Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 6, No.1, June : 2025

GENERATIVE AI WITH BLOCKCHAIN TECHNOLOGY

Shivam K. Yadao, Students, Department of Computer Science Engineering
Dr.Manisha Pise, Guide, Department of Computer Science Engineering
Rajiv Gandhi College of Engineering Research and Technology, Chandrapur, Maharashtra, India
Tanay R. Tiwari, Students, Department of Computer Science Engineering
Pranjal Zode, Students, Department of Computer Science Engineering
Shubham B. Vaidya, Students, Department of Computer Science Engineering
Sayyad Anasali, Students, Department of Computer Science Engineering

ABSTRACT:
Generative AI has revolutionized content creation across various media formats, including text, images, speech, presentations, and videos. However, challenges such as content ownership, security, and authenticity persist. This paper presents a blockchain-integrated AI system capable of generating multimodal content through five modules: Text-to-Image, Text-to-Speech, Text-to-PPT, Text-to-Video, and Text-to-Text. The system employs deep learning models such as Stable Diffusion, Tacotron2, GPT, and Wikipedia API-based knowledge retrieval to generate relevant content from text prompts. Blockchain technology is integrated to secure ownership, prevent unauthorized modifications, and ensure transparency. This study aims to develop an efficient and scalable AI-powered tool for content generation while preserving authenticity and security Multimodal content generation has become an essential aspect of digital media, with applications spanning education, entertainment, and business presentations. Traditional content creation processes require significant time and expertise. This paper explores a novel AI-driven system that automates content generation through five core modules: Text-to-Image, Text-to-Speech, Text-to-PPT, Text-to-Video, and Text-to-Text. By leveraging deep learning techniques, the system can generate images, speech, videos, presentations, and text-based content from textual descriptions. To ensure authenticity and security, we integrate "Blockchain technology" for ownership verification and content protection. This approach enhances scalability, efficiency, and security, making AI-generated content more reliable for real-world applications.
Keywords: Generative AI, Google Gemini, Stable Diffusion, ZeroScope, Text-to-Speech (TTS), Wikipedia API, PowerPoint Automation, Multi-modal AI, Natural Language Processing (NLP), Text Generation, Image Synthesis, Video Generation, AI Presentation Builder, Diffusers, gTTS, Python-pptx, Interactive Content Generation, AI with Blockchain.

INTRODUCTION :

In recent years, Generative Artificial Intelligence (AI) has revolutionized the way digital content is created, enabling machines to produce human-like text, realistic images, synthesized speech, and even short video clips. This project harnesses the power of state-of-the-art generative models to build an interactive platform where users can input any topic or prompt and receive dynamically generated content in multiple formats—including text, images, audio, video, and presentation slides.The primary objective of this system is to simplify content creation and knowledge dissemination by automating the generation of multimedia outputs. It uses Google Gemini for natural language understanding and text generation, Stable Diffusion for AI-driven image synthesis, ZeroScope for text-to-video generation, and advanced TTS models for audio narration. Wikipedia's API is also integrated to extract reliable information for generating structured PowerPoint presentations.To enhance security, content authenticity, and data integrity, the system is being extended with blockchain-based functionality. Although this blockchain module is still under development, the goal is to implement decentralized verification and tamper-proof records for all AI-generated outputs, addressing concerns about trust,

plagiarism, and content misuse. This platform not only demonstrates the potential of multi-modal generative AI in education, research, and media but also paves the way for future applications where security and trust in AI-generated content are paramount.

LITERATURE:

Author	Year	Technology Used	Key Finding	Advantage	Disadvantage
OpenAI (Brown et al.)	2023	GPT-4, Transformer Models	GPT-4 demonstrated the ability to generate high-quality, coherent text and code across diverse applications, including text generation for automation.	State-of-the-art language generation, adaptable to various industries	Requires vast computational resources for training and operation
Ramesh et al. (OpenAI)	2022	DALL·E 2, Diffusion Models	DALL·E 2 generated highly realistic images from text prompts, showing impressive fidelity in image-text association.	High creativity in generating visually appealing images from text	Requires fine-tuning to avoid bias and ensure quality control in edge cases
Microsoft Research	2022	DeepSpeech, Tacotron 2, WaveNet	Significant improvements in speech synthesis accuracy and naturalness using Tacotron 2 and WaveNet.	Highly realistic speech generation, capable of handling various accents	Struggles with long-form content and occasional loss of coherence in audio output
Radford et al. (OpenAI)	2021	CLIP, Contrastive Language–Image Pre-training	CLIP efficiently linked image and text representations, improving image and text retrieval tasks.	High performance in zero-shot learning, flexibility in multiple tasks	Vulnerable to adversarial attacks and may inherit biases from training data
Protocol Labs	2020	IPFS (InterPlanetary File System), Decentralized Storage	IPFS enabled secure, decentralized storage of digital content, reducing dependency on centralized data storage.	Scalability and decentralized content storage	Complex for general users and requires improved integration with mainstream blockchain systems
Karras et al.	2019	StyleGAN, Generative Adversarial	Introduced a style-based architecture for GANs that	High-quality image generation,	Susceptible to mode collapse during training,

		Networks (GANs)	improved the control over generated images, especially in facial features.	increased control over feature generation	computationally expensive

PROPOSED WORK AND METHODOLOGY :

Data Collection:

The system is trained on various datasets to enhance its multimodal capabilities:

- **Text-to-Image:** COCO, LAION-5B (image-text datasets)
- **Text-to-Speech:** LJSpeech, LibriTTS (paired text-audio datasets)
- **Text-to-PPT:** Wikipedia API-based structured knowledge retrieval
- **Text-to-Video:** UCF-101, YouTube-8M (video generation datasets)
- **Text-to-Text:** BooksCorpus, Wikipedia, Common Crawl (text generation datasets)

Data Preprocessing

- **Text Processing:** Tokenization, stopword removal, lemmatization.
- **Image Processing:** Data augmentation (scaling, cropping, rotation).
- **Speech Processing:** Feature extraction (MFCCs, spectrograms).
- **Video Processing:** Frame extraction, caption alignment.

Model Architecture

- **Text-to-Image: Stable Diffusion** (diffusion-based image generation)
- **Text-to-Speech: Tacotron2-DDC + WaveGAN** (speech synthesis and enhancement)
- **Text-to-PPT: Wikipedia API + Python-pptx** (automatic slide generation)
- **Text-to-Video: MoCoGAN + Runway Gen-2** (AI-driven video synthesis)

I. Blockchain Integration

Blockchain technology ensures the security and traceability of AI-generated content through:

- **Smart Contracts:** Ethereum-based **Solidity smart contracts** to register content ownership and prevent forgery.
- **Decentralized Storage:** IPFS and **Filecoin** for secure storage of AI-generated assets.
- **Verification Mechanism:**
  - AI-generated content is **hashed and stored** on the blockchain.
  - Users can **validate ownership** and **track modifications** through blockchain records.
  - Content monetization is enabled via **tokenization (NFTs).**

MODEL EVALUATION:

To assess system performance, we employ:

1. **Text-to-Image:** Fréchet Inception Distance (FID), Inception Score (IS), and human evaluation.
2. **Text-to-Speech:** Mean Opinion Score (MOS), spectrogram similarity analysis.
3. **Text-to-PPT:** Content coherence, readability, and slide structure assessment.
4. **Text-to-Video:** Structural Similarity Index (SSIM), user feedback.
5. **Text-to-Text:** BLEU, ROUGE, and GPT-based text quality metrics.
6. **Blockchain Security:** Smart contract verification, transaction integrity analysis.

CONCLUSION:

Our AI-driven multimodal content generation system integrates **Generative AI** with **Blockchain technology** to ensure authenticity, ownership verification, and decentralized content management. By

leveraging **deep learning models** for creative automation and **blockchain security mechanisms** for content protection, this approach enables **scalable, efficient, and secure content generation.** Future research will focus on improving AI model robustness, expanding blockchain adoption for digital content security, and exploring novel applications in AI-driven digital marketplaces.

REFERENCES

[1] Brown, T. B., et al. "Language Models are Few-Shot Learners." Advances in Neural Information Processing Systems, 2020.
[2] Karras, T., Laine, S., and Aila, T. "A Style-Based Generator Architecture for Generative Adversarial Networks." IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.
[3] Radford, A., et al. "DALL·E: Creating Images from Text." OpenAI, 2021.
[4] Nakamoto, S. "Bitcoin: A Peer-to-Peer Electronic Cash System." Cryptography Mailing list, 2008.
[5] Protocol Labs. "InterPlanetary File System (IPFS)." 2020.
[6] Chen, X., et al. "MoCoGAN: Decomposing Motion and Content for Video Generation." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.