



# How to Harness AI for Killer Transcription

Neural networks drive quantum leap in accuracy and speed



# Executive Summary

You don't need a PhD in computer science to understand how Artificial Intelligence (AI) is revolutionizing automatic speech recognition (ASR).

ASR technology has come a long way in recent years, with significant advances in both accuracy and speed. One of the most promising developments in the field is the use of deep neural networks (DNNs) to power ASR engines.

In this ebook, we will explore the basics of DNNs and how they are used in ASR technology, including:

- What is driving the move to DNNs
- How DNNs improve accuracy
- Benefits to Total Cost of Ownership
- Maximizing Return on Investment



# AI is Everywhere

The final months of 2022 saw a veritable explosion of articles about AI in the media.

It seemed everywhere you looked, there were new stories about the latest AI breakthrough or the potential impacts of this rapidly evolving technology:

- A fantasy image painted by the AI image generation tool Midjourney beat human artists at a state fair.
- Social media accounts were populated with uncanny selfies created by Lensa.
- A company announced that its AI-powered chatbot would be the first to fight a traffic ticket in court.

AI is at a crossroads, and in the world of speech recognition, AI has a lot of upside.



# AI in Speech Rec

When it comes to speech recognition, AI is a more efficient process for translating audio utterances into readable text.

Traditionally, automatic speech recognizers have used preloaded dictionaries to try to match a spoken word with its text equivalent.

This worked well enough, particularly in interactive voice response systems with a limited set of possible inputs.

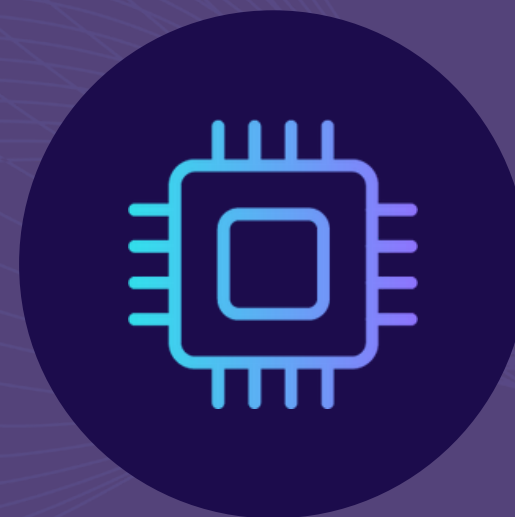
But a new AI speech recognition engine has just raised the standard for accuracy.





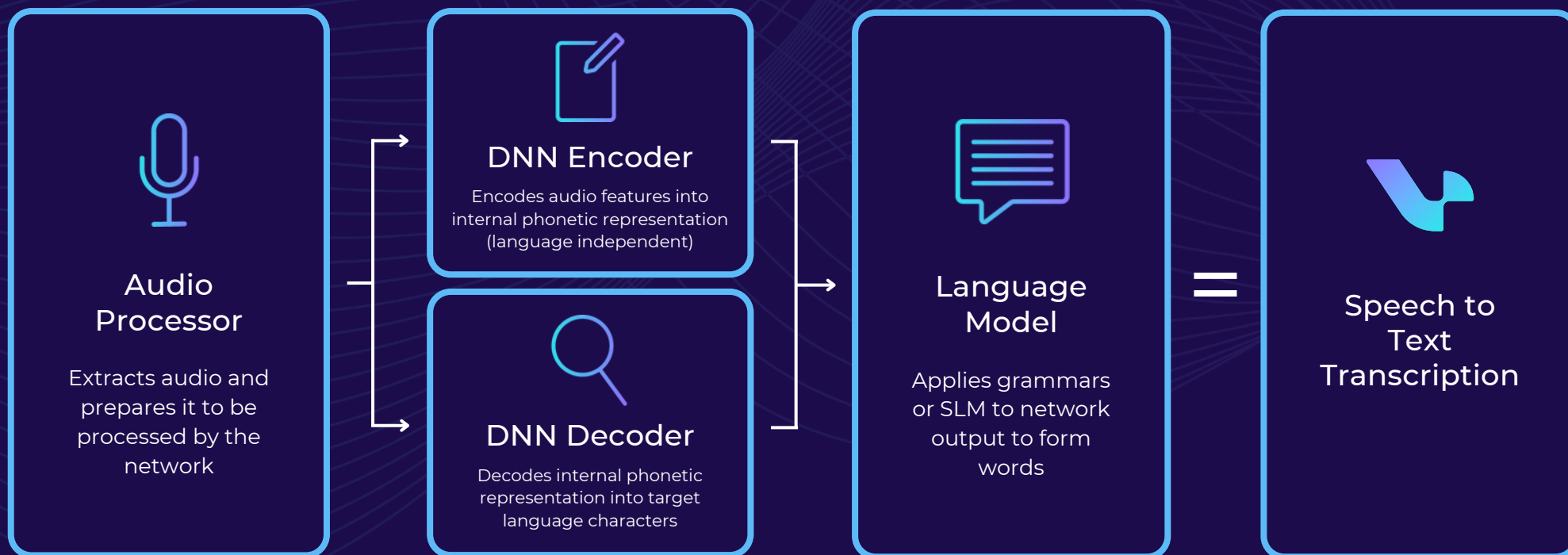
# Why AI is Winning

- ✓ Significantly improve accuracy
- ✓ Quickly learn new dialects and accents
- ✓ Easily add additional languages
- ✓ Handle lengthy transcription
- ✓ Deliver transcription faster
- ✓ Reduce footprint of storage and memory without sacrificing performance



# How It Works

At its simplest, this diagram shows how the new ASR engine uses deep neural networks to process speech:



# Lifecycle of a DNN

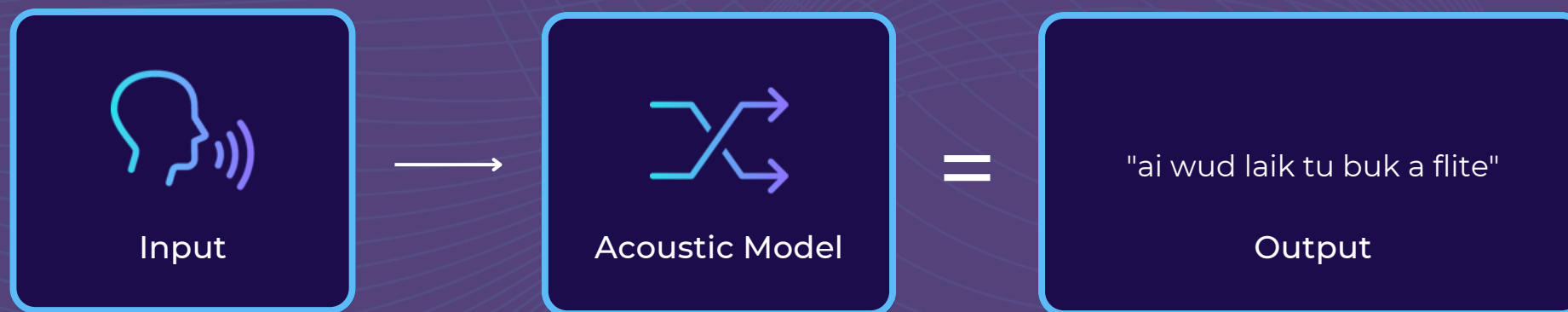
Implementing a deep neural network for language processing has 3 phases:

- 1.) **Training:** The neural network ingests a large sample of raw audio correctly matched to corresponding text. From this, the neural network can infer relationships and understand linguistic probabilities.
- 2.) **Testing:** The model is evaluated on test data that it has previously not seen during the training process. If the neural network scores high accuracy that means it has correctly interpreted the linguistic rules from the training data.
- 3.) **Transcribing:** The neural network goes live with customers.



# The Old Way of Doing ASR

This diagram shows how automatic speech recognition was traditionally done:



*It required a separate acoustic model for each dialect, and the pattern matching was grossly inefficient.*



# The New ASR Engine

Here you can see how the efficiency of acoustic modeling improves through the use of a deep neural network:



# The Louie Armstrong Problem

In the classic song “Let’s Call the Whole Thing Off,” jazz trumpeter Louie Armstrong comments on the differences in how we pronounce the same word:

*You like potato, and I like po-tah-to*

*You like tomato, and I like to-mah-to*

*Potato, potahto, tomato, tomahto*

*Let’s call the whole thing off*

Armstrong sang these lyrics long before the advent of automatic speech recognition, but he was presciently pointing to what would be a major problem for traditional ASR engines:

*How to account for all the different ways people from different regions pronounce the same word?*



# How DNN Solves It

A deep neural network leverages hundreds of thousands of hours of audio, covering every possible dialectal variation, in every possible environmental condition.

With so much variation in the training model, the neural network can recognize a word regardless of whether it's pronounced by an American, Englishman, or Australian.

Americans say tomato, while British say tomahto. The DNN can understand them both.

Louie Armstrong would approve.



# Tuning Out Noise

Because the DNN learns from whatever data you feed it, the neural network can also be trained for specific acoustic conditions.

By feeding the DNN data recorded in a noisy environment, it will learn to filter out the extraneous sounds and stay focused on the speech it is transcribing.

If all customer interactions come in the form of calls, for example, the DNN can be trained with a telephone dataset that screens out cellular interference and call center noise.

While this kind of fine-tuning is possible with DNNs, in most applications the general training is more than sufficient to get the job done.





# Domain-Specific Terms

In certain specialized fields – such as medical terminology – the DNN benefits from a more customized language model.

There are 2 ways to add domain-specific vocabulary:




- 1.) **Adding lists or phrases:** A typical use case would be adding a list of pharmaceutical drugs or specific restaurant dishes.
- 2.) **Adapt with domain-specific texts:** If there isn't enough data to fully train the new language model, you can use text to boost recognition of domain-specific audio.

In either case, we can assist in tuning the DNN engine to your specific needs.



# Performance Testing

We compared transcription accuracy across two datasets:

Dataset	 amazon Transcribe	 Google Cloud Speech API	 LumenVox®
Digits	94.3%	89.3%	98.5%
Phrases	71%	73.9%	84.6%

The new ASR engine achieved a tremendous performance advantage over Amazon and Google thanks to end-to-end DNN.

# Key Takeaways

- Legacy ASR systems may perform well enough in the short run, but deep learning will continue to outpace them and the gap will widen.
- Traditional ASR models have their performance capped at a threshold, beyond which additional training sees limited returns.
- Deep neural networks can learn to recognize the same word from different pronunciations and dialects, solving the Louie Armstrong problem.
- The new deep neural network ASR engine delivers a major improvement in transcription accuracy.



# Industry-Leading ASR

The LumenVox engine harnesses Deep Neural Networks to understand what customers are communicating, no matter how they speak. LumenVox's suite of services includes speech recognition, answering machine detection, automated transcription, and identity verification.

## Why LumenVox?

- ✓ Better ROI
- ✓ Lower TCO
- ✓ More Flexible
- ✓ More Accurate

## Looking to build next generation voice experiences?

Request a demo today to see how we can save your customers time and money.

Book a demo →