

Success Story

Building an Automatic Speech Recognition engine for AI-enabled legal tech software

Customer

A Nigerian legal technology solutions company

Country

Nigeria

Industry

LegalTech

About The Client

The client is a Nigerian legal technology solutions company that spearheads the development of innovative software products for seamless judgment delivery. The firm has also been providing a diverse catalog of digital solutions that assist 1000s of law students, legal practitioners, lecturers, and judges in their research work.

Technology Stack

ASR

 TensorFlow

Business Situation

Transcription of court case proceedings are essential for helping legal practitioners and judges in delivering trial outcomes accurately. Manual transcription requires very delicate work and careful attention to every single spoken detail during a particular deposition, which often leads to human errors in the final transcript.

Additionally, there is a considerable shift in accents while moving from region to region within Nigeria, and human transcribers are unlikely to be accustomed to all of these accents. So they would sometimes miss out entire sentences while listening and transcribing.

The legal technology company had the innovative idea of facilitating error-free transcription in Nigerian courts with an audio transcription software. The company chose to leverage Unthinkable's expertise in AI, and in enabling businesses to train machines to process large volumes of natural language data flawlessly. The Unthinkable team was required to ensure that the audio transcription software they would develop satisfies the following conditions:

- ✔ Enable the capability to transcribe streaming audio and video
- ✔ Automate the process of data augmentation into the appropriate format and size that would be consumable for the software
- ✔ Train the software to understand variations in accents and intonations in spoken English across regions in Nigeria
- ✔ Eliminate issues with noise that could lead to unnecessary aberrations in the transcripts
- ✔ Regularly maintain and train software solution based on new data sets collected from proceedings and judgements.

The Solution

When it comes to speech recognition technologies, the world is still figuring out the best AI strategies to do so with the cleanest possible methods. The Unthinkable AI team opted for the transfer learning-based Machine Learning (ML) model to develop an Automated Speech Recognition (ASR) engine for transcribing live court depositions.

The development of the ASR engine and training the machine to listen to audio and transcribe accurately involved a lengthy process. There was plenty of research and innovation that went into the conceptualization of this process as described in the following sections:

Automating Data Preparation

The ASR machine learning model and engine required thousands of hours of data – both audio of depositions and previously transcribed texts. The audio had to be mono-channel, with a 16 kHz sampling rate for the model to be trained to give more accurate outcomes. This preparation of data before being fed to the ASR engine was automated by our team, so that the engine was equipped to decipher several local Nigerian accents and nuances in the vocabulary.

Pre-Training For Data Augmentation

Once the data was fed into the ASR engine, next came the inclusion of optional modules for data augmentation and pre-training wherever required. Pre-training the ASR machine learning model meant fine-tuning its accuracy through transfer learning methodologies. So the audio files were in WAV format of several people with different linguistic nuances saying the same words. Data training was done using a set of 100,000+ words and linguistics samples. The audio had to be converted and the model was monitored for loss and accordingly fixed when needed.

Incremental Readability

The post-processing pipeline of the ASR machine learning model is quite complex in terms of the text produced and may occasionally have readability issues. An Inverse Text Normalization (ITN) mechanism was implemented to take care of delicate changes that improved the overall readability of the transcribed depositions.

Accurate Text Records

Audio of depositions can be transcribed with nearperfect accuracy. The ASR engine has been trained with so much data that the Word Error Rate (WER) and Character Error Rate (CER) of the audio transcription software never rises above 20%. Depositions, conversations, and hearings can be transcribed to the very letter with no redundant elements caused by noise.

The legal technology solutions company was able to use the final product to help hundreds of courts across Nigeria to automate the transcription of case proceedings, hearings, and depositions and achieve 80% accuracy of speech-to-text conversion. This made for greater accuracy in the resultant transcripts and ease in delivering judgments. This also reduced the TAT of documentation process by 35%. Unthinkable's innovative AI approach in developing the solution and swift turnaround time were extremely appreciated by the company.

The Impact

Is there a digital platform you want to build or take to the next level?

Setup a personalized consultation with our technology expert.

Let's Talk