

The Voice Report 2022 January 2022 Speechmatics.com



Contents

Executive Summary	3	
The History of Voice Technology		
Methodology and Demographics	5-6	
Content Statistics	7	
<u>Key Data Findings</u>	8	
<u>Headline Topic – Al Bias: Diversity,</u>		
Equity, & Inclusion	9	
An Introduction to AI Bias	10	
Where Does the Issue Lie?	11	
How We Perceive Al Bias	12	
Achieving Fairness in Al	13	
COVID-19: The Ongoing Pandemic	14	
Feature Development		
The Future of Voice Technology	15	
The Future of Features	16	
The Future of Languages	17	
The Future of On-Device Deployment	18	
The Future of Data Privacy and Security	19	

<u>Where the Future Lies: The Future</u>	
<u>is Self-Supervised</u>	20
Accuracy Remains Everything	21
The Ubiquity of Voicebots in 2022	22
The Silicon Shortage: A Focus on Software	23
Summary	24

"When you experience great voice apps, it makes tapping on an app so circa 2005. And so, I think all those experiences are in the process of being reinvented and you'll see a lot more of that moving forward."

Andy Jassy CEO of Amazon

Speechmatics The Voice Report 2022 2

Executive Summary

2021 was, in many ways, an even stranger year than the previous. As soon as we began to see something resembling normality returning, it disappeared again. But there were constants. In the Speechmatics Voice Report for 2022, we'll be concentrating on these consistent factors and delving deeper into the world that's evolved out of the COVID-19 pandemic.





One thing that is certain is that the deep-tech ecosystem continues to flourish. In the UK, investment in the deep-tech sector has grown significantly over the past five years, rising 291% to £2.3billion according to the <u>Annual Small Business Equity Tracker</u> from the British Business Bank.

Equally, across Europe, in Atomico's <u>State of European</u> <u>Tech 2021 Report</u>, total investments for 2021 will hit more than \$120billion. That's nearly three times the amount raised last year. Record-breaking levels of investment are flooding into the market to enable the rapid growth of incredible technology developments.

In this Voice Report we'll explore this continued boom for the industry and take an extensive dive into the development, expectations and trends we see going forward, with particular focus on what we see – and what our research shows – as being the future for voice technology. This Report contains key insights from industry experts, product specialists and machine learning engineers at the bleeding-edge of these technologies. They reveal their opinions and expectations of voice technology, the markets they believe this technology will influence, and the capabilities it will enable.

We'll also be taking the opportunity to look at Diversity, Equity, & Inclusion when it comes to voice technology. The much-discussed topic of Al Bias is something that's relevant to everyone in tech, and so we'll be asking how organizations perceive this problem and how they can tackle it head on.

Finally, we'll gather insight on the introduction of self-supervised learning in relation to removing human intervention in training data. And we'll be asking how aware people are of it as a step-change in voice technology.

2022 may bring more uncertainty. This Voice Report will guide you through it.



The History of Voice Technology

To celebrate just how far the voice industry has journeyed in the last 70 years, here's a brief look back on seven decades of progress – from the introduction of the first speech recognition to the groundbreaking technology of today.

1950s

Seventy years ago, in 1952, Bell Laboratories invented 'Audrey', the world's first speech recognition engine. It could understand single speakers and single digits even with limited memory and computational speed.

1960s

Ten years later at the 1962 World's Fair, IBM displayed the 'Shoebox'. Using a series of electrical impulses – and a measuring circuit to classify these impulses – it had the ability to recognize 16 words in English. 'Harpy' (funded by the US Department of Defense and worked on by IBM, Carnegie Mellon University and Stanford Research) was able to recognize 1,011 words, including entire sentences. It became the first system to successfully use a language model to determine sequences of words.

IBM developed a voice-activated typewriter named 'Tangora'. It had a vocabulary of over 20,000 words and used an IBM PC AT to recognize spoken words and put them to paper.

5

Computer processors and networking become faster and more prevalent leading to Dragon (Nuance) releasing Dragon Dictate, the first commercially available speech recognition product. By 1997, Dragon NaturallySpeaking could understand up to 100 words per minute. Voice recognition tools are added to smartphones by Google, including in 2009, the introduction of Mandarin Chinese. This leap forward uses data centers to provide the enormous amounts of computing power to match user voice queries with actual human speech.

By 2017, advancements in AI and deep learning models steadily progressed the recognition of conversational speech, helping Microsoft and others surpass humans in transcription accuracy tests.

Companies introduce self-supervised learning into their training, giving speech recognition engines the ability to learn from unstructured data across the web. This leads to even greater improvements in accuracy and a step toward Autonomous Speech Recognition.

Methodology and Demographics

For the 2022 Voice Report, Speechmatics reached out to a wide-range of industry experts to find out their views on both the current and future state of the voice industry. We wanted to hear from both Business Owners and Entry-Level workers, as well as everyone in between.

As a Global company we also wanted to include voices from all corners of the world and so this Report includes feedback from Europe, North America, Africa, Australasia/Oceania and Asia. Job titles of respondents have been self-described as ranging from CEOs to Software Developers. Industries include a wide scope of responses from sectors as diverse as Education, Legal, and Telecommunications.

We've collected data from small companies of under 50 employees, all the way up to companies who employ over 5000. They include a variety of B2B (business-to-business), B2C (business-to-consumer), G2C (government-to-consumer), B2G (business-to-government), and a mix of B2B and B2C.

A full breakdown of those who responded is opposite:





Breakdown of respondents by location



Breakdown of respondents by industry

31.2%	Technology
22.2%	Media and Entertainment
8.9%	Transcription Solutions
8.9%	Consultancies
6.7 %	Telecoms
4.4%	Healthcare and Life Sciences
4.4%	System Integrators
2.2%	Education
2.2%	Legal
2.2%	Banking, Financial Services, Insurance
6.7%	Other

Methodology and Demographics



Breakdown of respondents

Business-to-Business (B2B)

Business-to-Consumer (B2C)

Business-to-Government (B2G)

Government-to-Consumer (G2C)

Mixture B2B & B2C

by nature of business

66.7%

20.0%

8.9%

2.2%

2.2%



Breakdown of respondents

1-49

50 - 249

250 - 999

1,000+

by size of company

46.8%

33.4%

8.8%

11.0%



Breakdown of respondents by role

18.3%	Manager
13.7%	Engineer
11.0%	CEO
4.5%	Customer Manager
4.5%	COO
4.5%	President
2.3%	Head of Operations
2.3%	Marketing and Comms
_	

PHD

2.3% VP Business Development

34.3% Other

2.3%



Breakdown of respondents by seniority

- 51.1%
 Owner / Executive / C-Level

 20.0%
 Middle Management
- 15.6% Senior Management
- 11.1% Intermediate
- 2.2% Entry Level



Speechmatics The Voice Report 2022 6

Content Statistics

Before we delve deeper into the findings of our research and then take a look ahead to 2022 and beyond, here are some headline statistics from 2021 for the Voice Industry. These stats have been selected from, amongst others: Adobe, Forbes, Gartner, and PWC.



"In search today...when we're trying to understand what's in a video, the fact we could understand language deeper, would make a big difference. That's the primary purpose (of LaMDA). In conversing with an assistant, we can make conversations much more natural."

Sundar Pichai CEO of Google and Alphabet



44% of all respondents use a smart home voice assistant.

An estimated **8.4 billion** devices will use voice assistants by 2024.

– <u>Juniper</u>

45.2 million US adults used voice search to shop for a product at least once. - <u>Voicebot.Al Research 2021</u>

40% of technology users resist applications that have provided a poor experience. - <u>Gartner</u>

127 million Americans, 27 million Germans, and 19 million Brits use voice assistants in their cars.

- Voicebot.Al Research 2021

31.4% of US internet users used their smart speakers for answering general knowledge questions.

- Emarketer

The chatbot market value is projected to grow by **\$1.73 billion** during 2021-2025. – <u>Technavio</u>

92% of senior marketing executives say privacy is a fundamental part of the customer experience.

- Adobe Digital Trends 2021

Speech-to-text API market is set to grow at a Compound Annual Growth Rate of **19.2%** between 2021 and 2026.

- Chhattisgarhjournal

Key Data Findings

1 The consensus on the future of speech-to-text is **incremental advances** until we reach perfect transcripts.

2 Increased speaker diarization accuracy is the **most demanded** feature going into the next three years.

- 3 Cantonese is the **most often requested** "new" language, with Brazilian-Portuguese and Swiss-German in second place.
- 4 **Almost three-quarters** of those surveyed said they'd like to see on-device deployment of speech-to-text.
- 5 Data privacy and security is seen as a very high priority for **nearly 80%** of those asked.
- 6 Banking and Healthcare were named the two industries **most likely** to increase their need for voice technology in the continued COVID-19 pandemic.
 - **The majority** of respondents saw a shift to Autonomous Speech Recognition as the future of voice.

- 8 **Over 50%** of those asked knew what self-supervised learning was. Those that didn't, requested more information about it.
- 9 Speaker change, speaker diarization, and punctuation topped the list of the **most welcome** extra features.
- **Over half** of those interviewed have already successfully integrated voice technology into their business.
- **64%** of professionals said the COVID-19 pandemic hasn't affected their strategy for integrating voice technology.
- **12** Accent and dialect were seen as the **two factors** that prevent voice technology gaining full accuracy.
- **55%** of those asked believe the majority of voices are understood by Automatic Speech Recognition.
- 14 It's believed ASR is **more accurate** than it is inaccurate, but context was seen as a major factor.
- **15** Speechmatics was seen as the **most accurate** ASR provider. Google and Amazon came second and third.
- 16 English was perceived to be by far the language **best served** by speech-to-text.
- There was an **even split** in whether one gender was served better by ASR, with approximately one third unsure.
- **18 Nearly a third** of those asked have experienced AI Bias first-hand.

Headline Topic - Al Bias: Diversity, Equity, & Inclusion

In this section of the Voice Report 2022, we'll be taking the opportunity to look at Diversity, Equity, & Inclusion when it comes to voice technology. The much-discussed topic of Al Bias is something that's relevant to everyone in tech, and so we'll be asking how organizations can tackle this head on.

"Leveraged properly, voice-activated technology offers the significant ability to bypass disability or learning difficulties and create equality in an ever-increasing digital workforce and economy."

Mark Brown British Standards Institution (BSI) For this Voice Report we wanted to see how this topic was perceived across the industry. The results were simultaneously expected and surprising. We'll begin with the headline figure that 28.9% of those asked had experienced AI Bias first-hand when using voice recognition.



An Introduction to AI Bias

When negative headlines about artificial intelligence are shared around, it paints an unfavorable view of a technology that helps billions every day. One of the most familiar topics written about is Al Bias. In science and engineering, meanwhile, a bias is a systematic error. Statistical bias usually results from an unfair sampling of a population, or from an estimation process that does not give accurate results on average. As for voice in particular, we can see the perception of Al Bias in the following answer to the question, "Are most voices understood by speech recognition?"

While it's great for our industry that most people believe the majority of voices are understood by ASR, there's far too much opinion on the other side to feel comfortable that the Voice Industry is a place where AI Bias is not a problem. But before we can address it, we need to breakdown AI Bias further. There are many different types of Al Bias with five surfacing more often than others. These five are Algorithm Bias (issues with the instructions given to the program), Sample Bias (issues with the data itself - sometimes the dataset is too small or under representative), Prejudice Bias (an issue where real-world stereotypes are pulled into the system), Measurement Bias (issues with the accuracy of the data) and finally, Exclusion Bias (issues where integral factors are left out of the datasets).

These types of Al Bias are often interlinked. We can be guilty of exclusion bias because we have prejudices. Sample data can easily affect measurement bias. When it comes to speech recognition, just one of these biases can throw an algorithm off, leading to inaccurate results.

It's up to everyone to do their bit to address these imbalances.

Even though the positives vastly outweigh the negatives, it doesn't mean we shouldn't pay attention to where AI is falling short of Equity, Diversity, and Inclusion. AI Bias can be a problem and it's up to all of us to tackle it.

But what exactly do we mean by <u>"bias"</u>? Put simply, it's disproportionate weight in favor of or against an idea or thing, usually in a way that's closedminded, prejudicial, or unfair. Biases can be innate or learned and they can lead to some developing strong feelings for or against an individual, a group, or a belief.



Are most voices understood by voice recognition?

55.6% The majority of voices are understood
37.8% Too many voices are not understood
6.7% Hardly any voices are understood

Where Does the Issue Lie?

The first bit of Al Bias related feedback we'll look at from the Voice Report concerns where the respondents have found specific issues with ASR. Dialect and accent – which we all have in one way or another – is far away the standout issue that needs to be addressed.

> It's true that the more we're exposed to different ways of thinking and speaking, the more likely we are to understand them. And this is the same for machine learning. If we give the training models exposure to a different variety of voices, it should become familiar with them. While it isn't a cure-all fix, exposure is critical for reducing Al Bias.

> While data is not the only way to address Al Bias, it is a significant factor. Which is why self-supervised learning (already proving a success with the amount of data it can train on) is such a big factor to improving accuracy in ASR.

To see how familiar those within the industry were with the concept of self-supervised learning, we posed it as one of our questions in the Voice Report 2022. Just under half hadn't even heard of self-supervised learning, showing there was a large amount of education still to do in that area.



Which areas still need improving for ASR?

30.4%	Accent
21.2%	Dialect
14.1%	Ethnicity
11.1%	Location
7.1%	Age
5.1%	Gender
4.0%	Background Noise
4.0%	No improvement needed
3.0%	Other (including mix of languages, speaker identification and people with disabilities)



Are you familiar with self-supervised learning?

55.1%	Yes
44.9%	No

How We Perceive Al Bias

"I'm a white, educated American male with no accent and a somewhat nasal voice. Speech engines love me. But all speech engines will steer ambiguities towards a most-likely outcome. I get it; this can be interpreted as sexism and racism, but it simply reflects the society we live in, as does all datadependent AI machine learning."

Voice Report 2022 Response



The above quote was taken from our survey when asked for further detail about when someone has experienced AI Bias. It's an honest and revealing reply and one that speaks volumes about the complexities of this topic. But this kind of openness is necessary if we want to be able to discuss and dissect what AI Bias is. And how we can tackle it.



When we asked which language was best served by ASR the results were less than surprising. Almost 90% of those surveyed said English was their top answer.

"I've found speech recognition pretty good when talking to doctors and professors, who may speak in whole sentences. Working on consumer voice recognition has been much harder."

Voice Report 2022 Response



The quote opposite – also taken from the Voice Report 2022 – shows how crucial it is that voice technology is up to the task. Taken from a medical professional, we can see from their experience that, yes, educated voices are being heard, but quite often (and increasing with the pandemic) more and more voices will need to be understood. When it comes to something as important as health, inaccuracy can have fatal consequences.

Achieving Fairness in Al

"What does "fairness in AI" look like? Should a company of sufficient market size have to submit its algorithms for analysis to make sure it is not intentionally and unintentionally excluding disadvantaged groups? Are we prepared for the gremlins we may find? Companies working with voice should be asking themselves tough questions to ensure that our definition of 'accuracy' is as inclusive as possible."

Michael Tansini Product Manager at Speechmatics

Contents

As artificial intelligence continues to become an ever more present part of our daily lives, it's up to corporate organizations to be as proactive as they can be to ensure fairness. As products become increasingly productive and perform even better than expected, we must make smarter decisions. And those decisions have to start earlier in the process.

In a recent essay by the <u>World Economic Forum (WEF)</u>, Agbolade Omowole, CEO for Mascot IT Nigeria looked at strategies for mitigating fairness and non-discrimination risks and arrived at three conclusions. The first was around inclusive design and foreseeability. He suggests we look at race, gender, class, and culture at a design stage. Second, when it comes to user testing, groups should be as diverse as possible. His third recommendation was to perform STEEPV analysis to detect fairness and non-discrimination risks in practice.

One final suggestion from the WEF article was to have Al ethicists in development teams. While this may seem a step beyond for some, in many ways it can be the right thing to do both ethically and financially – with many companies saying making the right decisions early on has saved them money and time. If we plan ahead and ask questions about Al Bias – especially when it comes to accuracy and making sure all voices are heard – we'll start to see positive results across the board.

COVID-19: The Ongoing Pandemic

The COVID-19 pandemic has continued to have a major impact on the technology industry, with voice technology one of the many industries that have seen their usage surge as consumers and businesses find a way to adapt to the new world of lockdowns and living even more online. As we continue to find ways to live in a society defined by the pandemic, we asked which sectors people thought would see their use and application of voice technology significantly increased.

Nearly half of those asked believed Banking and Healthcare were the two industries most likely to increase their need for voice technology in the aftermath of COVID-19, with each achieving 13.9% of the vote. Consumer industries came out ahead of both Media and Entertainment, and Government. In last year's survey, Telecommunication was top. This year, it has fallen below a range of industries.



"Voice-activated technology is essential for nurses, doctors and other medical team members to do their jobs effectively. These frontline healthcare workers need their hands free to care for patients quickly and safely."

Brent Lang Vocera Communications A notable piece of data that came out of the Voice Report 2022 was seeing how business strategies for implementing a voice strategy was affected by COVID-19. When we asked this same question in 2021, 53% believed COVID-19 would have an effect. While this isn't conclusive, it would suggest that as a society we're beginning to find ways around the obstacles set by COVID-19. Or at the very least, we are learning to live with it and make changes that we know our integral to our business growth, rather than waiting for 'normality' to return.



Has your strategy for integrating voice technology been affected by the COVID-19 pandemic?

63.6% No 36.4% Yes

The Future of Voice Technology

One of the main questions we ask every year in the Voice Report, is where people see the future of voice technology going. The consensus as we enter 2022 is that we'll see incremental technological advances until we reach perfect transcripts.

This isn't a surprising response, with accuracy being the primary factor in this answer for the last few years. Looking back to the perceptions of accuracy, low word error rates are regarded to be the main definition of accuracy. As machine learning algorithms continue to evolve, it is likely that WER accuracy will reach 95%+ especially for commonly used languages like English.

Get free Developer access to the most powerful, inclusive, and accurate speech recognition ever released. Get your API token and <u>start your trial today</u>.

What does the future of speech recognition look like?





However, there is still significant work to do to achieve this due to the range of accents and dialects present in all languages and to deliver the same level of accuracy across them all. ASR providers and users need to be objective around how they conduct testing to ascertain these WER scores and understand what this means for them and their customers. In some cases, WER might also fall to the wayside when compared with other KPIs such as speed.

The second most selected response was that voice technology will be able to deal with robust and noisy environments with greater skill. It's no surprise that the ability to deal with noisy environment is a key factor in the future of speech recognition. Noise is a major factor that can impact accuracy. It was also highlighted as a risk that we could experience a downfall in audio quality due to the pandemic requiring the population to wear masks and other protective wear. Being unable to detect the words of a speaker due to noise has a direct impact on the outcome of the transcript, so the ability to reduce interference or deliver high-quality recognition even in challenging environments remains a top priority.

The Future of Features

"Given the current rate of development in the field of ASR, it's highly probable that machine-driven speech-to-text will surpass human transcription in accuracy. It already has surpassed it in speed. Especially in the field of media and entertainment. Hearing impaired subtitles will be fully generated by machine with punctuation, line breaks, profane word detection and marking. I also think there will be huge improvements in accuracy in the other NLP tasks like text summarization."

Voice Report 2022 Response

Increased speaker diarization accuracy is top of the list for hopes going into the next three years. Again, this could be attributed to the pandemic and the major ascendency of conference calling. To be able to detect who is speaking - and easily jump between the speakers – would have a huge benefit.

Speaker diarization is one of the most challenging elements of speech recognition. While speech and other audio characteristics are easy for the human brain to detect, this poses a challenge for automated systems due to the fluctuations in a single speaker's voice depending on their mood, hesitations, word emphasis, noise etc.

While speaker diarization exists today, it is still a key challenge that speech providers have not yet mastered. 2022 will likely see increased effort to improve speaker diarization to uplift use cases that benefit from being able to match a speaker with the words spoken.

Language identification was another key element respondents predicted would see an improvement in the coming year. Detecting the language of the speakers within a video or audio file automates the manual task of selecting the correct language pack to use to transcribe it.

By automating the language identification element of the transcription process, businesses can save time and human resource cost as well as unlock new information that would previously have been lost.

The full details of responses are opposite.



Which features would you like to see?

13.5%	Increased speaker diarization accuracy
10.7%	Non-speech detection
10.2%	Language model adaptation
9.3%	Language identification
7.9%	Translation
7.9%	Customer-specific language models
7.0%	Short-utterance accuracy
6.0%	Real-time transcription from cloud
5.6%	More languages
5.1%	More accurate number recognition
4.7%	Noise reduction
4.7%	Word alternatives available in output
2.8%	Audio file quality assessment
2.3%	Redaction
2.3%	Unsure

The Future of Languages

"For English, pinpoint accuracy is an expectation. We will see that expectation grow for non-English languages, not only in European languages but in East and South Asian languages too."

Michael Tansini Product Manager at Speechmatics

With so many of the "major" languages already very well-served by current speech-to-text, we wanted to see which languages those we surveyed thought were either underrepresented or completely lacking from most modern ASR engines.

Cantonese is the most often requested "new" language, with Brazilian-Portuguese and Swiss-German in second place. There was also a strong showing for Hebrew and Bahasa.

With so many languages having such a diverse range of accents within them, ASR will need to incorporate systems that are trained to be agnostic when it comes to deciphering and transcribing. That way those languages can better serve users and increase accuracy in the longer term.

13.3% 12.0% 12.0% 9.3% 5.3% 4.0% 2.7% 2.7% 1.3%

"The ability to distinguish between different users will help voice-activated tech bring more personalized content to each user, which will make it an effective marketing tool. Naturally, this will also apply to content children have access to, so certain restrictions will be necessary."

Sanjay Malhotra Clearbridge Mobile



Which language would you like to see in the next 3 years?

Cantonese Brazilian-Portuguese Swiss-German Hebrew Bahasa Tagalog Estonian Bahasa Malav Tamil French Canadian 1.3% 1.3% Irish (Gaelic) 1.3% African Languages 1.3% Icelandic 32.3% Unsure



The Future of On-Device Deployment

Almost three-quarters of those surveyed said they'd like to see on-device deployment of speech-to-text. On-premises ASR enables the transcription of latency or security-sensitive media in your own environment or within public cloud environments.



It enables users to keep their data secure within their own environments with no need for data to go into the cloud. On-premises deployments for voice technology are often done using virtual appliances or containers so they can be deployed effortlessly into existing technology stacks.

It can be hugely beneficial for industries such as Banking, Financial Services, and Compliance and regulatory challenges where customer data and voice data cannot leave the premises. In these cases, on-premises deployments are the best solution to preventing data breaches and avoiding risks associated with private cloud deployments.



Would you like to see on-device deployment of ASR?







The Future of Data Privacy and Security

Data privacy and security is seen as a very high priority for over three quarters of those asked. When it comes to speech-to-text, businesses, consumers and governments need to make sure that what has been said – and the data that's mined from what's been said – remains as safe as can be.

As sectors like Health and Finance become increasingly dependent on voice technology, it's no real surprise that data privacy and security is top of mind for most users. The idea that "someone is listening" has never been truer. People want to make security programmatic and cultural. They require comprehensive security that's continuously worked on and improved upon.

Robust and scalable security architecture, solutions, and operations will help people feel safe as we head into the future.

Is data privacy and security a high priority for you in 2022?

77.8% Yes

Where the Future Lies The Future is Self-Supervised

"Self-supervised pre-training is being utilized in many different forms to get rich representations for use in downstream tasks like ASR. Wav2Vec2.0 is an example of contractive predictive coding that leads to very impressive results."

John Hughes Head of Accuracy at Speechmatics



One of the biggest leaps forward for voice technology in 2021, was the introduction of self-supervised learning into training models. In the race to make data as accessible as possible, being able to step away from labeled data is key.

Producing labeled data has always been one of the most time-consuming elements of speech-to-text. For every audio file with an annotated transcript, there has, historically always been a human listening and writing down each word. This takes time, and time is money.

With self-supervised learning training on unlabeled data, the need to constantly label data is removed. <u>Speechmatics</u> have seen the audio data they're training on leap from 30,000 hours to a staggering 1.1 million hours. That figure will do nothing but rise.

As we look to the future, we can clearly see where there's more data to train on, there's a much greater chance of meaningful change across the industry – and more and more voices and languages can be heard. "Accuracy will continue to be the king measure for some time, but it's going to get harder for customers to evaluate the differences between providers. The core "easy" audio for English might all look similar, but accuracy comparisons on other accents/ languages/domains will start to differentiate the providers."

Stuart Wood Product Manager at Speechmatics

Accuracy Remains Everything



One of the biggest changes in voice in 2021 – linked strongly to our headline topic of Diversity, Equity, & Inclusion – was the breakthrough in accuracy for languages other than American English.

As voice technology continues to excel (and challenges such as surpassing human transcription are achieved) developers and engineers have started to cast their nets wider. By doing so, they have started to look for parity across accents and languages. This has led to huge reductions in the accuracy gap among differing demographics.

> Use cases for voice recognition have also continued to rise at a prolific rate. A large part of this can be linked to accuracy. As accuracy levels continue to reach the top end – and with the introduction of self-supervised learning already making huge movements in ASR – voice will become more and more integral to our daily lives.

> When it comes to use cases like dictating medical reports and training air traffic controllers, there's a margin of error that's slim to none. But when the best machine learning models are consistently outperforming their human counterparts in terms of accuracy, why would organizations choose anything other than the best technology?

The Ubiquity of Voicebots in 2022

As we enter 2022, many believe it will be the year of the "bot". It was only a little while back that **Gartner** predicted that, by this year, 70% of all consumer/business dialogue would be through artificial intelligent led chatbots. It was a good prediction, with the pandemic leading to more interactions online, and often without human interaction on one end. But are we really at the point where bots can mimic humans to the point we can hand over decision making controls in areas like Finance and Health?

Contents

Intelligent virtual assistants, virtual customer assistants, digital assistants; chatbots have a number of names and many more uses. According to <u>TechNavio</u>, the chatbot market value is projected to grow by \$1.73 billion at a CAGR of 29% during 2021-2025. And while COVID-19 has had an impact on the implementation for some, for many more it's been seen as a necessity in a post-pandemic world.

For a long period of time chatbots functioned purely as a Frequently Asked Questions stopgap. As technology has advanced and Al improved, their capacity to offer something other than limited scripted answers has meant they've become essential to almost every industry available.

In terms of progress of usage, we need look no further than voice shopping to see the change. In 2018, researchers found that 20.5 million US shoppers had used voice search at least once to find items they wanted to buy. According to data from <u>Voicebot.Al</u> that number rose by a massive 120% in 2021 to 45.2 million. It isn't just business-to-consumer relationships that will be helped by this technology, with businesses using chatbots and voicebots to help their employees get quicker results and deliver the right answer, in everything from HR to job applications.

Forbes recently argued that bots are the single biggest application of Al in any industry. The same publication also contested that over 50% of searches will utilize voice, meaning voicebots – just as much as chatbots – will be the tech breakthrough of 2022. With voice being by far the quickest form of communication, it makes absolute sense that voicebots will play an integral role in how we work and play.

The Silicon Shortage: A Focus on Software

There are a multitude of factors that have led to the global silicon shortage. While the COVID-19 pandemic was a huge factor, the history of the crisis goes back further. The trade dispute between China and the US has played its part, as has poor planning and even natural disasters in the form of typhoons in Taiwan. All-in-all it's been a perfect storm of problems.

Contents

While silicon metal's core components are plentiful within the Earth's crust, getting them out requires power. The kind of power traditionally generated from coal-burning electricity. As one of the main producers of silicon chips, China has had to make big decisions over whether it wants to continue using the same levels of energy to mine them – or reduce their carbon emissions.

In April 2021, automotive was the industry making the most headlines – with the auto industry pretty much stuck by the roadside due to the chip shortage – but there's been a knock-on effect for everyone that needs CPUs and GPUs. With more people working from home, more laptops have been needed. Without the silicon to supply the demand, industries the world over have been affected. And that includes the speech recognition industry.

Establishing meaningful change in machine learning and voice technology has always relied on three areas: improved algorithms, the amount of data available to the system and lastly, increased computing power.

While we've seen some huge breakthroughs in algorithms and – with self-supervised learning – some major advances in data collection, experts believe the stalling of computing power through lack of silicon could mean the majority of progress in 2022 will again have to come from software.

How long will the silicon shortage last? No-one really knows. The climate crisis continues to make the future appear uncertain and COVID-19 – in its many forms, from Delta to Omicron – shows no signs of relenting. Then there's the ongoing trade disputes between China and the US.

Perhaps then, the focus has to be on the software engineers, and the smarts they have in abundance, to really push us to the next level of development.



"ASR and NLP technologies are a valuable component for automating critical business workflows. If deployed wisely in human-in-the-loop applications, the continued evolution of these technologies should enable increasing efficiencies and improved business outcomes in any environment where human language processing is required."

Voice Report 2022 Response

The future looks increasingly bright for the global speech-to-text API market. According to **Research & Markets'** annual review, the industry is projected to grow from \$2.2 billion in 2021 to \$5.4 billion by 2026, at a Compound Annual Growth Rate (CAGR) of 19.2% during the forecast period.

This growth will be driven by a number of different factors, many tied to the continued global pandemic. It is believed by many in the industry that incremental improvements within speech-to-text will soon lead to near-perfect transcription. An increasing demand for Al-powered customer services and chatbots have all played a factor in growth too.

As we've seen within this Report, with this progress comes a greater sense of responsibility. Al Bias will continue to play an integral role in all our lives for years to come and the onus is on all those connected with the Voice Industry to play their part.

Every major technology publication and research company – from Forbes to <u>Gartner</u> – see voicebots as playing an increasingly integral role in our daily work and home lives in 2022.

In summation, in an unsteady global environment, voice technology continues to provide one of the safest bets for both those working in the industry and those investing in it.



Learn more, sign up to our newsletter.

24