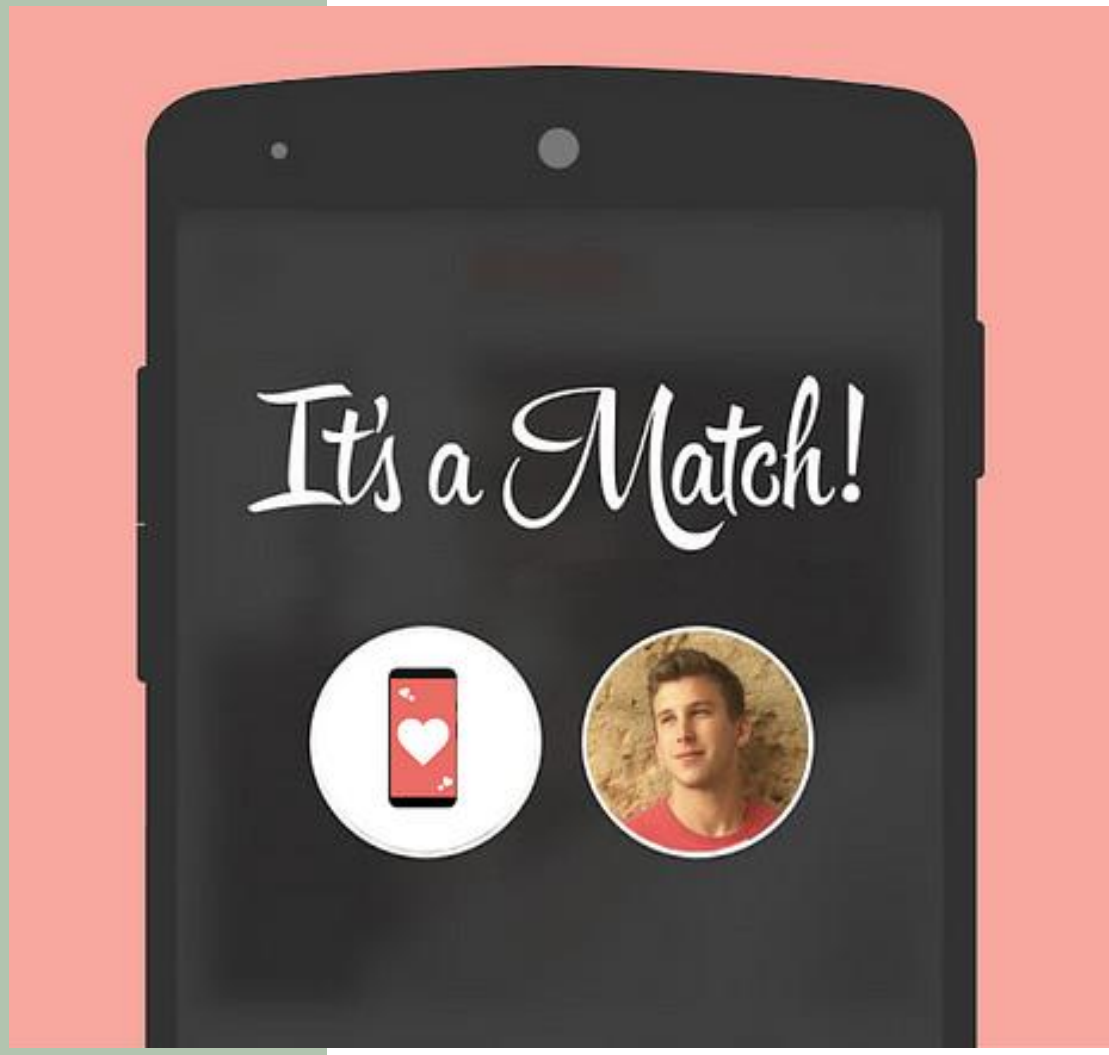
SAMMY SWIPE

Your Match! Your Vibe! Your Story

Group 6
DATA 245 ML Technology
Prof. Shih Yu Chang

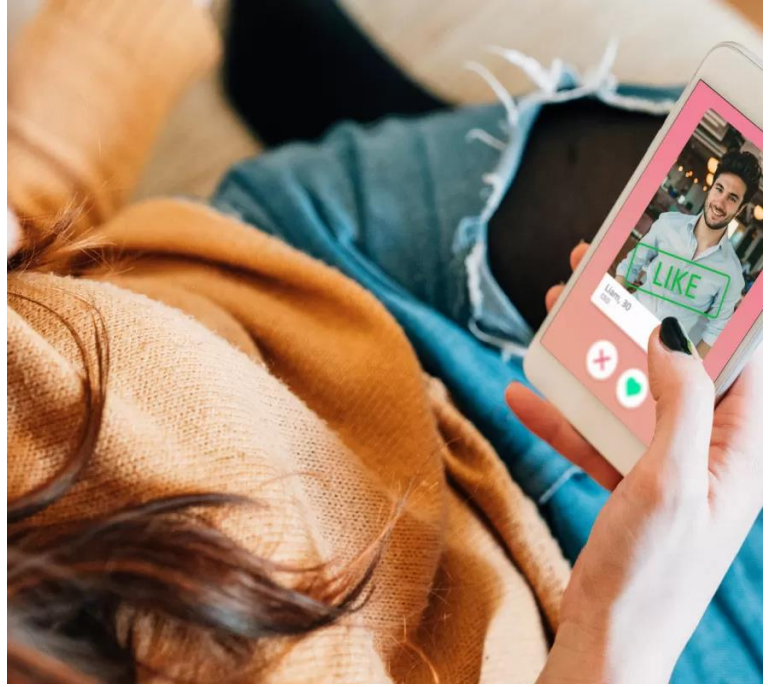


CONTENT OVERVIEW

- Problem Statement
- Objective
- Challenges & Our Approach
- Proposed Solution
- Data Understanding & Exploration
- Feature Engineering
- Model Building
- Evaluation & Metrics
- Deployment
- Future Direction

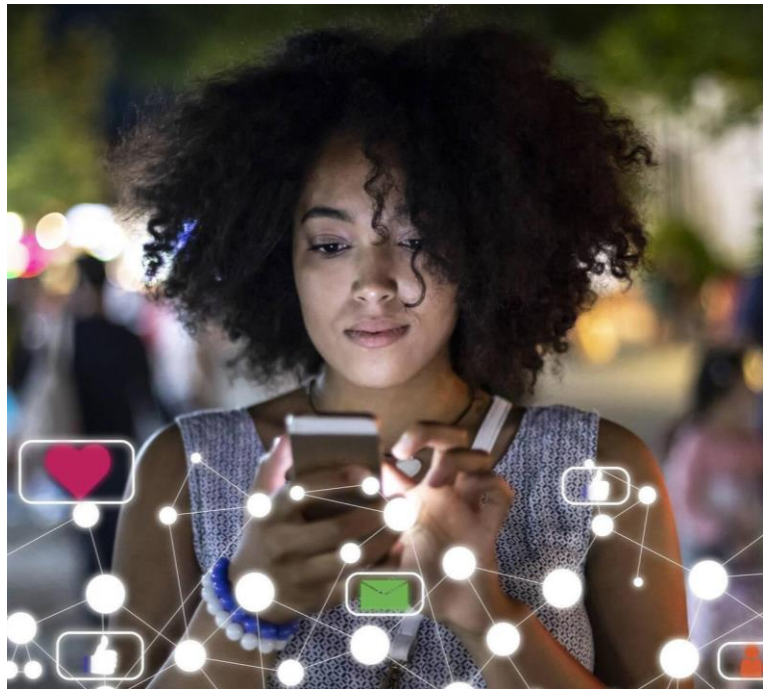
PROBLEM STATEMENT

While over one-third of adults use dating apps, most users spend more time swiping than forming meaningful relationships. This is because traditional matching systems focus on superficial engagement metrics rather than deep compatibility signals. There is a pressing need to rethink matchmaking using more advanced, emotionally intelligent, and behavior-aware approaches.



Key Issues with Traditional Systems

- Prioritize swipe-based engagement over genuine compatibility
- Match users primarily based on proximity and static preferences (age, gender, interests)
- Lack deeper personalization using emotional or intent-based data
- Minimal use of advanced algorithms like graph theory or psychological profiling
- Result in lower match quality and user satisfaction

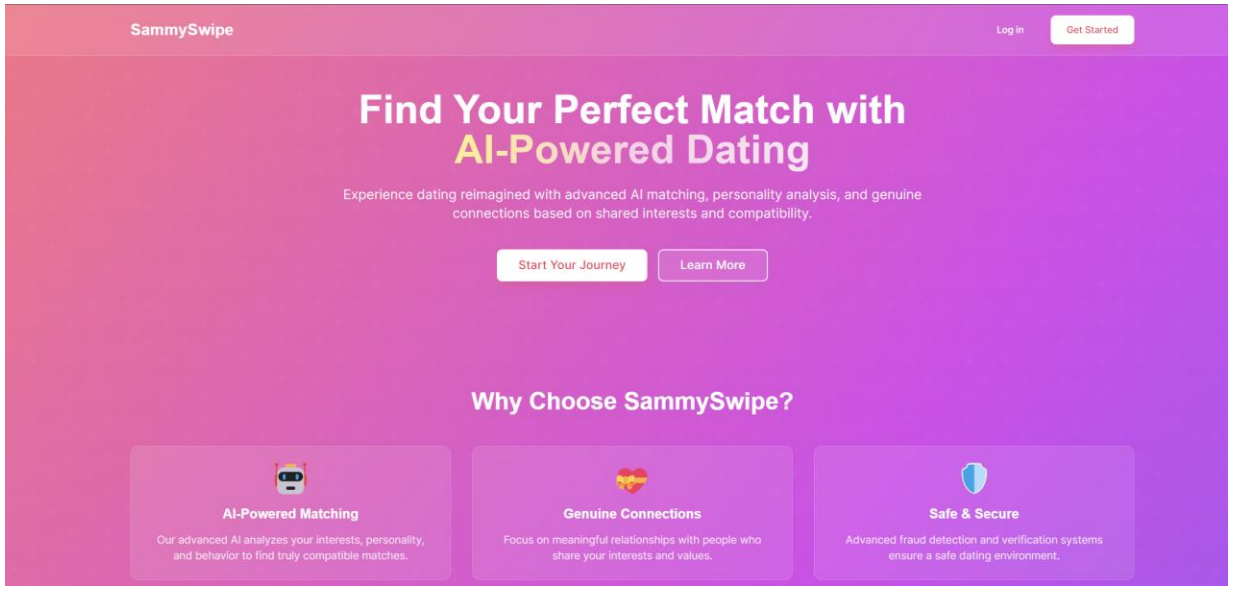


OBJECTIVE

To revolutionize the online dating experience by moving beyond superficial swiping and enabling **genuine, compatible connections** through intelligent matchmaking.

Key Goals:

- Reduce time spent on low-value interactions
- Improve match quality using advanced behavioral and emotional signals
- Leverage graph-based models for meaningful relational context
- Empower users with deeper, trust-driven recommendations
- Create a more emotionally resonant and purpose-driven dating platform



CHALLENGES & OUR APPROACH

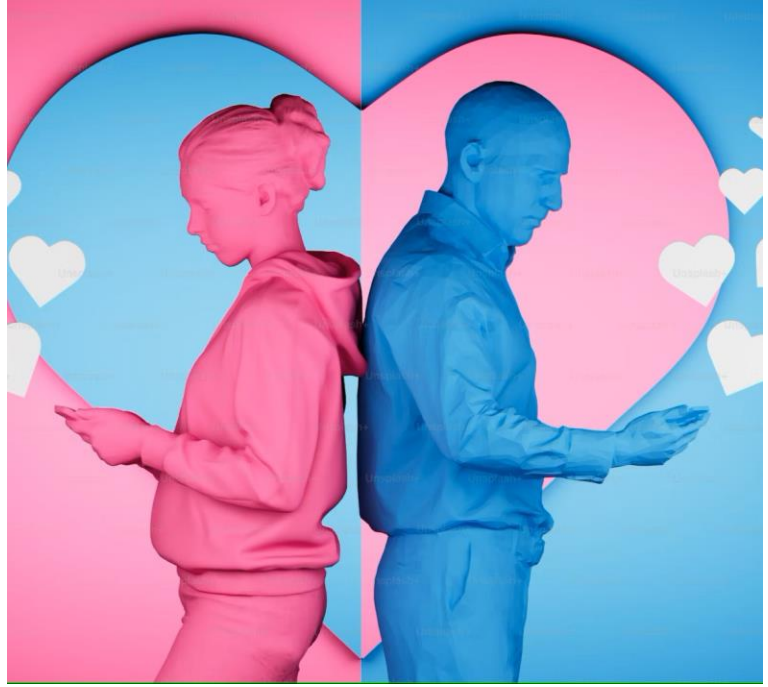
Despite high user engagement, dating platforms fail to convert swipes into meaningful matches due to shallow algorithms and outdated data assumptions

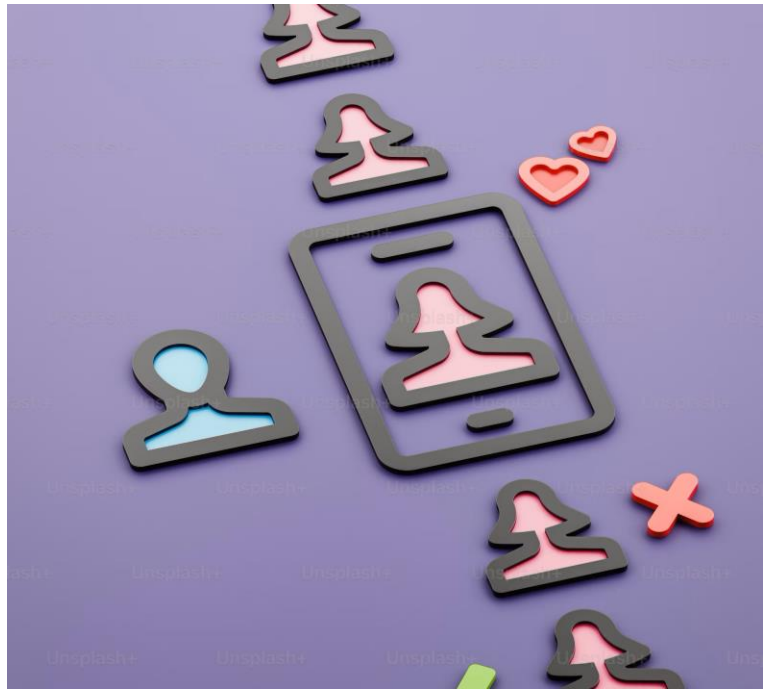
Challenges

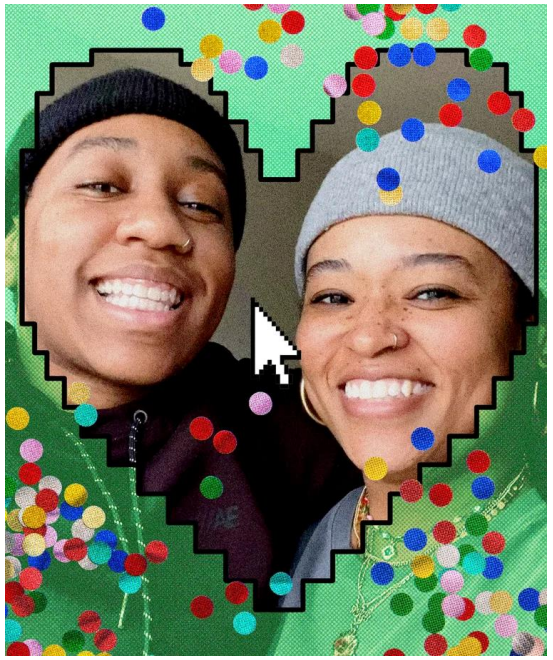
- Overemphasis on visual or location-based preferences
- Low signal-to-noise ratio in user interactions
- Lack of emotional or intent-based personalization
- Absence of social context and trust cues
- User fatigue due to repetitive and aimless swiping

Our Approach

- Implement graph-based recommendation engine to understand relational context
- Integrate intent modeling and emotional profiling to enrich compatibility scoring
- Analyze deep behavioral data (not just swipes) for true match potential
- Continuously refine models using feedback and evolving user behavior







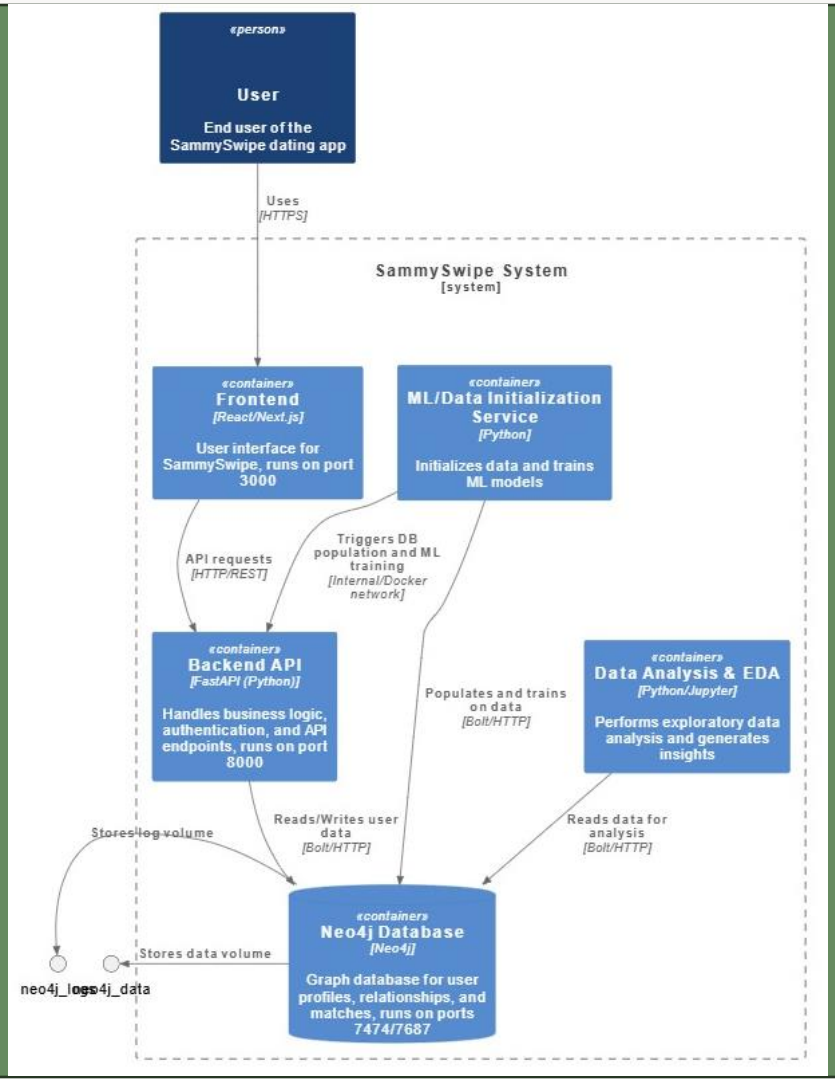
PROPOSED SOLUTION

SammySwipe redefines digital dating by combining behavioral science, adaptive algorithms, and social graph intelligence to surface high-quality, emotionally resonant matches.

Key Components of Our Solution

- **Interest Modelling**
  → Leverage user-declared hobbies, Instagram/Twitter likes, and in-app activity to model authentic interests.
- **Adaptive Weighting**
  → Dynamically adjust compatibility scoring based on age, gender, and cultural cohort—ensuring relevance and personalization at scale.
- **Personality Profiling**
  → Integrate a lightweight neuroticism-based scale to improve emotional alignment between matches.
- **Graph Database (Neo4j)**
  → Use Neo4j for fast, relational traversal of social graphs—enabling mutual connections, friend-of-friend insights, and context-aware recommendations.

SYSTEM ARCHITECTURE



DATA UNDERSTANDING & EXPLORATION

RICH DATASET

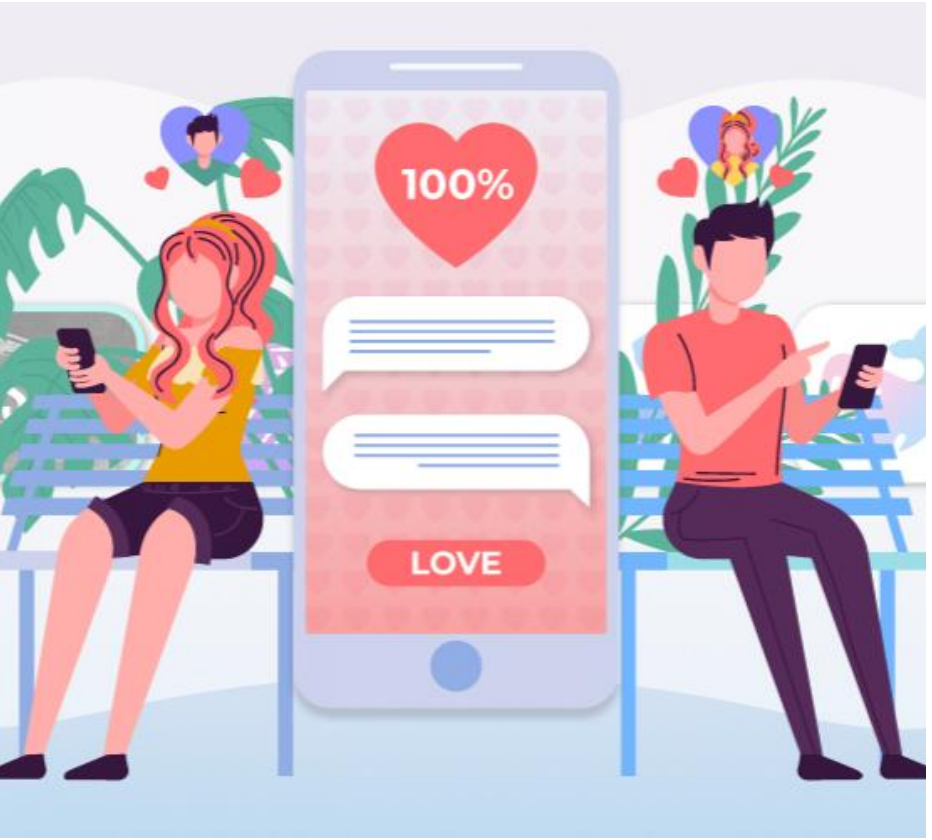
Ok Cupid Dataset

DATA CLEANING

Rigorous preprocessing to ensure data quality and consistency.

FEATURE ENGINEERING

Creating new features to capture complex cricket dynamics.



DATA UNDERSTANDING & EXPLORATION

User Onboarding Forms
→ Real-time data on age, gender, location, and interests (~2KB/user)

Behavioural Telemetry (Data Streams)
→ ~45 events/user/day including swipes, messages, profile edits

Social Media APIs (OAuth)
→ ~20KB/user of liked content and pages to enrich interest graphs

Moderation Labels (ML)
→ Sparse but crucial signals for spam, fraud, and offensive behavior

Insights:

- Interests follow a power-law distribution — top 15 interests cover 60% of users
- Users within 25 km are 3.4× more likely to initiate chats within 24 hours
- High neuroticism (top 25%) correlates with 12% shorter message lengths
- Friendship graph shows small-world properties (avg path ≈ 5.2)

FEATURE ENGINEERING

Key Engineered Features

Interest Vectors (TF-IDF + SVD)

→ Extract user interests from profile text & social data

→ 3,000-dimensional TF-IDF reduced to 256 via Truncated SVD

Geo Features

→ Calculate haversine distance and timezone difference

→ Enables location-aware matching and timezone coordination

Engagement Signals

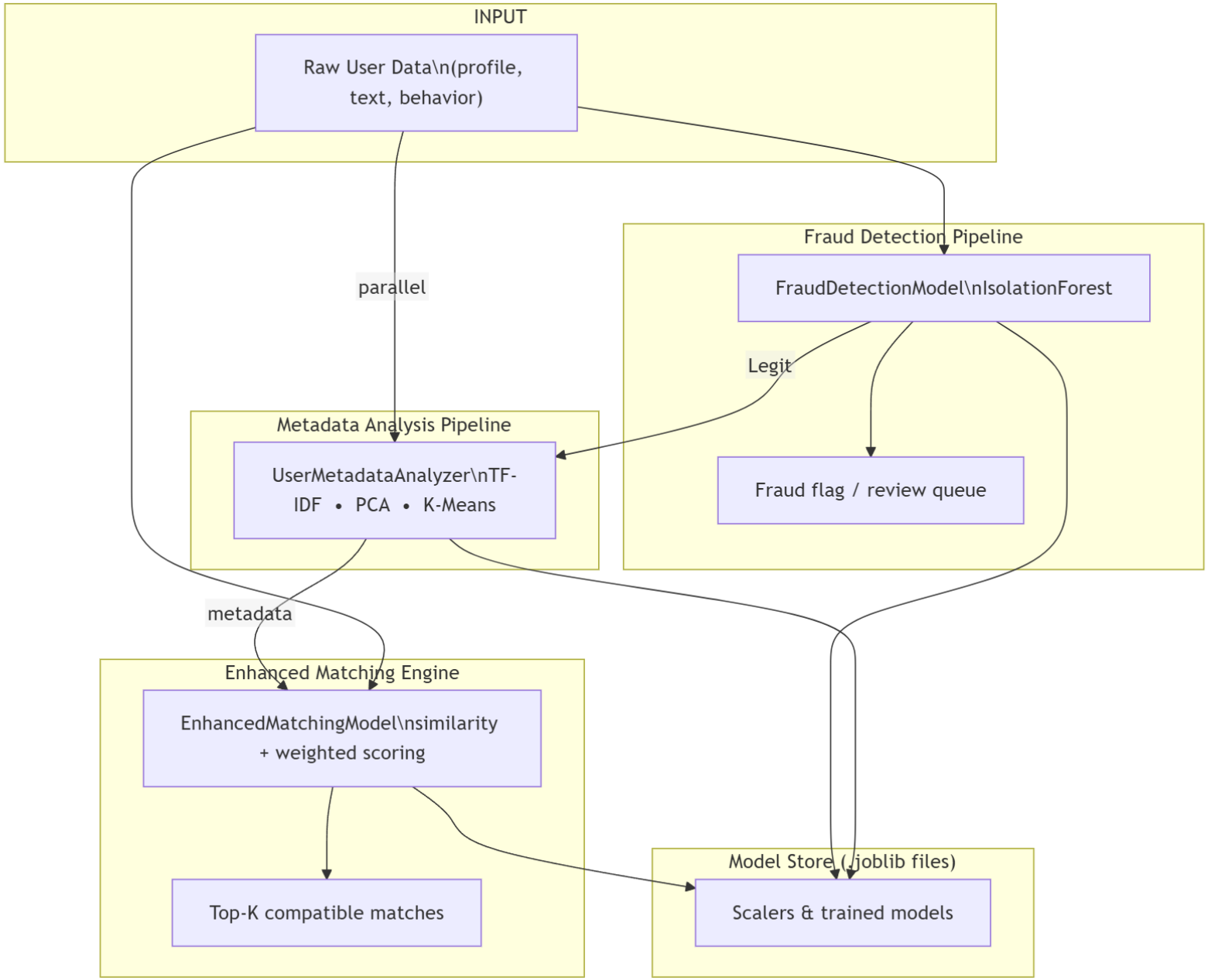
→ Apply a capped logarithmic scale to daily in-app actions

→ Helps capture meaningful user activity without skew from power users

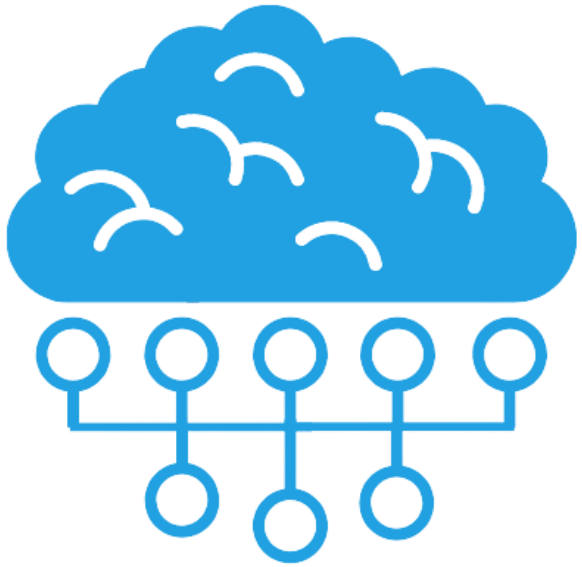
Trait Embeddings

→ Use Big Five personality traits and polynomial interactions

→ Enables emotional and psychological compatibility modeling


INPUT
Raw User Data\n(profile, text, behavior)
parallel
Fraud Detection Pipeline
FraudDetectionModel\nIsolationForest
Legit
Metadata Analysis Pipeline
UserMetadataAnalyzer\nTF-IDF • PCA • K-Means
Fraud flag / review queue
metadata
Enhanced Matching Engine
EnhancedMatchingModel\nsimilarity + weighted scoring
Top-K compatible matches
Model Store (.joblib files)
Scalers & trained models

MODEL BUILDING



Model Suite
Enhanced Matching Model
→ Algo: Gradient Boosting
→ Purpose: Compatibility score based on 32 engineered features

Cluster Assigner
→ Algo: Mini-Batch K-Means (k=20)
→ Purpose: User segmentation via TF-IDF, geo, engagement vectors

User Meta Data Analyser
→ Algo: Isolation Forest
→ Purpose: Detect anomalies in device usage and login patterns

Graph-Powered Insights
→ MATCHED, LIKED, and FRIENDS_WITH relationships modelled as edges in Neo4j
→ Properties include interaction score, timestamps, and match status
→ Cypher queries perform real-time scoring (e.g., Jaccard similarity, mutual intent)

EVALUATION & METRICS

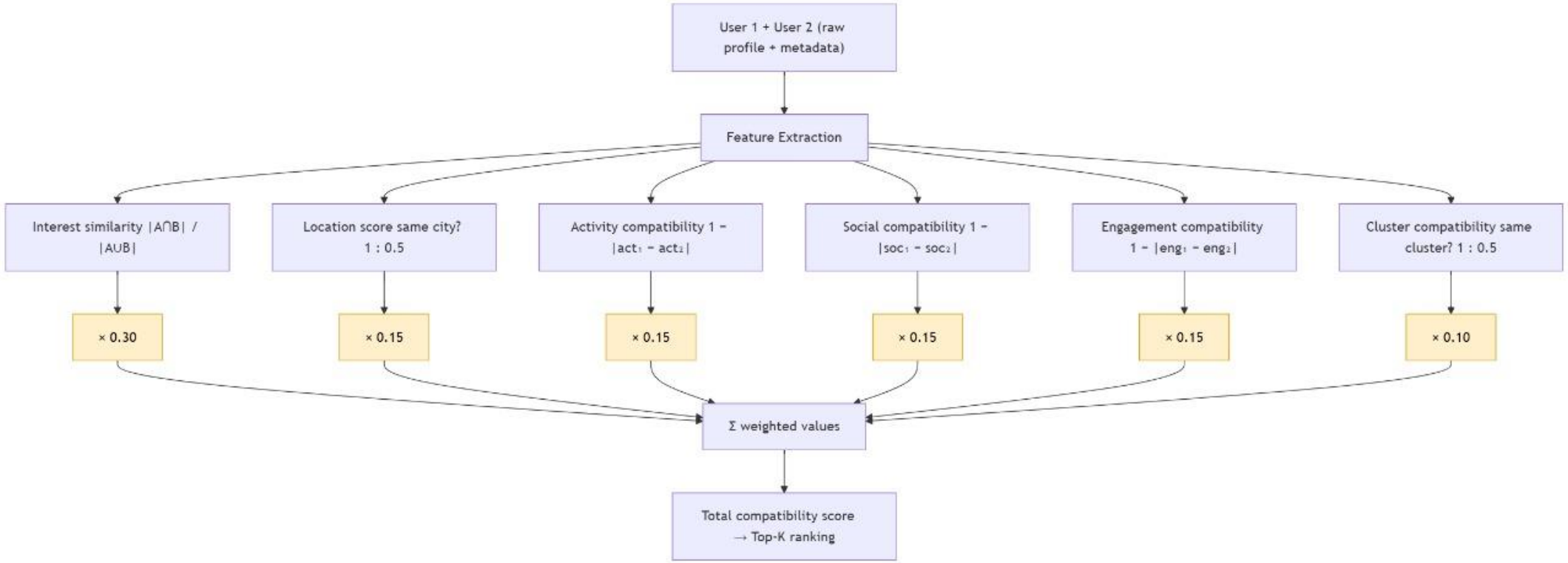
Match Quality
Precision@10 = 0.72
→ 7 out of top 10 recommended matches were relevant

Message Quality Boost
→ Median message exchange length increased by 21% over baseline — indicating more engaging conversations

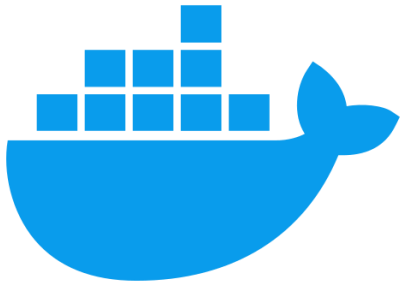
System Performance
Latency:
→ Backend response time at p95 = 88 ms

Graph Write Throughput
→ Neo4j handles 3.2k relationships/sec on a 2-core instance — enabling high-throughput matchmaking at scale

EVALUATION & METRICS



DEPLOYMENT



Deployment & Operations

Our deployment strategy ensures high scalability, modularity, and CI-driven automation across services—from model inference to frontend APIs.

Containerization & Orchestration

- Docker Compose for local development; AWS ECS for production deployment
- FastAPI: 4 replicas, 2 cores each – handles API traffic
- ML Service: 2 replicas, 4 cores each – serves compatibility predictions
- Neo4j: 2 replicas, 8 cores each – handles graph queries and updates

CI/CD Flow

- Pull Request merge triggers GitHub Actions
- Executes tests + static code analysis
- Builds Docker image and pushes to AWS ECS

SammySwipe

Log in | Get Started

Find Your Perfect Match with AI-Powered Dating

Experience dating reimagined with advanced AI matching, personality analysis, and genuine connections based on shared interests and compatibility.

Start Your Journey | Learn More

Why Choose SammySwipe?

AI-Powered Matching

Our advanced AI analyzes your interests, personality, and behavior to find truly compatible matches.

Genuine Connections

Focus on meaningful relationships with people who share your interests and values.

Safe & Secure

Advanced fraud detection and verification systems ensure a safe dating environment.

SammySwipe

Create an Account

Join SammySwipe and find your perfect match

Email

Username

Full Name

Password

Confirm Password

Gender

Select Gender

Birth Date

mm/dd/yyyy

Location

City, Country

Bio

Tell us about yourself...

Profile Photo

Choose Photo

No file chosen

Create Account

SammySwipe

Welcome back

Sign in to your account to continue

Email

Password

Sign in

Enter as Super Admin

Super admin mode bypasses all authentication

Forgot your password?

Don't have an account? Sign up

SammySwipe Matches Chat Profile Dashboard SUPER ADMIN MODE Super Admin Log out

Discover

Super Admin Access
All routes unlocked

- Profile
- Discover
- Matches
- Chat
- Media
- Video Chat

User 290

50% Match

Sample bio for user 290. Interests include Technology, Fitness.

COMMON INTERESTS

Discover new interests together!

Skip Like

Super Admin Access
All routes unlocked

Profile
Discover
Matches
Chat
Media
Video Chat

People You've Liked

These are people you've liked from the Discover page. They haven't responded yet.

Refresh

User 377, NaN

75% Match Pending

Sample bio for user 377. Interests include Reading, Technology.

INTERESTS

Reading Technology

Liked on Invalid Date

User 290, NaN

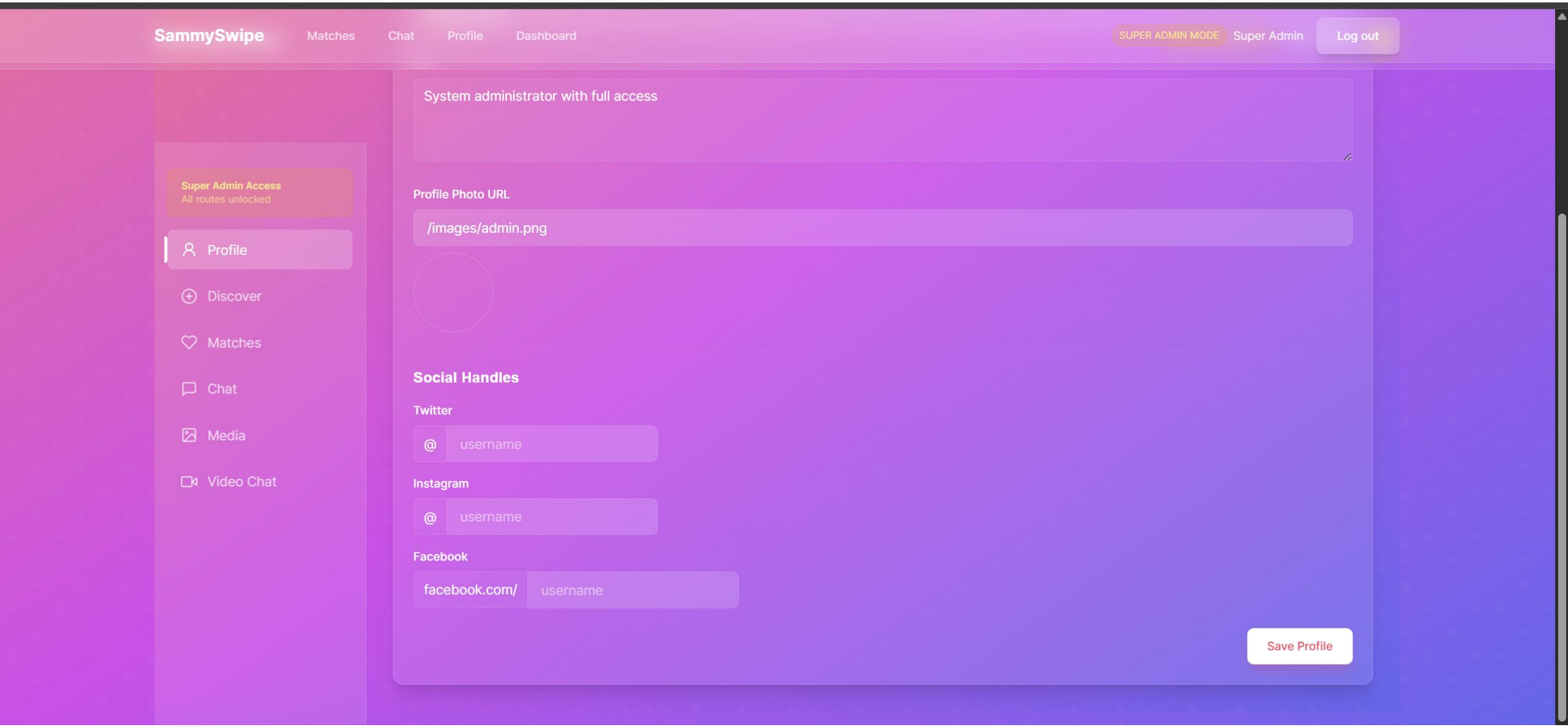
75% Match Pending

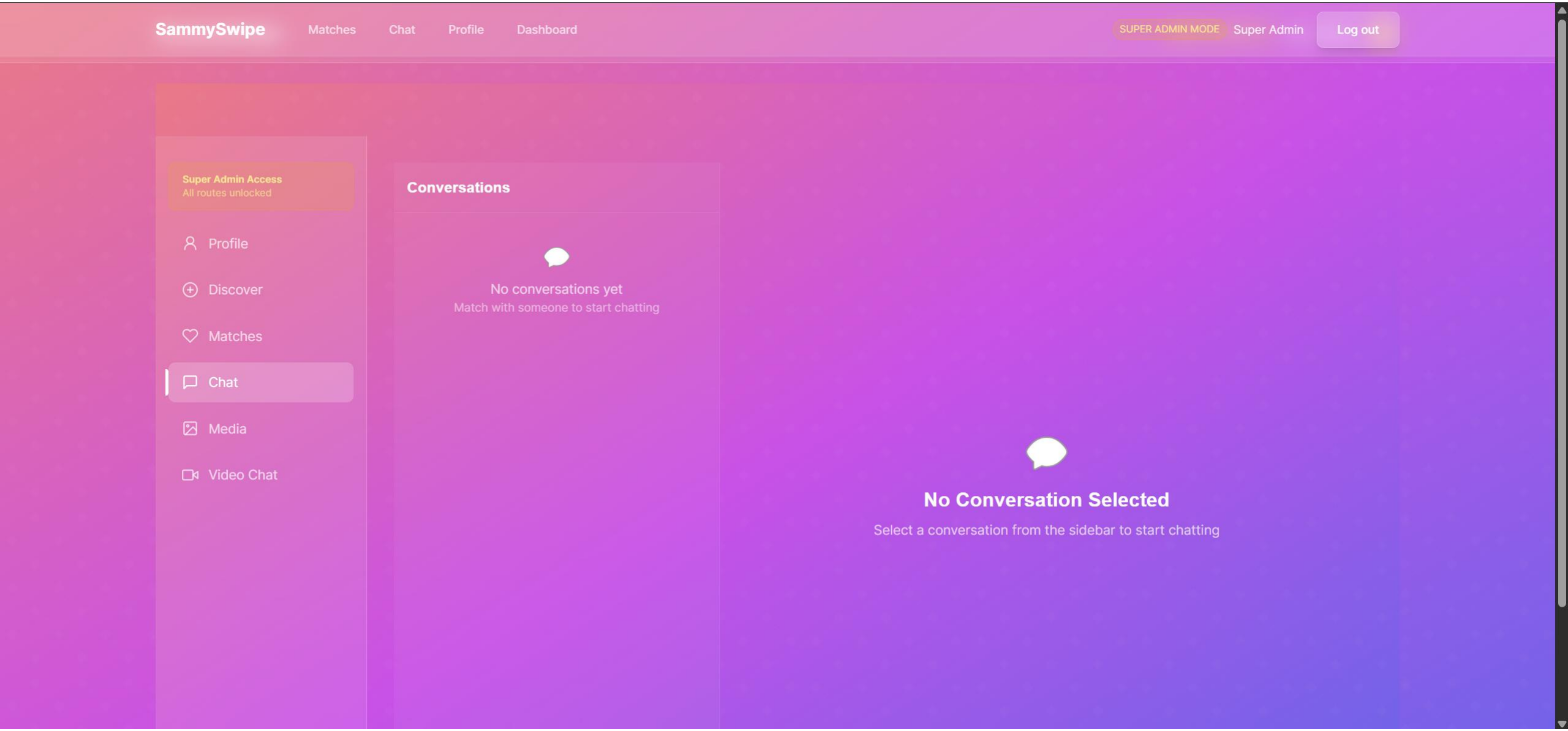
Sample bio for user 290. Interests include Technology, Fitness.

INTERESTS

Technology Fitness

Liked on Invalid Date


SammySwipe
Matches
Chat
Profile
Dashboard
SUPER ADMIN MODE
Super Admin
Log out
System administrator with full access
Super Admin Access
All routes unlocked
Profile
Discover
Matches
Chat
Media
Video Chat
Profile Photo URL
/images/admin.png
Social Handles
Twitter
@
username
Instagram
@
username
Facebook
facebook.com/
username
Save Profile


SammySwipe
Matches
Chat
Profile
Dashboard
SUPER ADMIN MODE
Super Admin
Log out
Super Admin Access
All routes unlocked
Profile
Discover
Matches
Chat
Media
Video Chat
Conversations
No conversations yet
Match with someone to start chatting
No Conversation Selected
Select a conversation from the sidebar to start chatting

FUTURE DIRECTION

Privacy
→ Current: Interest data stored in plain text
→ Planned: AES-GCM field-level encryption

Diversity
→ Current: Binary gender options only
→ Planned: Support for gender spectrum and pronouns

Cold-start Problem
→ Current: Sparse data for new users
→ Planned: Zero-shot interest inference using LLM embeddings

Mobile Experience
→ Current: Only Progressive Web App
→ Planned: Native iOS & Android apps

THANK YOU

Time for Q&A