

TÓM TẮT

Khóa luận tập trung giải quyết bài toán xây dựng hệ thống giới thiệu kết bạn dựa trên sự tương đồng về tính cách, đồng thời đảm bảo quyền riêng tư cho dữ liệu người dùng thông qua mô hình xử lý trực tiếp trên thiết bị (on-device processing). Trong bối cảnh các mạng xã hội hiện đại đang đối mặt với nhiều thách thức về bảo mật thông tin nhạy cảm, đề tài đề xuất một quy trình hoàn chỉnh từ thu thập dữ liệu trả lời câu hỏi của mô hình tính cách Big-5, chuyển đổi sang không gian vector thu gọn bằng Phân tích Thành phần chính (PCA-4), cho đến việc lưu trữ an toàn bằng chuẩn mã hóa AES-256-GCM thông qua các hàm thực thi biên (Edge Functions). Hệ thống sử dụng thuật toán giới thiệu lai kết hợp giữa độ tương đồng tính cách, hành vi xã giao qua điểm số ELO và nhúng ngữ nghĩa (semantic embedding) cho sở thích cá nhân. Kết quả thực nghiệm trên hệ thống thật cho thấy mô hình PCA-4 giữ được hơn 90% phương sai dữ liệu gốc, cho phép xác định các cặp người dùng tương đồng với độ chính xác cao (0.9999). Hiệu năng hệ thống được tối ưu hóa đạt mức trễ chấp nhận được với trung bình 1.8 giây cho quy trình đăng ký và 2.3 giây cho việc tạo danh sách giới thiệu, đồng thời khẳng định tính khả thi của kiến trúc “Quyền riêng tư theo thiết kế” trong việc bảo vệ dữ liệu nhạy cảm mà vẫn duy trì hiệu quả kết nối cộng đồng.

Từ khóa: Twins, Big-5, PCA, Quyền riêng tư, Hệ thống giới thiệu, Mã hóa dữ liệu

