CONTENT GURU

Engagement Made Easy®

Whitepaper

"Did You Get All That?" The Three Thousand Year History of AI Transcription

3400 BCE
saw the invention of the first form of written transcription.[1]

1952
saw the invention of the first computer capable of recognising human speech.[2]

20,000
words could be recongised by transcription software by 1986.[3]

<5%
error rate for transcription software by 2017.[4]

Executive Summary

In this whitepaper, we discuss how AI can enhance the customer journey at every stage; before, during, and after interactions.

1-2
The First Three Thousand Years

3
The First Steps

4
Neural Networks

5
Generative AI

6
The Story of Together Money

7
Content Guru: Your CX AI Partner

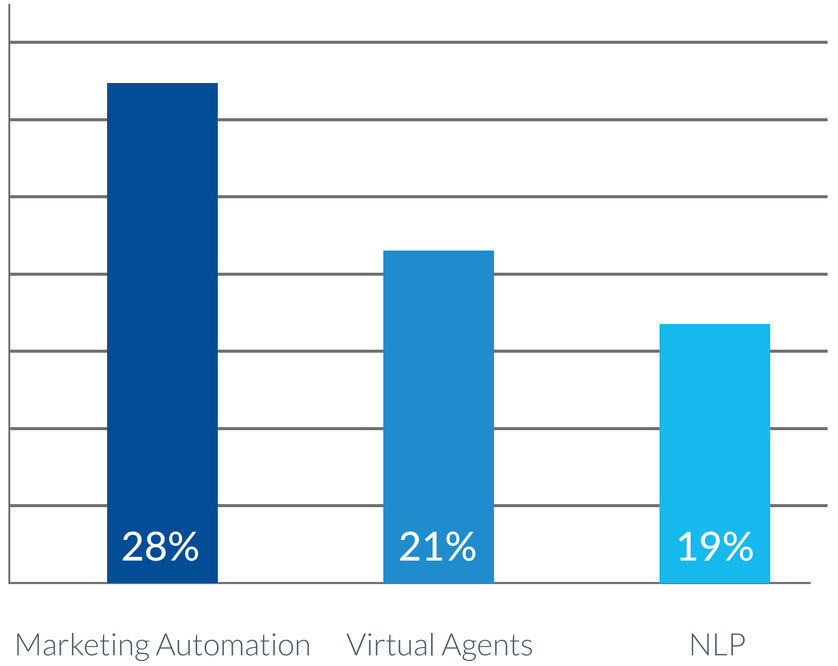
Transcription: The First Three Thousand Years

Today, AI speech recognition and transcription are ubiquitous.

It's in our homes, in our smartphones, even in our TVs and our cars. The ability to interact with a device by speaking aloud is expected. But it wasn't always that way. Creating a computer capable of recognizing and processing human speech was a challenge that preoccupied some of the brightest minds in computer science for decades. The story of how AI speech recognition was finally cracked reveals much about the nature of the AI landscape today, and how we can best adapt to an uncertain future.

Today, AI speech recognition is becoming more and more essential to enterprises and public-sector organizations. Now that transcription accuracy levels are high enough and consistent enough to justify mission-critical use cases, enterprises are looking for ways to integrate AI transcription technologies that are both cost-effective and efficient. This is particularly relevant to the realm of customer communications: customer communications represents the single largest area of AI adoption across the economy, with nearly 30% of all AI applications being customer-focused in some way, the largest single use-case by volume.[5] The future of AI is in CX.

To understand the future, it's useful to know where we came from. In this whitepaper, we examine the three thousand-year history of transcription, from its earliest days, all the way through the computing revolution of the twentieth century, to the potential of generative AI. We examine the unfolding impact of generative AI, and what that means for business efficiency in real terms. Plus, we take a look towards the future and examine ways in which organizations can capitalize on the opportunity.



AI adoption by use-case (US Census Bureau).[6]

Transcription: The First Three Thousand Years

Monasteries full of monks copying out illuminated manuscripts; court reporters dutifully tapping at typewriters; in the past, the history of transcription was a history of hard work. Like so much of our world, that all changed at the dawn of the information age. Transcription is how we turn one form of information into another. It's about taking the spoken word and transforming it into an encoded form of data (written script, binary code) to increase its lifespan. When the first written languages emerged, thousands of years in the past, this was their intention. Effective transcription has been a human fixation for millennia.

Writing made transcription possible; it didn't make it easy. Accurately reproducing a text could be the work of a lifetime, and receiving dictation of spoken words required highly specialized training just to keep up. The proper management of information requires a dedicated class of scribes and clerks. As the industrial revolution transformed manual labor, and as the first computers transformed computational methods, people began to wonder, how might computers transform transcription?

First steps: From Audrey to Markov

The story of automated transcription begins at Bell Labs. Founded in 1925 as the research division of the Bell Telephone Company (later to become AT&), by the 1950s, Bell Labs had a long history of discovery and the Nobel prizes to prove it. In 1952, the lab produced a device called AUDREY (short for Automatic Digit Recognizer) that could identify numbers spoken aloud. This device could only recognize the ten basic digits. Its accuracy approached 90%, but only with the voice of its original inventor. Competitors like IBM were producing their own versions. Researchers were grappling with the continuous talk problem; that is, how to get computers to recognize words without having the speaker having to pause between each.

This early phase of development, lasting from the ‘50s through to the mid-‘70s, saw two competing visions of speech recognition emerge. One equipped a computer with a pre-existing record of word sounds, in the form of waveforms. The computer then took these waveforms and compared them with the words spoken to it by the operator, to make an educated guess as to which word was meant. This, obviously, could only take the computer so far. With over one million words in the English language, providing a standard waveform for every single one would be prohibitively laborious; and that’s not accounting for homophones, contractions, and the ongoing evolution of language and accent.

The second strategy went back to first principles. By encoding certain linguistic ideas in a device, the computer could then decode a spoken sentence according to how likely certain combinations of words were. If a sentence was complete gibberish, for instance, it was unlikely to be what the speaker had intended. An understanding of grammar was essential to automated transcription. This basic principle of language by probability still underpins language transcription and generation technologies today.

It was in the mid-70s that computer scientists began to use Hidden Markov Models (HMMs) to guess the probabilities of certain words or sounds. HMMs allowed for faster computation, and overcame the continuous talk problem; transcription could now take place without the speaker needing to pause between words. This development combined with a similar technology, called Beam Search, to appraise each part of a sentence, suggest possible translations, and then discard the unlikely outcomes. Automated transcription technologies were able to recognize more and more words: from 1,000 in 1976 to 20,000 in 1984.

The Rise of Neural Networks

Speech recognition technologies were developed in concert with Natural Language Processing (NLP). The strategies used for encoding linguistic rules had initially been developed to aid in the processing of written speech, for uses such as translation and segmentation. Speech recognition researchers were often operating several years behind their NLP counterparts, as they faced the additional challenge of applying NLP innovations to the spoken word.

The 1980s saw the rise of the personal computer. The big commercial tech companies were interested in deploying speech recognition in their products. For this rollout to work, however, speech recognition needed to be user-independent. No longer did computer scientists have the luxury of training their systems on a single speaker, in laboratory conditions. Transcription systems had to be workable anywhere and with any user.

The fundamentals of the HMM and the Beam-Search model (referred to together as a 'Trigram model') would continue to be central to transcription technologies. With the invention of neural networks in the late '80s and early '90s, these probability models could be trained to even greater levels of accuracy. Neural networks allowed researchers to improve both the recognition of the individual words and the statistical processing of the overall sentence. This made it possible to reliably incorporate voice recognition into consumer devices.

Depending on the device, this did not require too much complexity. Devices like Amazon's Alexa only need to respond to a limited set of commands. In the home, this was all well and good. But in a commercial context, this simply wasn't enough. The most complex business use cases required a dramatic technological upgrade; the level of accuracy demanded by the market cost a huge amount of processing power.

The invention of Deep Learning Neural Networks, as a result of more powerful computers and new methods of processing, allows an even higher level of accuracy. By 2017, speech recognition systems had an average error rate of only 5%; matching human levels of performance.[7] The performance revolution made possible by Deep Neural Networks is still in progress; transcription technologies are only getting more and more accurate.

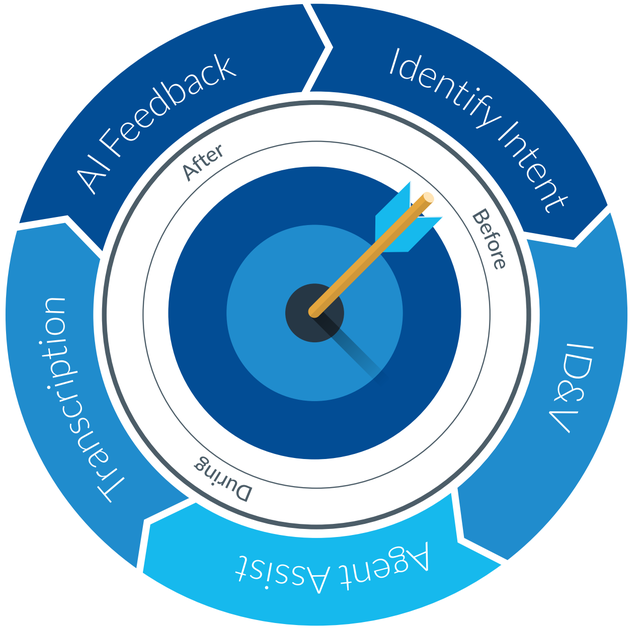
The Generative AI Revolution and the Future of Transcription

Though modern Large Language Models (LLMs) are light-years more advanced than the first trigram models, they operate on the same principles. They assign probabilities to words and phrases based on their understanding of the prompt, according to pre-existing variables. A modern LLM can have millions upon millions of variables, and the reasons behind their blistering performance aren't always clear even to their designers.[8]

Given these shared principles, it's no surprise that generative AI has proved adept at transcribing video and audio content, and creating summaries of that content for easy review. Unlike other Natural Language Processing technologies, however, since generative AI is not intrinsically limited by the audio it's meant to transcribe, it's more at risk of going off-piste; 'hallucinating' information that wasn't part of its original instructions, and may be untrue. For the most reliable transcription, stick to tried and tested Natural Language Processing. That's not to say, however, that generative AI has no role to play alongside speech recognition technologies.

Because no matter how accurate a transcription of speech is, if no one reads it, it's wasted effort. In fast-paced enterprise environments, efficiency determines the usefulness of a technology. This is especially true with a transcript of recorded speech; when written down, human speech comes looping, convoluted, and repetitive; in general, a challenge to understand. That's where generative AI becomes important; it forms a mutualistic relationship with specialized speech recognition technologies, ensuring accuracy and efficiency.

Generative AI can function as a summarization tool. Accurate transcripts of speech can be entered into an LLM to be processed and summarized into fluent, readable text. In this way, transcripts become useful in an enterprise context. The accuracy and reliability of modern AI transcription are paired with the flexibility of generative AI technologies to process data in ways that are both reliable and practical for human use. The potential use cases are manifold; particularly for business functions that hinge on spoken communication: that is, customer communications.



At every stage of the customer journey, Generative AI works to ensure that your CX is on the mark.

AI-Powered Customer Communications: The Story of Together Money

The customer communications wing of any public-facing business is one of the essential business functions. Without customers, businesses fail. If those customers can't reliably get in touch, they'll go to the competition. Creating an outstanding Customer Experience (CX) is essential to the continued success of any organization. CX poses a double challenge, however; many customer-facing organizations face higher levels of customer contact than they can effectively manage. That's not to mention the amount of administrative work that goes into every customer interaction; key customer details have to be drawn from voice interactions and inputted into systems of record; any data not recorded is lost, with serious quality (and potentially legal) implications. Answering every incoming contact at the highest level of quality, and ensuring accurate management of post-call data, poses a serious dilemma.

This was the problem faced by Together Money, a UK-based property finance organization, responsible for over £230 million in property development loans every month. Together Money faced a huge influx in incoming customer communication and found that post-call 'wrap' data-entry tasks were taking up far too much time. As innovative, AI early adopters, Together Money went looking for an AI-powered solution to their problem.

They found one in Content Guru's portfolio of AI solutions and in particular our transcription and summarization tool. This tool leverages market-leading AI transcription (from Google DialogFlow, Microsoft, and Jabra, amongst others) to accurately transcribe customer calls. These transcriptions are then submitted to an LLM for natural language analysis. This LLM (chosen by the customer) draws out key details according to a structured logic. The customer data, their problem, the proposed solution, and any post-call actions are identified and uploaded into systems of record automatically. All the human agent needs to do is review these entries for accuracy, and approve them.

The impact of this technology was transformative for Together Money. At first, AI transcription and summarization saved them 60 seconds of wrap-time per interaction. As Together Money got used to their new solution and improved their generative AI prompting, this number rose to 280 seconds saved per interaction. With all relevant data being recorded automatically, Together Money's agents were no longer at risk of having to repeat vital information, saving a further 17 seconds of interaction time. This success drove further AI investment; Together Money expanded their AI usage by a factor of nine, going from 20 user licenses in July 2024, to 175 in October.

The story of Together Money is a case study that provides valuable insight into the potential uses of AI speech recognition and generative AI summarization. The secret to Together Money's approach is in their AI partnership; by working with Content Guru, Together Money was able to deploy the perfect AI for their particular use case, drawing on different vendors flexibly to meet their needs. If history teaches us anything, it is that AI changes quickly. Your organization needs a partner who can help you navigate change.

Content Guru: Your CX AI Partner

Over three thousand years, methods for transcribing speech were limited to what the human ear could hear, and the human hand write. In the last fifty years, that has transformed completely. For the first time, we can automate away the most time-consuming of information management tasks. The twenty-first century has seen an information revolution; a revolution we're still living in the midst of. With the explosion of new, generative AI technologies, it's never been more crucial to establish firm technological foundations. More than new technologies, your organization needs an AI partner, to guide you through a period of radical change.

Through brain®, Content Guru handles the underlying complexity of AI technologies for its users. brain democratizes AI, making the best of the technology accessible, without tying you to an individual vendor. Regardless of size or sector, Content Guru makes AI adoption easy.

brain works as an AI orchestration layer, linking AI functionality seamlessly to the normal operation of your CX estate. Sitting alongside the cloud-native storm® solution, brain is constantly updated with the newest technology, as soon as it becomes available. brain gives your organization access to best-in-class AI capabilities such as Google Dialogflow, Azure, and IBM Watson, as well as generative AI systems like ChatGPT.

AI is revolutionizing the customer journey. To deliver first-class experiences for customers and agents, businesses must be integrating new technologies into their CX. Content Guru stands ready to make your AI transformation easy.

Ready to take the next step?

Get in touch, and take your first step into an AI-powered future. Provide us a few details, and a member of our expert team will be in touch within 24 hours.

Get in Touch

Want to learn more?

Discover our work with Together Money, and learn about how our AI solutions transformed their contact center.

Learn More

Endnotes

[1] Wikipedia

[2] Computer History Museum

[3] Ibid

[4] Ibid

[5] Content Guru

[6] US Census Bureau

[7] Computer History Museum

[8] Arxiv.org

CONTENT GURU

Engagement Made Easy®

contentguru.com

+[44] (0) 1344 852 350

JSB1024