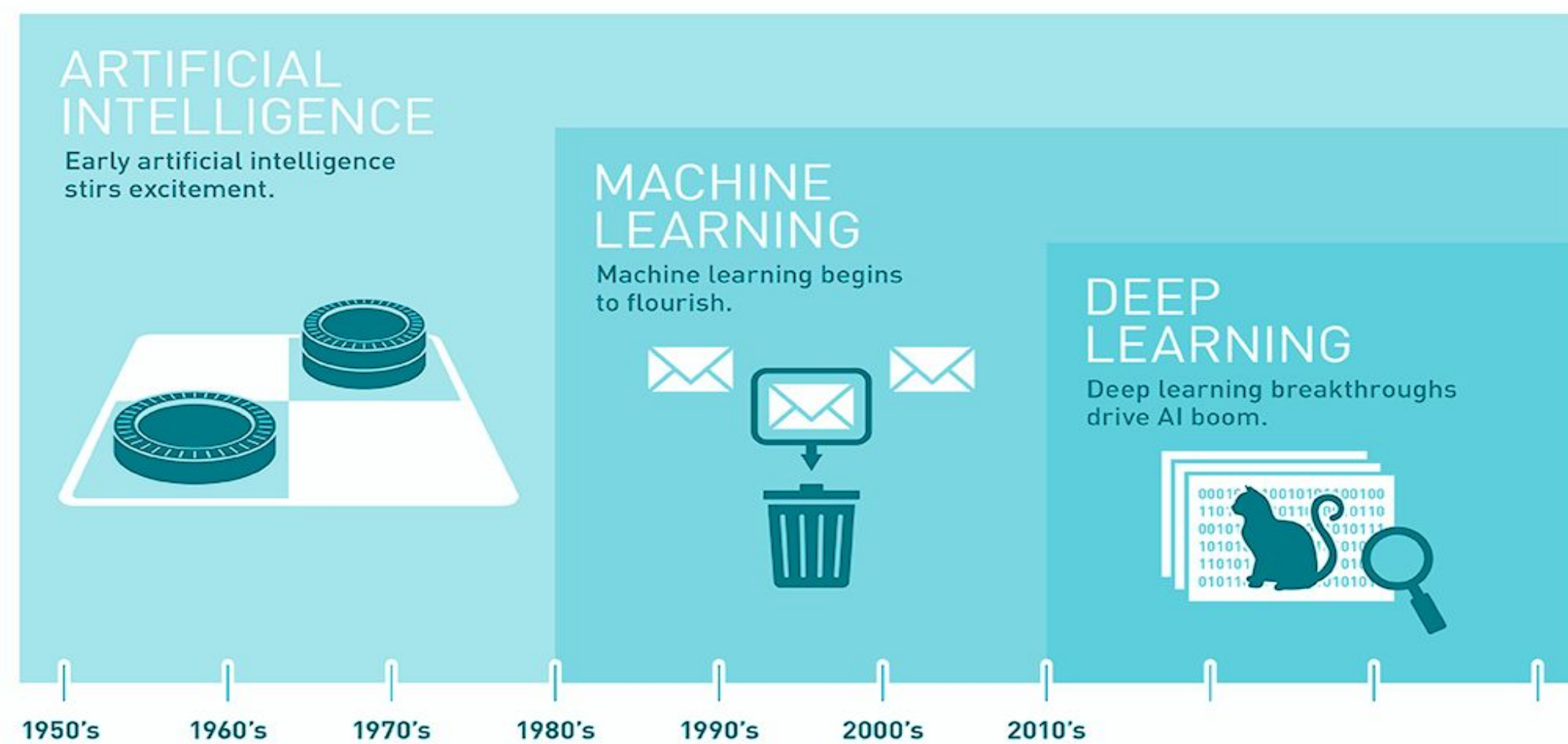


The Project

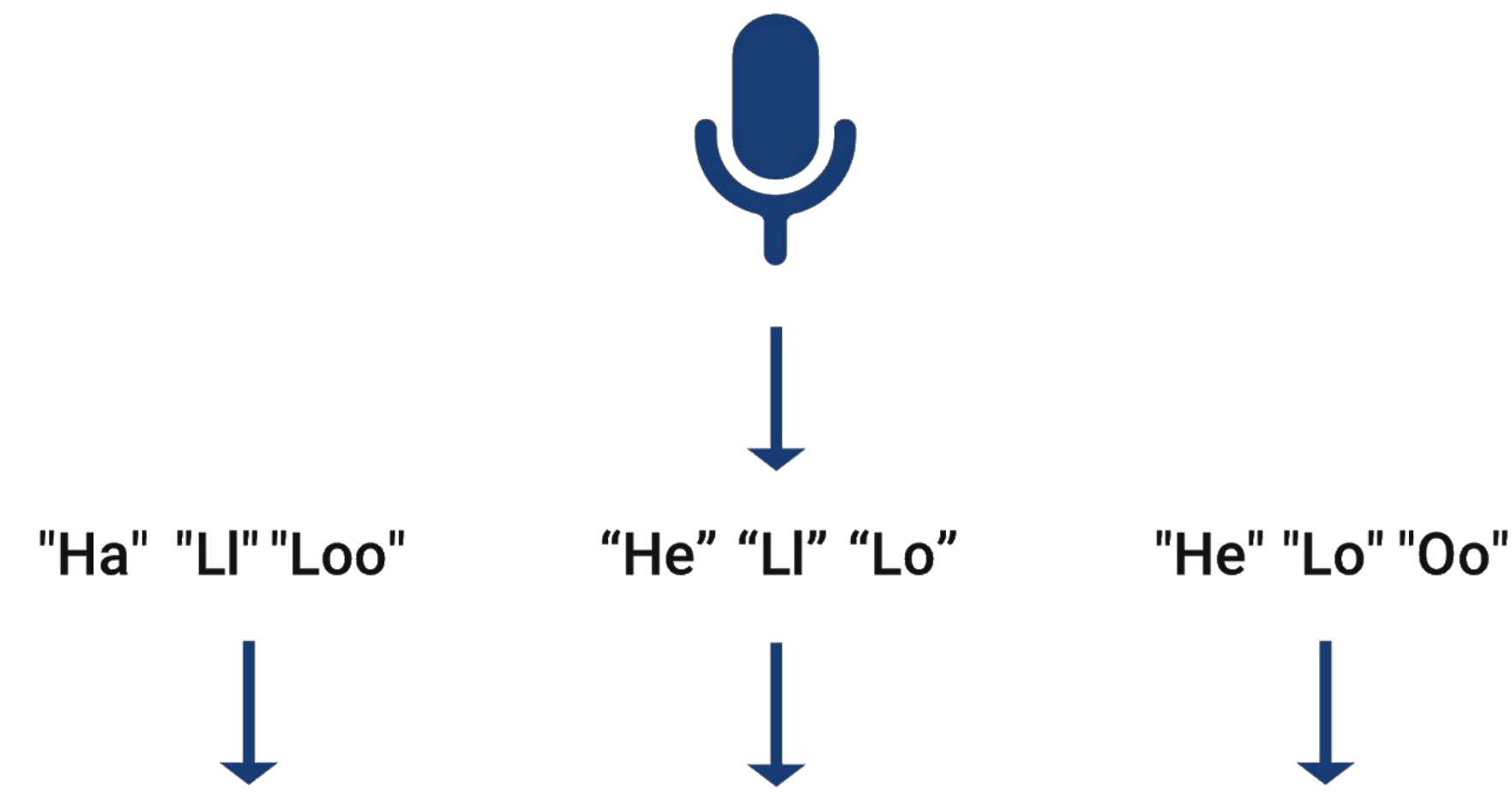
We partnered with Veritone to innovate a new voice recognition engine that can achieve 93% accuracy under any condition making anyone audible anywhere. To achieve this level of accuracy we trained a neural network to recognize phonemes, the building blocks of languages, under noisy conditions.



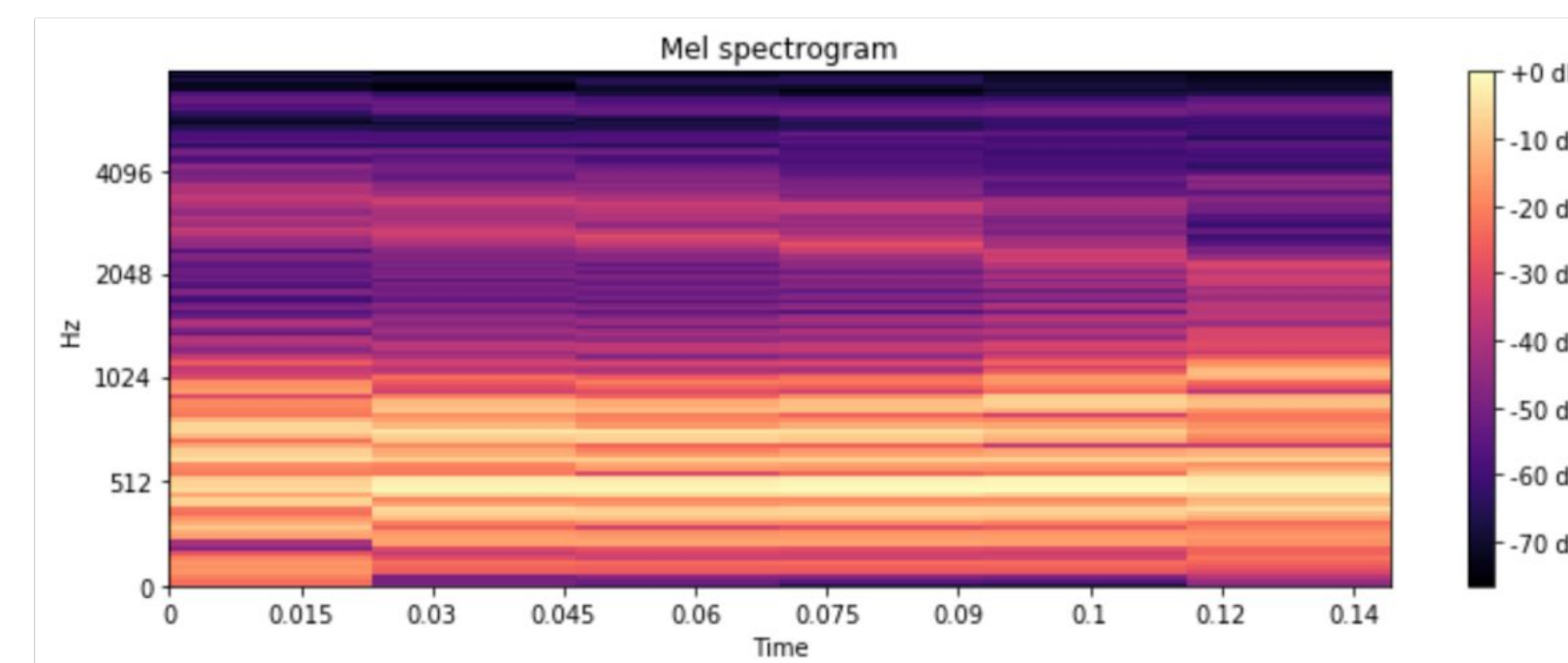
Why Deep Learning?

Deep learning is a subset of machine learning, where networks learn from patterns in unstructured data. Similar to the human brain, deep learning adopts neural networks to extract high dimensional data using multiple layers. Voice recognition is a difficult technology to tackle due to the high variability of audio data. Using deep learning, we can break the threshold to create a cutting-edge product.

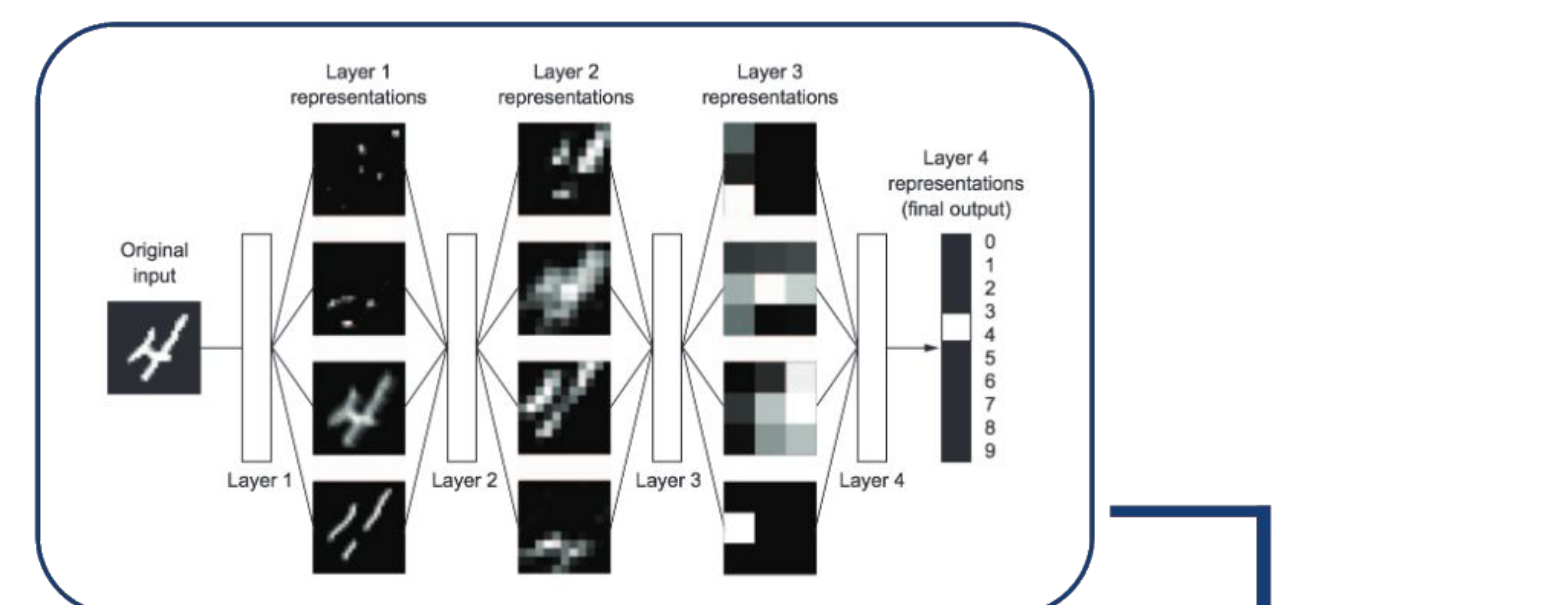
Inputs



Phoneme



Neural Network



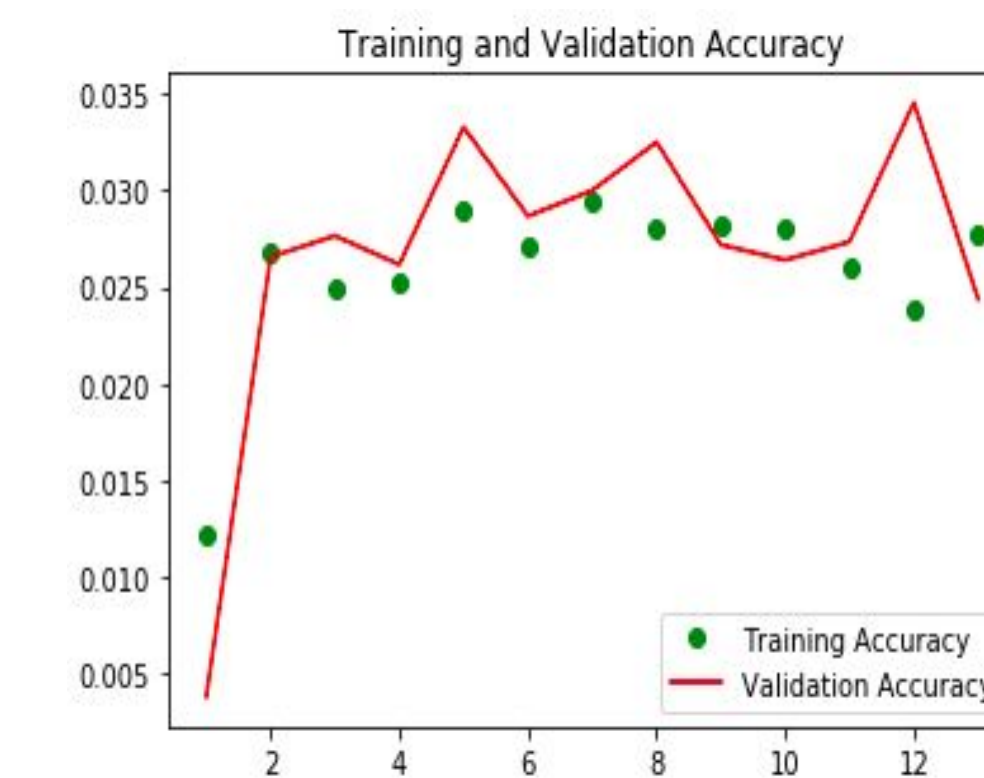
Output

"Hello!"

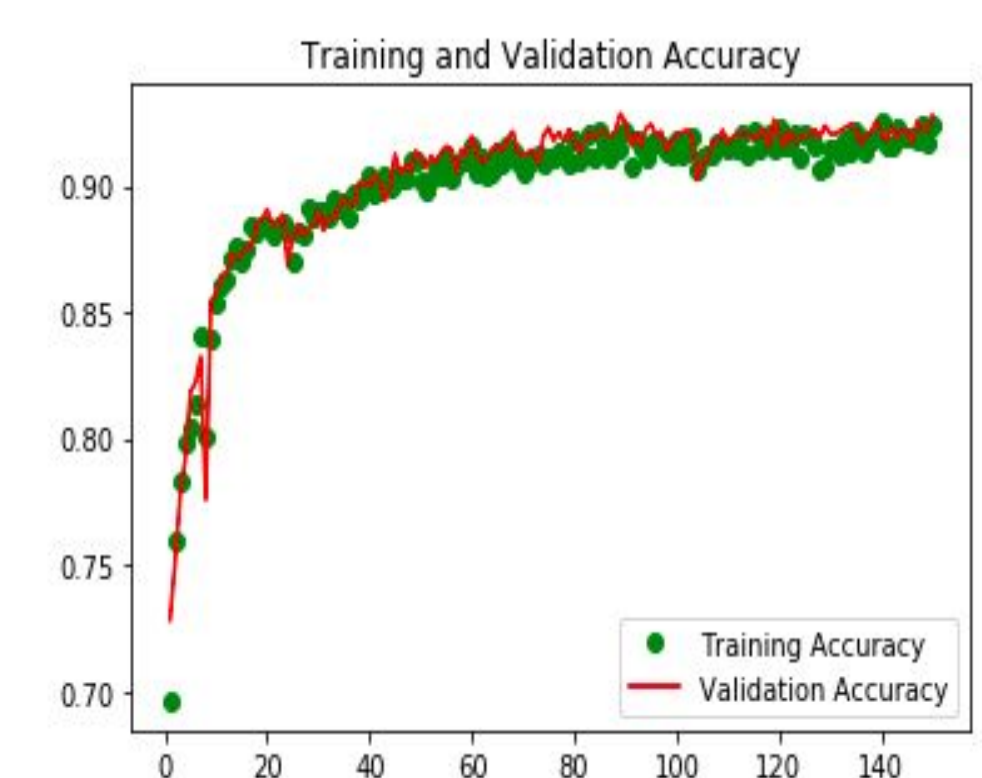
Our Solution

Our solution is to start with the building blocks of languages called phonemes. We slice the audio files into many short intervals containing one phoneme each and convert them into spectrograms. Then, we design and train our artificial neural network, which is evaluated based on accuracy, precision and recall.

Before



After



Results

The neural network works with 93% - 99% accuracy on different phonemes. After working on the project, we have gained some knowledge of deep learning concepts and how to construct a simple neural network using open source libraries. We will utilize what we learned in the project to further expand the applications of deep learning algorithms in other fields.