

AI and Software Quality: Trends and Executive Insights

Why building trust in AI is essential to
ensuring end-to-end software quality



CONTENTS

Executive summary	3
Key findings	4
Everyone is integrating AI, and many are doing so in untested waters	5
Is AI-augmented testing the solution to emerging challenges?	7
Will testing still require humans in the future?	9
Investment in AI-augmented testing on the rise	11
How AI can be used to validate AI applications for increased trust	12
Building trust in AI with Leapwork	13

Executive summary

AI has transitioned from an emerging technology to business-critical. Companies across industries have rapidly integrated AI applications into their operations to enhance efficiency, innovation, and customer engagement. However, as AI becomes more deeply embedded in these processes, its limitations and the challenges it introduces are increasingly apparent.

This report delves into these challenges, particularly in the modern customer journey, where AI applications like chatbots are pervasive, and where end-to-end quality is more important than ever.

While AI is widely adopted across sectors, concerns about its reliability, accuracy, and overall effectiveness continue growing. These concerns have been amplified by high-profile system failures over the past year, exposing critical vulnerabilities in software systems. As a result, businesses are increasingly re-evaluating how technologies are tested, maintained, and trusted to deliver the expected quality consistently.

At the same time, companies are increasingly using AI-augmented testing tools to ensure applications perform as expected across the customer journey.

This raises several questions: Can we fully trust AI to address concerns about the very technology it is meant to test? And how will AI-augmented tools impact the role of humans in testing and shape the Quality Assurance teams of the future?

Leapwork spoke to 401 senior and technical professionals in the US and UK, half of which were C-Suite executives, to understand two critical components of the new AI era in the context of the digital enterprise:

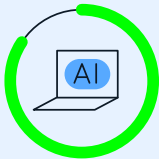
- 1 Testing of AI:** How businesses will build trust in the AI applications they are integrating
- 2 Testing with AI:** How businesses can improve the efficiency and effectiveness of their software testing

This report provides decision-makers with a comprehensive overview of the current state of AI and software quality, offering essential insights and solutions for businesses to adapt and consistently deliver exceptional user and customer experiences at scale, now that AI is a critical part of the equation.

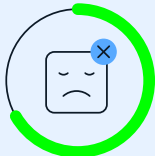
Key findings

TESTING OF AI

AI applications pose opportunities but also risks across the business journey



There is widespread adoption of AI, with **85% of organizations** having integrated AI applications into their tech stacks in the past year.

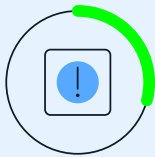


However, **68% of those** have already encountered performance, accuracy, and reliability issues.

There is a critical need for AI testing



78% of companies see testing AI as essential, which aligns with the majority already encountering significant issues.



However, **30% don't believe their current testing processes are up to the task** of ensuring reliable AI apps, indicating a substantial gap in effective AI testing practices.

TESTING WITH AI

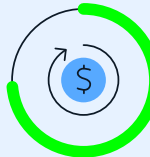
Businesses are turning to AI-augmented testing tools to solve quality challenges



Only **16%** believe their current testing practices are efficient.



Recognizing these inefficiencies, **79% of companies** have already adopted AI-augmented testing tools, and 64% C-Suites trust their results – technical teams trust them even more (72%).



74% plan to further invest in these tools within the next year.

The continued need for human oversight in AI-driven QA



Despite advancements in AI-augmented testing tools, **68% of C-Suite executives** believe human validation will continue to be essential for ensuring quality across complex systems.



The adoption of AI-augmented tools is not replacing human roles but transforming them. **53% of C-suite executives** report an increase in new positions requiring AI expertise, emphasizing the evolving nature of quality assurance in the AI era.

Everyone is integrating AI, and many are doing so in untested waters

Amid a year of outages that have taken social media platforms offline, stopped fast food deliveries, stranded passengers at airports, ceased bank operations and postal deliveries, global conditions are ripe for a potentially rocky AI revolution.

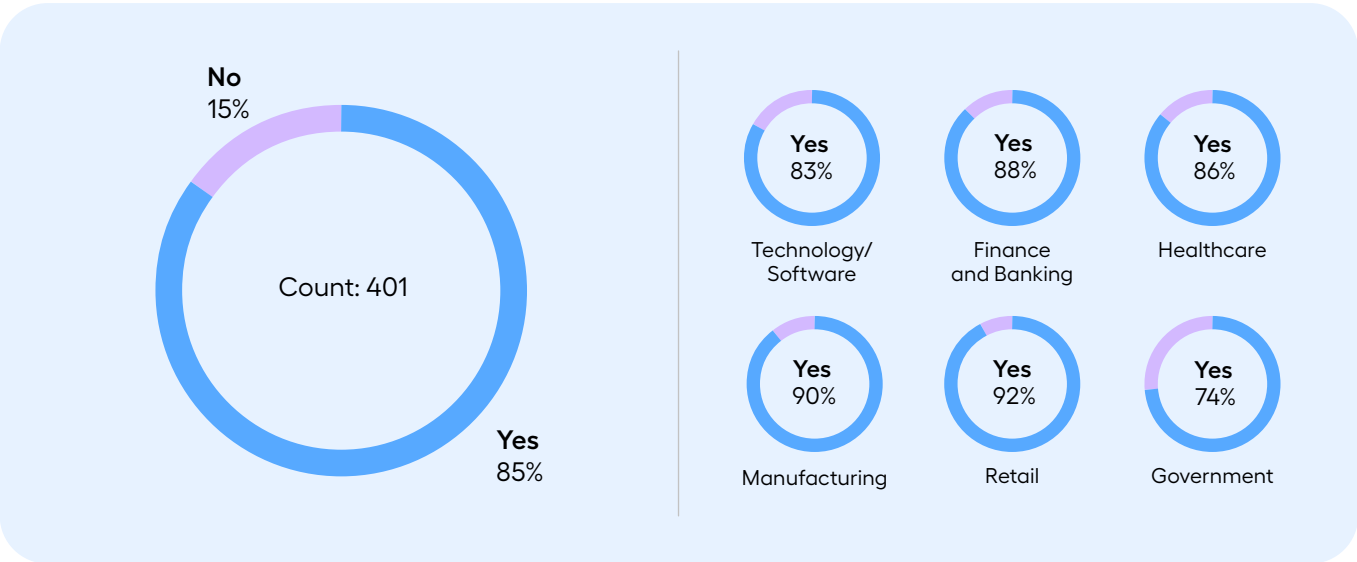
With 85% of companies having integrated AI into their tech stack in the past year, there's cause for concern that the number of IT failures is on the rise. Already, 68% of companies have faced performance and reliability issues with their AI applications.

“There have been too many outages this year alone, many of which affected millions of customers for big brands. We’ve been given a wake-up call no one can ignore,” says Christian Brink Frederiksen, CEO of Leapwork. *“What makes digital infrastructure today so tricky to test is the copious amount of complex, interconnected applications. A tiny error in one application could have a monumental cascading effect and shut down businesses.”*

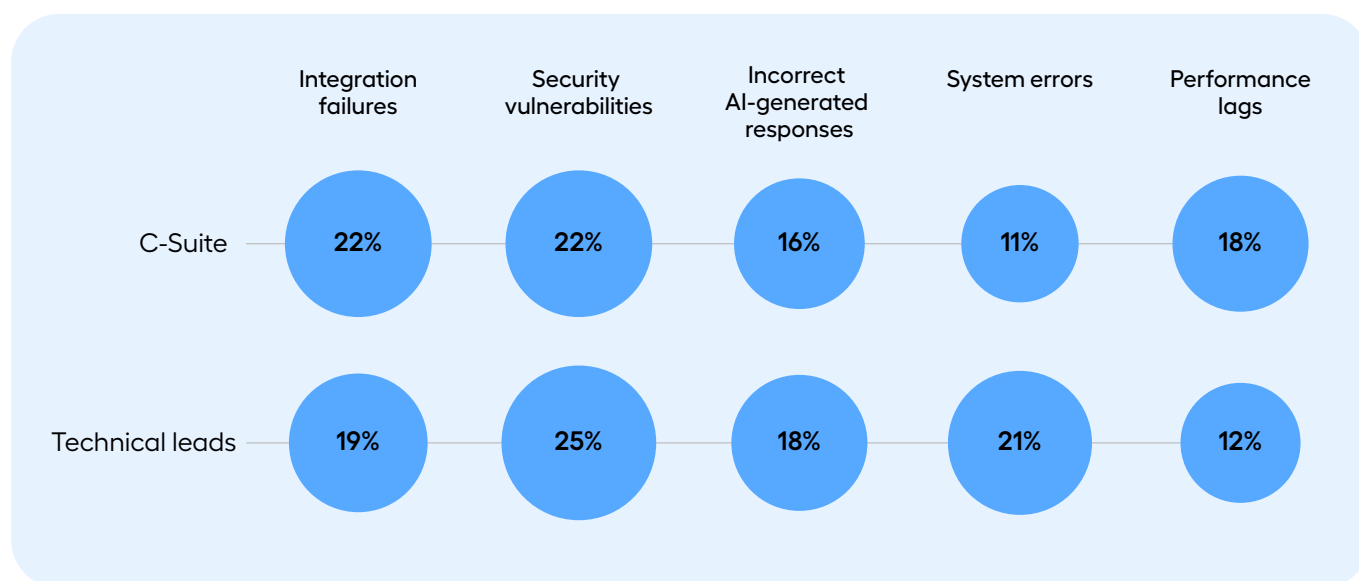
Like any software, there is growing recognition that AI applications need thorough testing to prevent business disruptions. An overwhelming 78% of companies agree the AI apps need better testing; in fact, 77% of C-Suite executives said testing is critical to ensuring their performance, accuracy, and reliability.

“For all its advancements, AI has limitations, and I think people are coming around to that fact pretty quickly,” says Robert Salesas, CTO at Leapwork. *“The rapid automation enabled by AI can dramatically increase output, but without thorough testing, this could also lead to more software vulnerabilities, especially in untested applications. It makes sense that C-Suite executives would be especially sensitive to this because of the implications for customer experience and negative publicity.”*

Have you adopted and integrated AI applications into your technology stack in the last year?



What is the most common bug or issue you've encountered with your AI applications?



There's an opportunity here for cross-industry collaboration to ensure more testing tools are up to scratch for the challenges of the modern world where AI apps are more and more widespread."

For now, AI has limitations, and integration failure in particular was the most popular issue cited by 22% of C-Suite executives.

Integrating AI apps is a problem for companies because of three main reasons: a resistance to change within the company (20%), the difficulty of managing the rapid pace of AI advancements and updates (19%), and crucially, inconsistent performance and reliability of AI applications (19%).

Unfortunately, sizable gaps remain for existing testing resources and practices. For starters, 24% do not have a dedicated team or individual responsible for testing AI apps, and 26% do not have a commercial testing platform. Nearly a third (30%) say outright that they do not believe their current testing processes can ensure reliable AI apps.

Despite the urgent need for reliable AI, only 16% of companies believe their testing processes are efficient.

This reveals a troubling gap in quality assurance that raises an important question: **If today's testing methods are falling short, how can the industry ensure that AI delivers on the promised benefits?**

Is AI-augmented testing the solution to emerging challenges?

As companies grapple with the inherent bugs and limitations of AI, a consideration emerges: **the potential of AI-augmented testing tools to effectively tackle the unique challenges posed by AI applications.**

While AI-augmented testing tools are gaining traction across industries, their true potential in enhancing AI reliability and performance is yet to be fully realized. Leapwork’s findings reveal that the trust is there - 68% of overall respondents say they trust the results that AI-augmented tools provide - but isn’t there a paradox in using AI to address concerns about the very technology it is meant to test? This is a crucial aspect we’ll explore further in this report, as leveraging these tools could be key to mitigating the risks associated with AI integration.

One thing is clear: **To overcome the shortcomings of existing testing processes, many organizations have**

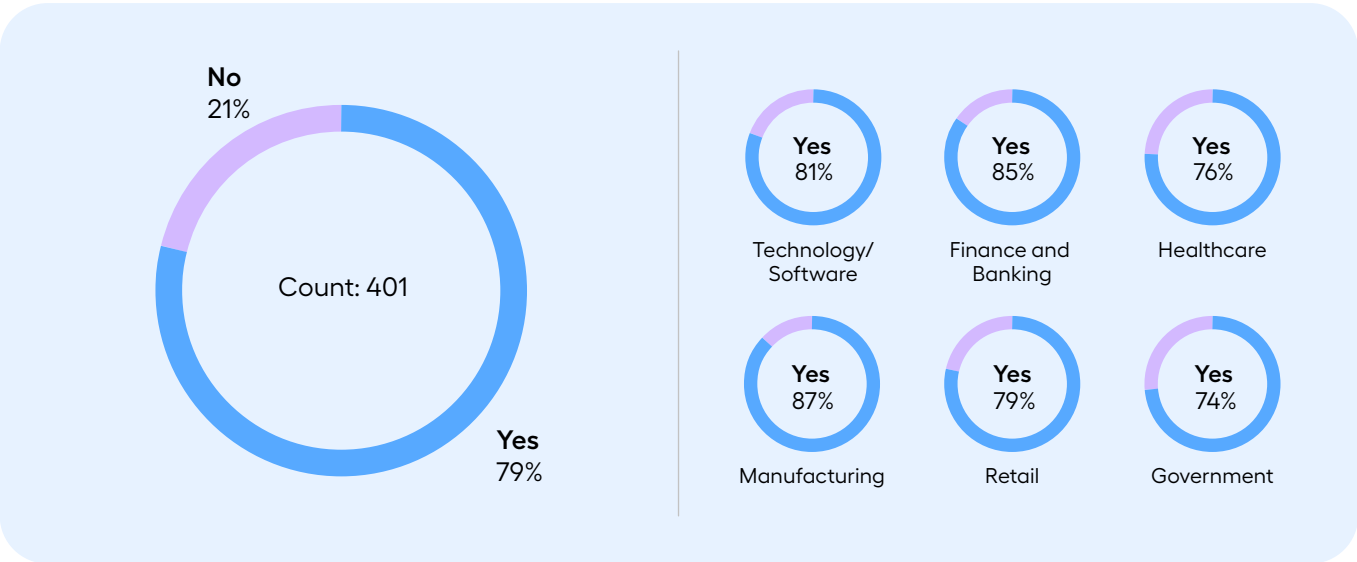
started adopting AI-augmented testing tools – an overwhelming 79% now use them.

The fact that so many also trust the results of these tools indicates an understanding amongst early adopters of their possibilities and limitations. Especially as trust is even higher amongst technical leaders (72% vs. 64% for C-Suite).

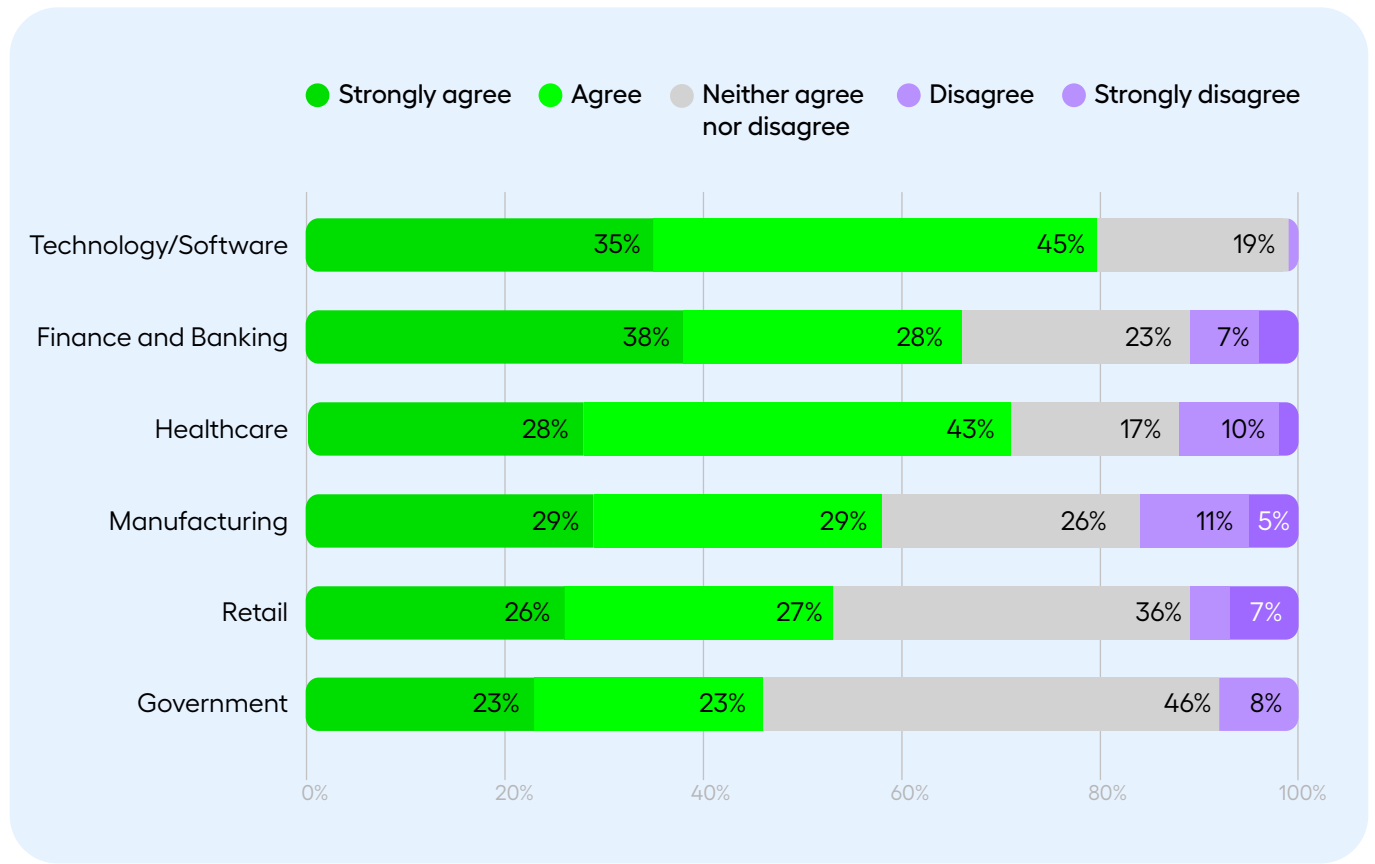
But a closer look at the survey results reveals gaps between industries: **there’s significantly more trust placed in AI-augmented testing within the technology industry (80%) than in retail (53%).**

“With retail, it’s easy to think about the mega retailers and forget about the smaller boutique vendors who have less familiarity with AI,” says Salesas.

Do your testing processes currently incorporate AI-augmented testing tools?



I trust results that AI-augmented testing tools provide



“Not every retailer is tech-first like Amazon, and there’s likely a cultural gap at play here: tech companies are at the forefront of AI development and implementation, which means they get first dibs on the talent who is more likely to have a deeper understanding of the tools’ capabilities and limitations. On a practical level, there are also still many retail operations that rely on older systems that don’t integrate seamlessly with AI-augmented testing tools, and the stakes of failures are high when any errors can directly impact customer satisfaction and sales. Retail environments themselves can be enormously diverse and complex – and the customers even more diverse – which might be giving professionals in the industry pause about trusting AI’s ability to test all situations accurately.”

Still, as trust in these tools grows, it naturally raises a broader reflection on how adoption of AI-augmented testing tools will affect the role of human testers in the long term.

Will testing still require humans in the future?

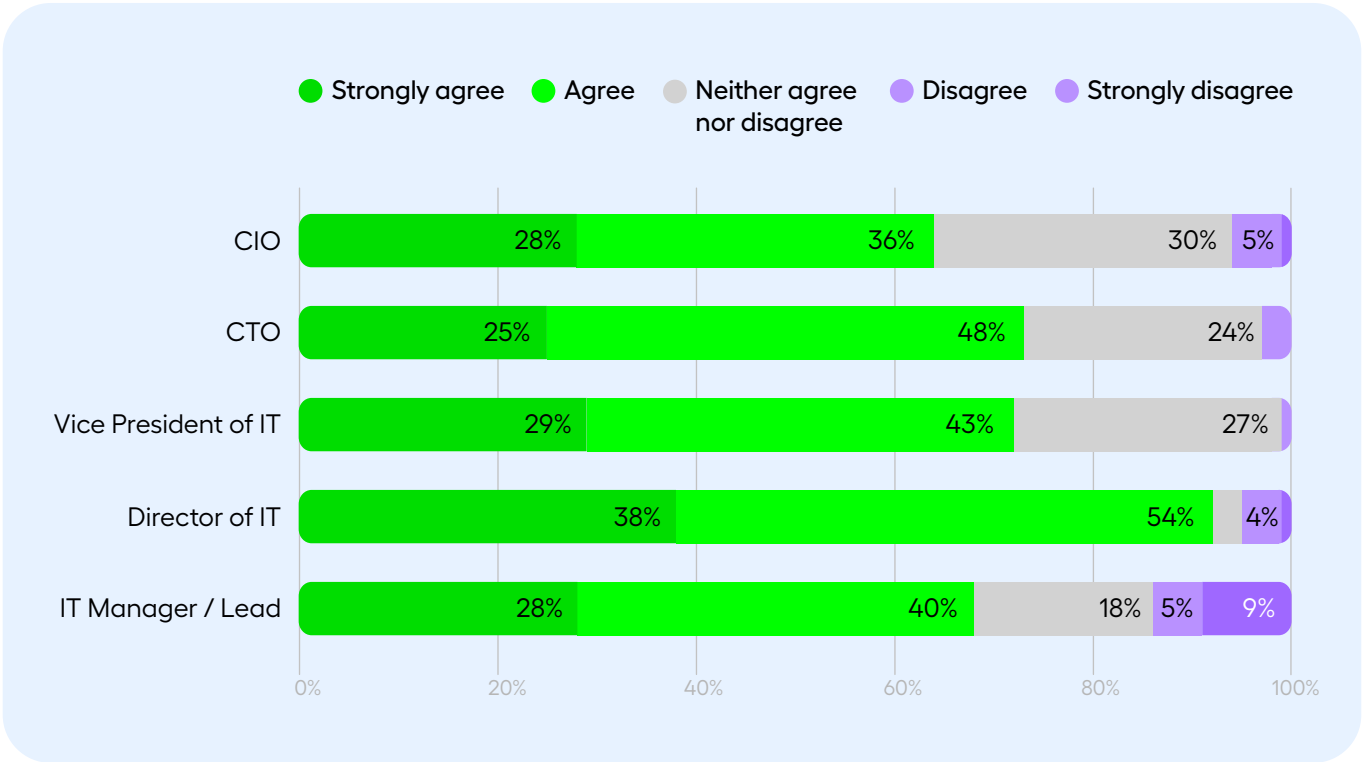
Like every industry and trade impacted by AI, QA teams now face the question, ‘what will happen to humans?’ Leapwork’s findings suggest humans are unlikely to disappear from the testing equation anytime soon. In fact, over two-thirds of C-Suite executives (68%) believe that testing will need human validation for the foreseeable future, and almost every single IT Director (92%) agrees.

“There’s always going to be some variation in how technical teams and C-Suite executives perceive the need for human validation,” says Salesas. “For IT teams, there’s a natural concern about job security as AI tools evolve, but the focus should be on how

these tools can enhance their roles rather than replace them. On the other hand, C-Suite executives are looking at business operations more broadly, with optimism about how technology can improve efficiency. While their perspectives may differ, both groups agree that human input will remain a critical part of the testing process.”

“I believe that the synergy between AI and human expertise represents a transformative partnership in software testing. AI tools can significantly enhance efficiency, allowing technical teams to focus on innovation and ideation rather than the repetitive details of testing in an increasingly complex software

I believe that testing will continue to need human validation for the foreseeable future



environment. However, no matter how advanced the tools become, the principle of requiring human oversight and independent review will always be essential to ensure accuracy and reliability.”

Interestingly, one sector is an outlier: manufacturing. While most respondents still believe in human validation, the number is a lot lower (56%) than in other sectors like technology/software (79%), finance and banking (83%), and healthcare (85%).

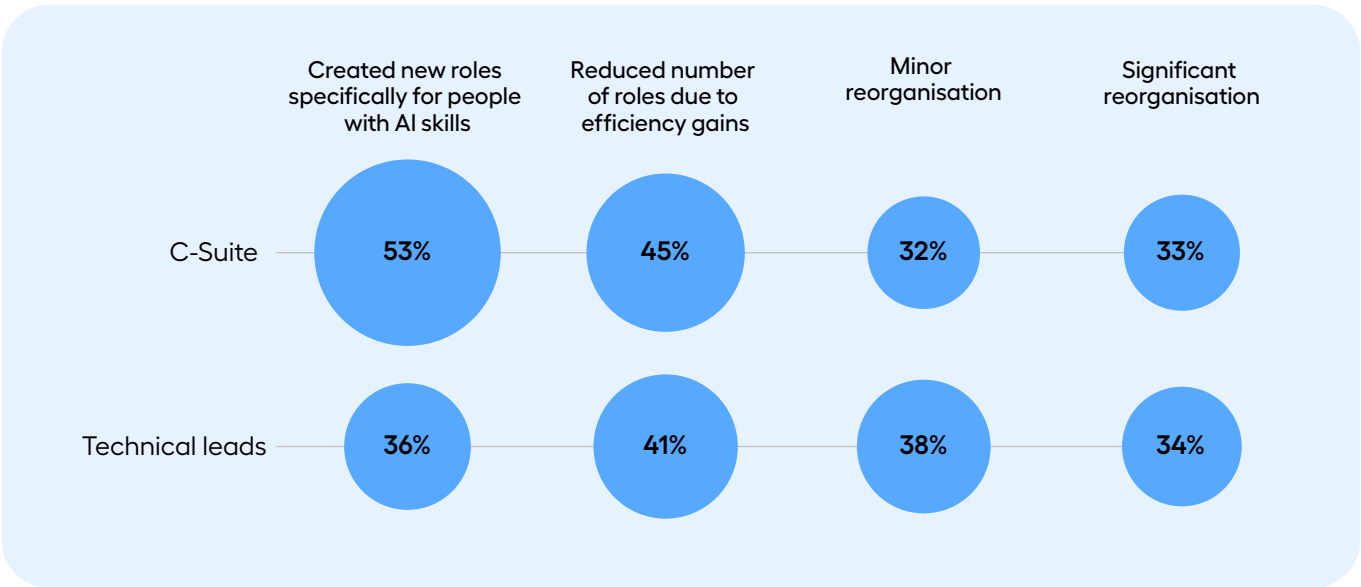
“Manufacturing is centered today around maximizing automation – it’s all about standardization, repetitive processes, efficiency, and a strong desire to keep costs low. This could explain why respondents in the sector perceive less need for human intervention. It speaks to a range of different priorities, regulations, environments, and operational characteristics of these sectors. Finance and healthcare come with considerably strict and unique compliance and safety requirements that may be nudging those sectors towards a stronger preference for human supervision.”

Whether or not AI-augmented testing will directly impact headcounts is a question senior leaders are still debating. Nearly half (45%) of overall respondents say AI-augmented testing has created

new roles specifically for people with AI skills, but almost as many (43%) have seen a reduction in roles due to efficiency gains. C-Suite executives are much more bullish, though, than technical leaders.

Over half of C-Suite executives (53%) say the tools have increased the number of new roles compared to just over a third (36%) of technical leaders. These perceptions also vary based on sector: 52% of respondents in healthcare are seeing new jobs created compared to only 36% in government. About 57% of respondents in manufacturing report a reduction in jobs.

How has the introduction of AI in testing changed your team’s structure and roles?



Investment in AI-augmented testing on the rise

No matter how AI impacts human testing roles, the trend is clear: **AI-augmented testing tools continue to gain popularity. Most organizations (74%) foresee increased investment in AI-augmented testing tools in their organization in the next year.** The degree to which they agree varies from role to role. For example, most CTOs (77%) expect increased investment into the tools, but far fewer CIOs (58%) do.

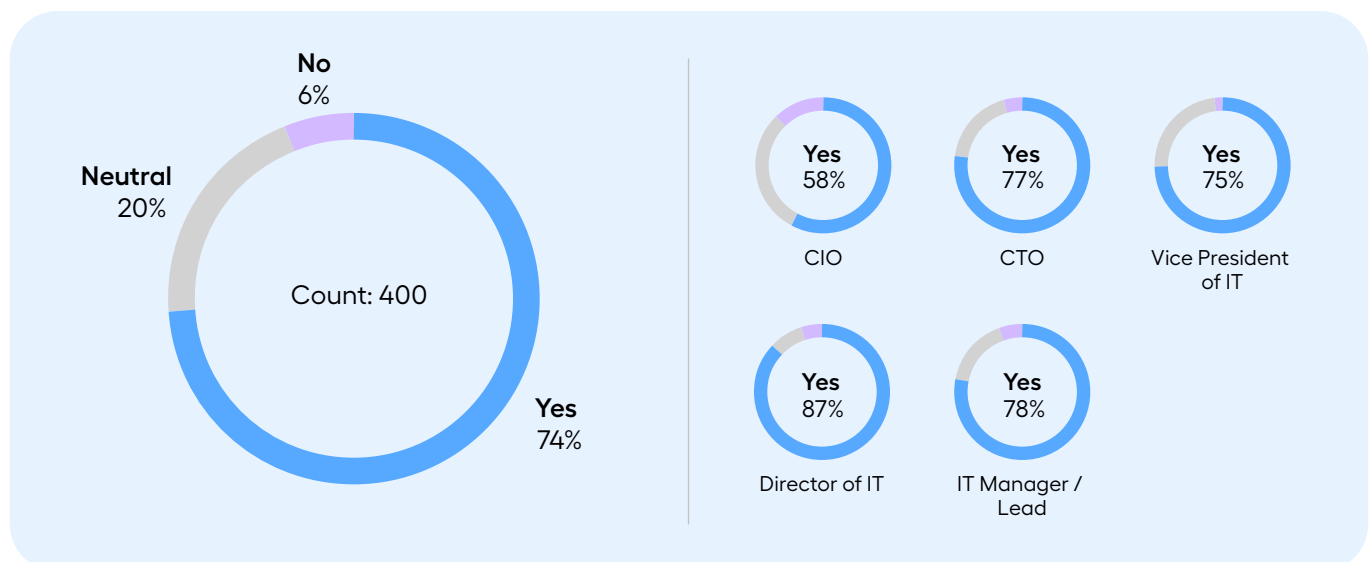
“Generally speaking, CTOs tend to focus much more on emerging technologies and how to apply them to the business. That goes a long way to explain why they might advocate more strongly for AI-augmented tools and AI in general. They’re thinking about the long-term vision and staying competitive in that future.” Says Salesas. “CIOs, on the other hand, are more about the day-to-day IT operations – particularly the short-to-medium term. They manage a far larger remit of operations and competing priorities. This might make CIOs more reluctant to take any big leaps on tools that

may require further skills and financial investment to fully take advantage of.”

“I believe that a critical component to swaying executives who are unconvinced about the value of AI-augmented testing will be to present tools that are intuitive to use, making them accessible not just for technical teams but also for business users. When the skills gap is so wide, you can’t afford for things to be difficult to adopt.”

This leads us back to the pressing question: If easy-to-learn AI-augmented testing tools are the answer to buggy AI apps, is there a paradox in relying on the same technology that’s already causing issues to test itself? To resolve this potential paradox lies in understanding that AI-augmented testing tools are not simply another layer of AI—In the right context, they can be used to address specific shortcomings of AI systems by providing rigorous, unbiased assessments of their performance.

I foresee an increased investment in AI-augmented testing tools in my organization over the next year



How AI can be used to validate AI applications for increased trust

This is where Leapwork's approach comes into play. Leapwork's approach to using AI to test AI is highly effective and designed to build trust - not just in your AI applications, but across all the applications that make up your business processes.

Understanding how AI can validate AI

Leapwork's AI capabilities are designed to assess how well generative AI handles specific tasks and responds to user-defined prompts. It does this by comparing the AI-generated responses against predefined, human-crafted expectations. This ensures that the outputs of your AI applications are consistent, accurate, and aligned with your intended results.

This process helps to catch hallucinations, where AI might generate incorrect or nonsensical responses. By comparing these outputs against independent data, Leapwork acts like a second pair of eyes, objectively verifying the AI's responses without the biases the original AI might introduce.

For instance, consider an AI chatbot designed to answer user queries. If asked, "Do you offer international shipping?"—a question that could be phrased in various ways such as "Can you ship to other countries?" or "Do you deliver overseas?"—the AI Validate block ensures that every response, which might also vary linguistically, still accurately conveys the correct answer: "Yes, we offer international shipping to select countries." By automating this validation process, Leapwork helps you avoid the pitfalls of AI unreliability, providing a robust framework for ensuring that your AI applications deliver the right results every time.

Building trust in AI with Leapwork

Leapwork's AI capabilities

AI Validate

AI Validate compares AI-generated responses with expected outcomes to ensure consistency and accuracy, catching errors like hallucinations and verifying that outputs align with intended results.

AI Transform

AI Transform reduces test creation effort and cost by standardizing input data formats. This allows for quick content translation, simplifying the testing process with efficient text manipulation.

AI Extract

AI Extract reduces the effort needed to test generative AI and unstructured text use cases by automating the extraction and formatting of data from API responses, improving data accuracy in systems such as CRM.

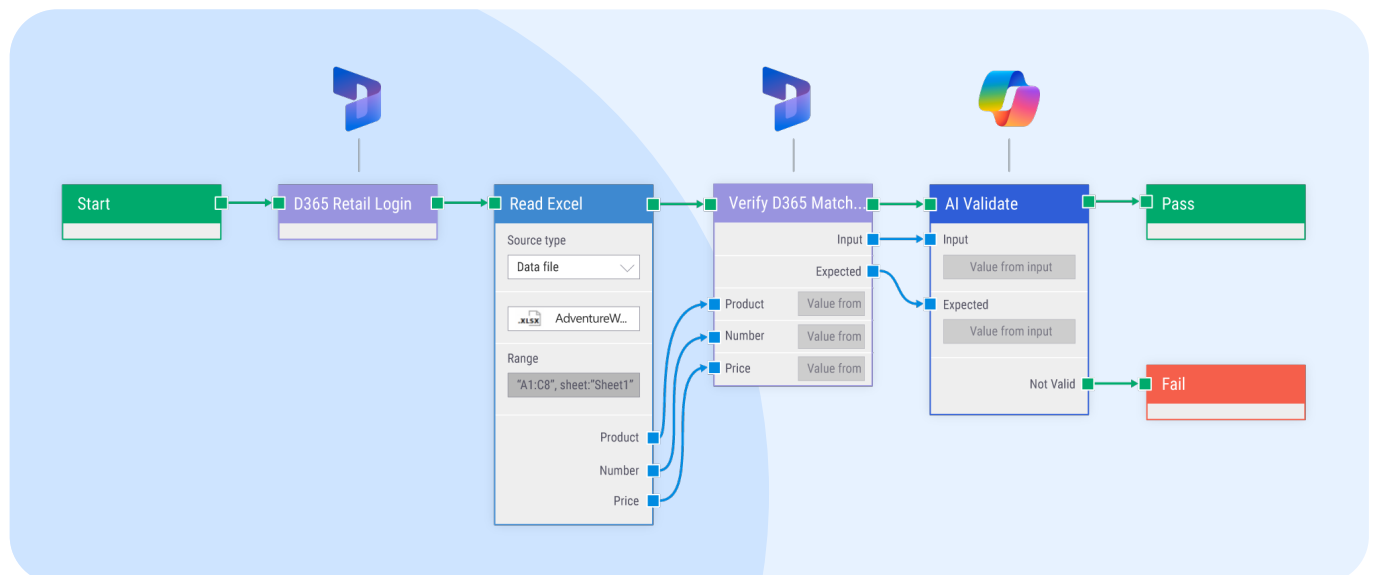
AI Generate

AI Generate creates realistic and varied datasets. This allows for comprehensive testing that accurately reflects the environments your AI systems will operate in, ensuring they perform reliably under diverse conditions.

The Leapwork Test Automation Platform

Leapwork's approach to AI-augmented testing is essential for ensuring the reliability of your AI applications. But the capabilities of Leapwork go far beyond testing AI applications. Leapwork helps you deliver better outcomes for your business by ensuring that your software meets the highest standards of quality and reliability. Leapwork's end-to-end test automation platform ensures that every aspect of the customer journey is validated, by testing across all your business applications. This comprehensive approach guarantees that your software, including newly integrated AI apps, deliver reliable and high-quality experiences consistently.

➔ [Book a demo](#) today and discover how Leapwork can drive end-to-end quality software in your business.



AI and Software Quality: Trends and Executive Insights

Report methodology

The research was conducted by Censuswide who gathered responses from 401 respondents across US and UK organizations. These included 201 C-Suite executives (CTO/CIOs) and 200 technical leads (including VP of IT, Director of IT, IT Manager, software engineering leads, QA test manager/director).

The respondents were surveyed across sectors that include technology/software, finance and banking, healthcare, manufacturing, retail, and government. The respondents were aged 18-55+. 63% of the organizations were large (501-5000 employees), whereas 37% were enterprise (5001+ employees).