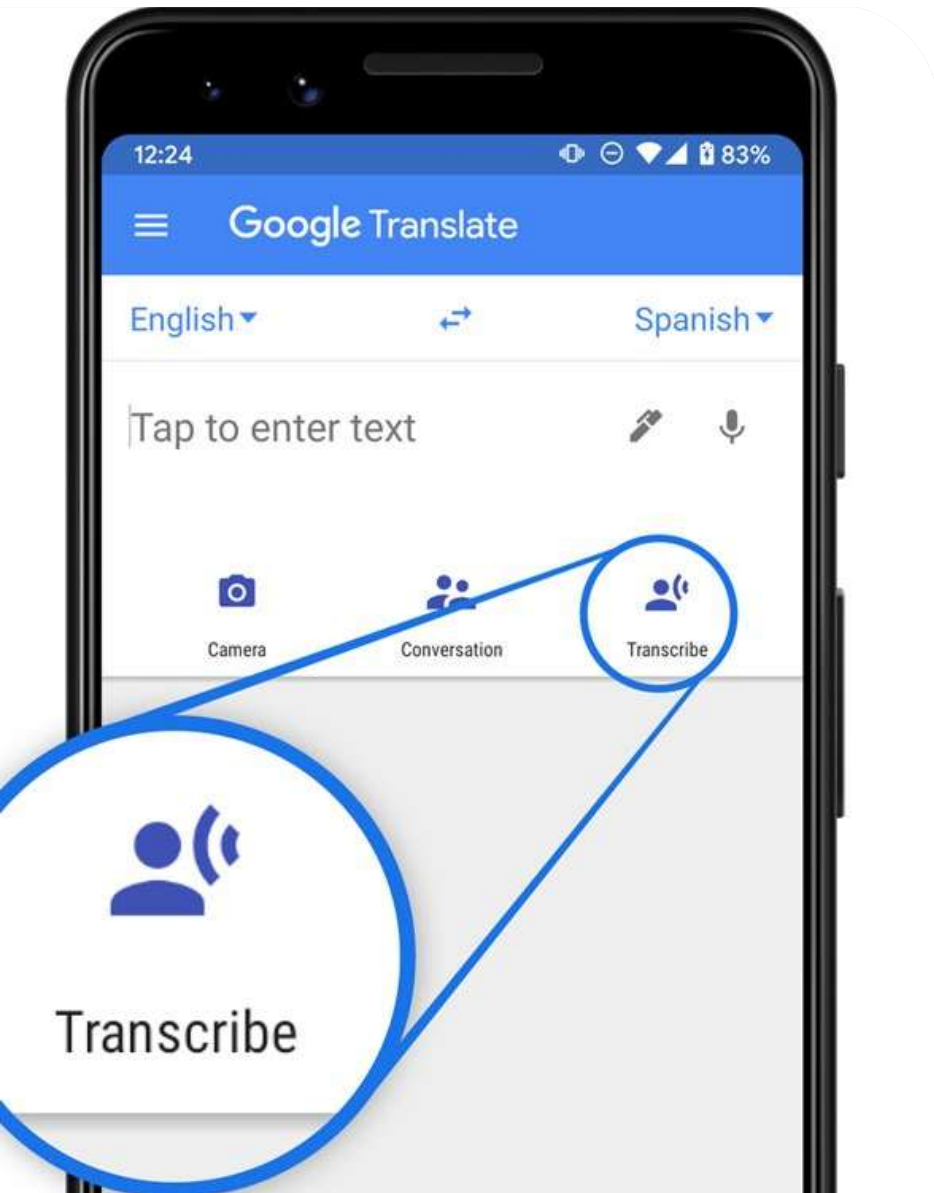# ROMINUTE

Your AI-Powered Assistant
For Meetings and Conference Rooms

Soroush Gooran

Department of Computer Engineering

Sharif University of Technology

Live Translate / Transcribe

Products:

- Google Translate
- Speechlogger
- Krisp
- Vowel
- Notta
- Scribbl
- Whispering Tiger

ROMINUTE

Meeting Room Minutes

Introduction:

- Transcription or translation of any incoming audio or video stream in face-to-face or virtual meetings

- Support for more than a hundred living languages

- Can be used on different platforms and on different devices

- Provide a summary of the talks by identifying the speaker

- Can be used in web browser, meeting room monitors and various kinds of electronic devices

- Support for various accents and dialects of the Persian language

ROMINUTE

Features:

- Support for multiple languages
- Integration into hybrid meeting
- Customizable plugins and extensions
- Detection of users' voice activity
- Dual display in text and audio
- Mobile and desktop applications
- Contextual and topic-oriented translation
- Transcription services
- Summary of meetings
- Sentiment analysis of conversations
- Custom target accent and voice selection
- LLM, ASR, TTS, Translation, VAD, OCR, …Services

ROMINUTE

Advantages:

- Multilingual
- High accuracy
- Realtime
- Scalable
- Modular
- User friendly
- Cross-Platform

ROMINUTE
State-of-the-Art Machine Learning Models

- **Runs 100% locally on your machine.** (Once A.I. Models are downloaded, no further internet connection is required)
- **Speech recognition, translation and transcription**
    - OpenAI's Whisper, Supports ~98 languages
    - Meta's Seamless M4T, multi modal, Supports ~101 languages
    - Microsoft's Speech T5, English only
    - NVIDIA's NeMo
    - Wav2Vec 2.0
- **Text translation**
    - LID [Language Identification] (Supports 200 languages)
    - NLLB-200 (single model, Supporting 200 languages, high accuracy)
    - M2M-100 (single model, Supporting 100 languages, high accuracy)
    - Seamless M4T (single model, multi modal, Supporting ~101 languages)

ROMINUTE
State-of-the-Art Machine Learning Models (continued)

- **OCR** [Optical Character Recognition] (to capture game images and translate in-game text)
  - EasyOCR (Supports 80+ languages)
- **TTS** [Text-to-Speech] (Read out transcriptions/translations)
  - Silero
- **VAD** [Voice Activity Detection]
  - Silero-VAD
- **RVC** [Retrieval-based Voice Conversion] (Convert your voice, the voice in audio files or from Text-to-Speech)
  - RVC (Using the RVC STS Plugin)
- **LLM** [Large language model] (Continuation of text. automatic answer generation etc.) Proof of concept
  - FLAN-T5, GPT-J, Bloomz etc. (Using the Plugin for LLM)
- **And more using other Plugins...**

ROMINUTE
Gantt chart timeline (Phase 1)

فاز	عنوان اصلی	توضیحات	زمان (۲۰ روز)
فاز ۱ توسعه سرویس‌های پایه ۲ ماه	تکمیل برنامه ریزی	جستجوی روش‌های نوین و تکمیل مطالعات اولیه تکمیل تیم فنی تهیه منابع سخت افزاری و نرم افزاری لازم	
	توسعه و بهبود سرویس بازشناسی گفتار	تهیه دادگان بازشناسی گفتار آموزش مدل بازشناسی گفتار	
	توسعه و بهبود سرویس تبدیل متن به گفتار	تهیه دادگان گفتاری سنتز گفتار توسعه سرویس G2P آموزش مدل TTS	
	توسعه سرویس ترجمه ماشینی	توسعه سرویس ترجمه ماشینی مبتنی بر داده اختصاصی	
	توسعه سرویس پرسش و پاسخ صوتی	ادغام و یکپارچه سازی سرویس ها	
	توسعه سرویس حسابداری کاربران	تعریف سطوح کاربری	
	توسعه پنل کاربری	طراحی رابط کاربری اتصال به سرویس حسابداری کاربران	

ROMINUTE
Gantt chart timeline (Phase 2)

زمان (۲۰ روز)	توضیحات	عنوان اصلی	فاز
	توسعه سرویس خلاصه ساز متنی بر پایه LLM	توسعه سرویس خلاصه ساز متنی	
	توسعه سرویس تبدیل گفتار محاوره‌ای به رسمی بر پایه LLM	توسعه سرویس تبدیل گفتار محاوره‌ای به رسمی	
	۱. آرام ۲. شاد ۳. غمگین ۴. عصبانی	توسعه سرویس تحلیل احساس گفتاری	
	• تنظیم صدای مترجم صوتی هوشمند • تنظیم سرعت بیان مترجم صوتی هوشمند • تنظیم بازه زمانی نمایش زیرنویس و ترجمه • انتخاب زبان مبدا و مقصد	توسعه پنل مدیریتی	فاز ۲ توسعه سرویس های تکمیلی ۲ ماه
	امکان به روزرسانی سامانه امکان افزودن پلاگین امکان مانیتورینگ لحظه‌ای از عملکرد سامانه و تعداد کاربران آنلاین	افزودن قابلیت‌ها	
	تهیه راهنمای استفاده کاربر تهیه مستندات استفاده از API برنامه تهیه مستندات مربوط به معماری سرویس‌ها	مستند سازی	
	توسعه چارچوب تست دقت مدل ها تست کارایی مداوم تست عملکردی تست استرس ایجاد بسته های نصبی داکرایز کردن سرویس‌ها نصب و استقرار سامانه	تست و تحویل محصول نهایی	

References Link:

- https://www.interprefy.com/solutions/event-ai-speech-translator
- https://www.wordly.ai/real-time-translation
- https://speechlogger.com/
- https://krisp.ai/
- https://deepgram.com/ai-apps/vowel
- https://www.notta.ai/en
- https://www.scribbl.co/
- https://github.com/Sharrnah/whispering
- OpenAI **https://github.com/openai/whisper**
- Awexander **https://github.com/Awexander/audioWhisper**
- Blake **https://github.com/mallorbc/whisper_mic**
- Meta (LID, NLLB-200, M2M-100) **https://ai.facebook.com/blog/nllb-200-high-quality-machine-translation/**
- Meta (Seamless M4T) **https://github.com/facebookresearch/seamless_communication**
- faster-whisper **https://github.com/guillaumekln/faster-whisper**
- EasyOCR **https://github.com/jaidedai/easyocr**
- Silero (TTS, VAD) **https://github.com/snakers4/silero-models**

LEXEME AI INNOVATION HUB

Sponsored By ASR Gooyash Pardaz Co.

LEXEME AI INNOVATION HUB

The Lexeme AI Innovation Hub (LIH) serves as a pivotal arm of research and development for ASR Gooyash Pardaz company.

ASR Gooyesh Pardaz is a pioneer knowledge-based enterprise offering a wide variety of products in the field of Artificial Intelligence and Signal Processing. It was established in 2003 with the technical support of a group of professors and students from Sharif University of Technology.

PROJECTS

- **Shenasa:** Speaker recognition software

- **Ariana:** Text-to-speech engine

- **Nevisa:** Speech recognition system

- **Jooya:** Keyword spotting service

- **Vira:** Speaker voice conversion service

- **Dana:** Intelligent Assistant for businesses

- **Shiva:** Speech Enhancement Service

- **Niusha:** Intelligent Call Center Service

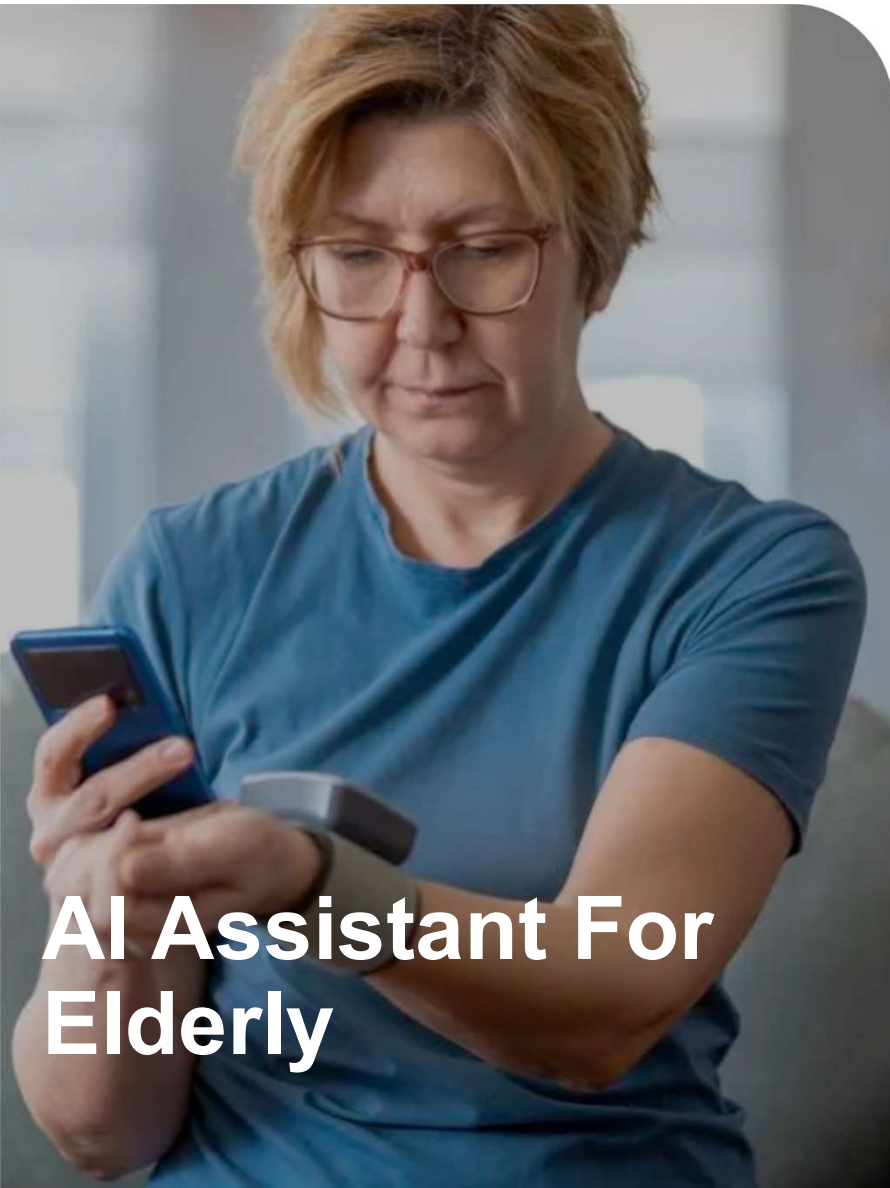- **Rominute:** Live Translate/Transcribe

OTHER SERVICES

- Semantic Similarity Service
- Formality Style Transfer Service
- Text Summarization Service
- Persian Text Spelling and Grammar Correction Service
- Speaker Diarization Service
- Speaker Recognition Service
- Machine Translation Service
- Persian OCR Service
- Speaker Voice Conversion Service

LEXEMEAI MISSIONS

- Definition and implementation of master company projects

- Contracting with organizations and implementing joint projects

- Carrying out scientific and research activities

- Conducting educational activities and providing mentoring services

- Providing support services in the form of a startup studio, startup accelerator, and co-working space

- Developing working relationships with technology companies in the Sharif Innovation District for the growth of the center

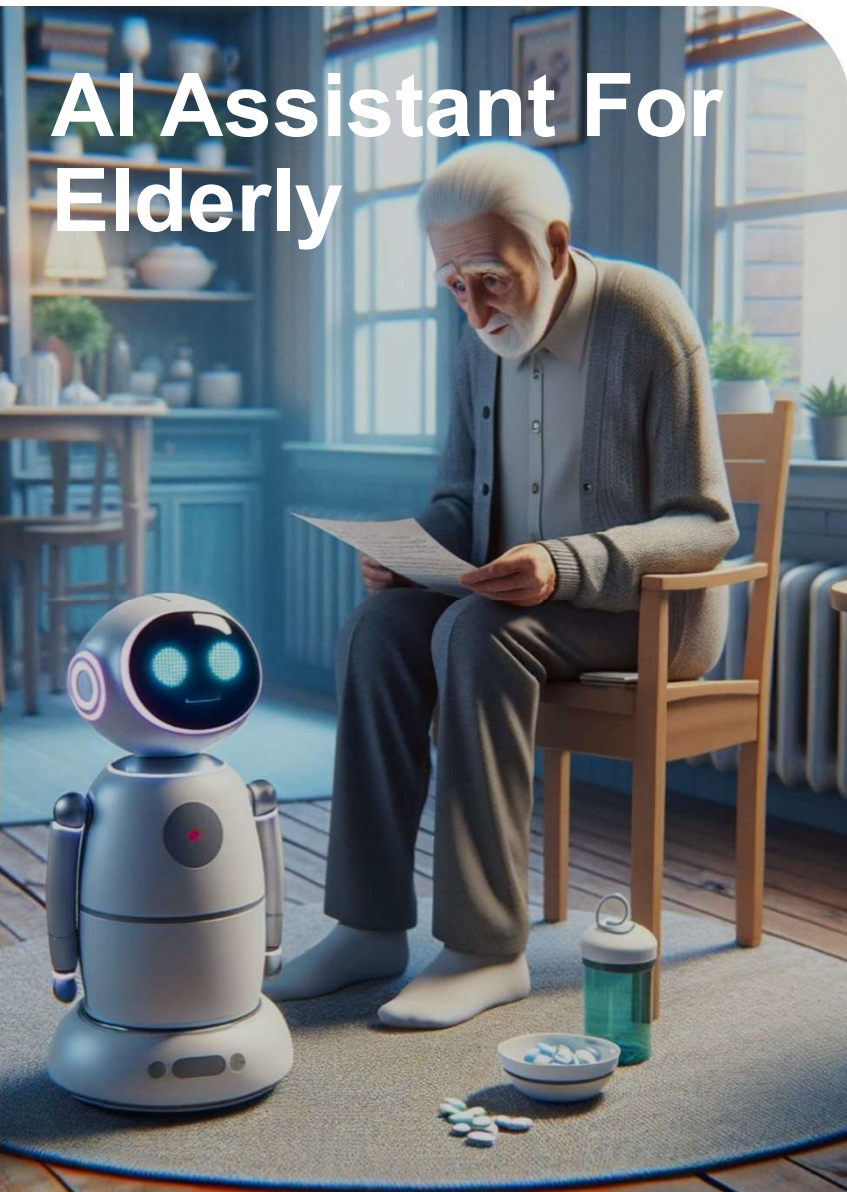- Networking and developing relationships with university laboratories

Objectives:

- Health monitoring

- Independent life

- Cognitive aids

- Social participation

- Support services

- Smart access to emergency services

AI Assistant For Elderly

AI Assistant For Elderly

Suggested Solutions:

- Voice assistant

- Wearable assistant

- Mobile assistant

- Health assistant

- House assistant

- Companion assistant

**Self Care
AI Assistant**

Benefits:

- Enhanced Patient Engagement

- Streamlined Administrative Tasks

- 24/7 Availability

- Improved Accuracy

- Cost-Efficiency

- Telemedicine Support

- Personalized Healthcare

- Quick Access to Medical Information

- Improved Data Security

- Enhanced Communication

Features:

- Medical answering service

- Prescription drug support

- Complaints and grievances

- Health risk assessments

- Medical record retrieval

- Appointment setting

- Patient satisfaction surveys

Niusha
Medical Call Center

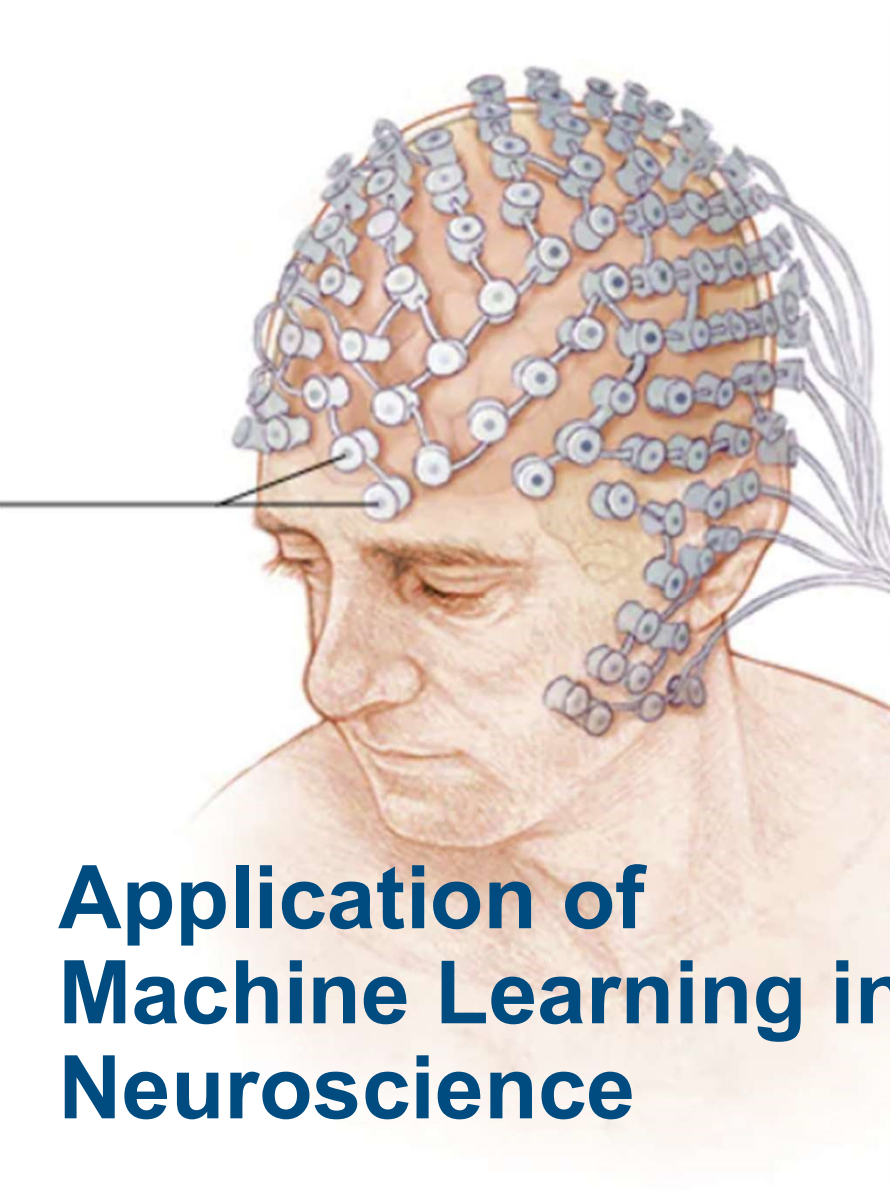# Live Translate / Transcribe

Features:

- Runs 100% locally on your machine.

- Speech recognition, translation and transcription

- OpenAI's Whisper, Supports ~98 languages

- LID [Language Identification] (Supports 200 languages)

- NLLB-200 (single model, Supporting 200 languages, high accuracy)

- TTS [Text-to-Speech] (Read out transcriptions/translations)

Cooperation with
Iran Autism Association

The Iran Autism Association (IAA) is a non-governmental, non-profit organization established in 2014 by specialists, experts, stakeholders, people with autism, and their families. Its mission is to improve the quality of life and empower people with autism.
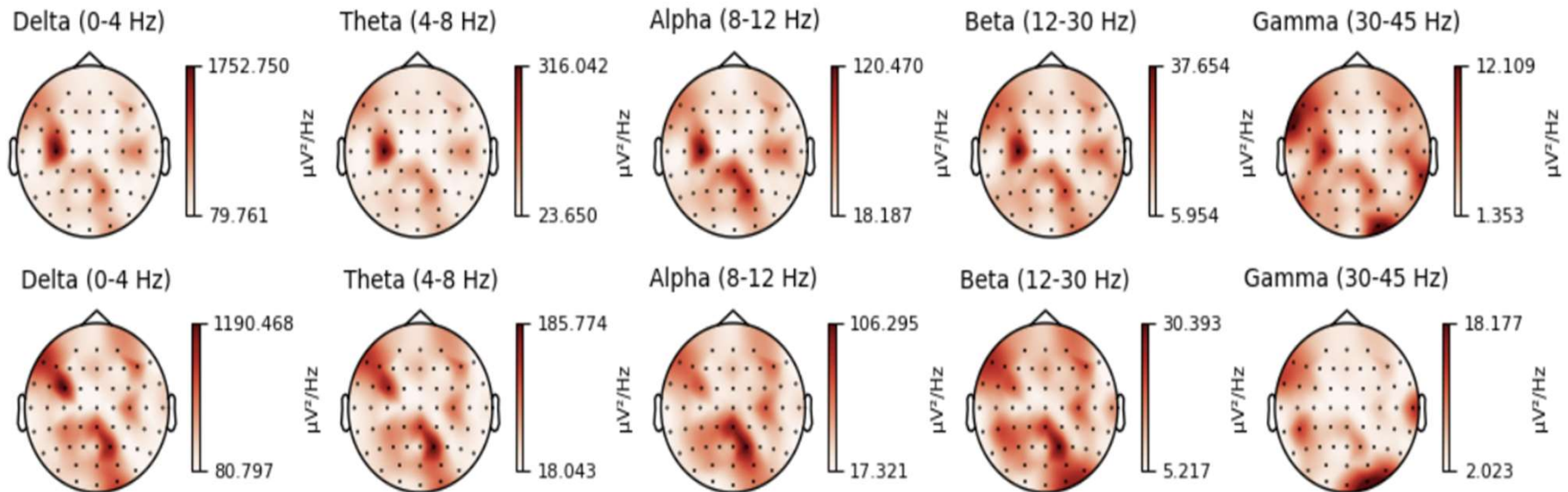
Application of Machine Learning in Neuroscience

- Reconstruction of Imagined Speech

- Interdisciplinary Collaboration

- BCI

- Computational Neuroscience

Reconstruction of Imagined Speech

- Record EEG signals

- Non-Invasive method

- Utilize machine learning for pattern recognition

Thank you for your attention