Digital Doppelganger: AI-Powered Conversational Personality Simulation

Sonali Pakhmode, Chaitanya Haryan, Satyam Prajapati, Khandu Sontakke, Gaurav Kalsait

Department of Information Technology, Vasantdada Patil Pratishthan's College of Engineering and Visual Arts, Sion, Mumbai, India

DOI: https://dx.doi.org/10.51584/IJRIAS.2026.11030015

Received: 04 March 2026; Accepted: 10 March 2026; Published: 28 March 2026

ABSTRACT

This study about the project presents the design and development of an AI-based Digital Doppelganger aimed at modeling and simulating an individual's personality, communication patterns, and cognitive characteristics. The proposed system integrates natural language processing techniques, similarity analysis, and adaptive learning mechanisms inspired by transformer-based architectures such as LLaMA2. This approach prioritize consistent personality, emotional depth, and contextual awareness over simple task execution or information retrieval. By continuously learning from user interactions and real-time data, the system is able to reproduce human-like tone, intent, and linguistic behavior while preserving coherence across conversations. The Results show the unique personality of an individual by its digital twin. Experimental results indicate strong performance in intent recognition and contextual relevance, suggesting that the Digital Doppelganger framework can be effectively applied in personalized virtual assistants, digital identity representation, and research on advanced human–AI interaction.

Index Terms - Digital Doppelganger, Artificial Intelligence, Digital twin, Personality Simulation, NLP, Machine Learning, Adaptive Learning, LLaMA2.

INTRODUCTION

In recent years, significant progress in artificial intelligence and deep learning has reshaped the ability of machines to model and interpret human behavior. Advances in transformer-based models such as GPT and LLaMA2 have made it possible for systems to capture intricate language patterns, emotional nuances, and contextual dependencies with greater precision. As a result, modern AI applications are no longer limited to basic task automation but are increasingly capable of supporting adaptive and expressive forms of communication.

Leveraging these advances, a Digital Doppelganger emerges as an intelligent system that mirrors a person's personality , communication style, and thinking traits. By integrating natural language processing, similarity analysis, and adaptive learning mechanisms, the proposed system facilitates context-aware interactions that improve over time. This approach addresses the limitations of traditional rule-based or static chatbots and represents a meaningful step toward personalized and emotionally responsive human–AI interaction.

LITERATURE REVIEW

Recent research has explored the creation of digital twins and AI-driven personality modeling. Smith and Kumar [1] explored large language model–driven doppelgänger systems capable of generating human-like responses, while Patel et al. [2] examined the advantages and limitations of digital twin frameworks for enabling real-time contextual learning. A comprehensive review by Tanaka and Gomez [3] classified digital twin applications across multiple domains, highlighting their growing relevance beyond industrial systems.

More recent work has extended digital twin concepts toward human-centered AI. Mehta and Rao [4] and Das and Singh [6] focused on adaptive and cloud-based learning architectures to support personalized AI behavior.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue III March 2026

Wang [5] investigated the application of digital doppelgängers in educational environments, emphasizing their role in individualized learning experiences. In parallel, Green and Thomas [7] and Keller [8] discussed ethical, social, and identity-related concerns associated with AI-driven personality replication. Together, these studies lay a solid basis for the proposed Digital Doppelganger model, focusing on personality awareness, contextual insight, and adaptive interaction.

System Architecture

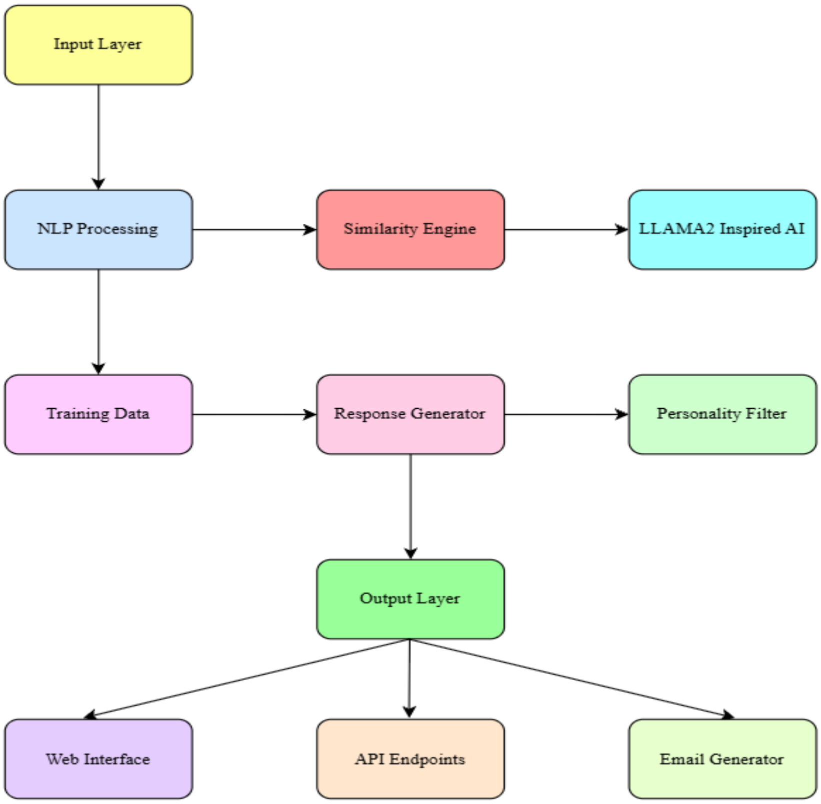

Fig. 1: Digital Doppelganger System Architecture

Digital Doppelganger system follows a three-layer design comprising the frontend, backend, and data layers. This modular design supports scalability, maintainability, and efficient interaction between system components.

A. Frontend Layer:

The frontend is developed using HTML5, CSS3, and JavaScript and provides an interactive user interface for chat and email-based communication. A glassmorphism design is used to improve visual clarity and user engagement. The interface is fully responsive, ensuring compatibility across multiple devices, and incorporates real-time interaction features to support seamless communication.

B. Backend Layer:

The backend is implemented in Python and serves as the core processing unit of the system. It provides RESTful APIs manage conversation requests, generate emails, and monitor system health. Each API endpoint interacts with the NLP processing pipeline to generate adaptive responses based on detected user intent and contextual information.

C. Data Layer:

The data layer is responsible for storing interaction logs in JSON format. These logs create an ever-growing learning corpus that allows the system to adapt in real time. By analyzing historical interaction data, the model incrementally improves response accuracy, coherence, and consistency over time.

Figure 1 illustrates the high-level system architecture, showing the flow of data from user input through NLP processing to adaptive response generation.

METHODOLOGY AND ALGORITHMS

The methodology of the Digital Doppelganger system is based on four core components: natural language processing, similarity calculation, intent classification, and adaptive learning.

A. NLP Pipeline: Incoming text is first cleaned and tokenized, after which TF-IDF–based vector representations are generated. These vectors are then passed to the similarity calculation module for further analysis.

B. Similarity Calculation Engine: This component combines cosine similarity, Jaccard index, and keyword-based weighting to improve matching accuracy. Domain specific keywords such as identity, academic, and technology are given higher weights to improve relevance. The final similarity score is used to select the most appropriate response template.

C. Intent Classification System: Intents are identified using pattern recognition and probabilistic estimation and categorized into domains such as identity, project, academic, and greeting. Response generation is guided by confidence thresholds, with adaptive learning triggered when confidence drops below 0.3.

D. Learning Subsystem: The system regularly saves newly identified interaction patterns after every five interactions. Unrecognized queries are analyzed and incorporated into future dataset updates, enabling continuous refinement of system performance.

E. Dataset Collection and Preprocessing : The dataset used in this study consists of 768 conversational pairs designed to simulate real-world human interaction. The data was curated through a combination of manually constructed dialogues and context-aware synthetic generation to ensure diversity in tone, intent, and subject matter. Each conversation pair includes a user query and a corresponding response aligned with a predefined personality profile. The dataset covers multiple domains, including personal identity, academic discussions, project-related queries, and general conversational exchanges. Preprocessing steps were applied to enhance data quality, including text normalization, removal of stop words, tokenization, and lowercasing. Additionally, noise such as punctuation inconsistencies and redundant spacing was eliminated. These steps ensured that the input data was structured and suitable for vectorization using TF-IDF techniques.

Implementation And Results

A set of 768 carefully selected conversation pairs was used for training and evaluation. Response accuracy varied across confidence levels, with high-confidence responses achieving 97% accuracy, medium-confidence responses 87%, and low-confidence responses 72%. Intent detection recorded a precision of 90.4% and a recall of 88.7%.

As illustrated in Fig. 2, the training results show consistent performance gains, with accuracy rising and loss decreasing across epochs. The analysis also demonstrates improved response authenticity with larger training datasets and an average similarity score of 0.796, suggesting strong alignment between generated and expected responses.

Performance evaluation further showed sub-second response times across 100 simulated sessions, confirming system efficiency. Additionally, the interface shown in Fig. 3 enables intuitive, real-time interaction, strengthening the system’s focus on usability and natural communication.

The performance of the system was evaluated using standard metrics including accuracy, precision, and recall. Accuracy was calculated as the ratio of correctly predicted responses to the total number of predictions. Precision measures the proportion of relevant responses among the predicted outputs, while recall indicates the system’s ability to retrieve relevant responses from the dataset.

These metrics were computed using a test subset derived from the original dataset, ensuring that evaluation was performed on unseen data. The results demonstrate that the system achieves higher accuracy at increased confidence levels, indicating reliable intent classification and response generation.

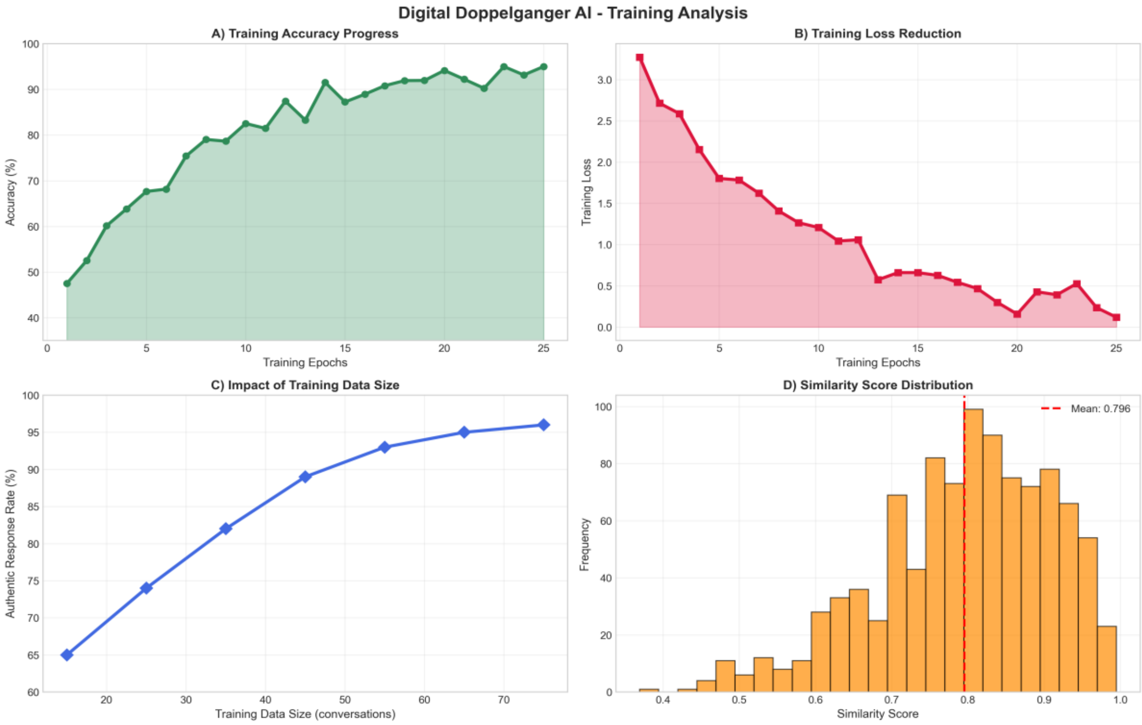

Fig. 2: Training Analysis of Digital Doppelganger

Fig. 3: Digital Doppelganger User Interface

DISCUSSION AND PERFORMANCE EVALUATION

Model	Data Size	5-point	3-point
Surveyed LLM	1,000	0.30	0.38
Surveyed LLM	5,000	0.46	0.65

Proposed Digital Doppelganger	1,000	0.30	0.42
Proposed Digital Doppelganger	5,000	0.67	0.80

Comparison Table : Other LLM vs Digital Doppelganger

This table compares the performance of the surveyed LLM and the proposed Digital Doppelganger at different data sizes. Although both models improve with larger datasets, the Digital Doppelganger achieves higher scores, especially at 5,000 samples - indicating superior response quality and personalization. Overall, the results highlight the effectiveness of the proposed approach in capturing user-specific behavior.

Future Enhancements

Future work will focus on enhancing the system by incorporating deep learning techniques, transformer-based models, and multimodal interaction features. Adding voice-based communication and image recognition is expected to expand the system's interaction capabilities. In parallel, ethical aspects such as user privacy, data ownership, and responsible personality replication will be carefully addressed to ensure transparent and trustworthy AI development.

CONCLUSION

This paper presented the conceptualization, design, and implementation of the Digital Doppelganger, an AI-driven framework for personality simulation. By combining natural language processing, similarity computation, and adaptive learning, the system supports context-aware, personality-consistent interactions. Experimental results demonstrate that the proposed approach effectively supports human-like communication while maintaining computational efficiency. Overall, this work contributes toward the development of personalized AI systems and advances research in human–AI identity modeling.

REFERENCES

1. J. Smith and A. Kumar, "LLM-Based Doppelgänger Models for Human-Like Survey Simulations," IEEE Xplore, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/10758652
2. R. Patel, M. Lee, and T. Chen, "Digital Twin: Benefits, Use Cases, Challenges & Opportunities," ScienceDirect, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S277266222300005X
3. K. Tanaka and L. Gomez, "Digital Twins: A Systematic Review via Topic Modeling," ResearchGate, 2024. [Online]. Available: https://www.researchgate.net/publication/365881019
4. N. Mehta and S. Rao, "AI-Driven Digital Doppelgangers for Virtual Personas," IRJMETS, vol. 3, Mar. 2025. [Online]. Available: https://www.irjmets.com/uploadedfiles/paper/issue_3_march_2025/69488/final/fin_irjmets1742300736.pdf
5. N. Wang, "Learning by Explaining to a Digital Doppelganger," University of Southern California, 2018. [Online]. Available: https://people.ict.usc.edu/~nwang/PDF/ITS2018-Wang.pdf
6. P. Das and R. Singh, "Human Digital Twin: A Comprehensive Survey," Journal of Cloud Computing, 2024. [Online]. Available: https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-024-00691-z
7. M. Green and L. Thomas, "Ethical & Societal Implications of Pre-Mortem AI Clones," arXiv preprint, 2025. [Online]. Available: https://arxiv.org/pdf/2502.21248
8. S. Keller, "Digital Doppelgängers and Lifespan Extension: What Matters?," AI & Society Journal, 2024. [Online]. Available: https://doi.org/10.1080/15265161.2024.2416133