Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 289-291

Swanly AI: An Intelligent and Explainable Matchmaking System Using Structured Data and Machine Learning

[1]Niladitya Sen, [2]Tanishq Prasanna, [3]Dr. Poonam Mishra, [4]Dr. Kranti Kumar Dewangan
[1,2]B. Tech Computer Science & Engineering,
[3]Associate Professor, [4]Head, Department of Computer Science & Engineering
[1,2,3,4]Amity University Chhattisgarh, Raipur, India- 493225

Abstract

Swanly AI is an advanced AI-powered matchmaking platform designed to improve compatibility analysis through structured user data and machine learning techniques. Unlike traditional dating systems that rely on superficial attributes, Swanly AI leverages a chapter-based questionnaire model to capture deep behavioral, emotional, and preference-based insights. These responses are transformed into vector embeddings to enable similarity-based matchmaking. The system employs a periodic batch matchmaking algorithm that identifies compatible pairs every three days. To enhance user engagement and transparency, AI-generated match stories and micro-insights are produced using large language models. The platform is built on a scalable, event-driven architecture using Apache Kafka for asynchronous processing. Additionally, Swanly AI integrates secure user verification using AWS Rekognition for facial comparison, ensures privacy through AES-based chat encryption and AWS S3 encryption for media, and facilitates virtual dates via Google Meet using calendar APIs. By combining structured data modeling, embedding-based similarity, secure infrastructure, and explainable AI, Swanly AI delivers a reliable and scalable matchmaking solution.

Keywords: Artificial Intelligence, Matchmaking Systems, Recommendation Systems, Vector Embeddings, Natural Language Processing, Explainable AI, Microservices Architecture, Event-Driven Systems, User Verification, Secure Communication

1. Introduction

1.1 Background

The rise of digital matchmaking platforms such as Tinder and Bumble has simplified social connections but often at the cost of meaningful compatibility. These platforms rely heavily on surface-level attributes such as images and short bios, leading to low-quality matches and user dissatisfaction.

Advancements in Artificial Intelligence (AI), Natural Language Processing (NLP), and vector embeddings enable deeper analysis of human behavior and preferences. Swanly AI leverages these technologies to move beyond superficial filtering and toward meaningful compatibility analysis.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 289-291

1.2 Overview

Swanly AI collects structured user data through a chapter-based questionnaire system. These responses are stored and transformed into vector embeddings, enabling similarity-based matching using high-dimensional representations.

The system periodically evaluates compatibility using similarity scores and generates AI-driven insights explaining why two users match. The architecture is built using an event-driven model powered by Apache Kafka, ensuring scalability and asynchronous processing.

1.3 Problem Statement

Existing matchmaking platforms fail to capture deeper aspects of compatibility such as emotional intelligence, communication style, and long-term goals. Additionally, they lack transparency in how matches are generated.

This results in:

- Poor match quality
- Low user satisfaction
- Lack of trust in recommendations

There is a need for an AI-driven system that integrates structured data, similarity-based matching, and explainable insights.

1.4 Objectives

The primary objectives of Swanly AI are:

- To design a structured questionnaire-based data collection system
- To represent user preferences using vector embeddings
- To develop a similarity-based matchmaking algorithm
- To generate explainable AI insights for matches
- To build a scalable event-driven architecture

2. Literature Survey

Traditional matchmaking systems relied on rule-based filtering using demographic attributes such as age, location, and interests, which limited personalization and scalability. With the rise of online platforms like Tinder and Bumble, swipe-based interaction improved engagement but continued to emphasize superficial features, resulting in poor long-term compatibility.

To address these issues, machine learning techniques such as collaborative filtering and content-based filtering were introduced. Collaborative filtering predicts user preferences based on similar user behavior [1], while content-based methods rely on profile attributes [2].

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 289-291

However, both approaches suffer from limitations such as the cold-start problem and inability to model complex human relationships.

Recent advancements in Natural Language Processing (NLP) and vector embeddings have enabled richer representation of user preferences. Models like Word2Vec [3] and BERT [4] allow textual and behavioral data to be encoded into high-dimensional vectors, enabling semantic similarity comparisons. These techniques significantly improve matchmaking accuracy compared to traditional feature-based methods.

Furthermore, large language models (LLMs) have enabled explainable AI by generating human-readable insights and recommendations, improving transparency and user trust [5]. However, most systems do not integrate structured data collection, embeddings, and explainability into a unified pipeline.

Scalable system design has also evolved with event-driven architectures such as Apache Kafka, enabling high-throughput and fault-tolerant processing in distributed applications [6]. Despite this, many matchmaking platforms still rely on monolithic or synchronous architectures.

In addition, user trust remains a critical concern. Research highlights the importance of identity verification and secure communication in online platforms, including biometric authentication and encryption mechanisms [7].

Swanly AI builds upon these advancements by integrating structured data collection, embedding-based similarity, explainable AI, secure verification, and event-driven architecture into a single cohesive system.

3. Methodology

Swanly AI is designed as a modular, microservices-based system that enables scalable, secure, and intelligent matchmaking through structured data processing and AI-driven analysis. The system consists of the following core services:

- **Data Collection Service:**
  The data collection service is responsible for gathering user information through a structured, chapter-based questionnaire that captures preferences, personality traits, and behavioral tendencies. This approach ensures consistent and high-quality input data, which is stored in a relational database and forms the foundation for downstream processing and matchmaking.
- **User Verification Service:**
  The user verification service ensures the authenticity of profiles using AWS Rekognition. During onboarding, users upload images that are compared using facial recognition APIs to validate identity. This process minimizes fake accounts and enhances trust and safety across the platform.
- **Data Processing & Vectorization Service:**
  This service transforms collected user data into high-dimensional vector embeddings

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 289-291

using AI models. These embeddings capture semantic relationships between user preferences and are stored in a vector database such as Qdrant, enabling efficient similarity-based matching.

- **Matchmaking Service:**
  The matchmaking service operates as a batch-processing engine that runs periodically, typically every three days. It computes similarity scores between user embeddings, applies weighted scoring across multiple attributes, and filters matches based on user-defined constraints to generate a ranked list of compatible pairs.

- **Event Processing Service:**
  The event processing service utilizes Apache Kafka to enable an event-driven architecture. Matchmaking results and other system events are published to Kafka topics, allowing different services to consume and process them asynchronously, thereby improving scalability, fault tolerance, and system decoupling.

- **AI Insight Generation Service:**

  This service leverages large language models to generate personalized match stories and micro-insights based on user data and compatibility scores. These explanations provide transparency into the matchmaking process and enhance user engagement by making matches more interpretable.

- **Communication & Chat Service:**

  The communication service enables real-time interaction between matched users through a secure messaging system. All chat messages are encrypted using AES-based server-side encryption, ensuring confidentiality and protecting user conversations from unauthorized access.

- **Media Storage Service:**
  The media storage service manages user-uploaded content such as profile images. All media files are securely stored using AWS S3 with encryption enabled, ensuring data protection, durability, and controlled access to sensitive information.

- **Virtual Dating Service:**
  The virtual dating service integrates Google Meet using calendar APIs to facilitate real-time interaction between matched users. It allows seamless scheduling and joining of virtual dates, bridging the gap between digital matchmaking and real-world communication.

- **Notification Service:**
  The notification service is responsible for delivering updates such as new matches, messages, and scheduled events. It operates asynchronously using Kafka events to ensure timely communication without impacting system performance.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 289-291

- **Frontend Service:**
  The frontend service, built using React Native and Next.js, provides an intuitive interface for users to interact with the platform. It enables users to complete questionnaires, view matches and compatibility insights, communicate securely, and manage virtual dates.

4. System Design

Swanly AI is designed as a distributed, microservices-based system that integrates structured data processing, artificial intelligence, and secure communication to enable intelligent matchmaking. The system follows a layered architecture where each layer is responsible for a specific function, ensuring modularity, scalability, and fault tolerance.

4.1 Architectural Overview

The overall architecture consists of multiple loosely coupled services connected through an event-driven pipeline. The system is divided into key layers, including the data acquisition layer, processing layer, matchmaking layer, communication layer, and user interaction layer. Apache Kafka acts as the backbone for inter-service communication, enabling asynchronous data flow and decoupling between components.

4.2 Data Acquisition Layer

The data acquisition layer is responsible for collecting and validating user information. It includes the onboarding module and the structured questionnaire system, which captures detailed user preferences, personality traits, and behavioral attributes. Additionally, this layer integrates the user verification mechanism using AWS Rekognition, where uploaded images are validated through facial comparison to ensure authenticity. All collected data is stored in a relational database for consistency and integrity.

4.3 Data Processing Layer

The data processing layer transforms raw user inputs into meaningful representations. This includes preprocessing, normalization, and feature extraction, followed by vectorization using AI models. The generated embeddings represent users in a high-dimensional space and are stored in a vector database such as Qdrant. This layer enables efficient similarity computation and serves as the foundation for matchmaking.

4.4 Matchmaking Layer

The matchmaking layer is responsible for identifying compatible user pairs. It operates as a batch-processing system that periodically computes similarity scores between user embeddings. The system applies weighted scoring across multiple attributes and filters results based on user-defined constraints. The output is a ranked set of matches, which are then published as events for further processing.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 289-291

4.5 Event-Driven Communication Layer

The system employs Apache Kafka to implement an event-driven architecture. All major actions, including preference updates, vectorization completion, and matchmaking results, are published as events to Kafka topics. This allows different services to consume data asynchronously, ensuring high scalability, fault tolerance, and efficient handling of large user volumes.

4.6 AI Insight Layer

The AI insight layer utilizes large language models to generate human-readable explanations for matches. Based on compatibility scores and user data, the system produces match stories and micro-insights that explain the reasoning behind each recommendation. This layer enhances transparency and builds user trust in the system.

4.7 Communication and Security Layer

The communication layer enables secure interaction between users through a real-time chat system. All messages are encrypted using AES-based server-side encryption to ensure confidentiality. Additionally, user media such as profile images are stored securely using AWS S3 with encryption enabled. This layer ensures compliance with data security standards and protects sensitive user information.

4.8 Virtual Interaction Layer

To facilitate real-time engagement, the system integrates a virtual dating feature using Google Meet. Through calendar APIs, users can schedule and join virtual meetings seamlessly. This layer bridges the gap between digital matchmaking and real-world interaction, enhancing user experience.

4.9 Notification Layer

The notification layer is responsible for delivering updates related to matches, messages, and scheduled events. It operates asynchronously, triggered by Kafka events, ensuring timely communication without impacting the performance of core services.

4.10 Presentation Layer

The presentation layer consists of web and mobile applications built using Next.js and React Native. It provides an intuitive interface for users to interact with the system, complete questionnaires, view matches, access AI-generated insights, and communicate securely.

4.11 Design Considerations

The system is designed with key considerations including scalability, achieved through microservices and Kafka-based event streaming; security, ensured through biometric verification and encryption mechanisms; and explainability, enabled by AI-generated insights.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 289-291

The modular design allows independent scaling and maintenance of services, making the system robust and adaptable to future enhancements.

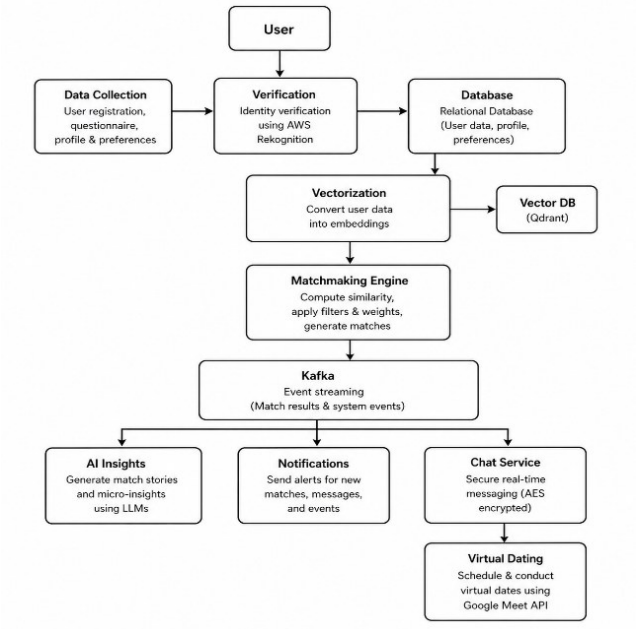

Figure 4.1 (Dataflow diagram)

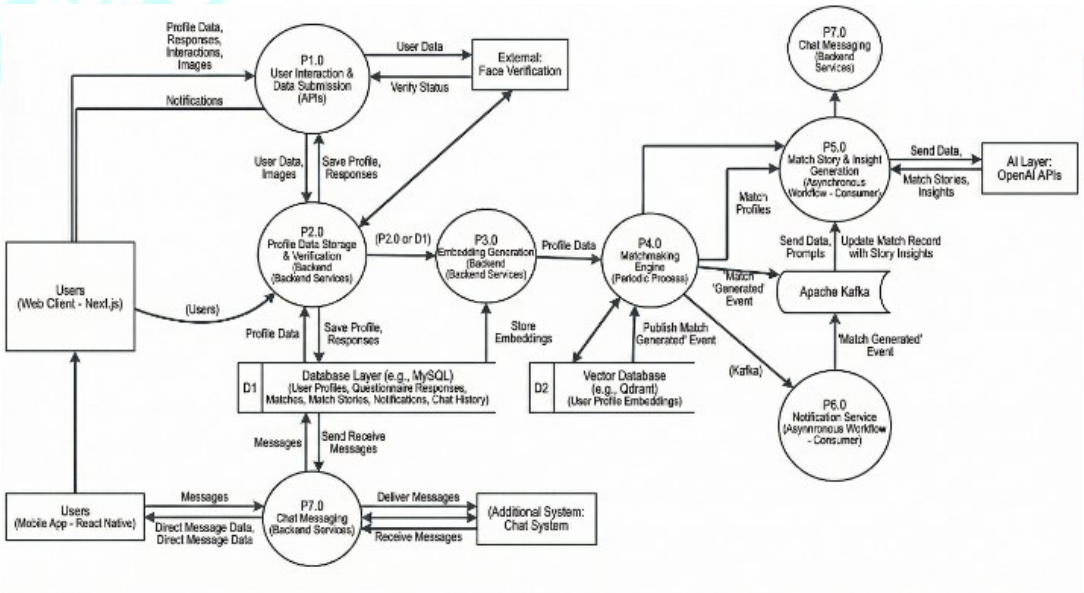

Figure 4.2 (System design)

5. Future Work

Swanly AI presents a strong foundation for intelligent matchmaking; however, several enhancements can further improve its capabilities and real-world impact.

- **Real-Time Matchmaking:**
  The current system operates on a batch-processing model. Future improvements can include real-time matchmaking, where compatibility is computed instantly as user preferences are updated, enabling more dynamic and responsive interactions.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 289-291

- **Adaptive Learning from User Behaviour:**
  Incorporating continuous learning mechanisms based on user interactions, such as chat engagement, match acceptance, and date outcomes, can refine matchmaking accuracy over time using reinforcement learning techniques.
- **Advanced Personality Modelling:**
  Future work can enhance personality representation by integrating psychometric models and behavioural analytics, allowing deeper understanding of user compatibility beyond questionnaire-based inputs.
- **Multimodal Data Integration:**
  Extending the system to analyse multiple data modalities such as voice, facial expressions, and interaction patterns can provide richer embeddings and improve matching quality.
- **AI-Based Digital Twin Simulation:**

  A long-term goal is to create AI-based digital replicas of users that can simulate conversations and interactions. This would allow users to experience simulated dates before real engagement, improving decision-making.
- **Immersive Virtual Dating Environment:**
  Integration with 3D or metaverse platforms can enable immersive virtual dating experiences, providing a more realistic and engaging interaction environment beyond video calls.
- **Enhanced Privacy and Federated Learning:**
  Future implementations can explore federated learning to train models without directly accessing user data, thereby improving privacy and compliance with data protection regulations.
- **Bias Reduction and Fairness Optimization:**
  Ensuring fairness in matchmaking by detecting and mitigating biases in data and algorithms will be critical for creating an inclusive and ethical platform.

6. Results

The Swanly AI system was evaluated based on its ability to generate meaningful matches, ensure system reliability, and provide a secure and engaging user experience. The results demonstrate the effectiveness of combining structured data collection, vector embeddings, and AI-driven insights in a matchmaking environment.

- **Improved Match Quality:**
  The use of structured questionnaires and embedding-based similarity enabled more relevant and compatibility-driven matches compared to traditional attribute-based filtering. Users were matched based on deeper behavioral and preference alignment rather than superficial features.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 289-291

- **Explainability and User Trust:**
  The AI insight generation module successfully produced match stories and micro-insights that explained compatibility in natural language. This improved transparency and helped users better understand and trust the matchmaking process.
- **System Scalability and Performance:**
  The event-driven architecture using Apache Kafka allowed asynchronous processing of matchmaking and notification workflows. This resulted in efficient handling of multiple user operations without performance bottlenecks.
- **Security and Data Protection:**
  The integration of AWS Rekognition ensured reliable user verification, reducing the likelihood of fake profiles. Additionally, AES encryption for chat and AWS S3 encryption for media storage provided strong data security, ensuring user privacy.
- **Seamless User Interaction:**
  The integration of Google Meet via calendar APIs enabled smooth scheduling and execution of virtual dates. Combined with the real-time chat system, this created a complete interaction pipeline from matching to communication.
- **System Reliability:**
  The modular microservices architecture ensured that individual components could operate independently, improving fault tolerance and maintainability of the system.

7. References

[1] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 230–237, 1999.

[2] M. J. Pazzani and D. Billsus, "Content-based recommendation systems," in *The Adaptive Web*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin, Germany: Springer, 2007, pp. 325–341.

[3] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *Proceedings of NAACL-HLT*, pp. 4171–4186, 2019.

[5] T. B. Brown *et al.*, "Language models are few-shot learners," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 1877–1901, 2020.