Real-Time Multilingual Translator with NLLB-200
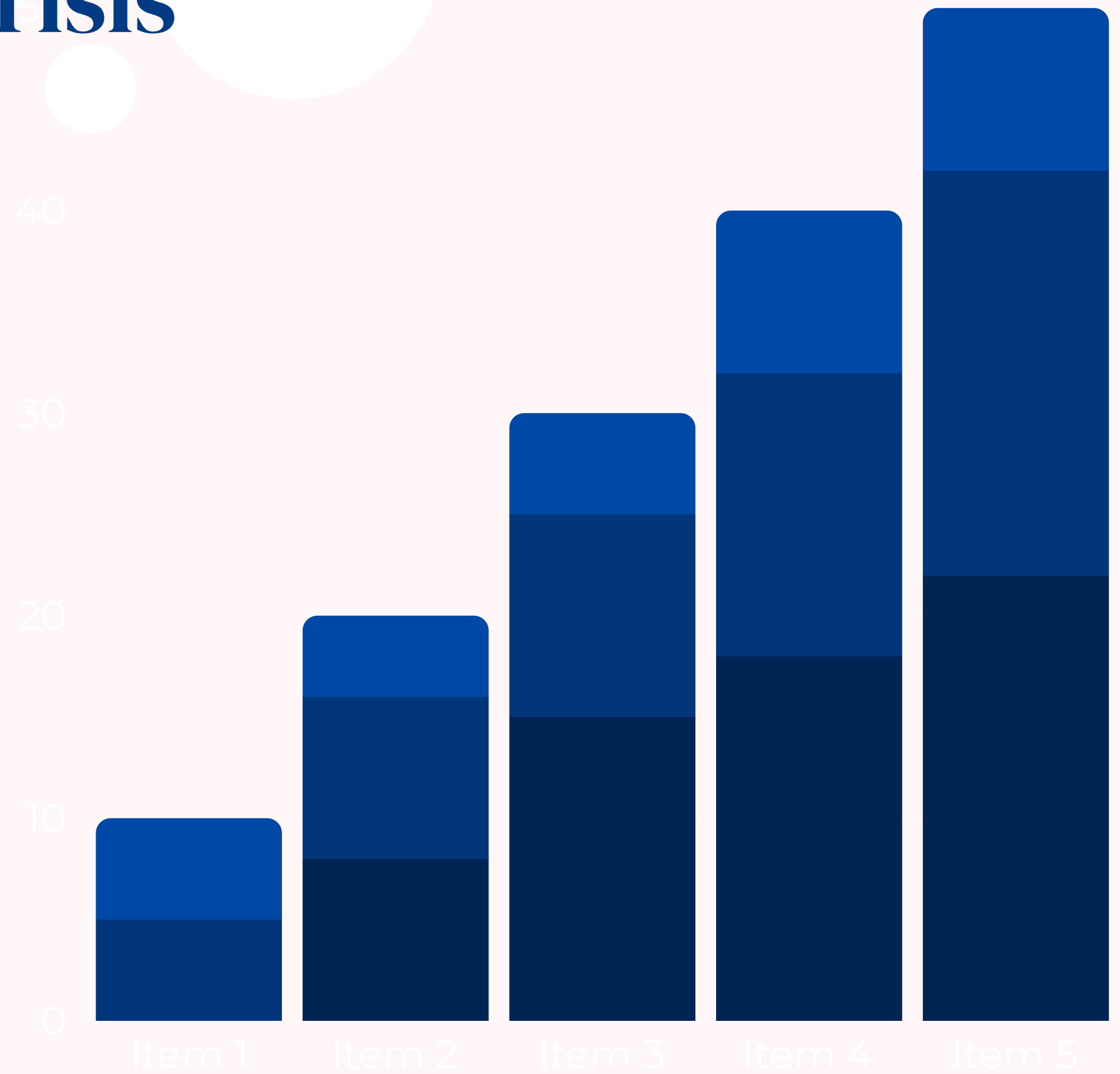
A Deep Learning Project for Seamless Cross-Language Communication

Presenters: Alex and Tami

Course: Deep Learning and Its Applications

Date: September, 2025

Institution: Ono Academic College

The Language Barrier Crisis

- Around 7,159 living languages (Ethnologue, 2025), but most people speak 1-2.
- Barriers in phone calls, video chats, messages, and global meetings.
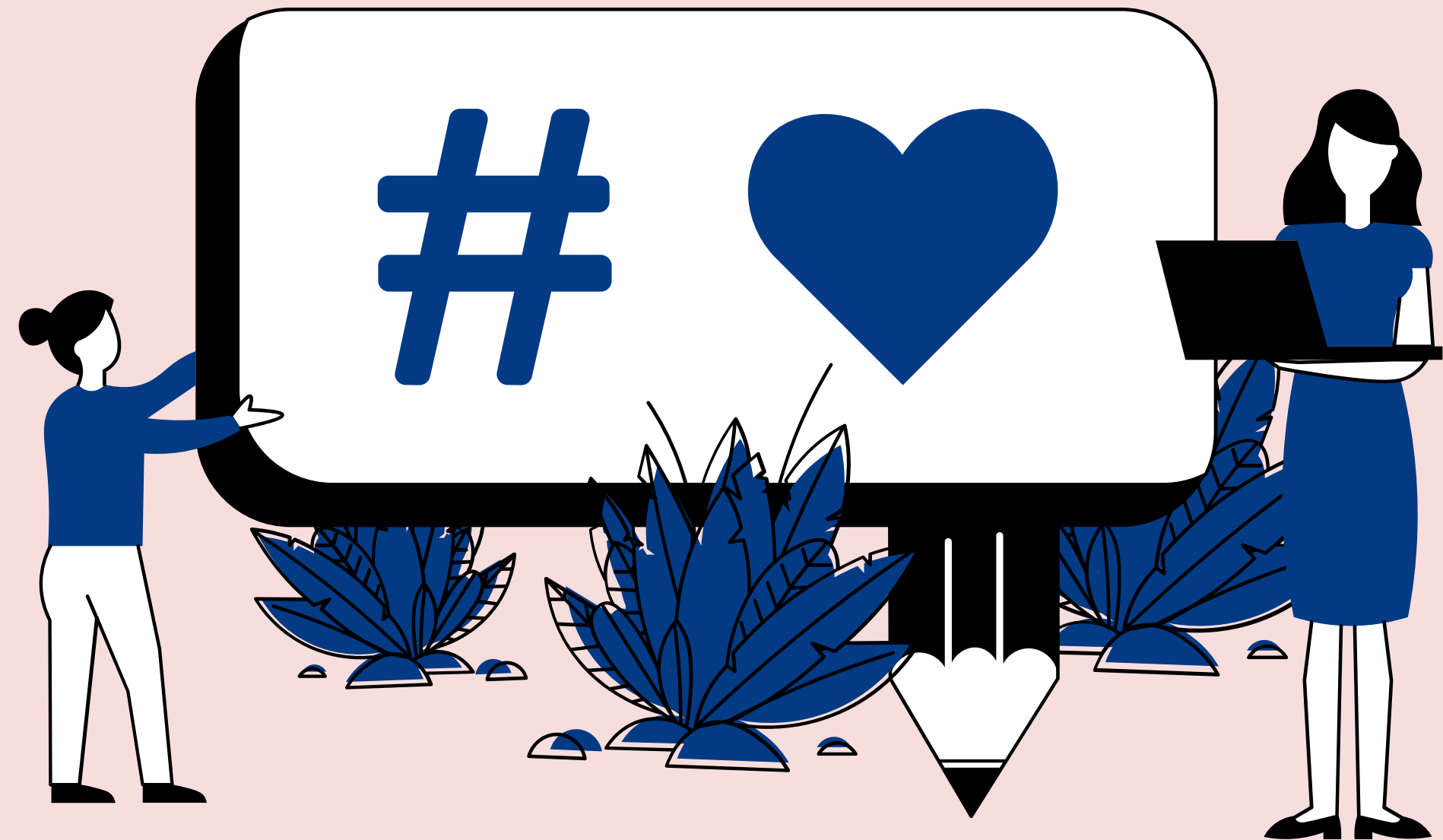- Impacts business, healthcare, education, and personal connections.

survey results

https://docs.google.com/forms/d/15E-CbHtdQRPs7BiZ-v9oNSgnP_QKuv6a1FX3L-Kx6GI/edit#responses

Bridging the Gap with LinguaLink

- LinguaLink: Real-time translation for text and planned speech.
- Uses deep learning for accuracy and natural audio.
- Bidirectional (e.g., EN ↔ HE), low latency, high accuracy.
- Applications: Business, healthcare, cross-cultural interactions; future enhancements planned.
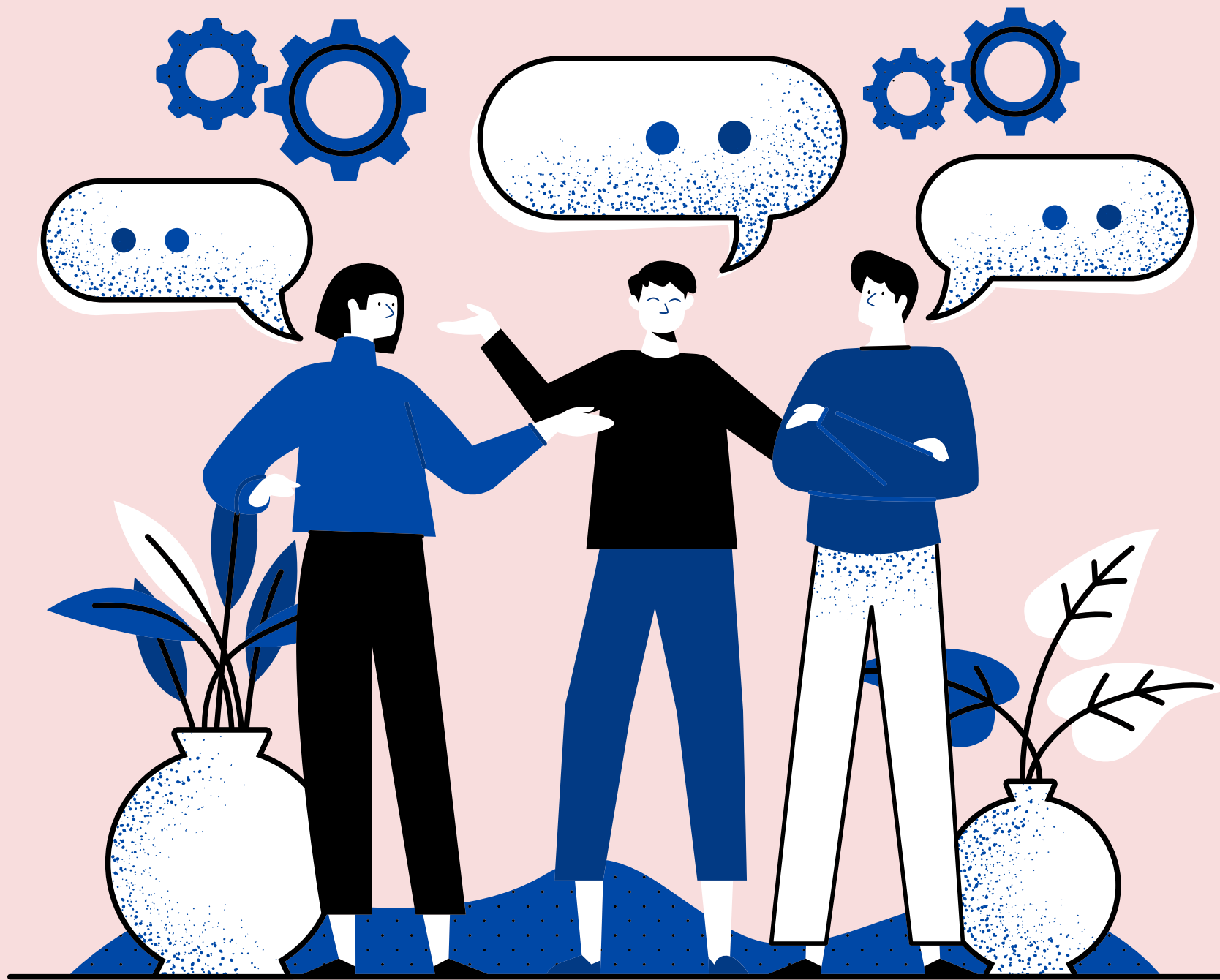- Future enhancements: Speech integration, expanded language support.

Project Overview

- Objective: Develop a real-time text and planned speech translator for multilingual conversations to address language barriers.
- Innovation: Leverages NLLB-200 (200+ languages) for high-accuracy translation in a WebSocket-based chat system.
- Scope: Supports English, Hebrew, Russian, Spanish, French, Arabic; designed to expand toward 200+ languages with future speech integration.

Technical Implementation

- Framework: FastAPI with WebSocket for real-time communication.
- Model: NLLB-200-distilled-600M (transformer-based, pre-trained by Facebook).
- Languages: Prototype supports 6 (EN, HE, RU, ES, FR, AR); model scalable to 200+.
- Environment: Developed in GitHub Codespaces with Python 3.12.
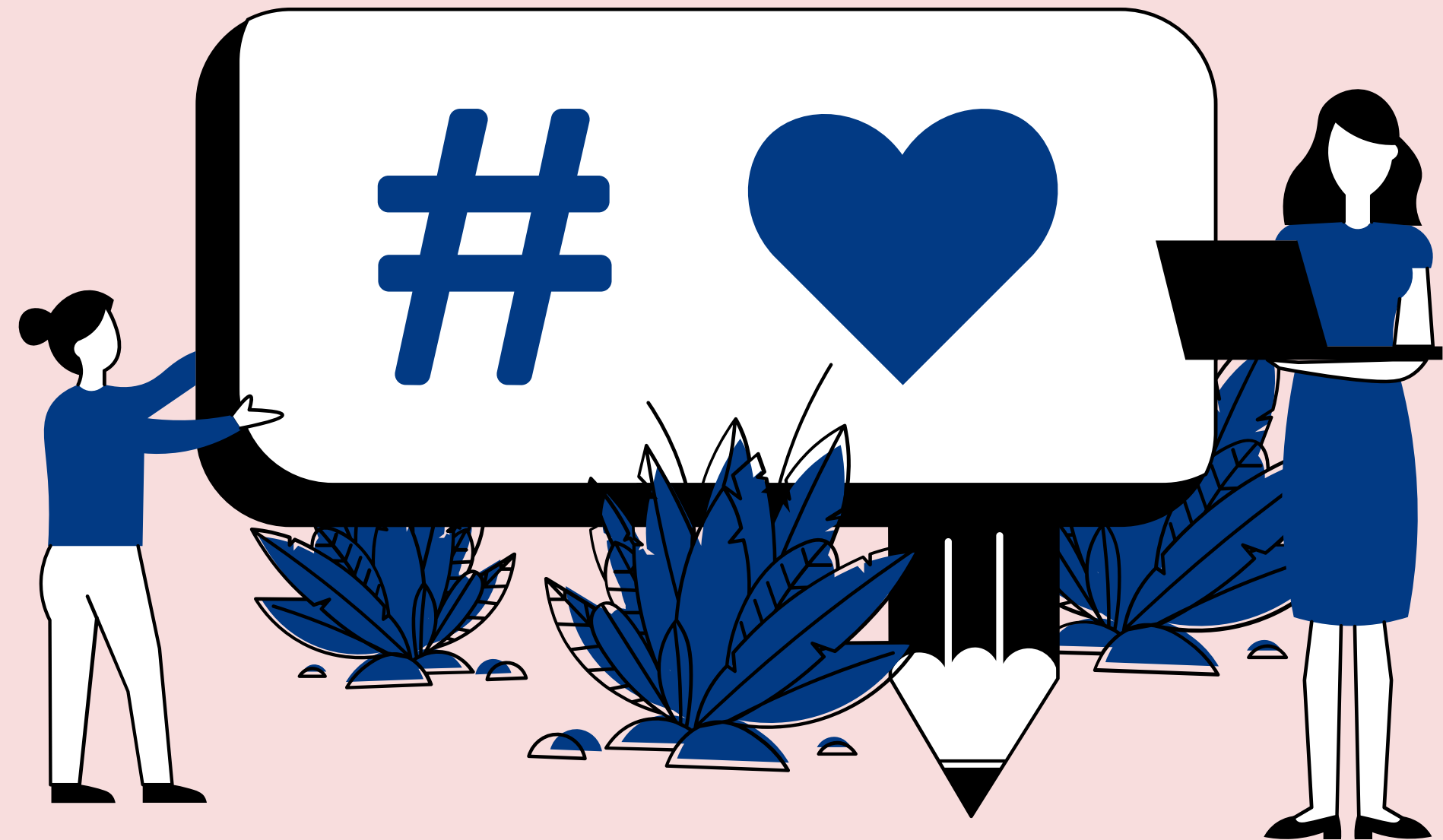
Features and Functionality

- Real-time pairing of users with different languages.
- Dynamic translation (e.g., EN ↔ HE, EN ↔ RU).
- Stress-tested with 10+ concurrent clients.
- Automatic partner reassignment when one disconnects.
- Planned speech-to-text and text-to-speech integration.
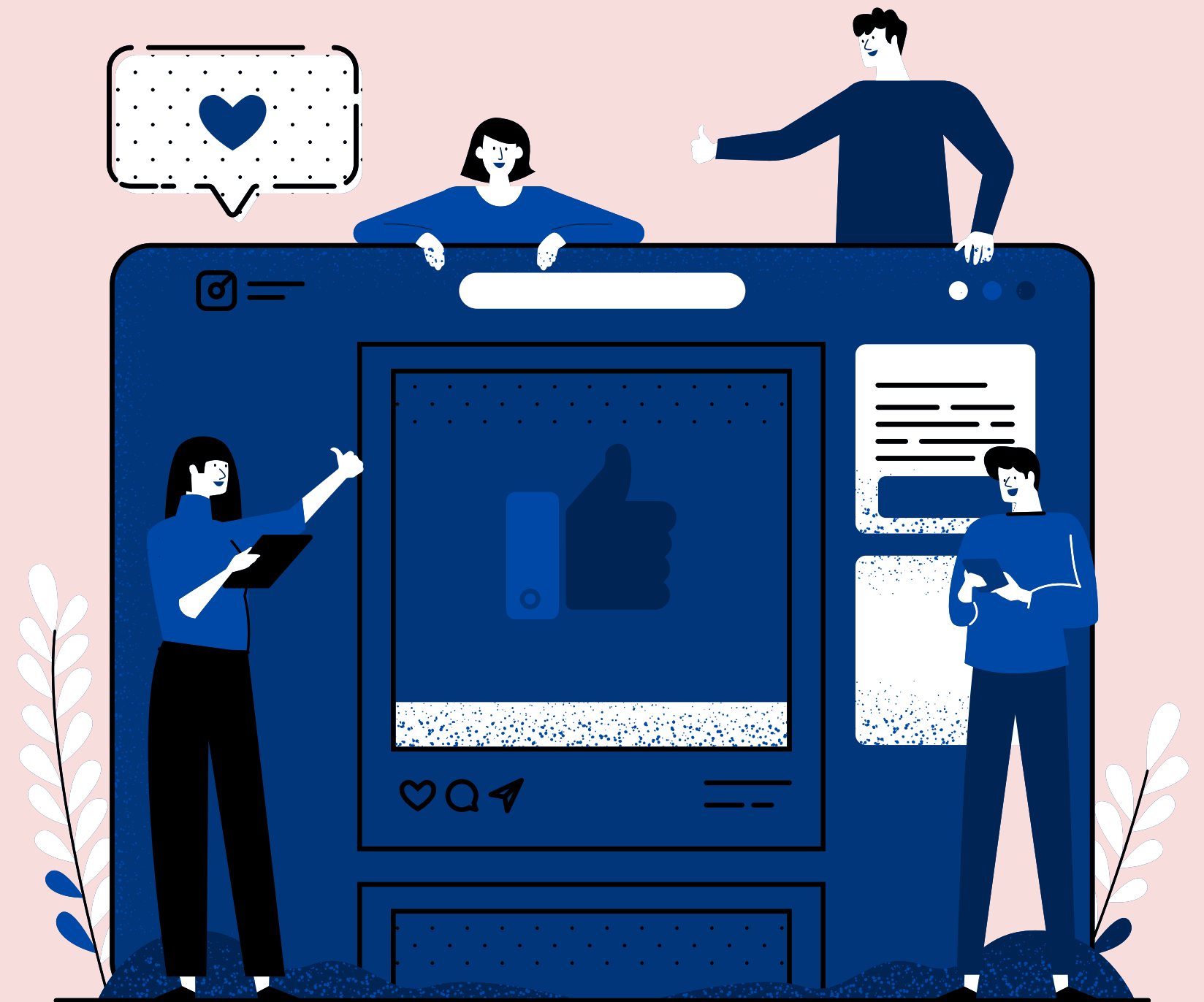
Quality Metrics and Testing

- Accuracy: BLEU score >0.35 for EN-HE pairs.
- Stability: Successfully handles 10 clients across 6 languages.
- Latency: <1 second per translation (preliminary).
- Tests: Basic, stress, and speech placeholder tests implemented.

Challenges and Solutions

- Challenge: Model loading and memory constraints in Codespaces (~4GB RAM).
- Solution: Optimized with NLLB-600M; fallback to 1.3B if needed.
- Challenge: Concurrent client handling.
- Solution: Thread-based stress testing with random delays.

Live Demonstration

https://github.com/saroatov2910/LinguaLink.git

https://github.com/Tams03/DL-Translator-Project.git

Conclusion and Future Directions

- Conclusion: Achieved a functional MVP with robust translation and testing.
- Future Work: Integrate speech (ASR/TTS), expand language support, and deploy on a public server.
- Acknowledgements: Thanks to [Lecturer/TA Names], course team, and GitHub Codespaces.

Questions & Answers