SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

أكاديمية طويق
TUWAIQ ACADEMY

Massively Multilingual Speech TTS

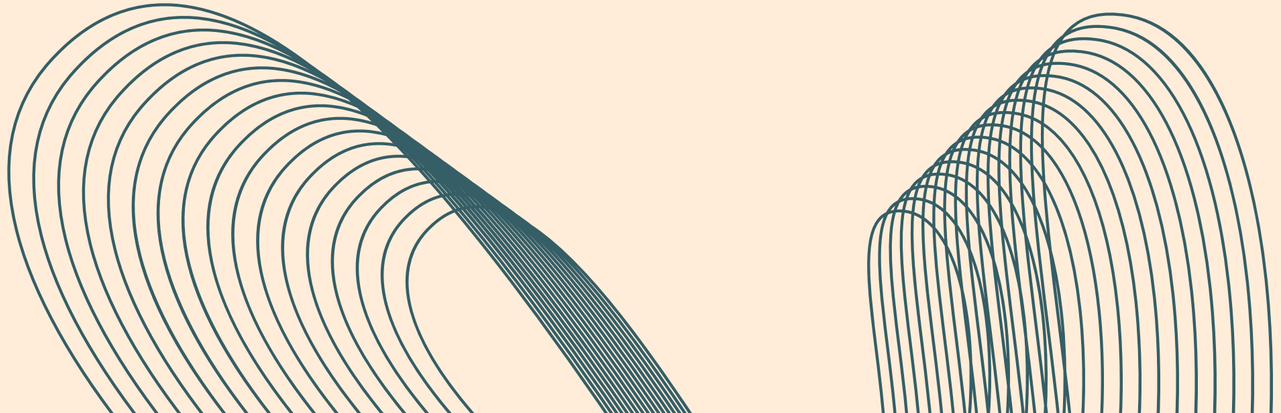
Mazen Alamri
Mohammed Alageel

SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

أكاديمية طويق
TUWAIQ ACADEMY

Model Overview

MMS-TTS is a self supervised model which is trained on unlabeled data over 1400 languages, it is a text-to-speech model that supports many languages, it is made of a conditional variational autoencoder (VAE) comprised of a posterior encoder, decoder, and conditional prior.



Applications

- Accessibility
- Navigation Systems
- Virtual Assistants
- Language Learning
- Many more

SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

أكاديمية طويق
TUWAIQ ACADEMY

Training Data

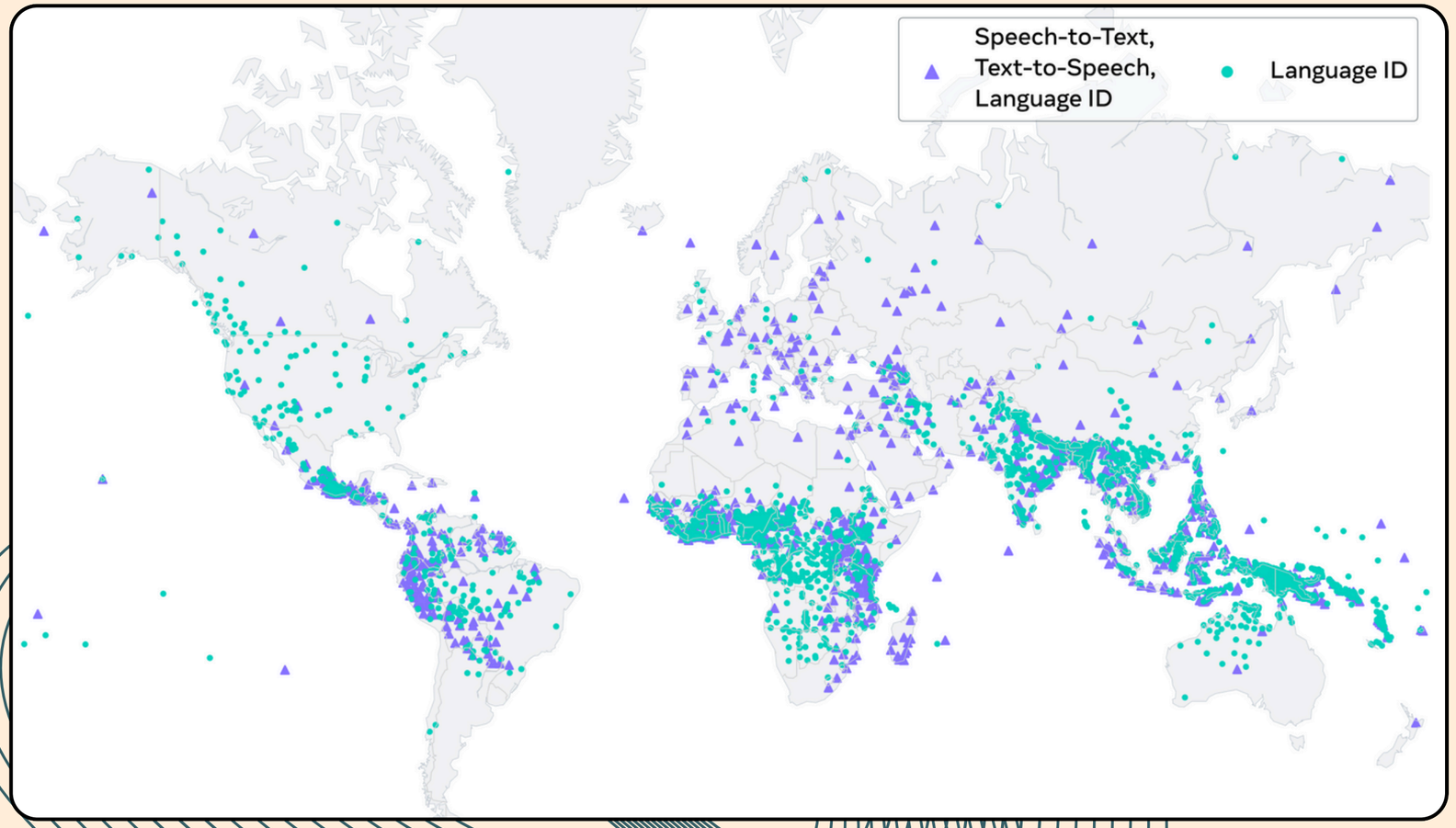
The pre-training data covers about 491K hours in 1,406 languages. This data is drawn from six training corpora with different characteristics.

- MMS-lab-U: 1,362 languages comprising 55K hours.
- Multilingual Librispech (MLS): 8 European languages of read books totaling 50K hours
- CommonVoice (CV): 89 languages totaling 8.8 hours of read Wikipedia text
- VoxLingua-107 (VL): 107 languages totaling 5.3K hours of YouTube content
- BABEL (BBL): 17 African and Asian languages totaling about 1K hours of conversational telephone data
- VoxPopuli (VP): 371K hours of unlabeled speech data in 23 languages derived from European Parliament event recordings

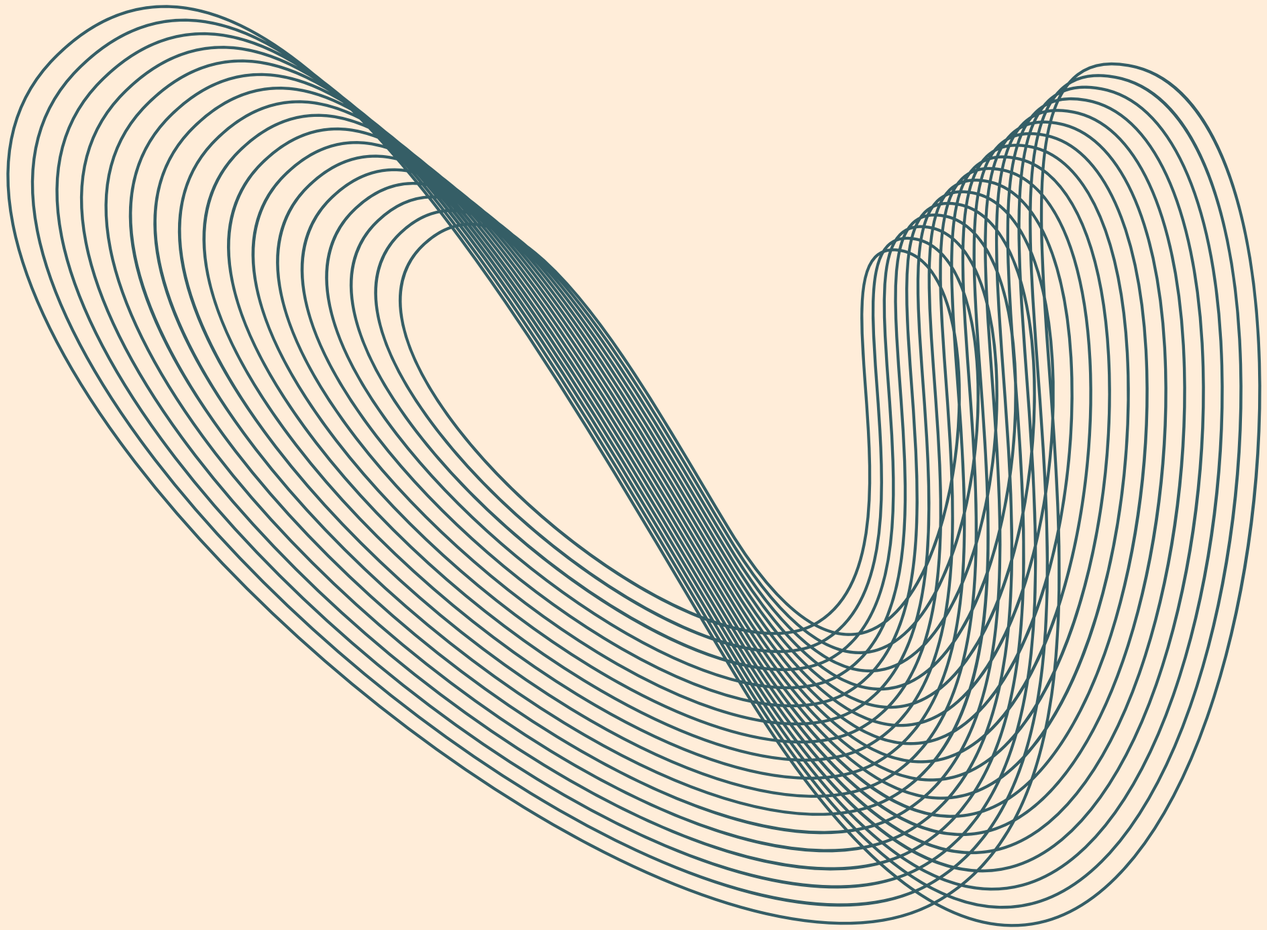
SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

أكاديمية طويق
TUWAIQ ACADEMY

Illustration of where the languages supported by MMS are spoken around the world: MMS models support speech-to-text and text-to-speech for 1,107 languages as well as language identification for 4,017 languages.



Example of usage



Example of usage

هَذِهِ تَجْرِبَةٌ. اَلسَّلَامُ عَلَيْكُمْ

SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority

أكاديمية طويق
TUWAIQ ACADEMY

Example of usage

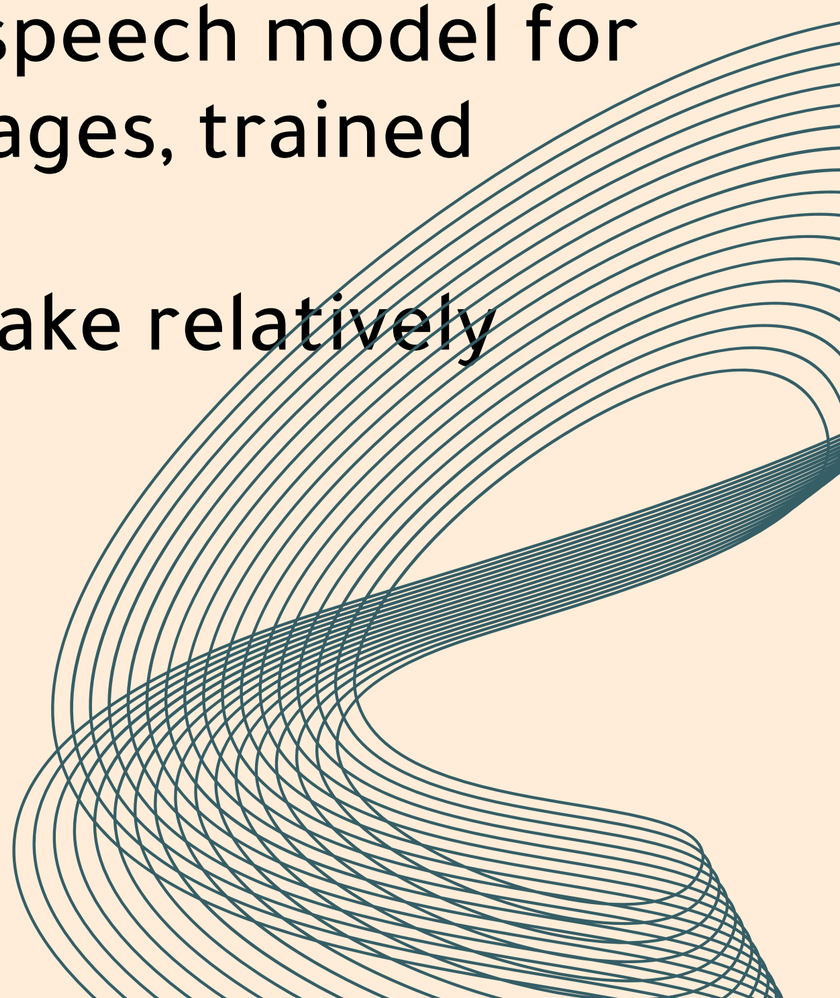
هاني الشافعي أفضل مدرب

SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

أكاديمية طويق
TUWAIQ ACADEMY

Conclusion

- MMS-TTS is a powerful multilingual speech model for using Text-To-Speech in many languages, trained with huge amounts of data
- We can use pre-trained models to make relatively accurate models in little time

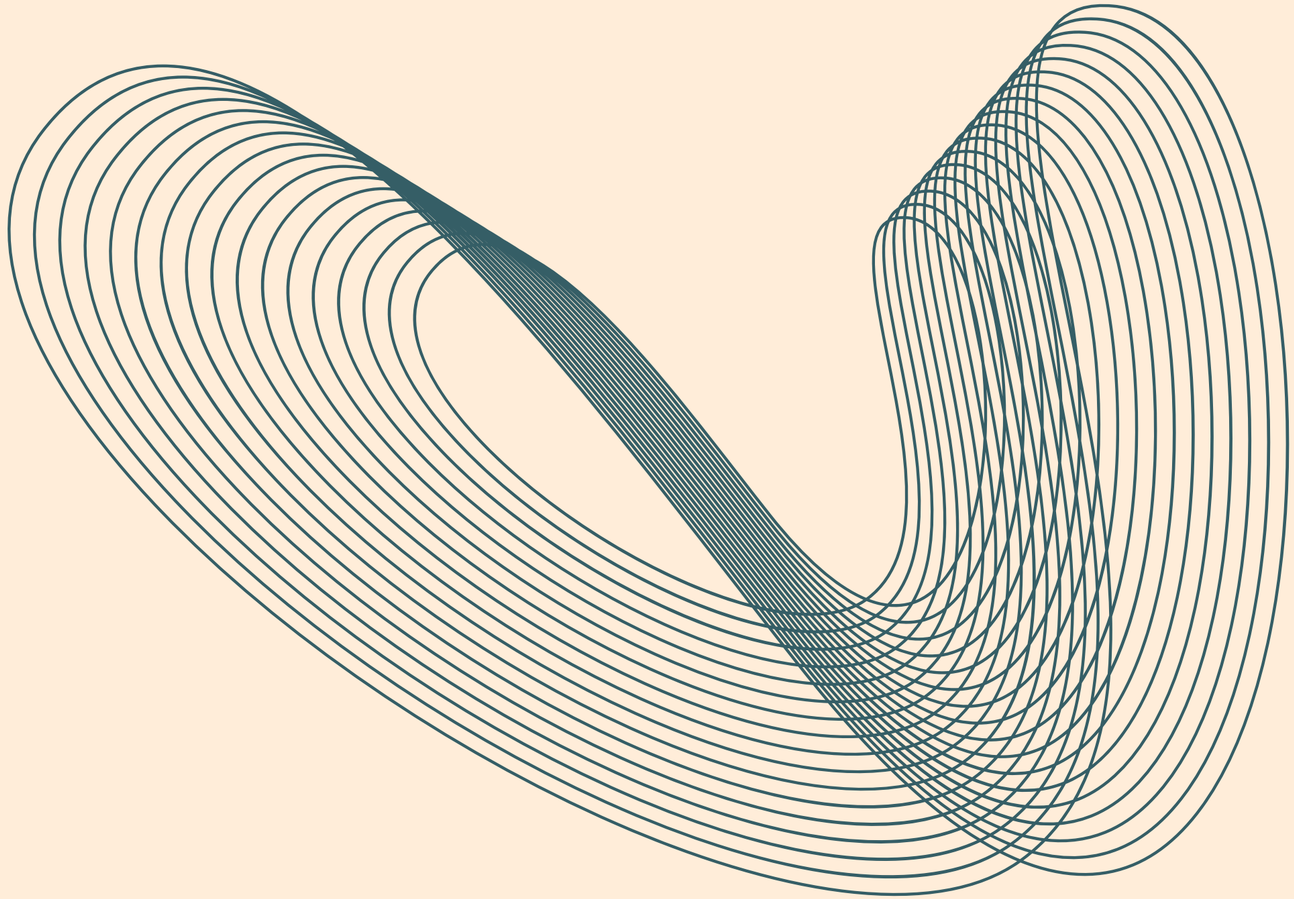


SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

أكاديمية طويق
TUWAIQ ACADEMY

References

- Model Paper
- Model Link (Hugging face)
- Model Link (Kaggle)
- Dataset Link



SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

أكاديمية طويق
TUWAIQ ACADEMY

Thanks for listening!