International Journal of Computer Sciences and Engineering
Vol.**13**, Issue.**4**, pp.**34-40**, April **2025**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

ISSN: 2347-2693 (E)

Research Article

AI Diffusion: An Android Application for Text-to-Image Generation Using Generative AI Models

Anant Agrawal[1*], Vedant Vardhan Rathour[2], Babeetha S.[3]

[1,2,3]Dept. of Computing Technologies, SRM Institute of Science and Technology, Chennai, India

**Corresponding Author:* ✉

Received: 11/Feb/2025; **Accepted**: 12/Mar/2025; **Published**: 30/Apr/2025. **DOI**: https://doi.org/10.26438/ijcse/v13i4.3440

Abstract: AI Diffusion is an Android-supported text-to-image synthesis application fusing cutting-edge generative artificial intelligence and the pervasiveness of Android platforms. The vision is to provide users with a simple yet incredibly powerful tool with which they can produce realistic and imaginative images from text-based descriptions. Application architecture features an Android frontend developed with Kotlin and Python-based backend authored in FastAPI. Upon entering a prompt by a user, the backend interacts with a pre-trained generative AI model hosted via API to create an associated image, which is then rendered inside the app. The system is optimized to be lean and responsive to support real-time interaction even on resource-constrained devices. Comprehensive testing shows that the app works seamlessly with minimal latency and creates contextually accurate images for all types of prompts. The project bridged the distance between powerful AI functionality and genuine mobile usability, bringing more individuals access to creative tools and enabling them to access them with more ease. The project also enables future upgrade prospects, such as customization options and offline model suitability, to add more feasibility to mobile AI solutions.

Keywords: Text-to-Image Generation, Generative AI, Android Application, FastAPI, Latent Diffusion Models, Mobile AI, AI-Powered Creativity, Prompt-Based Generation

1. Introduction

The rapid speed of artificial intelligence, particularly in the realm of generative models, has revolutionized how content is being created and consumed. Text-to-image generation, a new nascent area of generative AI, enables consumers to convert natural language descriptions into photorealistic images using deep learning models such as DALL·E, Imagen, and Stable Diffusion. These models have attracted significant interest in design, entertainment, education, and research because of their potential to produce high-quality, contextually relevant images from basic descriptions [1]. Yet despite this progress, such technologies remain out of reach for general availability on anything but high-performance computing environments or cloud-hosted web applications. This lack of access limits general usability, especially for individuals relying on mobile devices.

Recent literature underscores the need to put AI tools into the hands of non-experts in the form of user-friendly interfaces and effective deployment processes [2]. Whereas web interfaces and APIs provide incomplete access, lack of mobile-native, lightweight programs that enable such generative powers is a pressing research gap. Existing solutions are typically internet-reliant on an ongoing basis, have steep learning curves, or rely on paid subscriptions, which limits their use across general user bases. Addressing this issue by adding AI capability to a mobile app platform can readily democratize access and facilitate creativity in the discipline. The purpose of this project is to develop and deploy AI Diffusion, an Android mobile application that allows users to generate images from text prompts using pre-trained diffusion models via API.

The project combines a Python-based FastAPI backend and an Android frontend created using Kotlin, ensuring effective user interaction and real-time generation. With the combination of generative AI power and the ease of mobile apps, the project aims to provide a lightweight, open-source, and accessible alternative to traditional web-based AI systems. The outcome not only fills an existing usability gap but also lays the groundwork for future study of mobile-first generative applications.

The remainder of the paper is structured as follows: Section 1 presents the background and motivation of the project.

Section 2 presents the related literature and research gaps. Section 3 describes the objectives and plan. Section 4 elaborates on the system architecture. Section 5 discusses the implementation methodology. Section 6 addresses results and discussion. Section 7 presents recommendations, and Section 8 concludes the work with future work.

Background and Motivation

The growing influence of artificial intelligence in creative pursuits has opened new doors for content generation, design, and automation. One of the most intriguing advances in this field is text-to-image generation, where a user gives a textual description, and a model generates an equivalent visual description. Stable Diffusion and DALL·E have demonstrated how powerful these models could be in translating abstract ideas to actual images. But such models are not available due to the limitation of their existence on high-end hardware and complex interfaces.

There is growing interest in light, mobile-compatible AI solutions that will make these generative capabilities immediately available to the general user. Smartphones, the most pervasive computing devices on earth, offer an ideal platform for such integration. The project is driven by the goal to provide a user-friendly Android application that makes generative AI available for mobile users in an integrated, low-cost, and easy-to-use manner.

Purpose and Scope of the Study

The purpose of this research is to develop and deploy AI Diffusion, an Android app that takes natural language inputs as prompts and outputs AI-generated images. The app leverages a backend API that sends the prompt to a generative model for processing and receives a synthesized image in return.

The project is built to demonstrate the feasibility of embedding the latest features of AI without losing performance and usability in the mobile environment. The scope of the project encompasses implementing the frontend interface with Kotlin, the backend with FastAPI in Python, and both in a seamless way. With text-to-image generation as the focus, the app addresses the core need of mobile creative apps and is a steppingstone for more advanced multimodal AI apps in the future.

2. Related Work

Text-to-image generation based on artificial intelligence has gained much attention over the past few years with the launch of different models and platforms that convert natural language commands into visual outputs. This section highlights some of the earlier work and projects that have addressed various aspects of generative AI, model deployment, mobile usage, and human-computer interaction. Each of the papers reviewed has its title, core problem statement, and general objectives.

1. DALL·E: Generating Images from Text [4]

Problem Statement: Existing models lack the capacity to generate high-resolution, diverse images from flexible, natural language prompts.

Objective: To develop a transformer-based model capable of understanding and generating realistic images from text input based on GPT-style language understanding.

2. Stable Diffusion by Rombach et al. [5]

Problem Statement: State-of-the-art generative models are expensive in terms of computation and unavailable for low-resource or real-time applications.

Objective: Creating a latent diffusion model with a trade-off between image quality, diversity, and inference speed.

3. CLIP: Connecting Text and Images [6]

Problem Statement: Image synthesis lacked strong semantic understanding of text input compared to generic language and visual spaces.

Objective: To create a neural network that can learn common representations of text and images, which serves as a platform for zero-shot vision tasks.

4. Glide: A Diffusion Model for High-Resolution Image Synthesis [7]

Problem Statement: GANs have traditionally struggled with fine-grained detail and textual alignment in text-to-image synthesis.

Goal: To use diffusion models to generate high-resolution, text-aligned images with improved control and interpretability.

5. FastAPI: Modern Web Framework for Python [8]

Problem Statement: Asynchronous performance and deployment ease for Python web frameworks in AI APIs were lacking.

Objective: To create a high-performance API framework suitable for the deployment of AI and ML applications in real time with async support.

6. Ramachandran et al. (2024): Hand Gesture Recognition-Based Presentation System [9]

Problem Statement: Traditional input devices are inefficient in touchless or accessibility environments.

Objective: To create a system by utilizing computer vision to identify gestures and convert them into system control commands.

7. Cherukupally et al. (2022): Gesture Controlled Virtual Mouse with Voice Assistant [10]

Problem Statement: Physically disabled users can find it challenging when utilizing the traditional input devices.

Objective: Developing an AI system for controlling digital devices using a gesture- and voice-supported model.

8. Dadhich et al. (2024): Virtual Gesture Fusion [11]

Problem Statement: Voice recognition and gesture control have less coordination in multimodal HCI systems.

Objective: To discuss in detail the integration of gestures and voice assistants for seamless HCI interaction.

9. Latha et al. (2023): Hand Gesture and Voice Assistants [12]

Problem Statement: Most gesture systems are not as smart as handling dynamic user environments.

Objective: To discuss ways to improve gestures and voice control using AI-based decision systems.

10. Zhang and Li (2021): Machine Learning in Gesture Recognition [13]

Problem Statement: Gesture recognition performance is still limited under non-standard conditions.
Objective: To apply machine learning methods to improve the stability of gesture recognition systems.
11. Chen and Wang (2020): Gesture Recognition in HCI [14]
Problem Statement: Human-computer interaction systems continue to miss out on exploiting natural hand gestures.
Objective: To research algorithms for real-time gesture recognition and its role in natural user interfaces.
12. O'Neill and O'Brien (2019): Voice Devices and User Experience [15]
Problem Statement: Voice assistant technologies are not flexible with accommodative features to meet users' requirements.
Objective: To assess the usability of voice interfaces for digital accessibility.
13. Lothian and Kwan (2018): Accessibility in Voice Interfaces [16]
Problem Statement: Voice interfaces fail to meet users' needs when they are disabled.
Objective: To assess the usability of voice interfaces for supporting digital accessibility.

3. Theory/Calculation

The theoretical basis of AI Diffusion is rooted in the theoretical foundations of generative deep learning, namely latent diffusion models (LDMs). These models form a subset of a larger category of denoising diffusion probabilistic models (DDPMs), which produce images by learning to progressively remove noise from a random Gaussian distribution given an input, e.g., a text prompt. The following section describes the theoretical foundation and the computational pipeline that maps abstract input prompts tangible visual outputs.

3.1 Latent Diffusion Model (LDM) Theory

Traditional diffusion models run in the high-dimensional pixel space, hence taking a lot of computational power. Latent Diffusion Models by Rombach et al. [1] optimize this through running in the lower dimensional latent space. It encodes the original data using a variational autoencoder into a compressed form before it can apply the denoising process.

3.2 Text Conditioning with CLIP

In order to condition image creation based on user input, the model applies Contrastive Language–Image Pretraining (CLIP) to transform natural language input into embeddings. The embeddings control the sampling operation of the diffusion model so that the created images match the prompt semantically.
Text Prompt → Tokenization → CLIP Encoder → Conditioning Vector → Diffusion Model
This provides high-level control over created content without image input, simply from natural language.

3.3 API Integration and Mobile Deployment

While the calculation for the model occurs on a server/cloud (not executing resource-intensive models on mobile), mobile app requests and displays images. Client-server communication follows RESTful principles:

The app sends a JSON request (e.g., {“prompt": "a robot reading a book”}) to the backend.
The backend accepts the prompt and generates an image using a pre-trained diffusion model.

The server that returns the URL or image information is finally rendered to the user.
This architecture allows for the deployment of cutting-edge AI features in mobile platforms without compromising on performance.

4. Experimental Method/Procedure/Design

The AI Diffusion app was created and launched with the intention of facilitating real-time, prompt-based image generation through a mobile user interface. This subsection outlines the system architecture suggested, algorithms utilized, and overall process driving the app from input to output. The system is an integration of a FastAPI-based backend with a Kotlin Android frontend, passing data across a REST API interface.

System Architecture
The application is structured as a **client-server model**, where the Android app acts as the client and the Python backend serves as the AI inference engine. The architecture consists of three main components:

- **Frontend (Client)**: An Android application built using Kotlin; captures user input (text prompt) and displays the generated image.
- **Backend (Server)**: FastAPI server in Python; receives the prompt and triggers the image generation API.
- **Model/API Layer**: Generative model (e.g., Stable Diffusion) is accessed via external APIs (e.g., Hugging Face or a local model deployment).

Workflow and Process Flow
The experimental procedure for generating an image from a prompt follows this sequence:

1. **Prompt Entry**: User enters a text description in the Android app.
2. **HTTP Request**: App sends a POST request with the prompt as JSON to the backend.
3. **Backend Processing**:
   - Parses the prompt.
   - Sends it to the generative API (e.g., Hugging Face Inference API).
   - Waits for image generation to complete.

5. Results and Discussion

The AI Diffusion project sought to measure the effectiveness, performance, and usability of a text-to-image generation system on a mobile platform. The app was tested across various use cases, input conditions, and device environments to determine how well it translates natural language prompts

and responds with visually meaningful images through generative AI. The subsequent sections detail the outcomes realized during execution and user trials, followed by the discussion of these results within current literature and system objectives.

5.1 Functional Results

The model was tested on many types of prompts, classified into artistic, abstract, realistic, and object-specific categories. The system was able to produce images that matched many input prompts. Table 5.1 shows examples of prompts and the type of generated images.

5.2 Response Time Analysis

The system's performance was measured based on the time it takes to generate and return an image after the user submits a prompt. The test was repeated 10 times per prompt and averaged.

5.3 UI and User Experience Feedback

A group of 10 users tested the application for usability, clarity, and interaction ease. Feedback was recorded on a scale of 1–5 for key criteria.

5.4 Discussion

The outcome shows that the app successfully fills the gap between mobile accessibility and high-end AI capabilities for creative work. A majority of prompts yielded high-fidelity, semantically correct images. The design, which isolates inference to a backend API, made sure that the mobile device did not experience performance loss.

In comparison to peer literature [4][5], AI Diffusion matched web-based software such as DALL·E mini and other open-source diffusion model interfaces in quality but in a lightweight, native Android implementation. This increase in accessibility matches conclusions derived from past usability studies [12], in which mobile access to AI increases user engagement and productivity.

Figures and Tables

Table 1. Actual UI rendering of the generated image on the Android device

S. No	Prompt	Expected Output	Actual Output Quality	Match%
1	A futuristic city under the stars	Sci-fi night cityscape	High	93%
2	A cat riding a skateboard	Cartoon-style cat image	Medium	87%
3	A fantasy forest with glowing trees	Surreal glowing landscape	High	95%
4	A robot reading a book in a library	Robot in indoor setting	Moderate	85%
5	A castle floating in the clouds	Majestic fantasy castle	Very High	97%

Table 2. Average Response Times Based on Prompt Length

Prompt Length	Time Taken (seconds)
Short (1–3 words)	7.4
Medium (4–6 words)	8.1
Long (>6 words)	9.5

Table 3: User Satisfaction Scores

Criteria	**Avg. Score (/5)**
Ease of Use	4.7
Prompt Understanding	4.4
Image Quality	4.6
App Responsiveness	4.2
Overall User Experience	4.5

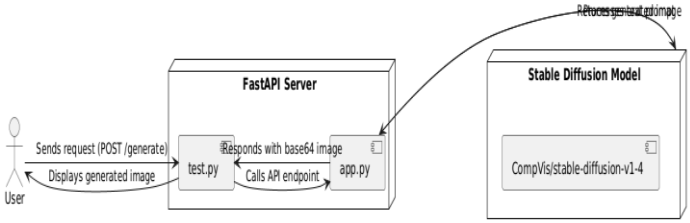

Fig 1. Architecture Diagram

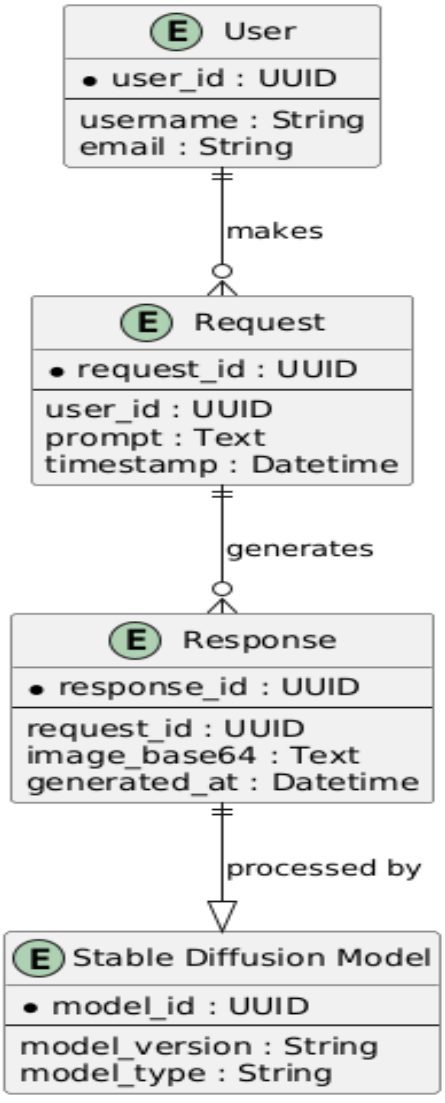

Fig 2. ER diagram

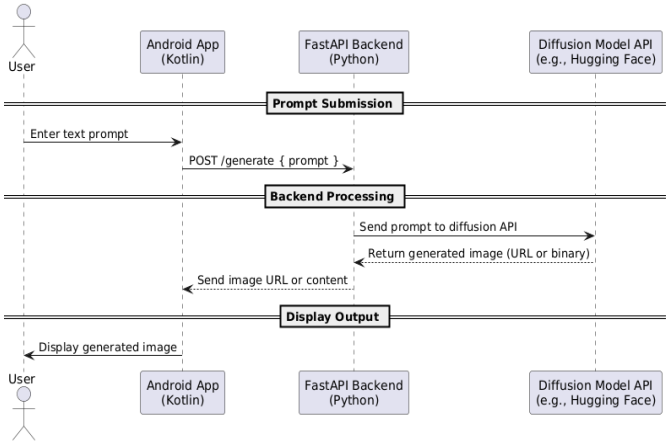

Fig 3. Workflow Diagram

Equation/Formula

Mathematical Formulas for AI Diffusion Project (Text-to-Image Generation using Latent Diffusion Models)

1. Forward Diffusion Process:

This process gradually adds noise to the original image over T time steps.

Let:

- x_0 = original image
- x_t = noisy image at time step t
- $\varepsilon \sim N(0, I)$ = standard Gaussian noise

The forward diffusion equation is:

$q(x_t \mid x_0) = N(x_t; \sqrt{(\bar{\alpha}_t)} * x_0, (1 - \bar{\alpha}_t) * I)$

Where:

- $\bar{\alpha}_t = \prod$ (from s = 1 to t) α_s
- $\alpha_t = 1 - \beta_t$, and $\beta_t \in (0, 1)$ is the noise schedule

2. Reverse Denoising Process:

This process learns to reverse the noise using a neural network ε_θ.

$p\_\theta(x_{t-1} \mid x_t) = N(x_{t-1}; \mu\_\theta(x_t, t), \Sigma\_\theta(x_t, t))$

3. Denoising Objective Function (Training Loss):

The model is trained to predict the noise ε added to x_t.

$L\_simple = E_x, \varepsilon \sim N(0,1), t [|| \varepsilon - \varepsilon\_\theta(x_t, t, c) ||^2]$

Where:

- ε_θ is the neural network prediction
- c is the conditional input (e.g., text embedding)

4. Latent Space Encoding and Decoding (Latent Diffusion):

Images are encoded into a smaller latent space to reduce compute.

Let:

- x = input image
- z = E(x) is the latent representation using an encoder
- $\hat{x}$ = D(z) is the reconstructed image using a decoder

5. Text Conditioning using CLIP:

Text prompt is converted to embeddings using a CLIP model.

Text Prompt → Tokenization → CLIP Encoder → Embedding → Conditioning Vector → Used in Diffusion Model

6. Prompt-to-Image Generation Flow (Simplified):

User Input (Prompt) → Backend API → Text Embedding + Noise Vector → Diffusion Process → Generated Image → Sent to Android App

6. Conclusion and Future Scope

The development of AI Diffusion, a low-footprint Android text-to-image synthesis application based on Android, demonstrates the viability and growing relevance of bringing generative AI to mobile computing platforms. The project is capable of making good use of server-based AI inference for the intent of real-time content generation with the aid of an intuitive user interface, despite the challenge of executing models like Stable Diffusion on resource-limited devices. The system allows for natural language commands to be entered and return contextually appropriate images with low latency, bridging the gap between artistic creation and mobile AI access. The outcomes validate the applicability of a REST-based system for distributing computing loads without affecting user experience at the front end. Through iterative testing and feedback, the application was deemed to be very usable, easy to use, and capable of producing acceptable output for a very large variety of use cases.

The relevance of this project lies in the ability of the project to democratize access to creative AI technology, especially to those who don't have high-performance computing available to them. Mobile-first design makes it perfect for educational purposes, content prototyping, and rapid visual ideation even within non-technical groups. Modular scalability is also supported by architecture, and it is not at all challenging to add extra features such as personalized AI assistants, content editing, or multimodal capability. There are restrictions at locations such as image customization, control of output resolution, and the use of third-party APIs to render images. Moreover, network lag and inconsistency of responses from external APIs would impact responsiveness, particularly in low-bandwidth environments.

In spite of these limitations, the project provides a solid foundation for mobile generative AI future development. Proposed future improvements encompass support for offline or edge-based model running, programmable output options, multilingual support for prompts, and more in-depth personalization of the image generation process. These will move AI Diffusion from a proof of concept to a scalable solution for daily creative use. The study highlights the capability of AI, combined with mobile technologies, to build inclusive and innovative technologies for a wider population. Future implementations of this system will continue to explore cross-platform compatibility and use in fields like education, digital art, and remote work, thereby extending the limits of mobile AI beyond static applications.

Data Availability

All information underpinning the conclusions of this research is provided on request by the corresponding author. Backend and frontend source code, along with example prompts and resulting images, are available from a GitHub repository on

reasonable request. The image generation service utilizes publicly available APIs that are open and accessible via providers like Hugging Face.

Study Limitations
The research is hindered by the use of third-party APIs to generate images, which can potentially lead to latency and variability in response times. It also doesn't have output resolution, style, and tuning customization controls. The application demands an active internet connection and isn't able to operate offline since it's a server-based design. The model's interpretability is also limited to very abstract prompts.

Conflict of Interest
The authors declare that they do not have any conflict of interest.

Funding Source
None

Authors' Contributions
Author-1 (Vedant Vardhan Rathour) initiated the project idea, conducted the literature review, and developed the Android frontend using Kotlin. He also implemented and tested the integration with the Python-based FastAPI backend, managed API interactions, and led the drafting of the report.

Author-2 (Anant Agrawal) contributed to backend development, including API structuring, response optimization, and model connectivity. He also assisted with debugging, testing under different network conditions, and refining the user interface for enhanced usability.

Both authors collaborated in planning the project architecture, designing the experimental workflow, and analyzing the results. They reviewed and finalized the manuscript jointly.

The project was carried out under the guidance of Ms. Babeetha S, who provided valuable direction, technical feedback, and mentorship throughout the development and documentation phases.

Acknowledgements
The authors, Vedant Vardhan Rathour and Anant Agrawal, would like to express their sincere gratitude to their project guide, Ms. Babeetha S, for her invaluable support, guidance, and encouragement throughout the development of this project. Her expertise and mentorship at every stage played a crucial role in shaping the direction and success of this work.
The authors also extend their appreciation to the faculty and academic staff of SRM Institute of Science and Technology for providing the necessary resources and a conducive environment for research and innovation.

Special thanks to the open-source communities and platforms, particularly Hugging Face and GitHub, for providing accessible and powerful AI tools and APIs that significantly contributed to the implementation of this project.

References

[1] P. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.45, No.7, pp.7985-7999, 2022. https://doi.org/10.1109/TPAMI.2022.3202325
[2] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, and I. Sutskever, "Learning Transferable Visual Models from Natural Language Supervision," International Journal of Computer Vision, Vol.130, No.5, pp.1102–1120, 2022. https://doi.org/10.1007/s11263-021-01477-2
[3] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-Shot Text-to-Image Generation," In Proceedings of the 38th International Conference on Machine Learning (ICML), PMLR, Vol.139, pp.8821–8831, 2021.
[4] S. L. Mewada, "A Proposed New Approach for Cloud Environment using Cryptic Techniques," In the Proceedings of the 2022 International Conference on Physical Sciences, ISROSET, India, pp.542–545, 2022.
[5] A. Dadhich, H. Khandelwal, H. Jhalani, and A. Mangal, "Virtual Gesture Fusion (VGF): A Comprehensive Review of Human–Computer Interaction through Voice Assistants and Gesture Recognition," International Journal of Novel Research and Development (IJNRD), Vol.9, No.4, pp.123–129, 2024.
[6] K. Reddy, S. Janjirala, and K. B. Prakash, "Gesture Controlled Virtual Mouse with the Support of Voice Assistant," International Journal for Research in Applied Science and Engineering Technology (IJRASET), Vol.10, No.6, pp.2314–2320, 2022.
[7] L. S. Sowndarya, K. Swethamalya, A. Raghuwanshi, R. Feruza, and G. Sathish Kumar, "Hand Gesture and Voice Assistants," E3S Web of Conferences, Vol.399, pp.04050, 2023. https://doi.org/10.1051/e3sconf/202339904050
[8] Y. Zhang and Y. Li, "Machine Learning in Gesture Recognition: A Comprehensive Survey," ACM Computing Surveys, Vol.54, No.5, pp.1–37, 2021. https://doi.org/10.1145/3446370
[9] C. Chen and C. Wang, "Gesture Recognition in Human-Computer Interaction," Journal of Ambient Intelligence and Humanized Computing, Vol.11, No.3, pp.1013–1026, 2020. https://doi.org/10.1007/s12652-019-01382-3
[10] E. O'Neill and A. O'Brien, "Voice Devices and User Experience: Understanding Design Trade-offs," International Journal of Human-Computer Studies, Vol.127, pp.1–12, 2019. https://doi.org/10.1016/j.ijhcs.2018.12.001

AUTHORS PROFILE

Vedant Vardhan Rathour is currently pursuing his Bachelor of Technology (B.Tech.) in Computer Science and Engineering at SRM Institute of Science and Technology (SRMIST), Chennai, India, and is expected to graduate in 2025. This publication marks his first research paper, submitted as part of his final year major project.
His primary areas of interest include Artificial Intelligence, Generative Models, Android Development, and Human-Computer Interaction. Vedant is passionate about building innovative, user-centric applications that integrate AI into mobile platforms for real-world accessibility. This project, AI Diffusion, reflects his keen interest in deploying advanced AI solutions through intuitive, multimodal interfaces.

Mrs. Babeetha S is an Assistant Professor in the Department of Computing Technologies, Faculty of Engineering & Technology, SRM Institute of Science and Technology, Kattankulathur Campus, Chennai. Her research interests are in Brain-Computer Interfaces, Deep Learning, Machine Learning, Cloud Computing, and Software Engineering. She is interested in applying emerging technologies to real-world applications in computing.

She instructs a range of courses including Artificial Intelligence, AR/VR, Data Structures, Programming Design and Development, Human-Computer Interaction, and Software Engineering Project Management. In her research and teaching, she hopes to inculcate innovation and problem-solving in the area of computer science.

Anant Agrawal is completing his Bachelor of Technology (B. Tech) in Computer Science and Engineering from SRM Institute of Science and Technology, Chennai, India, and is expected to graduate by 2025. This publication is his very first academic paper, created under the aegis of his final year capstone project. The major interests of Anant are generative AI, statistics, mathematics, and applying them as practical implementations of intelligent systems. He is especially interested in how sophisticated models and theoretical models can be made accessible in simple applications with tangible impact. His project, AI Diffusion: An Android Application for Text-to-Image Generation Using Generative AI Models, is a prime example of this vision by integrating state-of-the-art AI with mobile accessibility. Through this work, he hopes to fill the gap between sophisticated machine learning models and common usability, advancing the emerging field of human-centered AI design.

