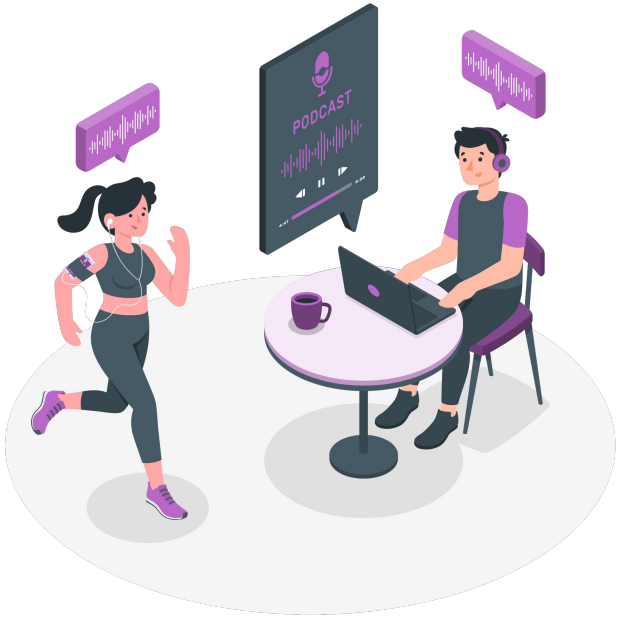
More than just a translator

Polina Zelenskaya
Ekaterina Maksimova
Ekaterina Urmanova
Evsey Antonovich
Daniyar Cherekbashev

About us

Our goal is to create perfect real-time speech-to-speech translation

We utilize existing ML models to their full potential



19 languages

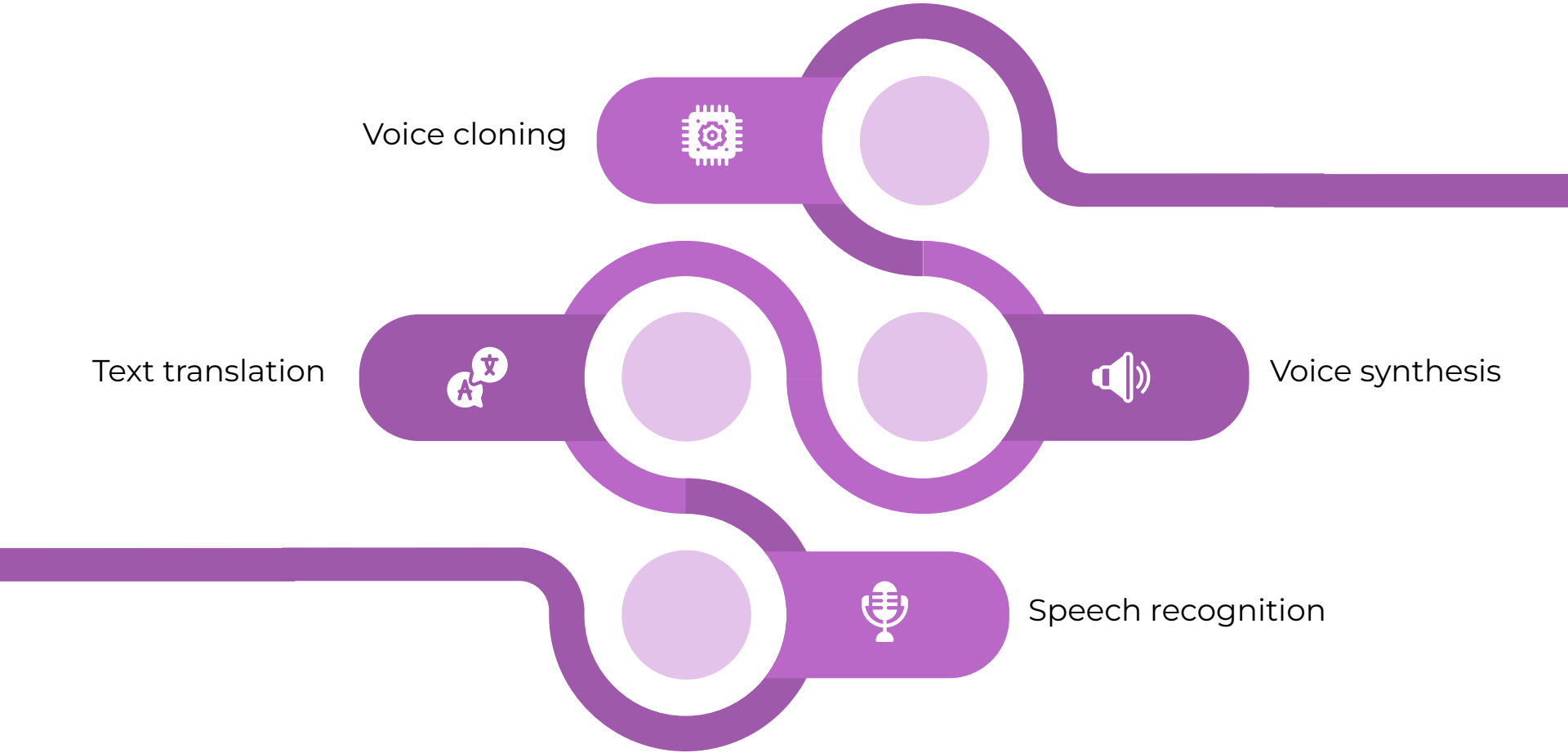
Already supported

1.68 seconds

Average delay

100+ models

Flawlessly interact with each other


Voice cloning
Text translation
Voice synthesis
Speech recognition

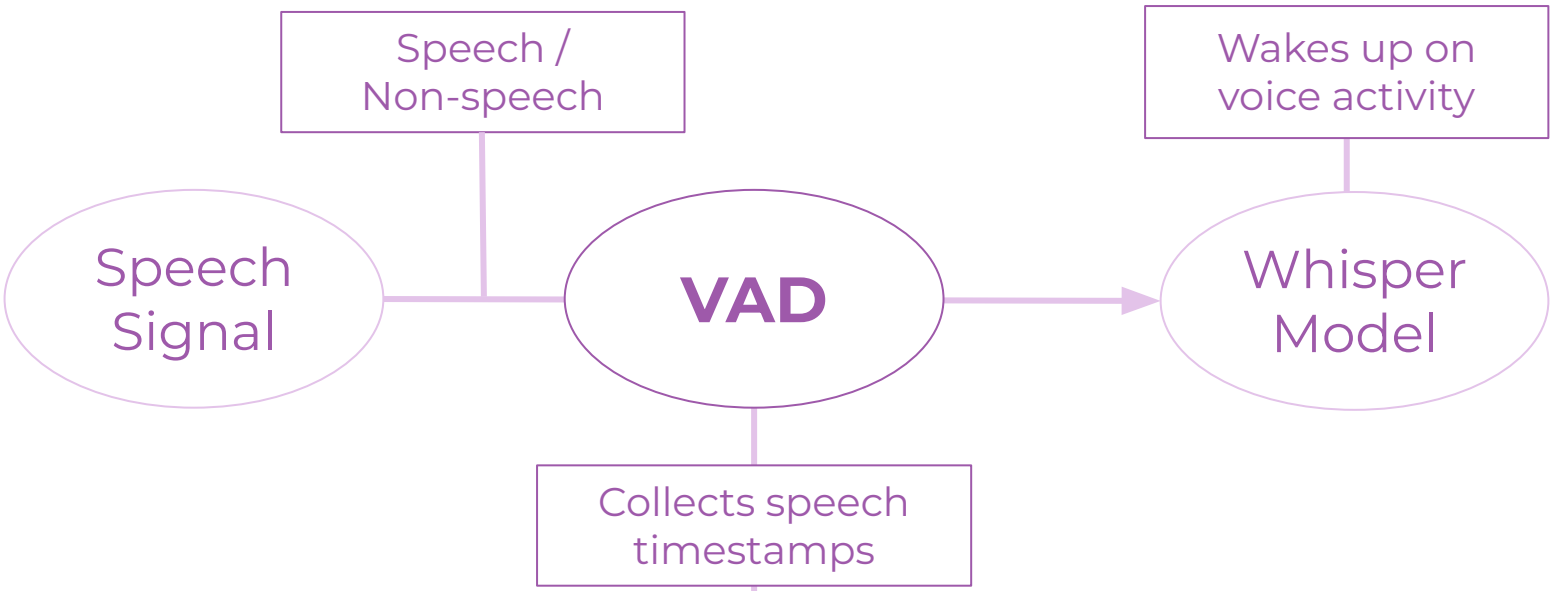
1

Speech-To-Text

Place where it all begins

How is audio processed?

Voice Activity Detection



Whisper Models Comparison

	Performance	Quality	Parameters
Tiny	Great	Very inaccurate	~39M
Medium	Great, on GPU	Accurate, but not perfect	~244M
Large-v2	Great, on decent GPU	Very accurate	~1550M

99 languages

Supported by all Whisper models

2

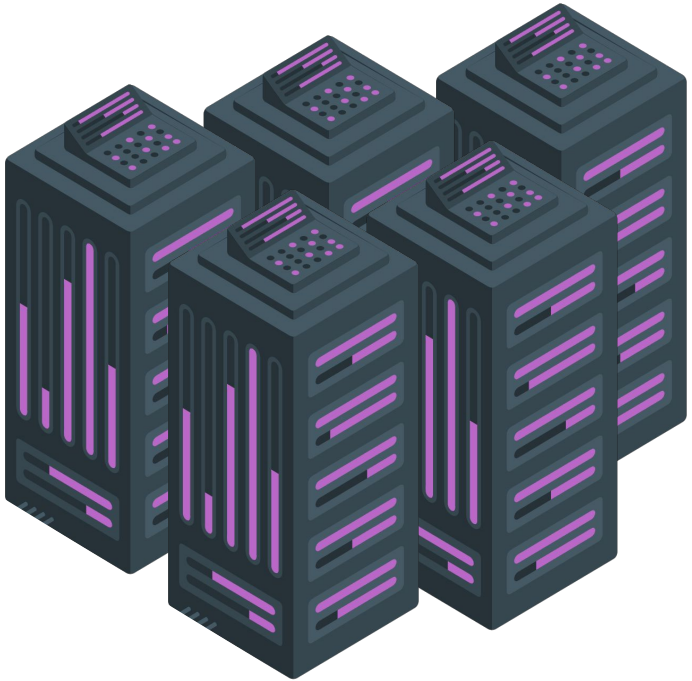
Text-To-Text

The place where translation happens

Translation Flow

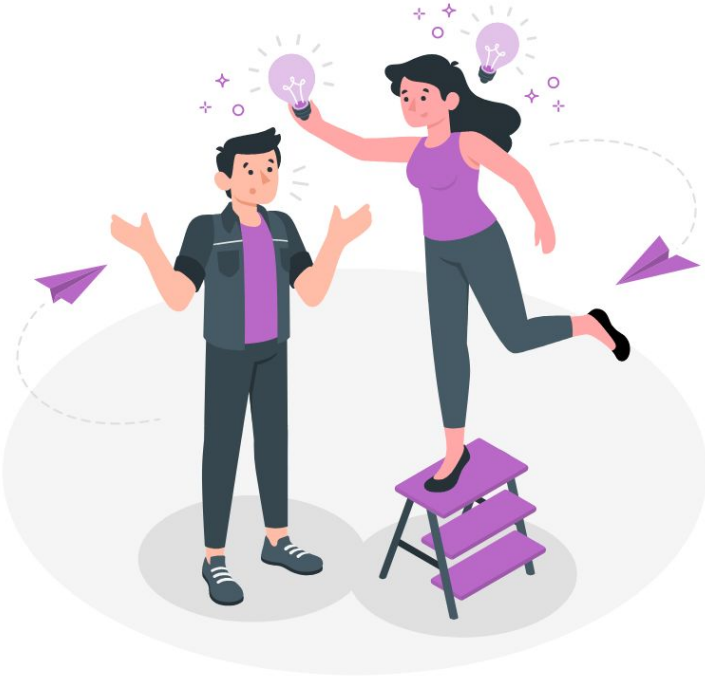
Input Language

Output Language



It’s free!

These models are publicly available
for commercial use

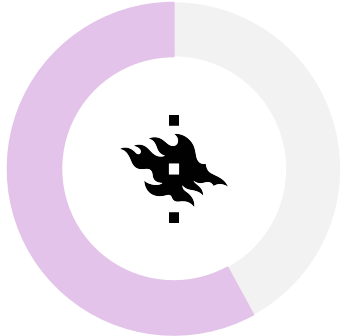


Human Like translation

50-60*

Average human

60+



Opus

***BLEU score - how human like text is**

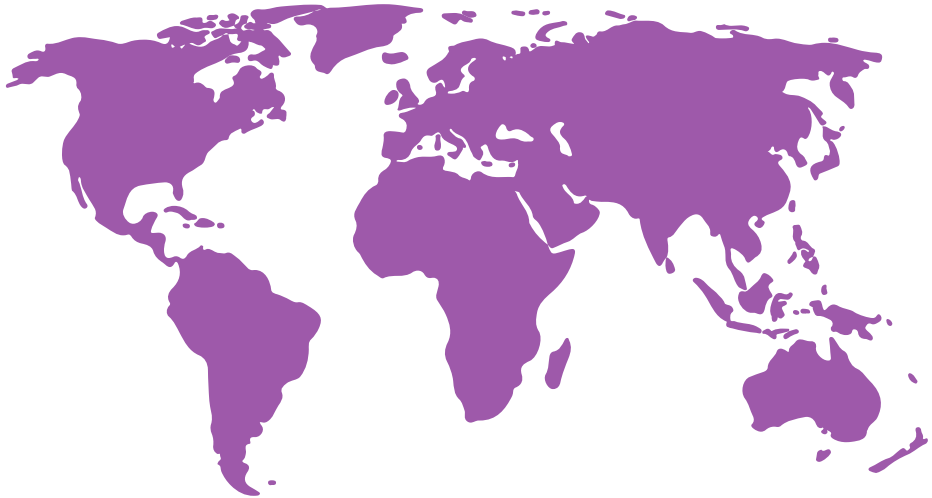
High variety of languages

1 **1440 models**

All publicly accessible

2 **230 languages**

Covering most of the Earth



Problems

1

Scalability

Translation server may not keep up with a large number of active rooms

2

No context

Models do not take into account previous messages when translating

3

Voice cloning

The cherry on top

Lack of research!

All already-build solutions didn’t fit. We need to create a real-time lightweight and customizable voice cloning pipeline

Decomposition

Piper

A fast, local neural text to speech system

FreeVC24

High-Quality, text-free, one-shot voice conversion

15 seconds

Of audio is sufficient to clone your voice

4

Backend

The one server to rule them all

Tech stack

Golang

Simple, yet capable; fast and reliable

Gin-gonic

Performant HTTP web framework

PostgreSQL

Advanced open source relational database

Gorilla

Widely-used websocket implementation

Data Security

JSON Web Tokens

A modern approach to access tokens, used by OAuth 2.0

bcrypt

A robust password hashing function, resistant to rainbow table and brute force attacks

Transport Layer Security

Industry standard encryption for the web

5

User Interface

How the user interacts with our app

UI tools

- Python for fast development and iteration
- Dear PyGui for quick and painless UI development
- Dear ImGui underneath for robust functionality and minimalistic design

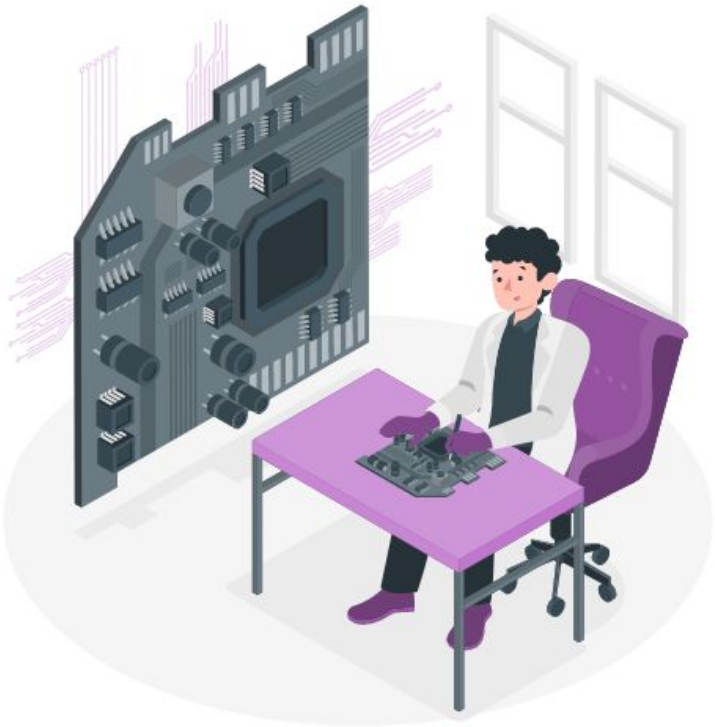
FUTURE

How are we planning to grow, develop and improve

Train Models

Right now models don’t use previous messages to translate new ones. We want to create models that can use context for better translation

In addition, we need a voice conversion model lighter than the current FreeVC24



Improve backend

Deploy our project to a cloud. Buy domain names. Expand functionality, improve modularity

Add support for streaming sound for users with no GPU

Fund & Launch

Reach out to investors and accelerators. Main goal - buy hardware for testing and development

Integrate with some services of Innopolis University

More than just a translator

Polina Zelenskaya
Ekaterina Maksimova
Ekaterina Urmanova
Evsey Antonovich
Daniyar Cherekbashev

Thanks!

Do you have any questions?

p.zelenskaya@innopolis.university
+7(908)992-93-96
tg: @cutefluffyfox

This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik** and illustrations by **Stories**

Links

1. https://huggingface.co/openai/whisper-large-v2
2. https://cdn.openai.com/papers/whisper.pdf
3. https://opus.nlpl.eu/Opus-MT/
4. https://huggingface.co/Helsinki-NLP/opus-mt-ru-en
5. https://github.com/rhasspy/piper/tree/master
6. https://github.com/OlaWod/FreeVC
7. https://capstone.innopolis.university/docs/groups/unifai