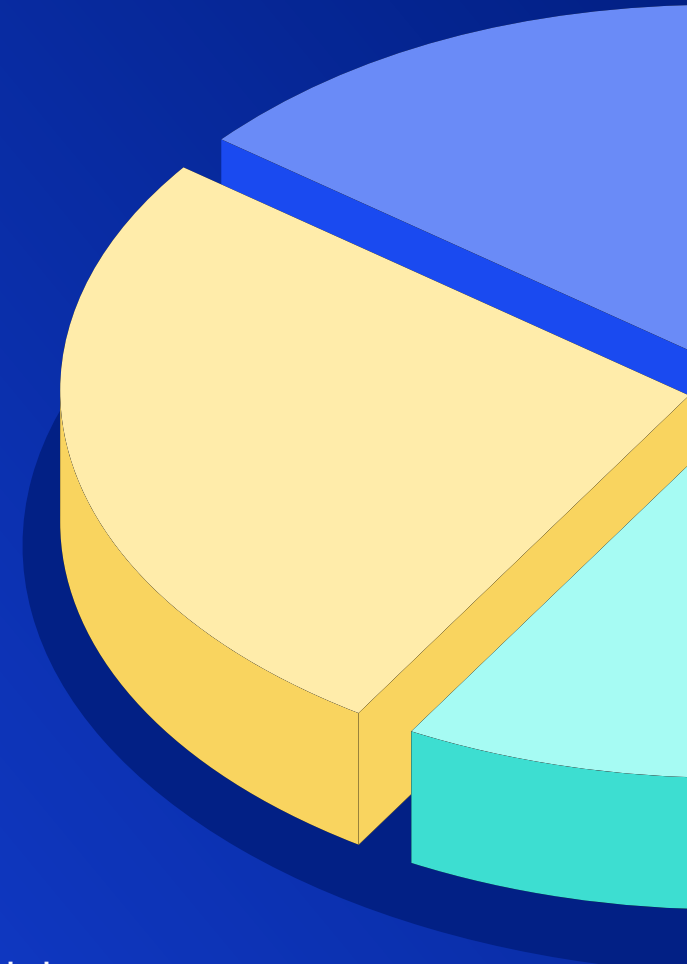


Navigating Automatic Speech Recognition

A survey-based overview
of ASR & captioning trends



2024 Survey Report

Conducted by independent research firm

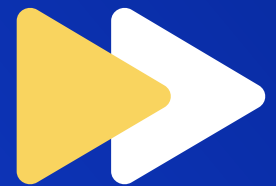


Global
SURVEYZ

Table of contents

▶ Introduction and key findings	3
▶ All report findings	8
What percentage of audio and video content is being captioned?	9
What are the most common use cases of ASR and captioning tools?	10
What types of tools and methods are being used to caption and/or transcribe?	11
How satisfied are users with the accuracy of their captioning and transcription tools?	12
How are free, built-in ASR tools used and perceived?	13
Which features are most important to ASR users?	14
What challenges are preventing greater usage?	15
What requires the most effort when using ASR, captioning and transcription tools?	16
Which improvements are likely to drive more usage?	17
What are the expectations for usage and investment in the coming year?	18
How many companies are exploring the use of Generative AI?	19
▶ Demographics	20
▶ About Verbit	23

Introduction and key findings



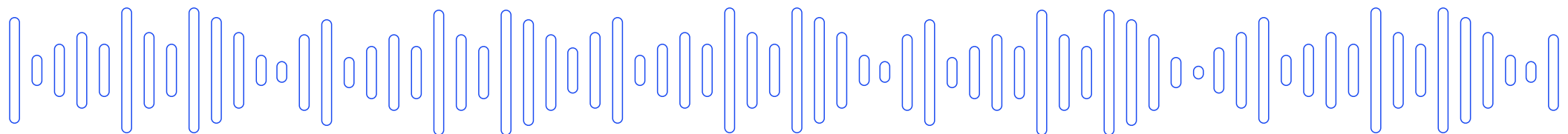
Introduction

Company leaders across industries are grasping the importance and benefits of captioning and transcribing their content and experiences. Advanced technologies, including Automatic Speech Recognition, are a key contributing factor for this growth in adoption. ASR technologies offer professionals simple, scalable and cost-efficient ways to caption and transcribe. As a result, the use cases are also expanding to a variety of business areas.

Professionals are exploring more ways to incorporate ASR technologies into their work. Many of the ASR tools currently on the market are free and built-in to other software they're already using. These tools are therefore among the most widely used. However, the accuracy of their output still varies and offers some risks. There is also little opportunity to provide feedback for improvement. This is a leading pain point for those who use ASR, especially for businesses and institutions that require high standards of accuracy.

Professionals who remain wary often opt for having these tasks performed solely by humans or through hybrid means. These more humanized, paid solutions do help them to get the accuracy they need, but at a price. As this report reveals, although accuracy remains a leading concern, there is still an increasing interest in using and investing in more advanced ASR tools - and in complementary Generative AI technology - to stay competitive.

This survey was designed to explore current attitudes on ASR technologies and their impact on market usage. It aimed to uncover opportunities for improvement and barriers that prevent businesses from increasing their ASR usage.



Methodology



To receive insight into current trends, perceptions on ASR and captioning tools, Verbit commissioned a survey of 200 business leaders across industries. This report was administered by Global Surveyz Research, an independent global research firm.



The survey is based on responses from senior executives (Director+) responsible for either DEIA efforts, captioning and ASR technologies or both. Respondents work in either entertainment and media, banking and finance, tech and telco, healthcare or pharma, business, advertising and marketing, automotive, retail, energy, food and beverage or hospitality.

Respondents hailed from companies in the US (33%), CA (33%) and UK (34%), ranging in size from 500-1K employees (50%) and 1K-5K employees (50%). The respondents were recruited through a global B2B research panel and invited via email to complete the survey, with all responses collected during February 2024. The answers to most of the non-numerical questions were randomized to prevent order bias.

Key findings

1 Respondents revealed a great need for improved accuracy in their ASR tools


47% of companies are dissatisfied with the accuracy of their current ASR technologies (Figure 4). This reality, coupled with the finding that 70% of companies are using free, built-in tools (Figure 3), suggests that a significant proportion of ASR technologies aren't cutting it. This finding represents an opportunity and a wakeup call for those not leveraging advanced ASR tools which offer ways to boost accuracy performance. ASR inaccuracies can harm experiences in a variety of use cases (Figure 1). However, inaccuracies may also present a legal challenge for those using it for accessibility accommodations where up to 99% accuracy levels are often required.

3 67% of businesses are captioning at least half of their audio and video content

55% of respondents indicated that they caption 50-75% of their content, while 12% caption 75-99% of their content (Figure 1). These figures indicate that businesses which are not yet leveraging ASR and captioning technology (or not using it enough) should prioritize it to remain competitive. As audiences and employees are exposed to captioning at greater frequency, they increasingly have come to expect it to be offered.

2 The use of ASR to improve business outcomes now exceeds accessibility use cases

Captioning was originally offered to accommodate people with disabilities. However, most companies are now using ASR and other captioning tools for a wide range of business applications, demonstrating the versatility and usefulness of the technology (Figure 2). The most popular uses include reaching greater audiences (52%), supporting internal company needs (50%) and creating transcripts (48%). Professionals are therefore looking at ASR beyond accessibility accommodations. It's proving to be a competitive asset in expanding reach and growth.



4 Companies are seeking more customization beyond what current ASR tools offer

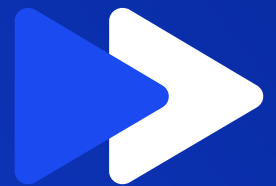
Current ASR, captioning and transcription tools still take too much manual effort to achieve satisfactory results. They're not built to be intuitive enough, given that the greatest effort in using them goes into understanding how to use them effectively (27%) (Figure 9). Others noted significant pain points in editing the output so that it's more accurate (21%), adding specifics to the output (21%) and integrating the tools with existing systems (18%). Most companies are using free, built-in tools for captioning and transcription (Figure 3), which allow for little to no customization. This reality indicates a greater need for ASR solutions which allow for more personalization. More customization features will lessen the manual efforts required to achieve the desired output.



5 49% of companies expect to increase their investment in ASR and captioning technologies

While 50% expect to at least maintain their current level of usage, 49% expect to invest more in ASR, captioning and transcription tools in the coming year (Figure 11). The pace of adoption of ASR and captioning tools is not slowing down, suggesting that those who do not strive to use and invest in effective technologies may find themselves at a disadvantage over those who do.

All report findings



More than half of all audio and video content is being captioned

The majority of audio and video content is currently being captioned by today's businesses. 55% of respondents indicate they caption at least half (50-75%) of their content, and 12% caption most (75-99%) of their content.

This figure highlights the existing expectations for business content and experiences to be offered with captions. The research also showcases how few businesses are not using captioning.

Those who are not taking advantage of captioning and ASR technologies can look to do so not only to meet competitive benchmarks, but reap the business benefits in boosting engagement, increasing reach, improving accessibility and repurposing content through transcripts (as seen in Figure 1).

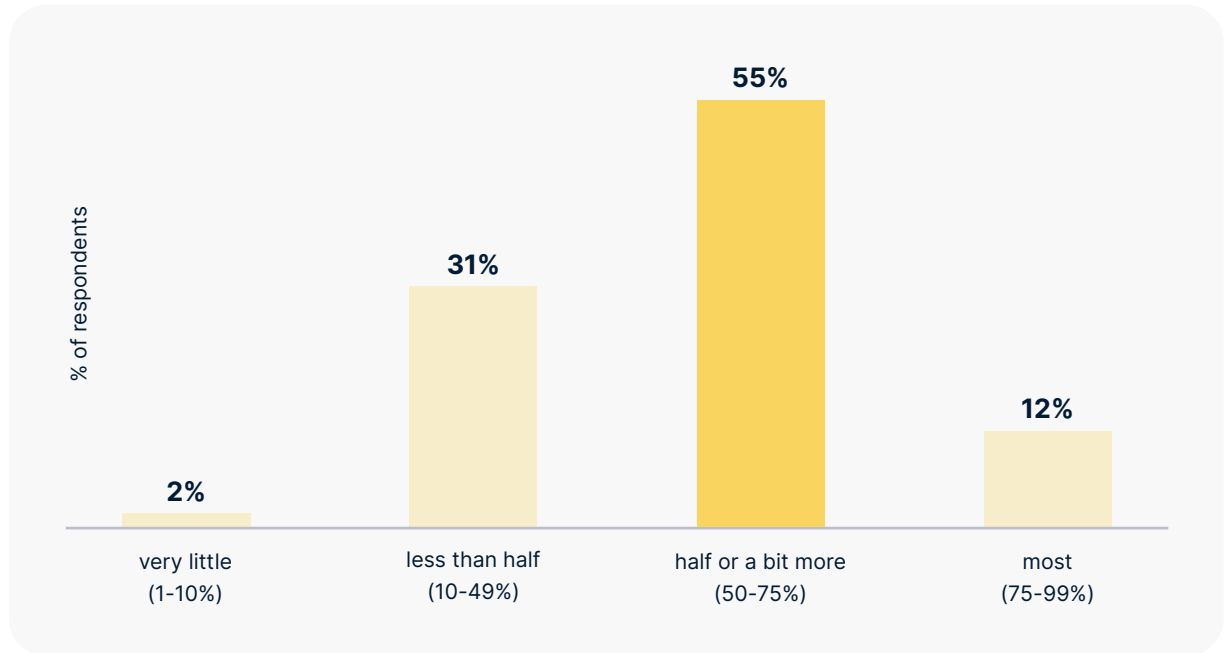


Figure 1: Percentage of Audio and Video Content Currently Captioned

ASR and captioning tools are used for a variety of use cases beyond accessibility

When asked how they use ASR and captioning tools, respondents revealed a wide variety of use cases, demonstrating the versatility and usefulness of the technology for a variety of purposes.

The top use cases include reaching greater audiences (52%), supporting internal company needs (50%) and creating transcripts (48%).

Interestingly, while captioning was originally deployed to accommodate people with disabilities, such as those who are Deaf or have hearing loss, the benefits of its usage have expanded significantly to improve business communication, engagement and audience reach. This data suggests greater business potential for using effective ASR, captioning and transcription technologies.

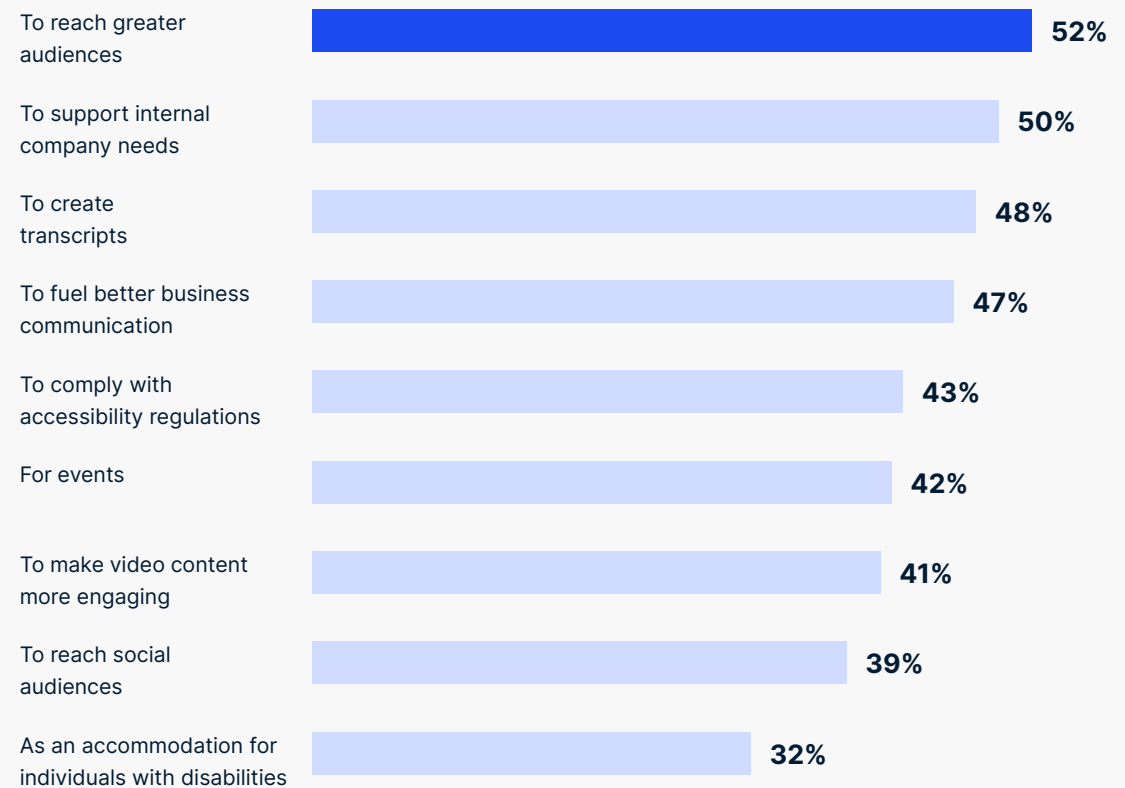


Figure 2: Use Cases for Using ASR or Other Captioning Tools

*Question allowed more than one answer and as a result, percentages add up to more than 100%

Most are using free, built-in ASR tools to caption and/or transcribe

When asked what types of tools they use to caption or transcribe their content, **most companies (70%) reported they are using free, built-in ASR tools.**

Yet the high usage of free, built-in ASR tools doesn't necessarily reflect user satisfaction. In fact, most of the companies using free, built-in tools report varied levels of accuracy (Figure 6), which means they are likely falling short of legislative standards for accessibility use across the US, UK and Canada.

Legislation, including The Americans with Disabilities Act (ADA), dictates that the accuracy rate should be at least 99%.

The respondents using paid, hybrid (49%) or paid, ASR-based services (46%), as seen in Figure 3, are more likely to be meeting these standards as they generally allow for more customization and human intervention to improve accuracy.

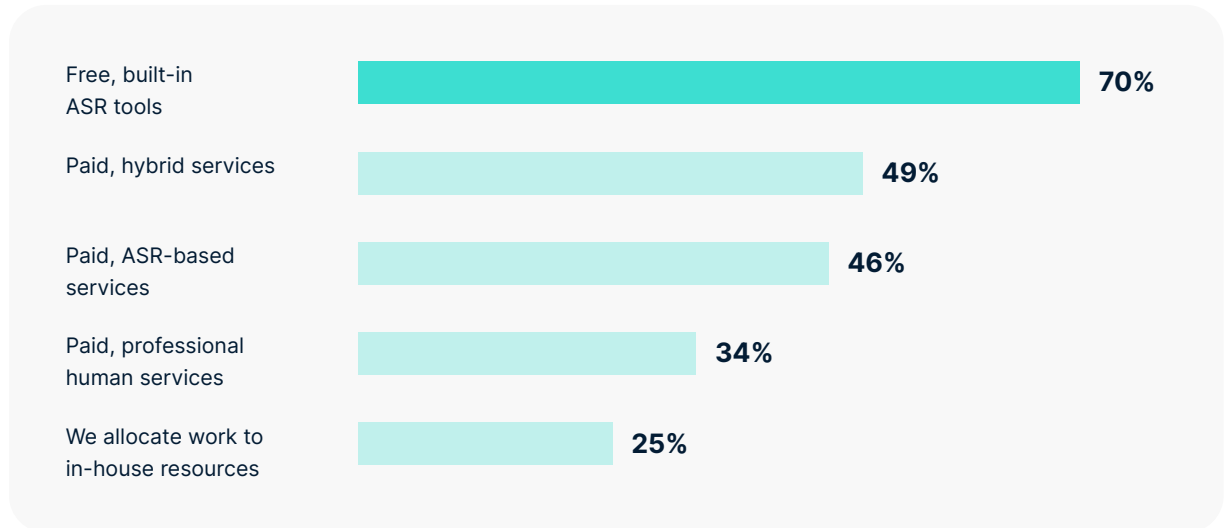


Figure 3: Tools Companies are Using to Caption and/or Transcribe

*Question allowed more than one answer. As a result, percentages add up to more than 100%

Almost half of business leaders are dissatisfied with the accuracy of their current captioning & transcription tools

When asked to rate the accuracy of the output of their current captioning and transcription tools, **47% reported that captioning accuracy varies.** They noted varied performance on identifying specific terms, such as names and more niche industry terms, correctly.

The fact that almost half of companies are dissatisfied when it comes to accuracy, coupled with the fact that most companies (70%) use free, built-in tools (as seen in Figure 3) suggests that a significant proportion of those tools simply isn't cutting it when it comes to nailing accuracy.

Regardless of the use case for ASR and captioning technology (Figure 2), companies are frustrated by inaccuracies. They cause them to set aside time to fix output errors, or worse, risk coming off unprofessional when using them for live events. As the need for effective ASR technology grows, this finding represents an opportunity (and a wakeup call) for those who are not yet leveraging more advanced ASR technologies which can be primed to boost their accuracy.

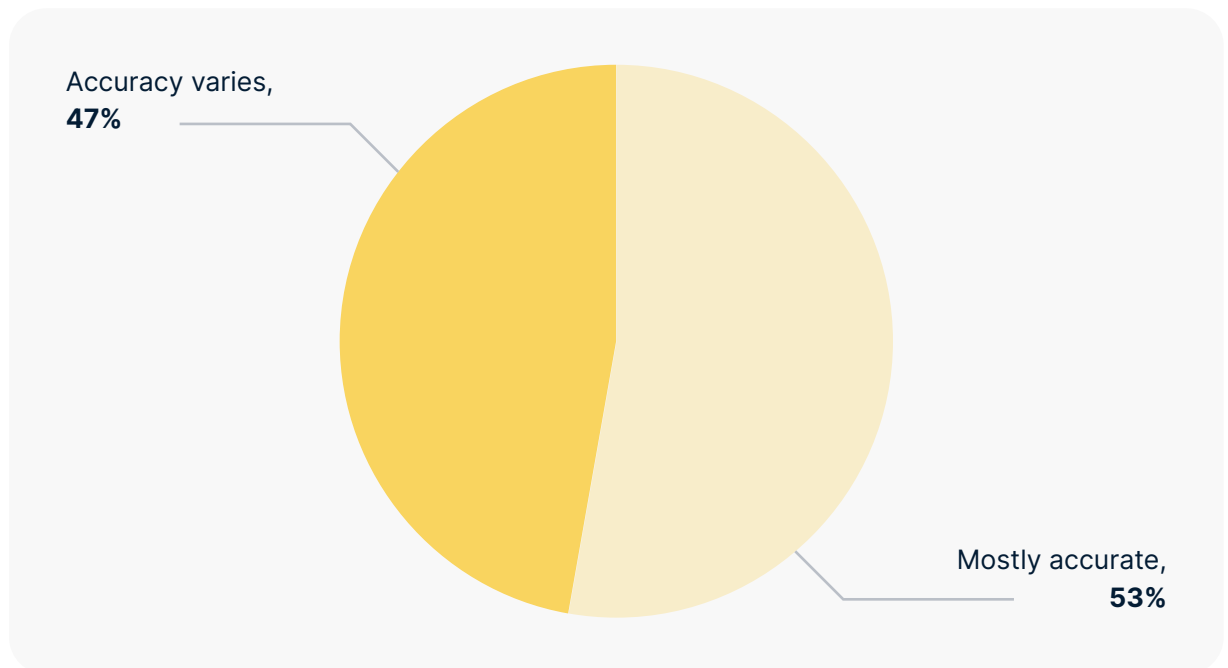


Figure 4: Accuracy of Current Captioning & Transcription Tools

Using free, built-in tools: Views on accuracy and usage for accessibility needs

When looking **more closely at the 70% of companies that are using free, built-in ASR tools** for captioning and/or transcription (as seen in Figure 3), we see that that a large majority of them, **(78%), use ASR to comply with accessibility regulations**. 55% of them use it for use cases outside of accessibility needs, as seen in Figure 5.

Also, more than half (55%) of those using built-in ASR tools report that they are dissatisfied with their accuracy, noting that specifics like names and terms are often displayed inaccurately. Only 45% find them to be mostly accurate, as seen in Figure 6.

This finding is worth noting, given that delivering accessibility effectively and in accordance with the law demands high accuracy. This data suggests there may be a gap between what companies actually need and the tools they are currently using to fulfill them.

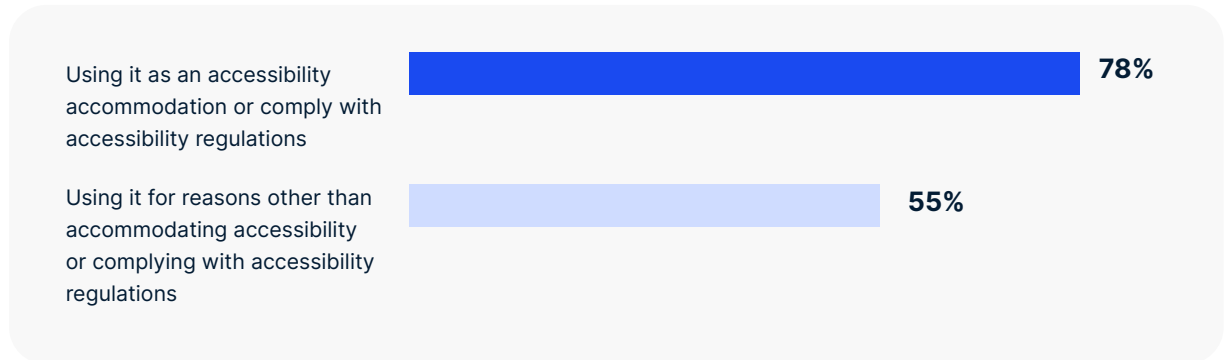


Figure 5: What the 70% of Respondents Using Free, Built-in ASR Tools Use Them For

*Question allowed more than one answer and as a result, percentages add up to more than 100%

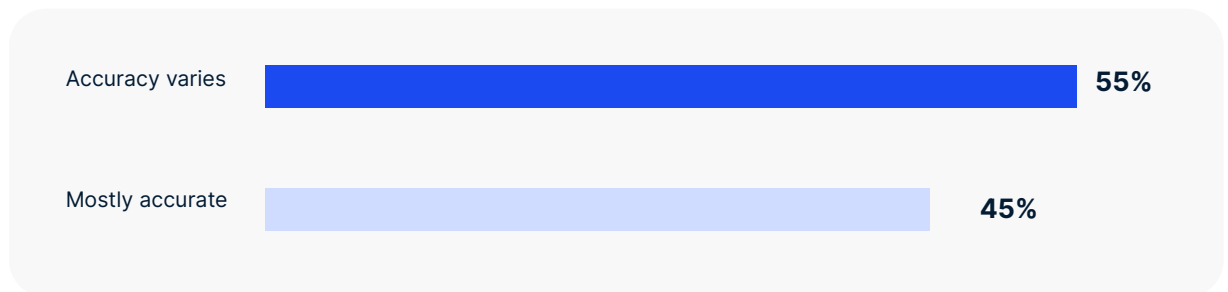


Figure 6: How the 70% of Respondents Using Free, Built-in ASR Tools View Their Accuracy

Speed, accuracy and scale reported as most important features for those using ASR

Respondents noted that the **top features important to them when using ASR were latency (93%), accuracy (90%) and scale (84%).**

It's worth noting that common ASR technologies do excel in their ability to offer companies scale and quick delivery of on-screen captions or transcript text. Yet many ASRs are not yet advanced enough to deliver on what's noted as additional important features like background noise and correct speaker identification on their own. Inability to perform on these specifics results in manual efforts and time needed to fix the outputs.

Additionally, 44% of respondents noted **custom vocabulary** as an important feature. Most common ASR technologies are not customizable with an 'upload terms' feature. Those who work in niche industries or have specific terminology important to them, are poised to use more advanced ASRs or hybrid technologies over free tools.

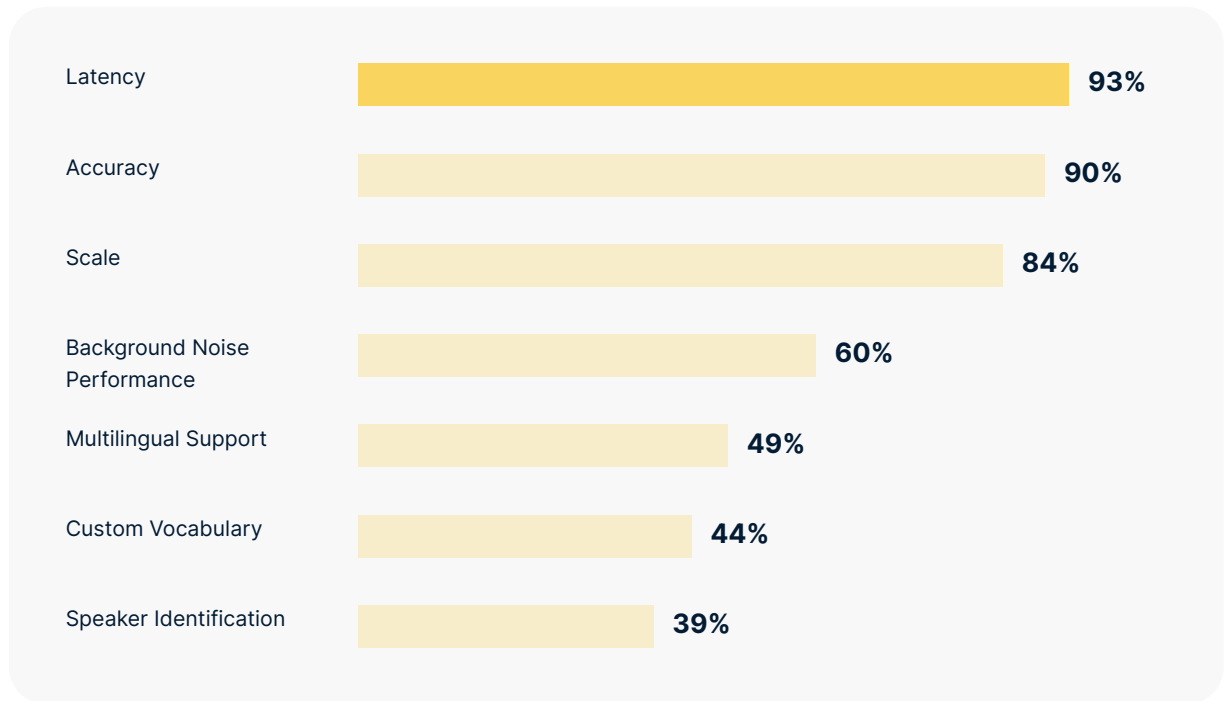


Figure 7: Most Important ASR Features

*Question allowed more than one answer and as a result, percentages add up to more than 100%

The challenges preventing greater usage of captioning and ASR technologies

What's holding companies back from increasing their use of captioning and ASR technologies? Many (49%) are still experimenting with these technologies and use cases and noted the lack of clarity around their value.

Interestingly, **data privacy and security are the second leading concern (48%) of respondents.**

Security is undoubtedly critical when captioning or transcribing sensitive or confidential information or conversations, but it's not guaranteed by many of the available ASR tools companies can use.

Not far behind are the inaccuracy of existing and/or free tools (39%) and the difficulty in using, setting up or integrating the tools into existing workflows (39%). It's easy to understand how these present a challenge to greater usage as they contribute to the manual effort needed, when the goal is to use ASR to automate the process in many cases.

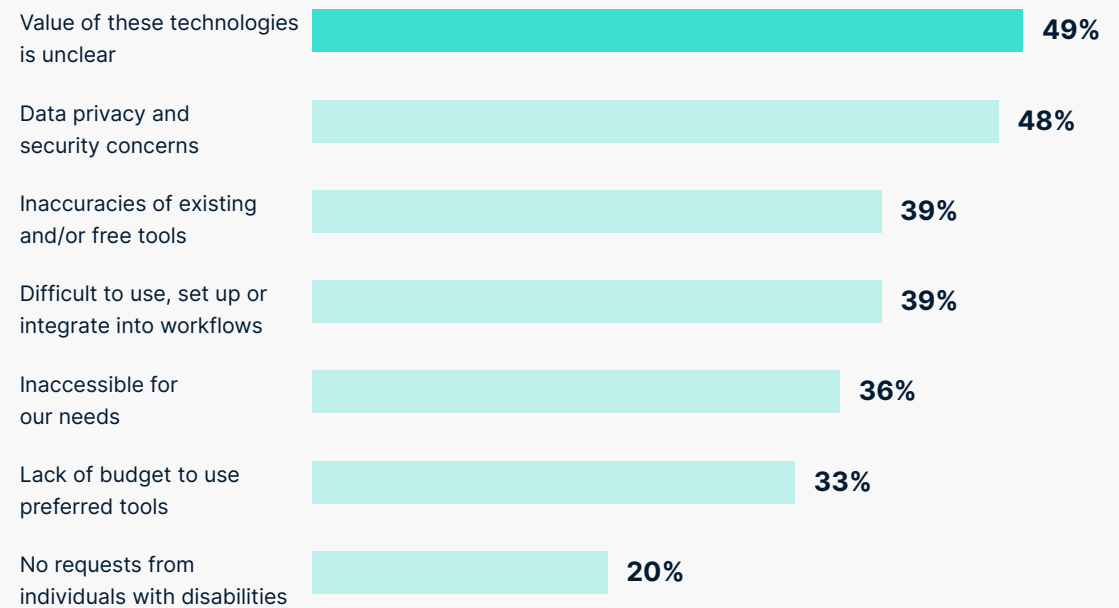


Figure 8: Top Challenges Preventing Increased Use of ASR Technologies

*Question allowed more than one answer and as a result, percentages add up to more than 100%

Current ASR, captioning & transcription tools aren't intuitive and require manual effort

Most respondents say that **understanding how to use ASR or captioning and transcription tools effectively (27%) requires the most effort**. Others noted editing the output so that it's more accurate (21%), adding specifics to the output (21%) and integrating the tools with existing systems (18%) also require work.

This data indicates that current tools require substantial manual effort to achieve satisfactory results and are most likely not sufficiently intuitive. Considering that most companies are using free, built-in tools for captioning and transcription (as seen in Figure 3), this is once again an indication of a notable shortcoming.

With many affordable paid and hybrid solutions on the market, it would seem that delaying the switch to better ASR tools is simply prolonging the need for manual effort and the sub-par results generated by less advanced tools.

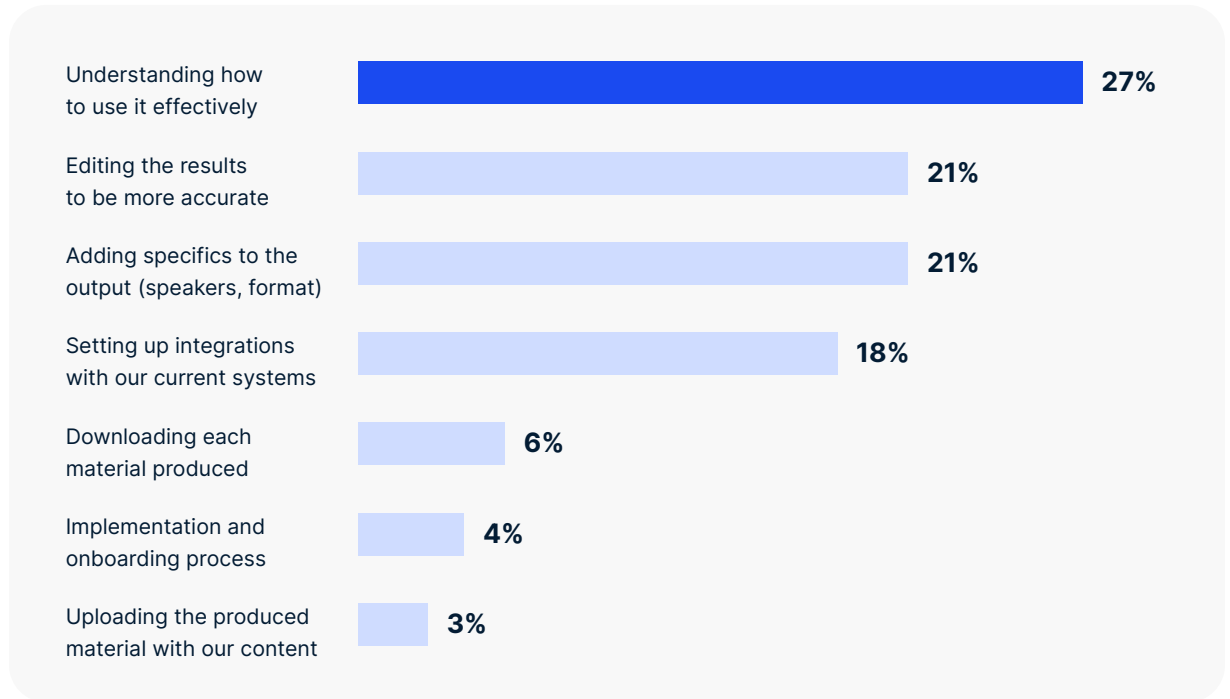


Figure 9: What Requires the Most Effort When Using ASR or Captioning and Transcription Tools

The biggest driver for greater usage of ASR is improved accuracy performance

Respondents noted which improvements would drive their companies to use ASR, captioning and transcription tools more often (Figure 10). **Higher accuracy performance (34%), faster end-to-end delivery process (30%) and more reliable security and privacy options (28%)** topped the list.

These findings were consistent with previous ones, such as accuracy being among the most highly regarded features in ASR tools (Figure 7) and inaccuracies being among the top challenges companies using ASR are faced with (Figure 8).

Additionally, a **greater proportion of those who use ASR for accessibility (40%) rated 'higher accuracy performance' as the leading improvement that would increase their usage.**

This statistic is lower (23%) for those who use ASR for other purposes, as seen in Figure 12. This reality indicates the potential gap between the importance of high accuracy when accommodating individuals with disabilities versus common ASR technologies' likely underperformance.

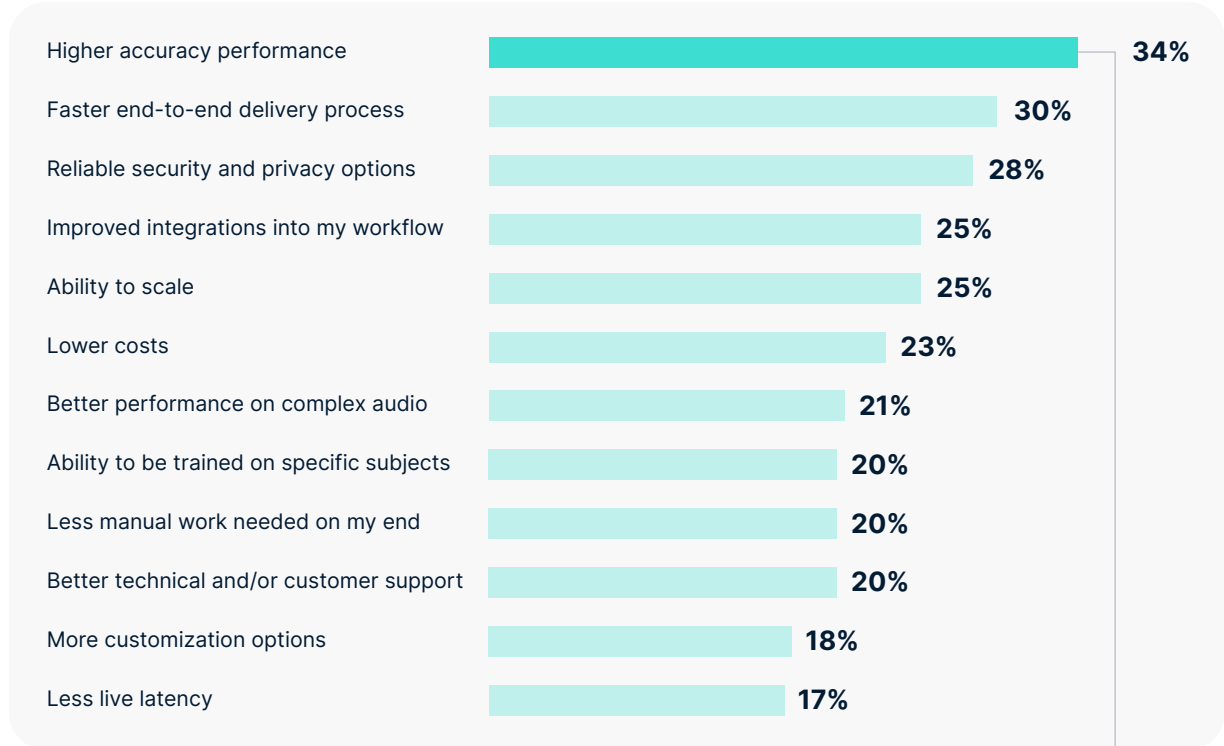


Figure 10: Improved accuracy performance would drive greater usage of ASR

*Question allowed more than one answer and as a result, percentages add up to more than 100%

40% of those using **ASR for accessibility and compliance** selected 'improved accuracy performance' as their **top driver for more potential usage**, compared to just **20%** who selected it and use **ASR for use cases outside of accessibility**.

99% of business leaders expect investment in captioning and ASR tools to remain steady or increase in the coming year

50% of respondents indicated that they expect themselves and their peers to maintain usage of ASR and captioning tools in the coming year. A close **49% indicated that investment in these tools will increase**, reaffirming that the pace of adoption of ASR and captioning tools is not slowing down.

This data suggests that those who are not investing in using ASR, captioning and transcription technology may find themselves lagging behind those who do.

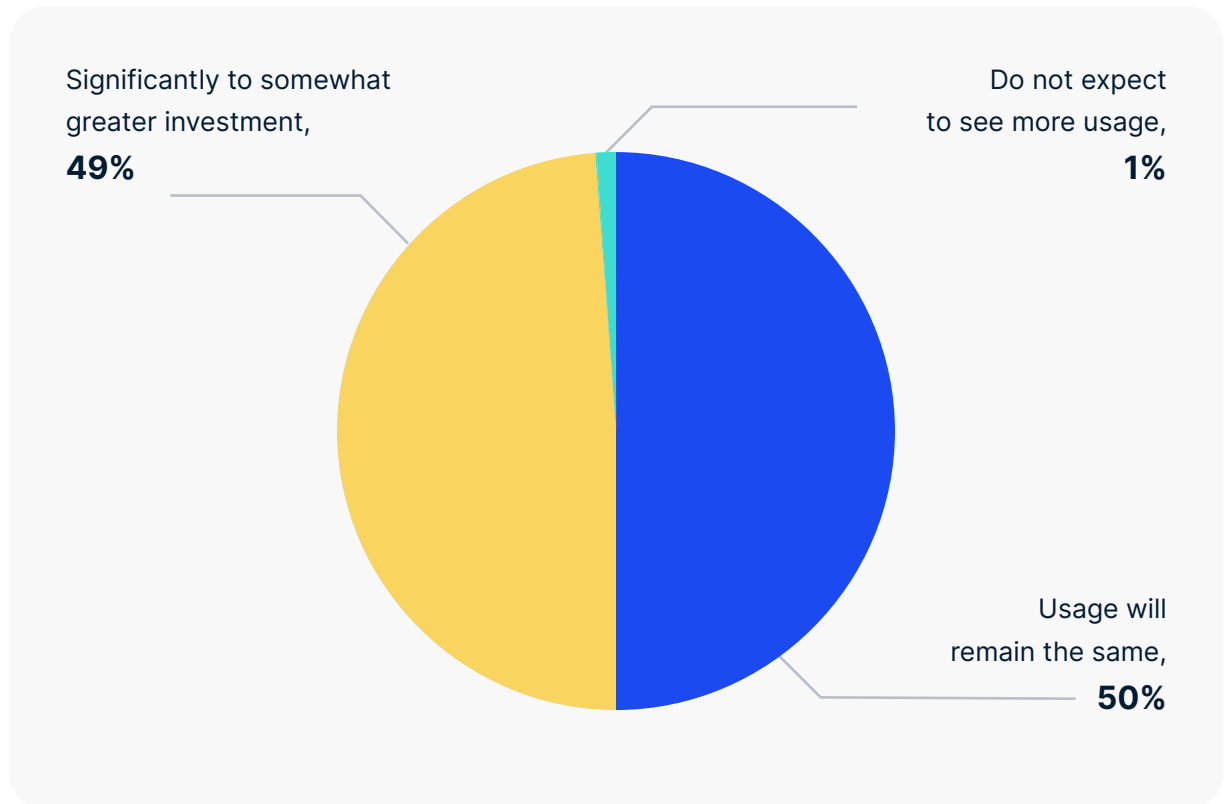


Figure 11: Expectations for Increased Usage of ASR and Captioning Tools in the Coming Year

92% of companies are exploring Generative AI in addition to ASR and captioning

Most companies (92%) are exploring the use of Generative AI in addition to using ASR and relevant captioning and transcription tools.

In a previous finding (Figure 8), half of respondents (49%) cited a lack of clarity on the value of increasing their ASR and captioning usage. However, as the significant majority experiment more with Gen AI's added value of insight extraction on top of ASR, that perception may shift.

Half of the companies (50%), for example, are exploring the use of Gen AI for creating summaries, 45% are looking into using it for content generation, 38% are exploring using it for extracting insights and analytics.

When it comes to the volume of content ASR and other captioning tools are capturing, Gen AI offers significant, time-saving value that shouldn't be overlooked.

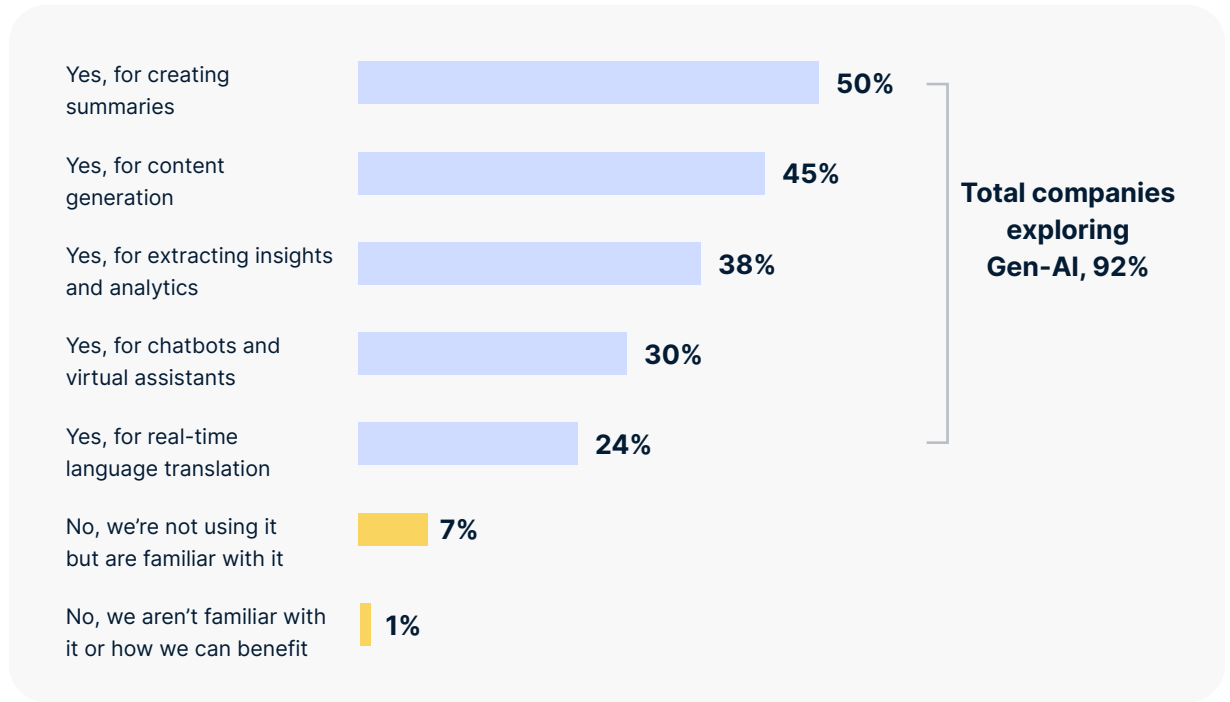
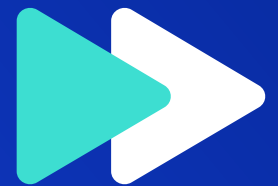


Figure 12: Companies Exploring the Use of Generative AI

*Question allowed more than one answer and as a result, percentages add up to more than 100%

Demographics



Country, industry, company size and job seniority of respondents

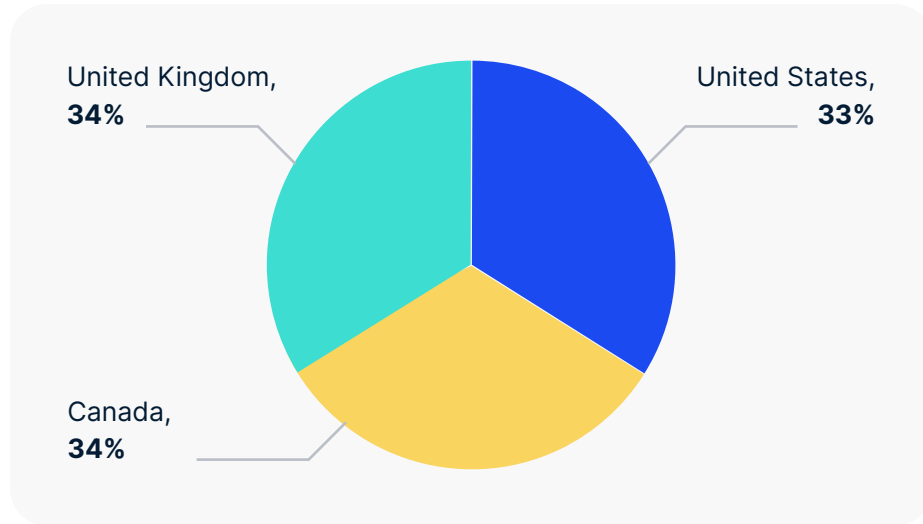


Figure 13: Country

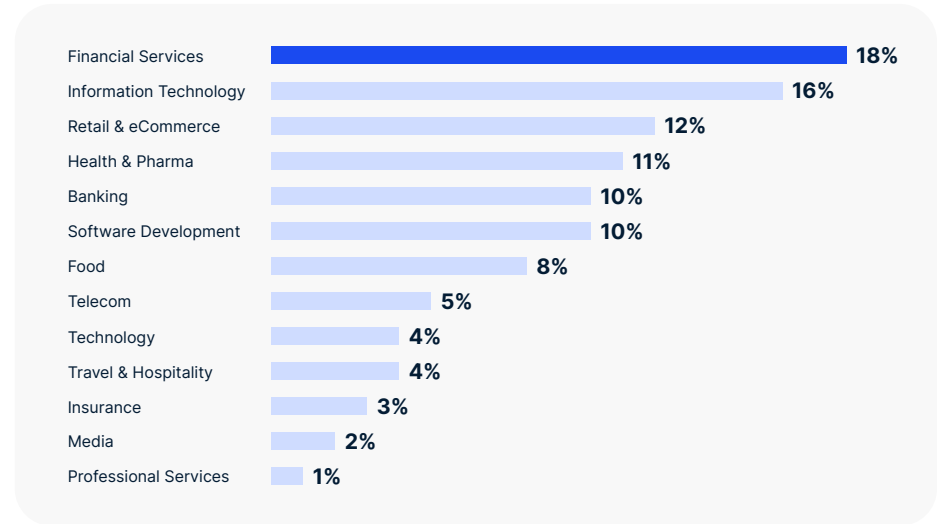


Figure 14: Industry

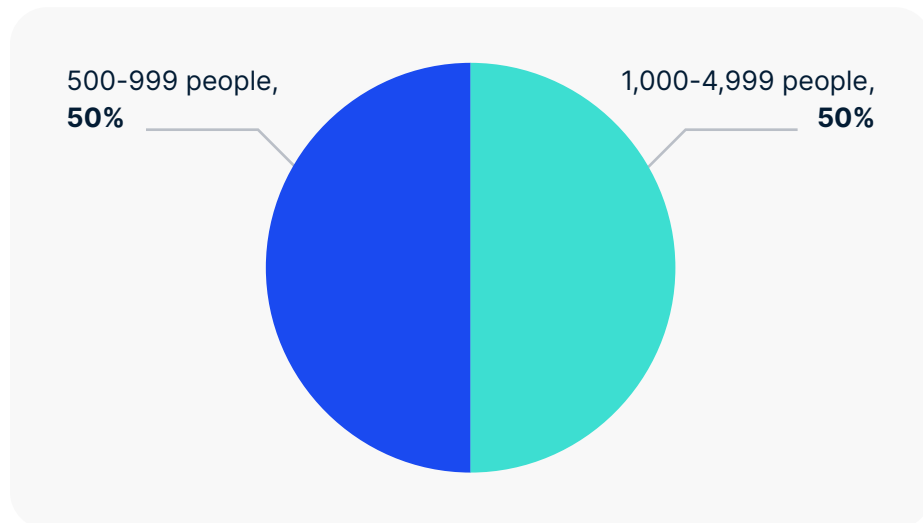


Figure 15: Company Size

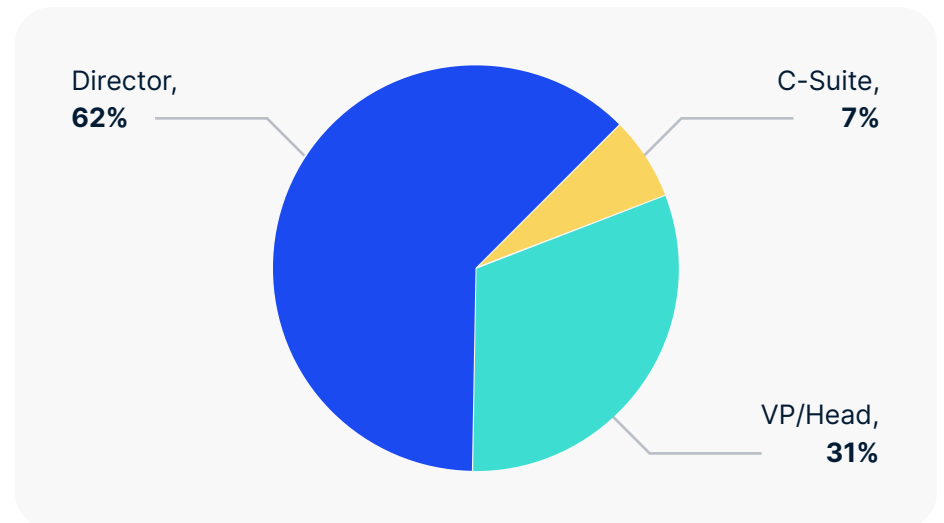


Figure 16: Job Seniority

Responsibilities and technologies used by respondents & their companies

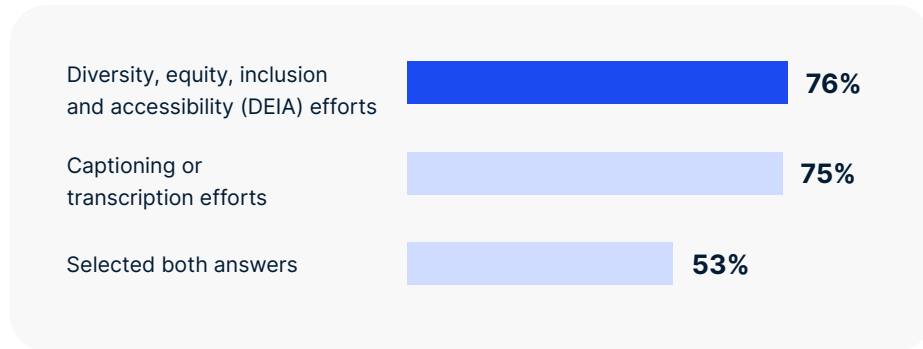


Figure 17: Responsibilities

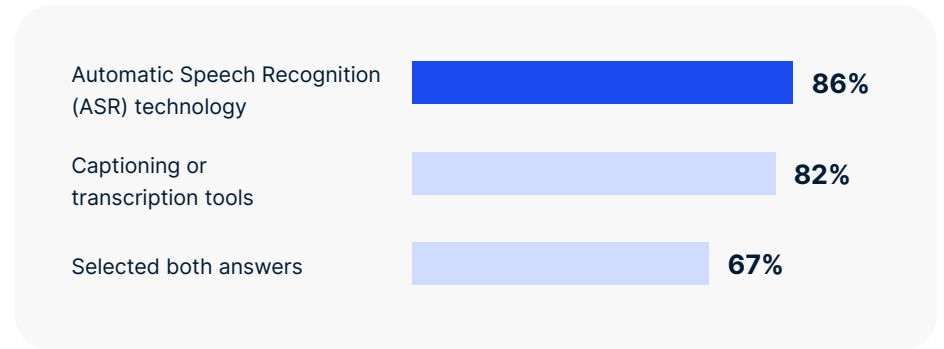


Figure 18: Technologies Companies Use

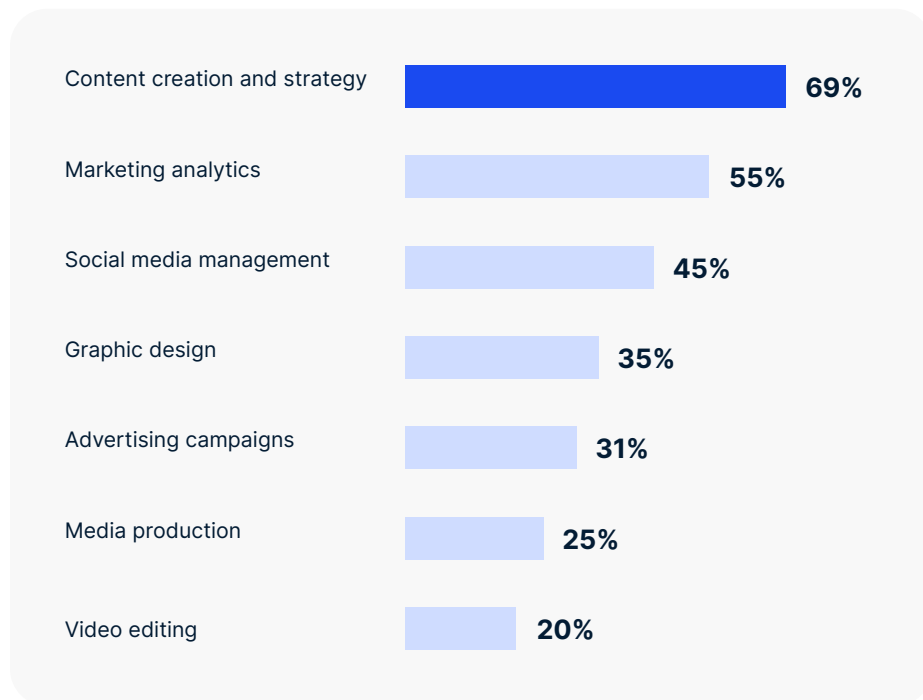


Figure 19: Additional role responsibilities

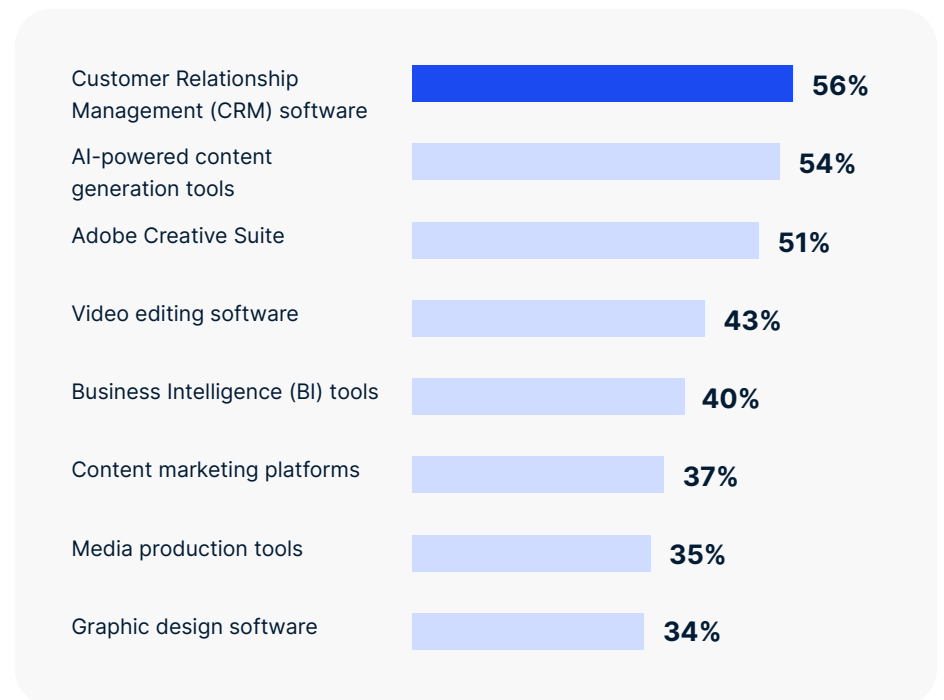


Figure 20: Additional technologies being used

About Verbit

More than 3,000 partners across the globe trust Verbit to put their words to work. Verbit delivers domain expertise to assist with captioning, transcription, audio description and translation needs across the corporate, media, education, legal and government sectors. As the leader in the \$30B transcription industry, Verbit's team continuously innovates to take its solutions further.

Verbit recently brought its proprietary ASR technology, Captivate™, to the market after witnessing existing gaps in ASR technologies that were not user-centric or accurate. With domain dictionaries, custom

term upload options and more, business leaders can leverage an ASR solution that is customizable and surpasses the accuracy of existing tools. Verbit also offers deep Generative AI-based insights with its Gen.V technology. With Verbit, customers can also apply these key insights wherever they are needed with helpful integrations.

Whether its greater audience reach, accessibility or efficiency you're after, Verbit's verbal intelligence platform is designed to help you make the most of every word captured.

Discover more about Verbit's
new ASR technology, 

Learn more >

For more information, please email press@verbit.ai or visit us on social.

