

# Synthetic Speech Generation and Detection



Teamproject



Maximilian Kreutner

11.02.2026

# Background

- **Large companies** generate voice content of public figures
- **Voice Cloning** is also possible on (powerful) consumer hardware



- [CosyVoice](#)
- [coqui-tts](#)
- [GPT-SoVITS](#)
- [F5-TTS](#)

---

CosyVoice 3: Towards In-the-wild Speech Generation  
via Scaling-up and Post-training

---



F5-TTS: A Fairytaler that Fakes Fluent and Faithful Speech  
with Flow Matching

- Voice Cloning is **more accessible than ever**

→ How hard is it to **clone a voice**

compared to **detecting cloned voices?**

# The Project

- We will use **SOTA open-source models** to clone voices
  - Collect voice data
  - Finetune models
  - Compare different methods
- Use the best model to clone **several** voices – e.g. team project members
- See if we can **detect** cloned voices
  - Human judgement
  - Simple Machine Learning algorithms on wavs
  - State of the art detection models

# General Information



- Responsible person: Maximilian Kreutner
- Email: [maximilian.kreutner@uni-mannheim.de](mailto:maximilian.kreutner@uni-mannheim.de)
- Language: English
- Duration: 6 Months
- Min/max number of participants: 3-5
- Prerequisites: strong python skills
- Applicable to MMDS: Yes

