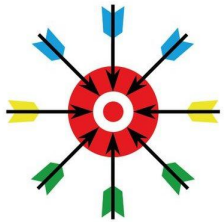
Voice technology: Opportunity for Africa

Presented by Samuel Rutunda
CTO, Digital Umuganda



Context: Africa

- Population:
  - 1.4 Billion (2022)
  - 2.4 Billion (2050)
- Languages:
  - Over 2000 languages
- Literacy:
  - ~ 67 % of adult pop.
  - <30% in some regions
- Culture:
  - Africa is mostly an oral culture

Context: Voice AI key terminologies

- ASR (Automatic speech recognition) also known as Speech To Text (STT)
  - Converts speech into text
  - Requires a large amount of data to build a good model
  - Prone to bias
- TTS (Text to Speech) also known as Speech Synthesis
  - Converts text into speech
  - Requires studio quality data (between 10 and 20 hours)
- MT (Machine Translation) also known as NMT (Neural Machine translation)
- Speech to Speech Translation (S2S Translation)
  - Translates speech in one language to speech in another language
- Natural Language Understanding (NLU)
  - Extract meaningful information from the text
  - These include intents, entity, response, Question answering

Context: Voice AI key technologies (contd.)

- Speech to text translation
  - Convert speech from one language to text in another language
- Speaker identification
  - Identify the identity of the speaker
  - Can be used to by banks for entering credentials
- Speaker diarization
  - Contains speech labelling, speaker identification and multi-speaker detection
- Conversational AI
  - A combination of STT, TTS and NLU
  - Also known as voice chatbot

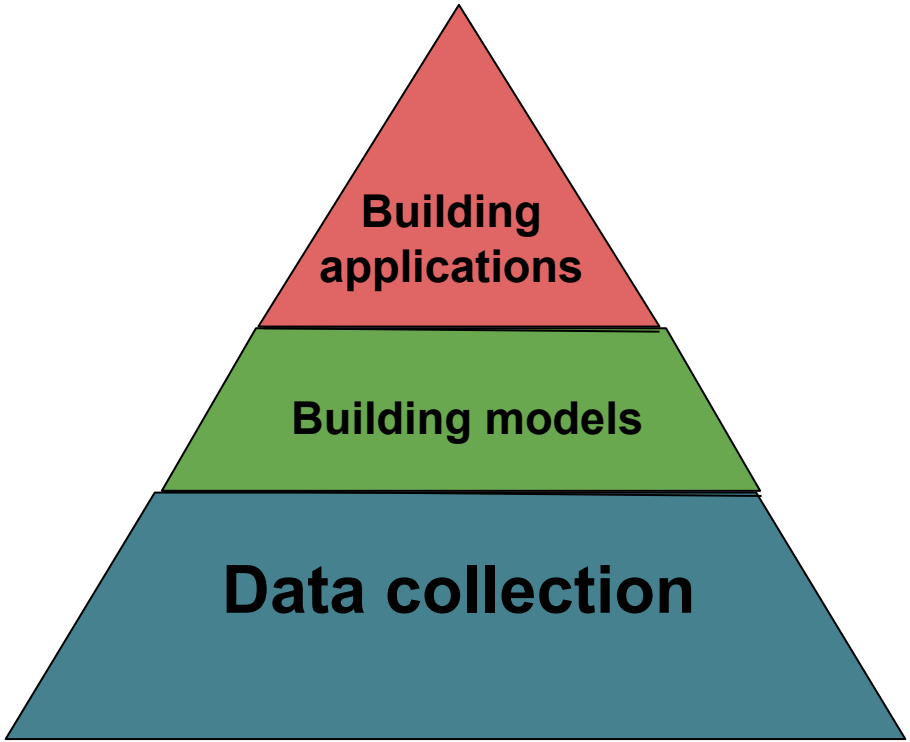
Context: Voice AI major breakthrough

Paper	Organization	Release date
Deepspeech end-to-end ASR model	Baidu	2015
Tacotron end-to-end TTS	Google	2017
Transformers	Google	2017
B.E.R.T.	Google	2018
Common voice platform	Mozilla	2017,2018
Wav2Vec	Facebook	2019
GPT-3	OpenAI	2020
Whisper	OpenAI	2022
No Language left behind	Facebook	2022
S2S Translation for a real-world unwritten language	Facebook	2022

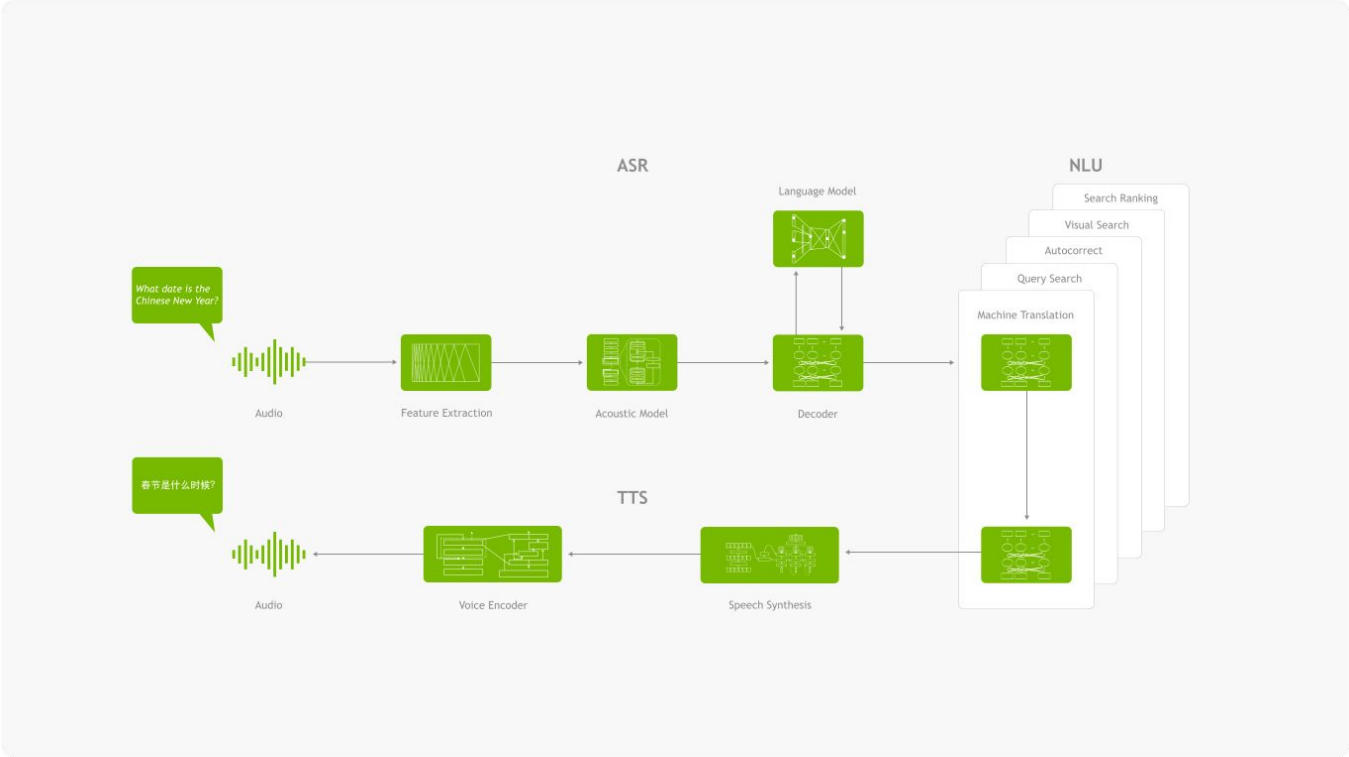
Applications of Voice AI

- **Voice chatbot**
  - Web voice chatbot
  - Call center voice chatbot
  - IVR automation (replace human Audio by TTS)
- **Voice transcription**
  - Meeting minutes transcription
  - Subtitles creation
- **Voice data analytics**
  - Keyword spotter
  - Sentiment analysis
  - Speaker identification
  - Speaker verification
- voice translations
  - Live translations
  - Inter-African E-commerce translation

ML process



Application diagram: conversational AI



Our work at Digital Umuganda

Data collection:

- **STT**: 2000 hours of Kinyarwanda dataset
- **TTS**: 17 hours of TTS studio recording
- **NMT**: 50,000 English to Kinyarwanda sentences
- **Text Classification**: 7,500 Kinyarwanda tweets classification

Next Step:

- **OD4All**: leverage our experience to do data collection for African languages

Our work: Models (contd.)

ASR Models

- Deepspeech Kinyarwanda model (done)
- Conformer Kinyarwanda model (in progress)
- Conformer Multilingual model (in progress)

TTS models

- Fastpitch Kinyarwanda model (done)
- Festvox based Kinyarwanda model (done)

NMT models

- Joel NMT kinyarwanda-english model (done)
- M2M100 kinyarwanda-english model (done)

Sentiment analysis

- Bert based twitter sentiment analysis (done)

Applications

Chatbot

- Covid text chatbot (done)
- Covid voice chatbot (in progress)

Translation

- Kinyarwanda-English voice translation (in progress)

Challenges of voice technologies for Africa

Tools are not necessarily built with African languages in mind

- Tokenization of Bantu language can be different from existing technologies
- Bantu language do have tones

Limited datasets

- AI always needs datasets, the more … the merrier

High cost

- Expensive cloud (Inference server $30/month for CPU vs $300/month for GPU)
- High cost for data collection
- Requires specialized skills

Opportunities

Interest in building AI infrastructure in Africa

- By for profit companies: Facebook, Google, Microsoft, Nvidia
- By non-profit: GIZ, ICDRC, Lacuna

Fast-paced discoveries

- NLP/Voice AI are currently trending and a lot of discoveries are taking place

Open Source tools

- Many open source tools, applications and models

Multiple supportive communities

- **In Africa**: Masakhane, Lanfrica
- **In the world:** Coqui, RASA, Label studio, Commonvoice
- **General platforms**: Huggingface, Twitter

Platform

Demo

- umuganda.digital

Datasets & Model

- Digital Umuganda Huggingface
- Mbaza nlp Huggingface

Applications

- Digital Umuganda Github
- Mbaza-nlp Github

Q&A

Thank you