

Egor Lakomkin

Research Scientist | Multimodal & Speech Foundation Models

Paris, France · Open to relocation · egor.lakomkin@gmail.com · +49 178 111 80 86 · Google Scholar

PROFILE

Research scientist with 10 years of experience building large-scale multimodal and speech foundation models across pretraining and post-training. At Meta, co-led Llama 4 Speech tokenization and pretraining design, as well as reinforcement-learning post-training; led the post-training team for speech-to-speech translation models deployed in Instagram AI dubbing. Now develops large-scale streaming speech LLMs at Gadium, spanning architecture, training, data, evaluation, and post-training.

EXPERIENCE

Gadium — Member of Technical Staff | Paris, France | 2026–Present

- Pretrain and post-train large-scale streaming speech LLMs.
- Develop post-training recipes and evaluation methodologies for naturalness, expressivity, and instruction following.
- Work across model architecture, large-scale training, data quality, and post-training.

Meta Superintelligence — Staff Research Scientist | Aachen, Germany | 2022–2026

- Contributed to speech pretraining for Llama 3; co-led speech tokenization and pretraining design for full-duplex Llama 4 Speech models.
- Co-led the reinforcement-learning workstream for Llama 4 Speech, developing training recipes for factuality, instruction following, and conversational quality.
- Designed and ran large-scale multimodal training and evaluation experiments on distributed GPU infrastructure.
- Led a team that post-trained speech-to-speech translation models deployed in Instagram AI dubbing.

Amazon Alexa — Applied Scientist | Aachen, Germany | 2019–2022

- Developed large-scale multilingual end-to-end speech recognition models.

University of Hamburg — Doctoral Researcher | Hamburg, Germany | 2016–2019

- Developed multimodal emotion and sentiment recognition models combining acoustic and linguistic signals; achieved state-of-the-art results on IEMOCAP by transferring knowledge from an end-to-end speech recognition network.
- Developed a low-latency continuous emotion recognition model using deep reinforcement learning, reducing latency by 50% while maintaining accuracy.

Amazon Alexa — Speech Scientist Intern | Cambridge, UK | 2018

- Developed a deep neural network model for accent recognition.

SELECTED PUBLICATIONS

Efficient Streaming LLM for Speech Recognition — ICASSP 2025

The Llama 3 Herd of Models — arXiv 2024

AudioChatLlama: Towards General-Purpose Speech Abilities for LLMs — NAACL 2024 (Long Papers)

End-to-End Speech Recognition Contextualization with Large Language Models — ICASSP 2024

TECHNICAL FOCUS

Foundation models: multimodal models and streaming speech LLMs; pretraining; post-training; reinforcement learning; speech tokenization; model evaluation

Engineering: large-scale distributed GPU training; data quality and evaluation pipelines

EDUCATION

University of Hamburg — PhD Candidate, Knowledge Technology Group | Defense pending

Dissertation: “Emotion Recognition for Human-Robot Interaction with Deep Neural Networks”

Bauman Moscow State Technical University — M.S. in Computer Science | 2011 · GPA 4.4/5.0

LANGUAGES

English (fluent) · German (advanced) · Russian (native) · French (beginner)