AI ADOPTION AND RISK REPORT

cyberhaven labs

Q2 2024

AI ADOPTION AND RISK REPORT

Table of contents

Introduction

Since ChatGPT launched in November 2022, generative AI has emerged as one of the fastest-adopted technologies in the workplace ever. But, as seen in past paradigm shifts like cloud computing, the productivity benefits of this new technology are balanced with new risks.

Similar to the early days of cloud adoption, workers are using AI tools ahead of IT departments formally buying them. The result is "shadow AI," employee usage of AI tools through personal accounts that are not sanctioned by—or even known to—the company.

Shadow AI carries inherent risks to the confidentiality and integrity of company data because AI systems typically assimilate the information provided to them, incorporating these inputs into their broader knowledge base. Companies have already noted that the substance of sensitive information workers put into ChatGPT has been inadvertently shared with external parties.

THIS ANALYSIS FROM CYBERHAVEN LABS IS BASED ON ACTUAL AI USAGE PATTERNS OF 3 MILLION WORKERS

Understanding how material generated by AI is used within organizations is equally important because this content may be inaccurate, inadvertently infringe on copyrights or trademarks, or include security vulnerabilities, especially concerning when AI-generated code is involved.

This analysis from Cyberhaven Labs is based on actual AI usage patterns of 3 million workers, providing an unprecedented lens through which we can examine the adoption and security implications of AI in the corporate environment.

Key Findings

01

AI USAGE IN THE WORKPLACE IS ACCELERATING

The amount of corporate data workers put into AI tools **increased 485%** from March 2023 to March 2024, and it is increasing exponentially.

AI usage is dominated by products from the big 3: OpenAI, Google, and Microsoft AI products account for **96.0% of AI usage** at work.

02

A LARGE AMOUNT OF SENSITIVE COMPANY DATA IS GOING TO "SHADOW AI" ACCOUNTS

82.8% of legal documents employees put into AI tools are going to non-corporate accounts, potentially exposing the information publicly.

Roughly half of source code **(50.8%)**, R&D materials **(55.3%)**, and HR and employees records **(49.0%)** put into AI are sent to non-corporate accounts.

03

END USERS ARE OUTPACING CORPORATE IT, FUELING USE OF RISKY "SHADOW AI" ACCOUNTS

73.8% of ChatGPT usage at work is through non-corporate accounts that, unlike enterprise versions, incorporate whatever you share into their public models.

The percentage of non-corporate accounts is even higher for Gemini **(94.4%)** and Bard **(95.9%)** usage.

04

BROADENING USAGE OF AI ENCOMPASSES MORE TYPES OF SENSITIVE COMPANY DATA

AI adoption is reaching new departments and use cases involving sensitive data: **27.4% of data** employees put into AI tools is sensitive, up from **10.7% a year ago**.

The top sensitive data type put into AI tools is customer support **(16.3% of sensitive data)**, which includes confidential information customers share in support tickets.
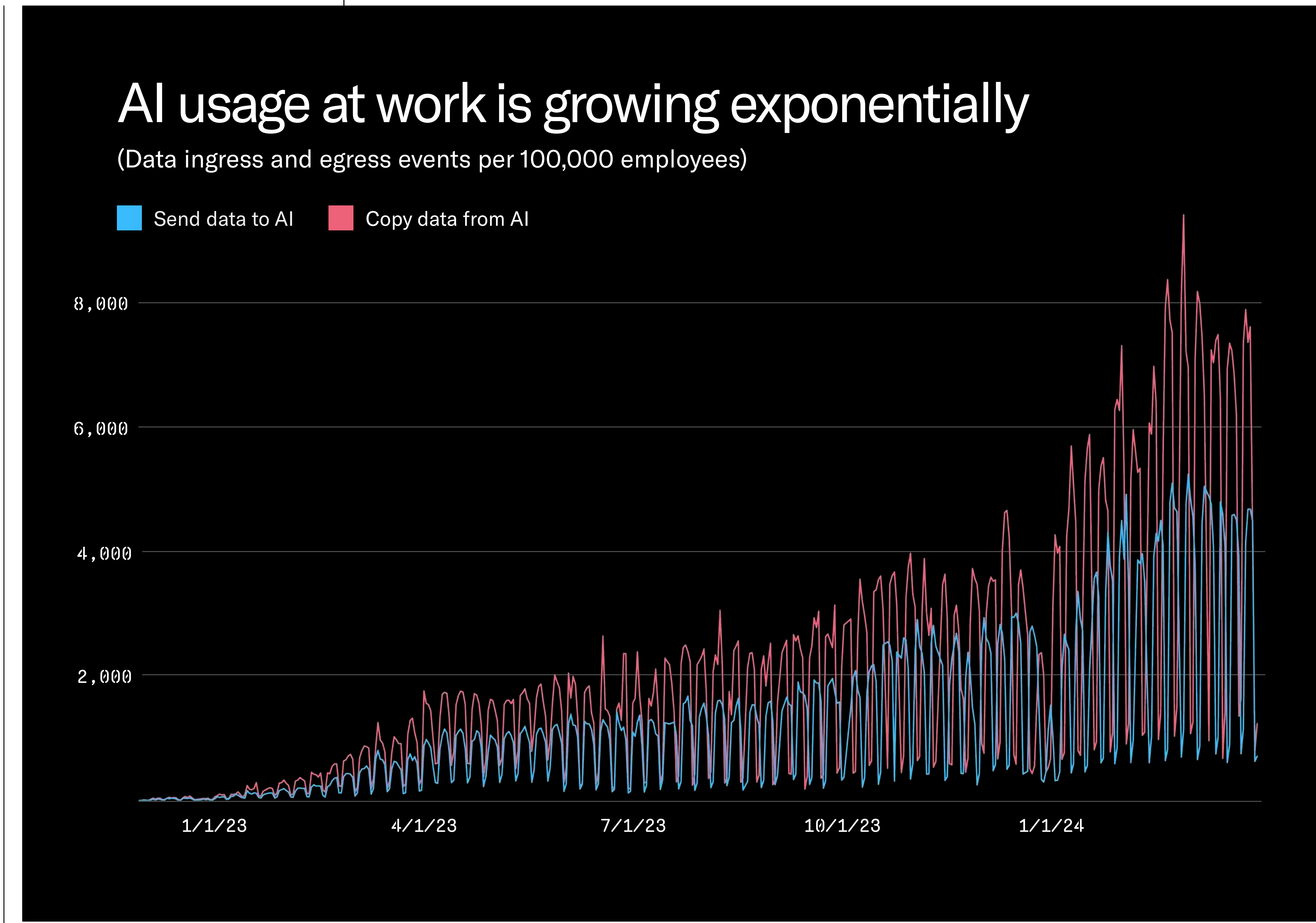
05

AI-GENERATED CONTENT IS BEING USED IN POTENTIALLY RISKY WAYS

3.4% of research and development materials created in March 2024 originated from AI tools, potentially creating risk if patented material was introduced.

3.2% of source code insertions are generated by AI outside of coding tools (which have enterprise approved coding copilots), potentially creating the risk of vulnerabilities.

Growth in AI usage and the top 25 AI tools

AI usage at work is increasing at an exponential rate and so is the amount of corporate data pasted or uploaded to AI tools. Between March 2023 and March 2024, the amount of corporate data workers put into AI tools increased by 485 percent. AI adoption has increased so much so fast that the average knowledge worker now puts more corporate data into AI tools *on a Saturday or Sunday* than they did during the middle of the work week a year ago.

AI usage at work is growing exponentially

(Data ingress and egress events per 100,000 employees)

■ Send data to AI   ■ Copy data from AI

Growth in AI usage
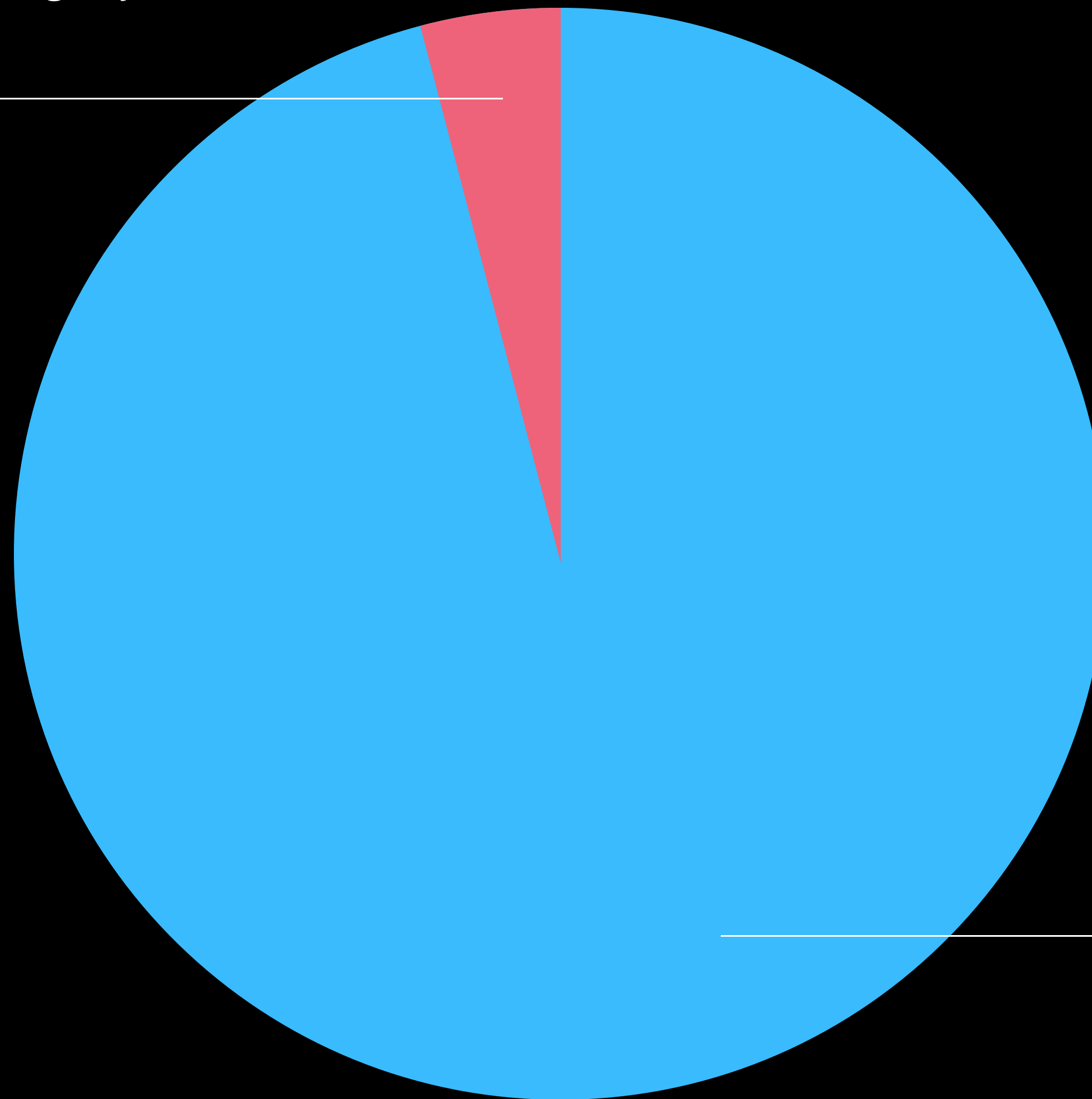# and the top 25 AI tools (cont.d)

ChatGPT, which ignited the AI boom in November of 2022, remains the most frequently used AI application in the workplace. A "big three" set of vendors has emerged over the past year to dominate the AI landscape: collectively, products from OpenAI, Google, and Microsoft account for 96.0% of AI usage at work. But we're still in the early stages of the AI revolution, and numerous startups also feature prominently on the top 25 AI tools list.

AI usage at work is dominated by products from the "big three" AI vendors

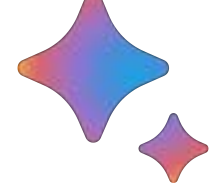(Percentage of AI usage by workers)

Everyone else

4,0%

OpenAI, Microsoft, Google

96%

The top 25 AI tools in the workplace

1	ChatGPT	
2	Bard	
3	Gemini	
4	Poe	
5	Perplexity	
6	Midjourney	
7	Microsoft Copilot	
8	Claude	
9	Wordtune	
10	Azure OpenAI Studio	
11	Otter.ai	
12	DALL-E	
13	Hugging Face	

14	Eleven Labs	
15	PDF.ai	
16	Writesonic	
17	DeepAI	
18	Jasper AI	
19	Copy.ai	
20	NaturalReader	
21	Phind	
22	Playcode	
23	Moveworks	
24	PopAi	
25	Synthesia	

AI adoption by industry

The industry with the greatest AI adoption is technology. Across every dimension, workers at tech firms use AI more than their counterparts in other fields. In March 2024, 23.6% of tech workers put corporate data into an AI tool. They also put data into AI tools more frequently (2,007,102 events per 100,000 employees) and copy data from AI tools more frequently (1,664,821 events per 100,000 employees) than workers in any other industry.

How many employees use AI tools
(By percentage of employees)

■ Send data to AI
■ Copy data from AI

Industry	Send data to AI	Copy data from AI
Finacial services	4.7%	4.1%
Healthcare	2.3%	2.2%
Manufacturing	0.6%	0.6%
Media and entertainment	5.2%	7.5%
Pharmaceuticals	2.8%	1.2%
Professional services	3.8%	4.4%
Retail	0.5%	1.1%
Technology	23.6%	11.2%
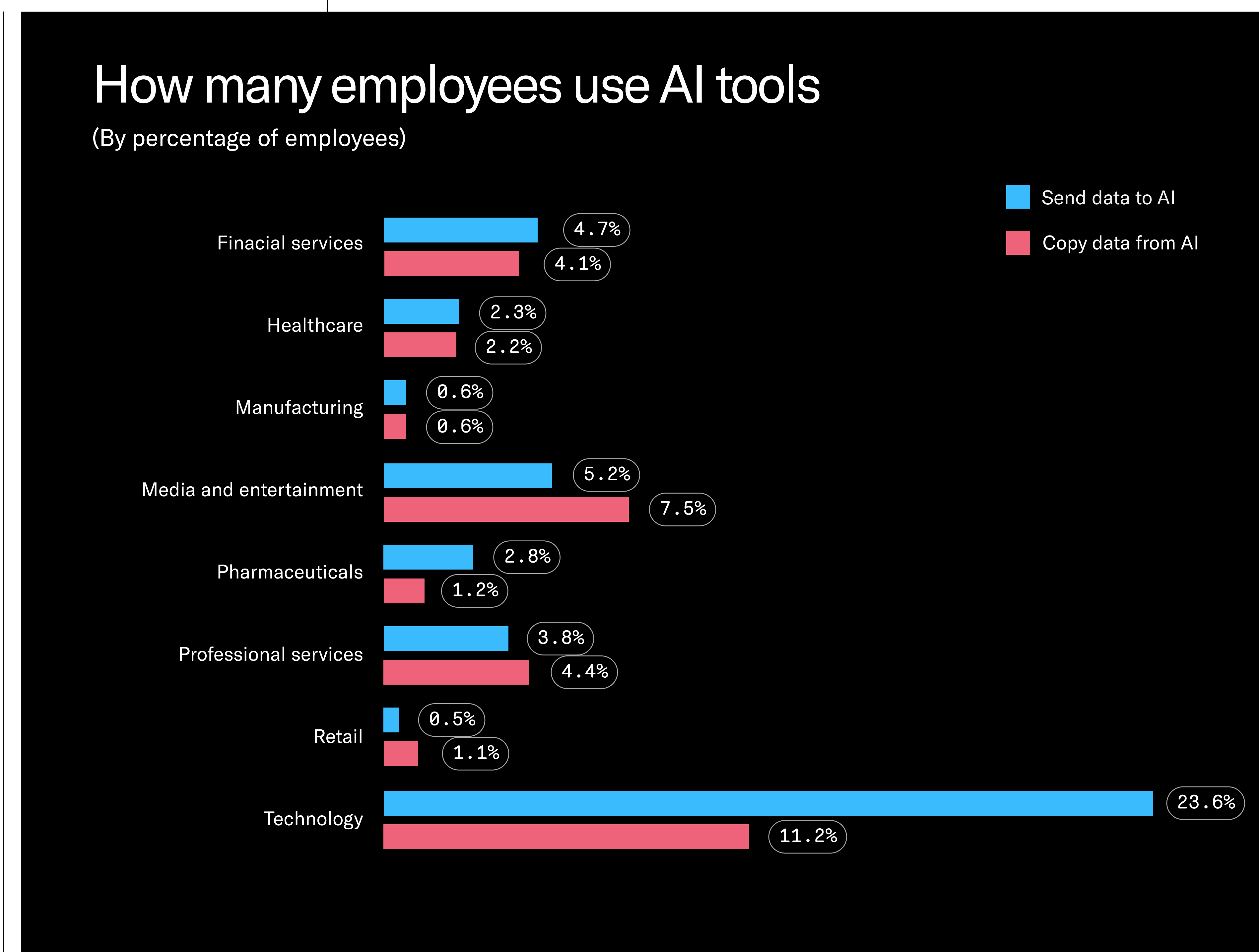
We're still early in the adoption curve with 5.2% of employees at media and entertainment firms, 4.7% of employees at financial firms, and 2.8% in pharmaceuticals and life sciences are putting company data in AI tools. Retail and manufacturing employees have much lower adoption rates of AI technology than other industries. Just 0.6% of workers at manufacturing firms and 0.5% of workers at retail firms put company data in AI in March 2024.

AI adoption by industry (cont.d)

Notably, workers in media and entertainment copy 261.2% more data from AI tools than they put into them. We found that in media, a growing amount of news content is generated using AI. We found a similar pattern in professional services (e.g. law firms, consulting firms, etc.) where workers copy 73.3% more data from AI tools than they put into them. Management consultants in particular are using AI to prepare presentation slides and other client-facing materials.

Usage volume of AI tools

(Events per 100,000 employees per month)

■ Send data to AI
■ Copy data from AI

Industry	Send data to AI	Copy data from AI
Finacial services	821,423	611,977
Healthcare	112,020	136,735
Manufacturing	28,026	58,615
Media and entertainment	442,487	1,598,366
Pharmaceuticals	212,425	65,789
Professional services	797,806	1,382,829
Retail	11,075	27,980
Technology	2,007,102	1,664,821

"Shadow AI" is growing faster than IT can keep up

"Shadow AI" is AI usage among workers that the company's IT and security function does not sanction or in some cases even know about.

OpenAI launched an enterprise version of ChatGPT in August 2023 with enhanced security safeguards and a guarantee that information from corporate customers won't be used to train models used by the general public (which can leak the substance of the training data). According to OpenAI, 80% of Fortune 500 companies have teams using ChatGPT with corporate accounts. However, we found a significant majority are still personal accounts.

Workplace AI usage by account type

(Percentage of AI usage by workers)

Tool	Personal	Corporate
ChatGPT	73.8%	26.2%
Gemini	94.4%	5.6%
Bard	95.9%	4.1%
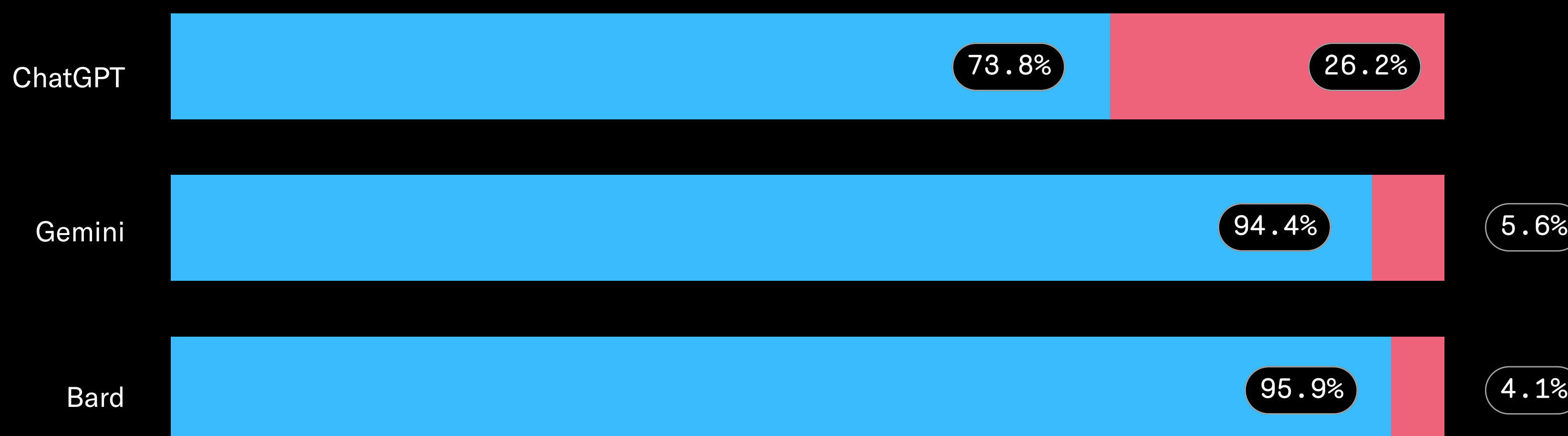
In the workplace, 73.8% of ChatGPT accounts are non-corporate ones that lack the security and privacy controls of ChatGPT Enterprise. Moreover, the percentage is even higher for Gemini (94.4%) and Bard (95.9%). As companies adopt enterprise versions of popular tools and for key use cases and migrate users to corporate accounts, more AI startups launch and end users are adopting new AI tools faster than IT can keep up, fueling continued growth in "shadow AI".

What corporate data workers put into AI tools

AI adoption is reaching new departments and use cases involving data that shouldn't be exposed outside the company. While the overall volume of corporate data bound of AI has increased significantly over the past year, the percentage of that data that is sensitive has risen significantly as well. In March 2024, 27.4% of corporate data employees put into AI tools was sensitive, up from 10.7% a year ago. The variety of this data has also grown.
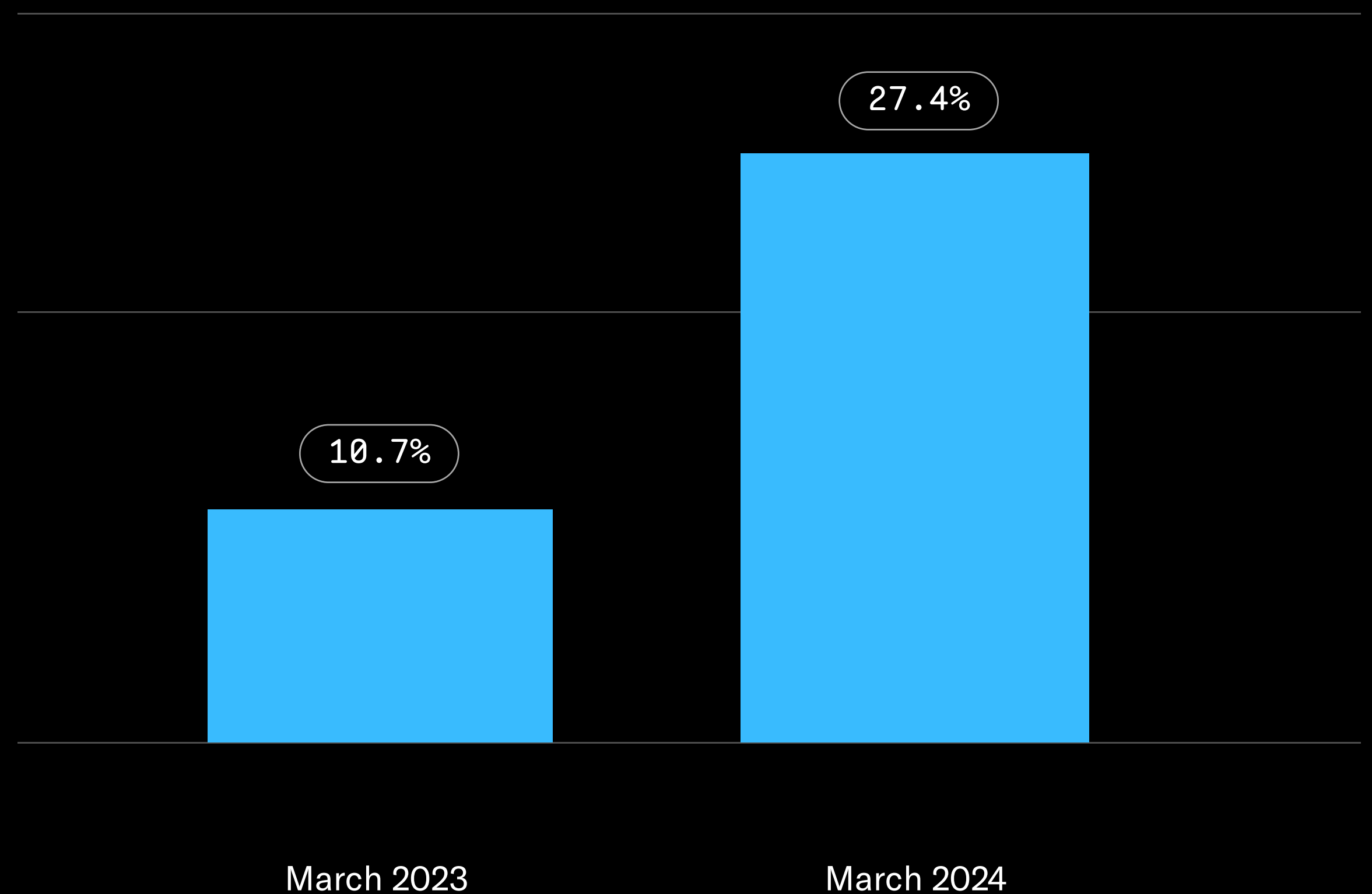
The top sensitive data type put into AI tools is customer support (16.3% of sensitive data), which includes confidential information customers share in support tickets. Source code comprises another 12.7% of sensitive data bound for AI. It's important to note this number does not include coding copilots built into developer environments; it represents source code a developer copies and pastes into a third-party AI tool, which is riskier than corporate sanctioned copilots.

Research and development material is 10.8% of sensitive data flowing to AI. This category includes results from R&D tests that in the wrong hands would provide insights to competitors working on competing products. Once new products get closer to launch, marketing teams begin to work on promotional materials. Unreleased marketing material forms 6.6% of data, and confidential internal communications also forms 6.6% of data.

HR and employee records are 3.9% of sensitive information going to AI, including confidential HR details such as employee compensation and medical issues. Financial documents are just 2.3% of sensitive data going into AI tools but this material can be especially damaging if it leaks since it includes unreleased financials for public companies.

Amount of corporate data sent to AI that is sensitive

(By volume of data)

March 2023	March 2024
10.7%	27.4%

Sensitive corporate data sent to AI by type

(By volume of data)

- Customer Support 16.3%
- Source Code 12.7%
- Research & Development 10.8
- Sales & Marketing 6.6%
- Internal Communications 6.6%
- HR & Employees Records 3.9%
- Legal 2.4%
- Financial Documents 2.3%
- Other 38.2%
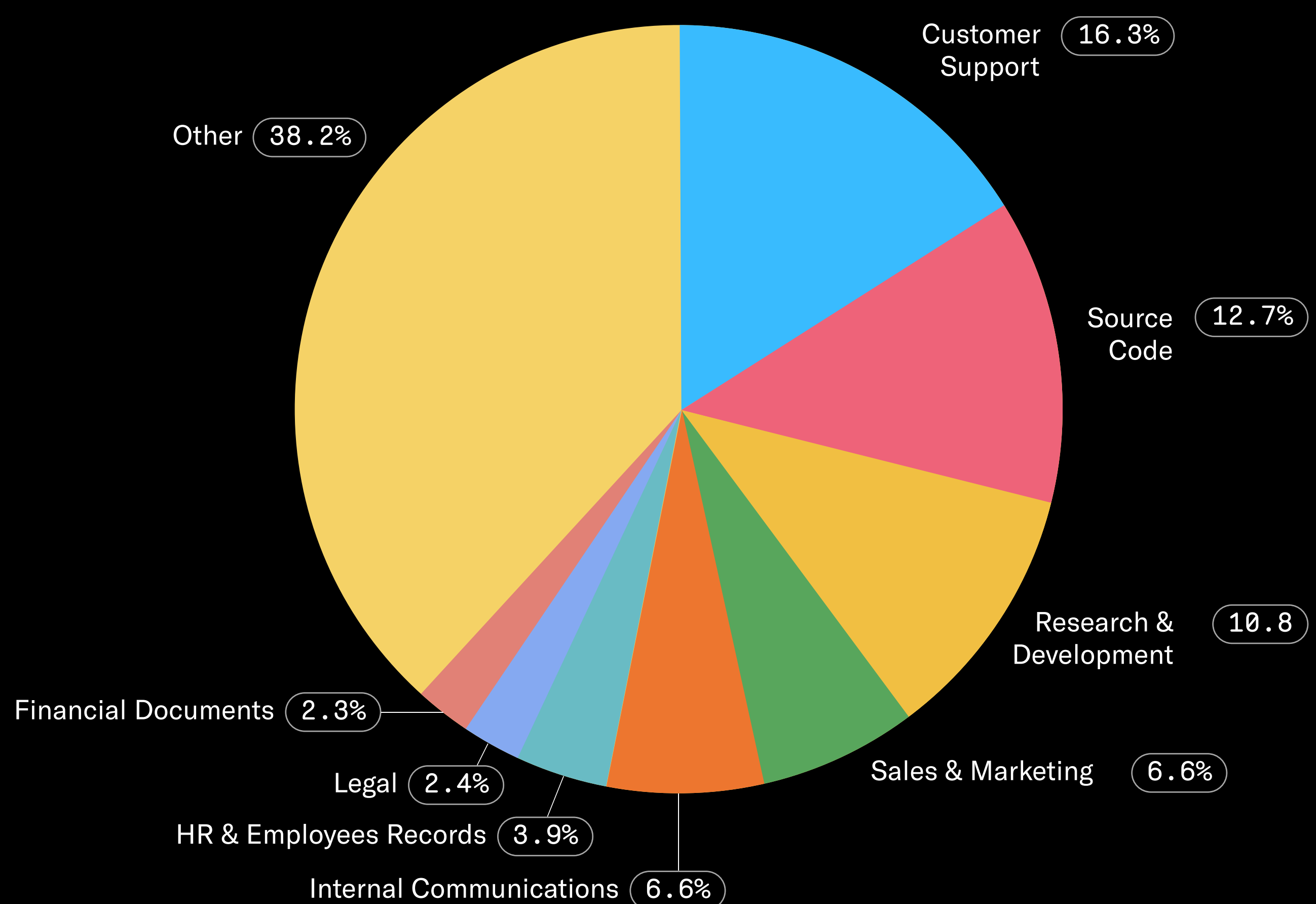
Sensitive data flows to "shadow AI"

Just how much sensitive data is going to these non-corporate AI accounts?

While legal documents comprise only 2.4% of sensitive corporate data going to AI tools, the vast majority (82.8%) are put into risky "shadow AI" accounts. Legal documents include drafts of acquisition or merger contracts for proposed deals, settlement agreements with confidential deal terms, and legal materials protected by attorney-client privilege.

Roughly half of source code (50.8%), research and development materials (55.3%), and HR and employees records (49.%) put into AI are going to non-corporate accounts.

Interestingly, we found that the most common form of sensitive company data going to AI tools is also most likely to be used with approved, corporate AI accounts. Just 8.5% of customer support information put into AI tools is going to non-corporate accounts.
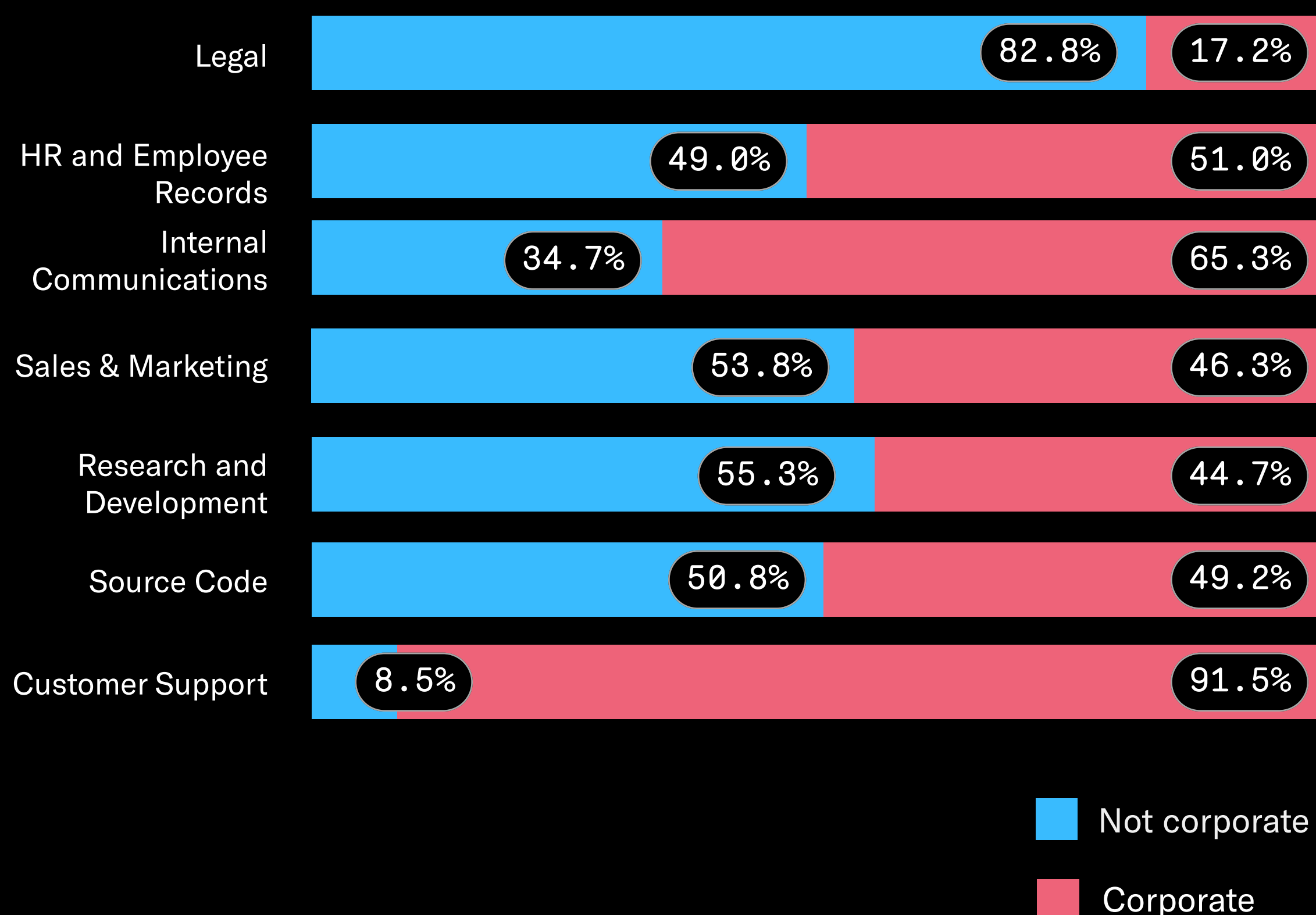
We found that 98.0% of the data entered into AI tools is pasted directly into interfaces such as chat windows, rather than being uploaded as discrete files. The prevalence of copy/paste highlights another security consideration: data protection mechanisms that focus on file origin or classification tags applied to files are not effective at protecting most data destined for AI.

Destination for sensitive data by AI account type

(By volume of data)

Account type	Not corporate	Corporate
Legal	82.8%	17.2%
HR and Employee Records	49.0%	51.0%
Internal Communications	34.7%	65.3%
Sales & Marketing	53.8%	46.3%
Research and Development	55.3%	44.7%
Source Code	50.8%	49.2%
Customer Support	8.5%	91.5%

■ Not corporate
■ Corporate

How data is inserted into AI tools

(By volume of data)

Upload 2.0%

Copy/Paste 98.0%
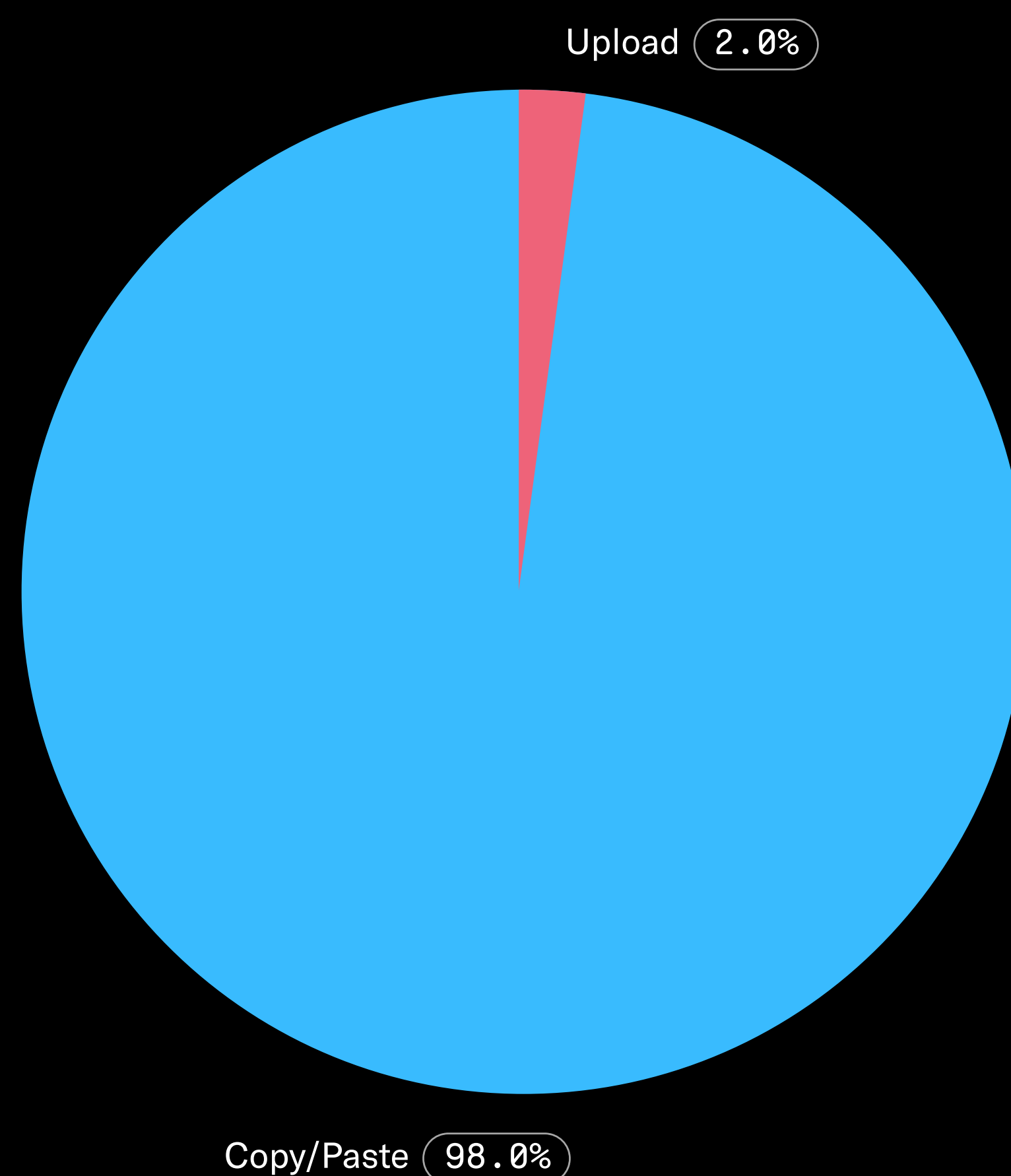
How AI-generated content is used at work

So far, this report has covered corporate data going to AI tools. It's equally important to look at where AI-generated material is being used within organizations—not only from a security and legal standpoint but also to understand the impact AI is having on different types of work. We measured for every category of data at a company, how much is being created using AI.

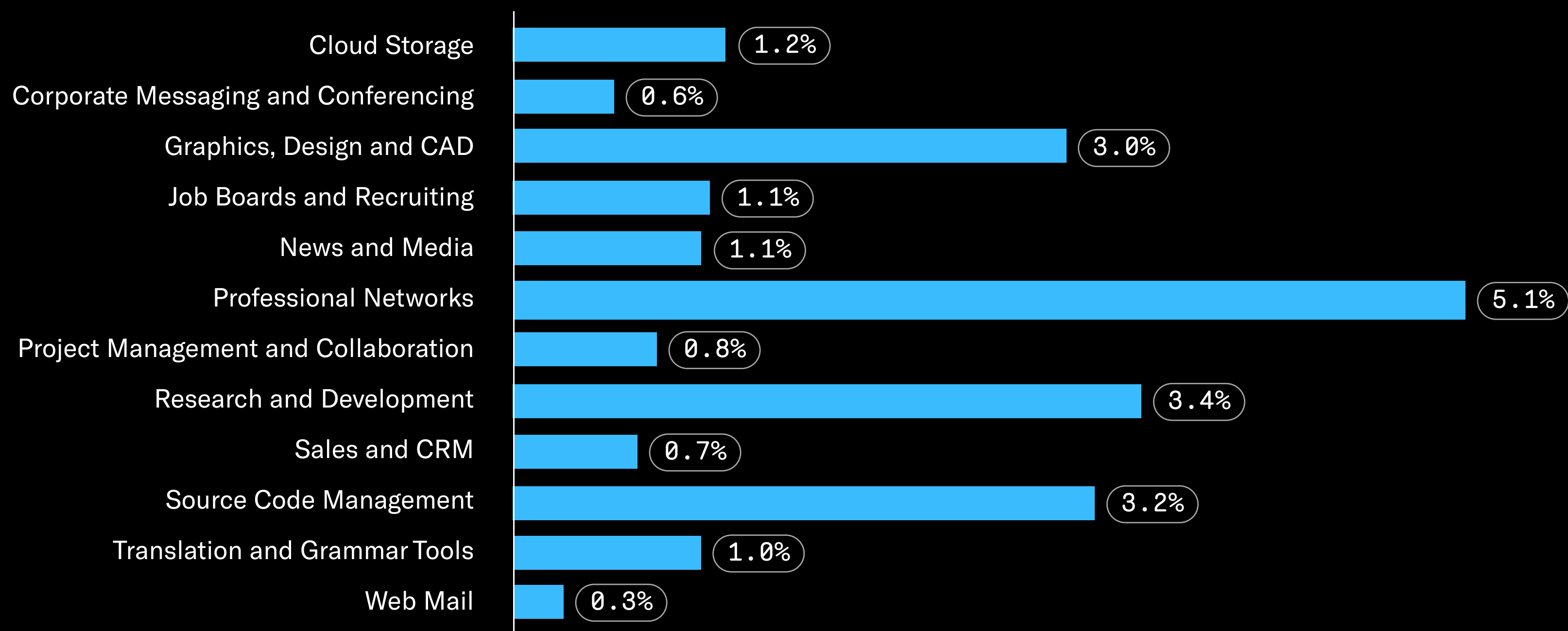From its origin in an AI tool, we traced where AI-generated content went within companies and found that the destination with the highest percentage of AI content is professional networks at 5.1%. In other words, roughly one in 20 LinkedIn posts employees create while at work are made using AI. However, other material from AI is being used in more business-critical ways.

How much of new data employees create originated in AI

(By volume of data)

Category	Percentage
Cloud Storage	1.2%
Corporate Messaging and Conferencing	0.6%
Graphics, Design and CAD	3.0%
Job Boards and Recruiting	1.1%
News and Media	1.1%
Professional Networks	5.1%
Project Management and Collaboration	0.8%
Research and Development	3.4%
Sales and CRM	0.7%
Source Code Management	3.2%
Translation and Grammar Tools	1.0%
Web Mail	0.3%

In March 2024, 3.4% of research and development materials were produced using AI, potentially creating risk if patented or inaccurate material was introduced.

How AI-generated
content is used at work (cont.d)

That was followed by 3.2% of source code insertions. Like in prior parts of this report this number only covers third-party AI tools not coding copilots built into development environments. The actual amount of source code being created with AI tools is likely much higher. But third-party AI tools often don't have the same controls to prevent malicious code.

We also found that 3.0% of graphics and design content originated from AI, which can be a problem since AI tools can output trademarked material.

A smaller but still meaningful amount of news and media material (1.1%), which includes written news articles, is made using AI. We found the same percentage for job boards and recruiting material, which includes online job descriptions, originating in AI tools.

cyberhaven

Cyberhaven is the AI-powered data security company revolutionizing how companies detect and stop the most critical insider threats to their most important data. Until now, data security products relied on manual rulesets and pre-defined policies that looked for keywords and specific user actions. Our AI technology analyzes billions of workflows to understand every piece of data within an organization, when it's at risk, and what's needed to protect it. It's like nothing that's come before and protects data like nothing else.

To learn more, visit **cyberhaven.com**