers that can target a spectrum of silicon platforms.

Beyond pure performance, system-level trade-offs involve energy consumption, heat dissipation, and device longevity, concerns that are especially acute in heritage contexts where equipment must operate inside protective enclosures near fragile artifacts. The total energy budget for a generation request includes not only the forward pass of the model but also the retrieval query against a local knowledge graph, the decoding of ontological identifiers, and the rendering of high-resolution output images. Research on the carbon footprint of AI underscores the importance of reporting and optimizing these end-to-end costs [17]. On edge devices that may be solar-powered and subject to dust, humidity, and temperature extremes, power-aware scheduling algorithms must dynamically adjust model precision, retrieval depth, and generation resolution to stay within instantaneous power caps while preserving the visual quality demanded by documentation standards. Sustainability thus becomes a first-class design constraint, not an afterthought.

4. Knowledge Integration and Model Adaptation Strategies

Embedding heritage knowledge into a text-to-image model on an edge platform requires a departure from monolithic end-to-end training in favor of modular, parameter-efficient adaptation. Full fine-tuning of a diffusion model on a heritage corpus is rarely viable on edge hardware and risks catastrophic forgetting of the broad visual grammar that the base model acquired during pretraining. Instead, lightweight adapter layers inserted into the cross-attention and self-attention blocks of the U-Net denoiser can map knowledge graph embeddings into conditioning signals while leaving the bulk of the pretrained weights frozen. These adapters can be trained on a relatively small number of culturally verified image–graph pairs using low-rank decomposition methods, enabling fast on-device personalization when a museum curator wishes to calibrate the generator to a newly discovered corpus of pottery shards. By decoupling knowledge encoding from the generative backbone, the system preserves the ability to swap in updated ontologies without retraining the entire model, an essential requirement for heritage knowledge that evolves as new archaeological evidence emerges.

Retrieval-augmented generation provides a complementary pathway for injecting temporally and spatially grounded references. Given a user prompt, a dual-encoder retrieval system queries a local vector index of certified heritage images and their associated CIDOC-CRM annotations, ranking candidates by a composite score that balances visual similarity, temporal compatibility, and source authority. The top-ranked references are embedded as visual tokens, concatenated with the text prompt embeddings, and fed into the denoising process. This approach mimics the workflow of a heritage illustrator who consults reference photographs and scholarly sketches before beginning a reconstruction, and it allows the system to produce outputs whose provenance is transparently linked to the retrieved materials. On edge platforms, the retrieval index must be carefully sharded and compressed; product quantization and graph-based approximate nearest neighbor indices can be hosted within the limited RAM of an embedded device while supporting millisecond-level queries for collections numbering in the hundreds of thousands of artifacts.

Hybrid prompting strategies further strengthen the alignment between generation and domain knowledge. Rather than relying on a single natural language string, a structured prompt format can combine a free-text description, a knowledge graph URI, and a style identifier drawn from a controlled thesaurus. The model's text encoder is fine-tuned using contrastive learning to align such structured prompts with the corresponding verified images, creating an embedding space where ontological proximity translates into visual similarity. This technique, allied with negative prompting that explicitly disallows certain anachronistic motifs, reduces the incidence of culturally implausible generations even when the underlying base model has been trained on noisy web data. The challenge lies in designing prompt templates that strike a balance between expressivity and complexity, such that field personnel without computer science backgrounds can construct effective queries using graphical interfaces that abstract the underlying knowledge graph operations. Human-in-the-loop evaluation should be embedded throughout the deployment lifecycle, with edge devices logging user corrections to continuously improve the knowledge-conditioning modules via federated reinforcement learning from preference signals.

5. Governance, Provenance, and Ethical Frameworks

The deployment of generative AI for cultural heritage preservation raises ethical questions that go beyond the typical concerns of bias and misuse in general-purpose image synthesis. Heritage objects are not merely aesthetic artifacts; they are carriers of identity, memory, and collective trauma for communities whose voices have historically been marginalized in institutional archives. When a text-to-image model generates a reconstruction of a sacred site or a ceremonial mask without community consent, or when it visually homogenizes diverse regional styles under a single "traditional" label, it commits a form of representational harm that purely accuracy-oriented metrics cannot capture. System architects must therefore embed governance mechanisms that extend from data collection and model training through generation and post-hoc auditing.

Provenance documentation is a foundational requirement. Drawing on the datasheet paradigm for datasets [14], every image generated by the system should be accompanied by a verifiable provenance record that enumerates the knowledge sources consulted, the retrieval scores, the adapter configurations, and the human approvals obtained. Blockchain-based content authenticity initiatives can provide tamper-evident logs, but they introduce energy overheads that conflict with edge sustainability goals. A more pragmatically suited approach for low-power devices involves signing provenance metadata with lightweight cryptographic hashes and storing the records in a decentralized content-addressed file system that synchronizes opportunistically when connectivity becomes available. Such records allow downstream historians, regulators, and community representatives to audit the generation process and, if necessary, contest outputs that misrepresent cultural facts.

Bias in generative models trained on imbalanced web corpora poses a direct threat to equitable heritage representation. Iconic monuments from Western Europe dominate public image datasets, while vernacular architecture, indigenous rock art, and oral-tradition performance practices are grossly underrepresented. If a knowledge-enhanced system is not carefully curated, the statistical priors of the base model will overwhelm the domain conditioning, producing outputs that default to familiar visual clichés. Mitigation strategies must operate at multiple levels: curating balanced fine-tuning corpora in partnership with source communities, applying bias-aware data augmentation, and implementing output filters that flag and suppress generations that exhibit known biases such as Eurocentric color palettes or homogenized facial features. The fairness metrics developed for natural language processing, originally aimed at word embeddings [19], must be translated into the visual domain through perceptual studies that assess cultural appropriateness rather than generic aesthetic quality. Moreover, the very definition of “fairness” in heritage reconstruction is contested, because prioritizing statistical parity across cultures may inadvertently flatten meaningful stylistic distinctions; this demands an iterative, deliberative process of metric design that involves ethicists, heritage scholars, and community custodians.

Institutional policy frameworks provide the scaffolding for these governance mechanisms. The UNESCO Charter on the Preservation of Digital Heritage [15] articulates principles of authenticity, accessibility, and stewardship that directly inform system requirements. Regional regulations such as the European Union’s AI Act increasingly subject high-risk AI systems, including those used in cultural sectors, to conformity assessments that mandate transparency, human oversight, and robustness. Edge-native architectures, by keeping data and computation local, offer a structural advantage in demonstrating compliance with data protection and sovereignty laws, but they also complicate centralized auditing because no single cloud operator can aggregate logs from a fleet of autonomous field devices. A federated governance model is therefore required, wherein each edge deployment maintains its own auditable records while participating in a consortium that defines shared knowledge graph updates, style thesauri, and bias benchmarks. The socio-technical design of such consortia must allocate decision-making authority to heritage institutions and communities, resisting the tendency for large technology vendors to dictate the terms of open-source heritage AI stacks.

6. Robustness and Sustainability in Long-Term Digital Preservation

Digital heritage preservation is measured in centuries, whereas typical machine learning system deployments are optimized for intervals of months or a few years. A knowledge-enhanced text-to-image system intended to serve a museum collection or a UNESCO World Heritage site must therefore be engineered for extreme longevity, confronting challenges of model decay, hardware obsolescence, and knowledge graph evolution. The robustness of the generation pipeline under distributional shift is a primary concern. As archaeological knowledge advances, the underlying ontologies and reference image databases are updated, rendering previously fine-tuned adapters stale. Unless the system supports backward-compatible incremental learning, its outputs will drift away from current scholarly consensus, eroding trust. Continual learning techniques that interleave rehearsal of foundational heritage examples with assimilation of new knowledge can be implemented on edge devices through replay buffers and elastic weight consolidation, but they require careful management of limited storage resources.

Hardware degradation and the eventual retirement of specific accelerator architectures introduce a preservation risk unique to digital systems. A generation artifact produced by a version of the system running on a particular edge AI accelerator may not be exactly reproducible after the hardware is decommissioned, given differences in numeric precision, quantization, and random seed handling. Unlike physical conservation, where the stability of materials is studied over decades, the bitwise reproducibility of AI-generated heritage images remains an open research topic. Emulation and virtualization layers that capture the complete software stack, including accelerator-specific runtime libraries, can be archived alongside the image outputs, enabling future scholars to regenerate and verify historical generations on emulated hardware. Such archival practices mirror the digital preservation strategies advocated by library and information science researchers for complex digital objects [21], and they impose a documentation burden on edge system developers who must maintain detailed manifests of their inference environments.

Energy sustainability is both an environmental and a practical necessity for heritage sites that operate off-grid. A system that generates high-resolution visualizations on demand must be powered by renewable sources with limited battery capacity. Research on the energy costs of AI has shown that generating a single image can consume orders of magnitude more energy than serving a text snippet, raising questions about the net benefit of on-site generation when compared with precomputed static visualizations [17]. Yet interactive generation offers significant educational and research value, as it allows site visitors and archaeologists to explore counterfactual "what-if" reconstructions (e.g., testing different lighting conditions or restoration hypotheses). A sustainable edge architecture can harness energy-aware scheduling that defers non-urgent generation tasks to periods of surplus solar generation, caches popular visualizations to avoid recomputation, and dynamically adapts the number of diffusion steps based on available battery reserves. Lifecycle analysis must extend beyond operational energy to include the embodied carbon of manufacturing and deploying edge hardware, informing procurement decisions that favor modular, repairable, and long-lived devices over disposable consumer-grade equipment.

Finally, the long-term accessibility of generated heritage content demands open, standards-based formats that outlast proprietary model ecosystems. Current practices of embedding generation metadata in non-standardized EXIF fields or relying on closed-source model APIs jeopardize the findability and interpretability of digital heritage assets for future generations. System architects should adopt the Semantic Web stack, representing generated images, their associated knowledge graph triples, provenance records, and model card information in linked data formats, ensuring that the knowledge-enhanced generation pipeline itself becomes part of the documented heritage rather than an opaque black box. This orientation transforms the edge platform from a transient computational tool into a durable component of the cultural record.

7. Conclusion

Knowledge-enhanced text-to-image generation on edge computing platforms represents a convergence of technical innovation and cultural responsibility. By integrating structured domain knowledge through ontologies, scene graphs, and verified retrieval, generation systems can produce historically grounded visual content that respects the evidentiary standards of heritage scholarship. Edge-native architectures, employing model compression, cooperative caching, federated adaptation, and emerging accelerator hardware, reconcile the computational demands of generative AI with the resource, latency, and sovereignty constraints of field and museum environments. The architectural trade-offs are multidimensional, spanning model fidelity against energy consumption, retrieval depth against storage limits, and local autonomy against federated governance. System design must therefore be approached as a socio-technical negotiation among computer scientists, heritage professionals, community representatives, and policy makers, not as a purely optimization-driven engineering exercise.

Governance frameworks for provenance, bias mitigation, and accessibility are not external supplements but integral system requirements that shape model selection, data partitioning, and interface design. The platform must be capable of generating not only visually compelling images but also auditable accounts of how each image was produced, which sources influenced it, and what steps were taken to guard against cultural stereotyping. Long-term sustainability demands that these capabilities persist across decades of ontological revision and hardware turnover, guided by digital preservation principles that treat the generative pipeline itself as a digital artifact worthy of curation. As edge AI accelerators diversify and their supply chains evolve, the heritage community has a unique opportunity to influence the development of open, verifiable, and ethically grounded generative infrastructure, ensuring that technology serves the pluralistic memory of humanity rather than amplifying a single dominant narrative. The research agenda emerging from this analysis calls for longitudinal field evaluations, community-centered metric design, and the formulation of interoperability standards that bind AI ecosystems to the enduring values of cultural heritage stewardship.

References

1. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In Proceedings of the 38th International Conference on Machine Learning (ICML).

2. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

3. Li, J., Li, D., Savarese, S., & Hoi, S. C. H. (2023). BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In Proceedings of the 40th International Conference on Machine Learning (ICML).

4. Gafni, O., Polyak, A., Ashual, O., Sheynin, S., Parikh, D., & Taigman, Y. (2022). Make-A-Scene: Scene-based text-to-image generation with human knowledge. In Proceedings of the European Conference on Computer Vision (ECCV).

5. Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A., & James, S. (2020). Machine learning for cultural heritage: A survey. Pattern Recognition Letters, 133, 102-108.

6. Wang, X., Han, Y., Leung, V. C. M., Niyato, D., Yan, X., & Chen, X. (2022). Convergence of edge computing and deep learning: A comprehensive survey. IEEE Communications Surveys & Tutorials, 22(2), 869-904.

7. Xu, D., Zhu, Y., Choy, C. B., & Fei-Fei, L. (2017). Scene graph generation by iterative message passing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

8. Doerr, M. (2003). The CIDOC conceptual reference model: An ontological approach to semantic interoperability of cultural heritage information. In Proceedings of the Conference on Artificial Intelligence and Cultural Heritage.

9. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning (ICML).

10. Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Adam, H., & Kalenichenko, D. (2018). Quantization and training of neural networks for efficient integer-arithmetic-only inference. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

11. Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

12. Yao, J., Han, T., & Ansari, N. (2019). On mobile edge caching. IEEE Communications Surveys & Tutorials, 21(3), 2525-2553.

13. McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS).

14. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daume III, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92.

15. UNESCO. (2015). Charter on the preservation of the digital heritage. Records of the General Conference, 38th session.

16. Chen, C., Wang, C., Li, Y., Wan, Z., Geng, M., Xiao, J., ... & Peng, Y. (2026). JuZhou 1.0 Technical Report: The First Edge-Native Text-to-Image Foundation Model Trained Entirely on China-Developed AI Accelerators. arXiv preprint arXiv:2606.28421.

17. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL).

18. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT).

19. Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In Advances in Neural Information Processing Systems (NIPS).

20. Hendrycks, D., & Dietterich, T. (2019). Benchmarking neural network robustness to common corruptions and perturbations. In Proceedings of the International Conference on Learning Representations (ICLR).

21. Lavoie, B. F., & Dempsey, L. (2004). Thirteen ways of looking at... digital preservation. D-Lib Magazine, 10(7/8).

22. Benzerki, M. A., Belkhir, L., & Elouatouat, N. (2020). Energy efficiency in edge AI: A review. IEEE Access, 8, 141224-141249.