



**7<sup>TH</sup> ANNUAL**

STATE OF

# Data Science

REPORT

**AI and Open Source at Work**



# Executive Summary

For the seventh consecutive year, Anaconda conducted our State of Data Science survey to surface insights, use cases, and trends across the data science, AI, and open-source community.

This year saw some intriguing developments among data scientists and innovators. A large majority are using AI techniques in their data science projects, such as generative AI, deep learning, and transformer models. As a result, fewer people are worried about AI overtaking their jobs, instead using it to help complement simple and laborious tasks while they focus on more innovative results.

Among the obstacles data scientists face are a feeling of unreadiness about rising challenges, such as learning new AI tools and government regulations and a lack of confidence in speaking up about security concerns. The data science community can start addressing both by encouraging and supporting learning, open dialogue, and collaboration internally and within the larger data science ecosystem.



# Table of Contents

**04**

Methodology

**05**

Demographic Overview

**08**

Open Source and AI Bring Value — And Challenges

**11**

The Community Must Collaborate for Innovation and Security

**14**

Work Styles of Data Scientists and IT Administrators

**17**

Supporting Education and Training

**19**

Key Takeaways



## Methodology

3,096 individuals representing 136 countries participated in our online survey, which was conducted from June to September 2024. Respondents came from the Anaconda email database, Anaconda.com, social media, and other sources, and respondents were invited to participate in a sweepstakes drawing as an incentive for completing the survey. Five winners were selected at random after the survey was complete. Respondents were divided into four tracks: IT workers, data

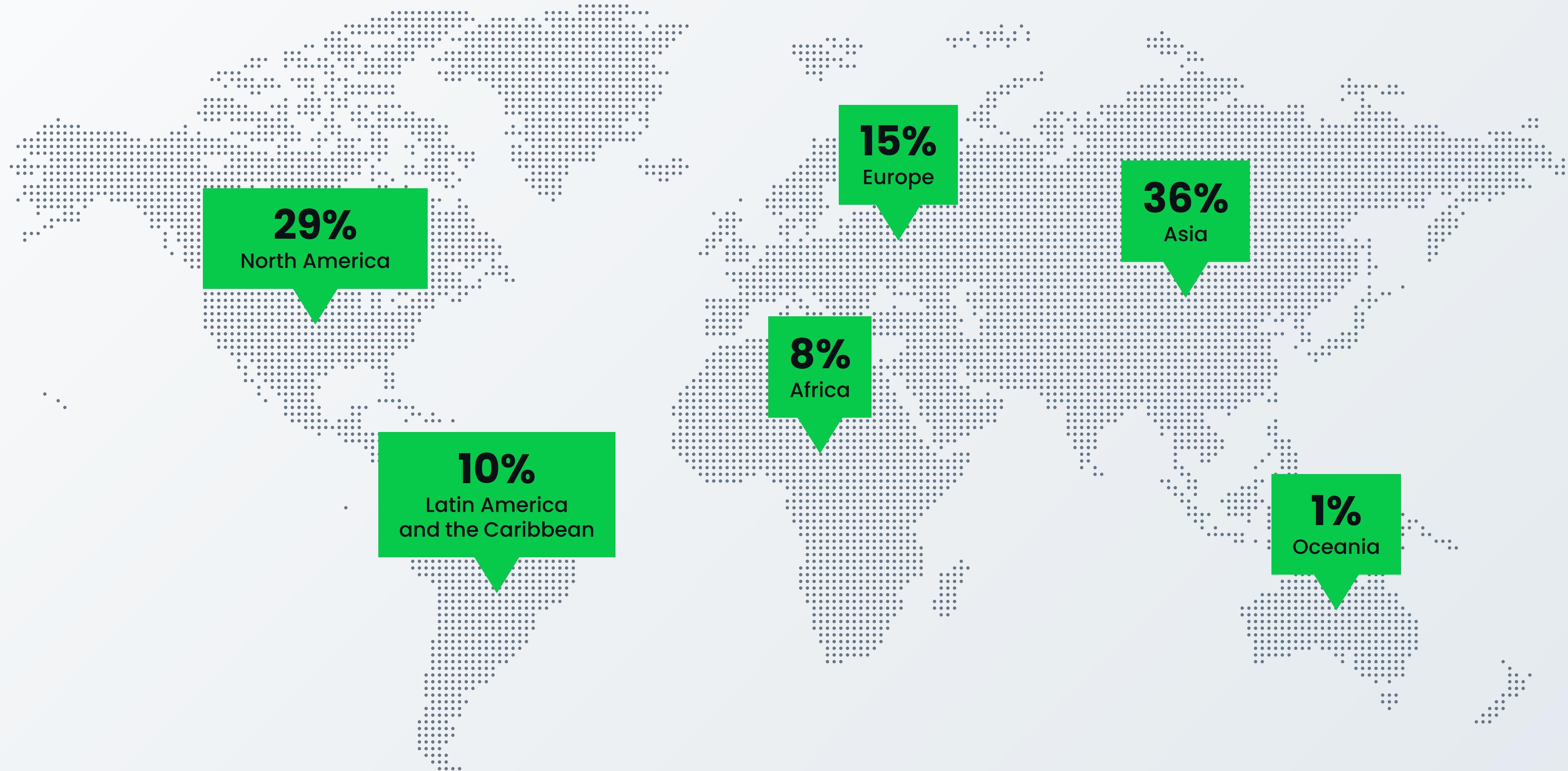
science practitioners, students, and researchers or university professors. All respondents were asked the same demographic questions, but some were unique to each track. In the report, we indicate whether responses come from the entirety of the respondents or a specific subset. All responses are self-reported.

Note: All percentages are rounded to the nearest whole percent. Due to rounding, some numbers may not equal 100.

# Demographic Overview

DEMOGRAPHIC

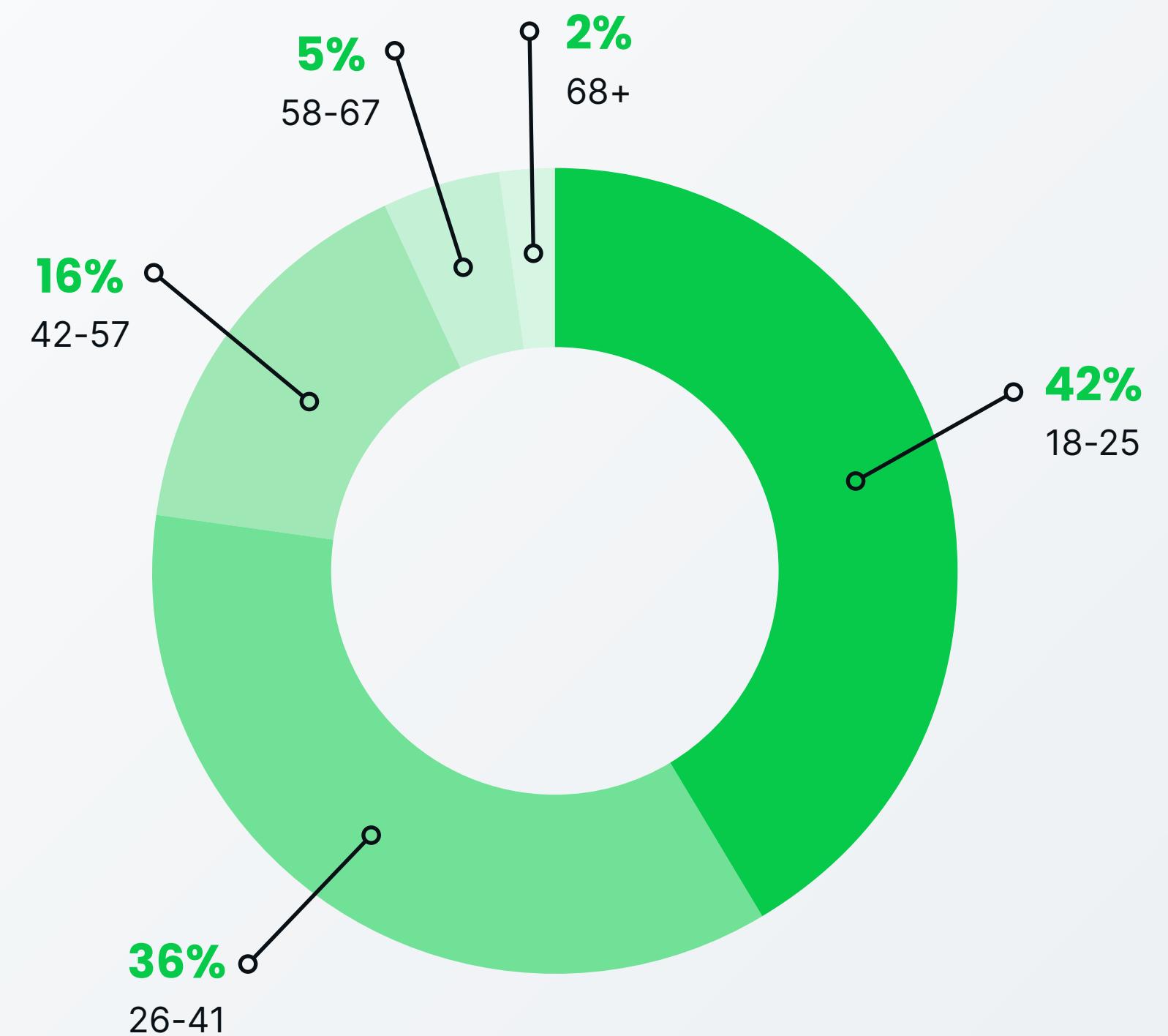
Country



## DEMOGRAPHIC OVERVIEW

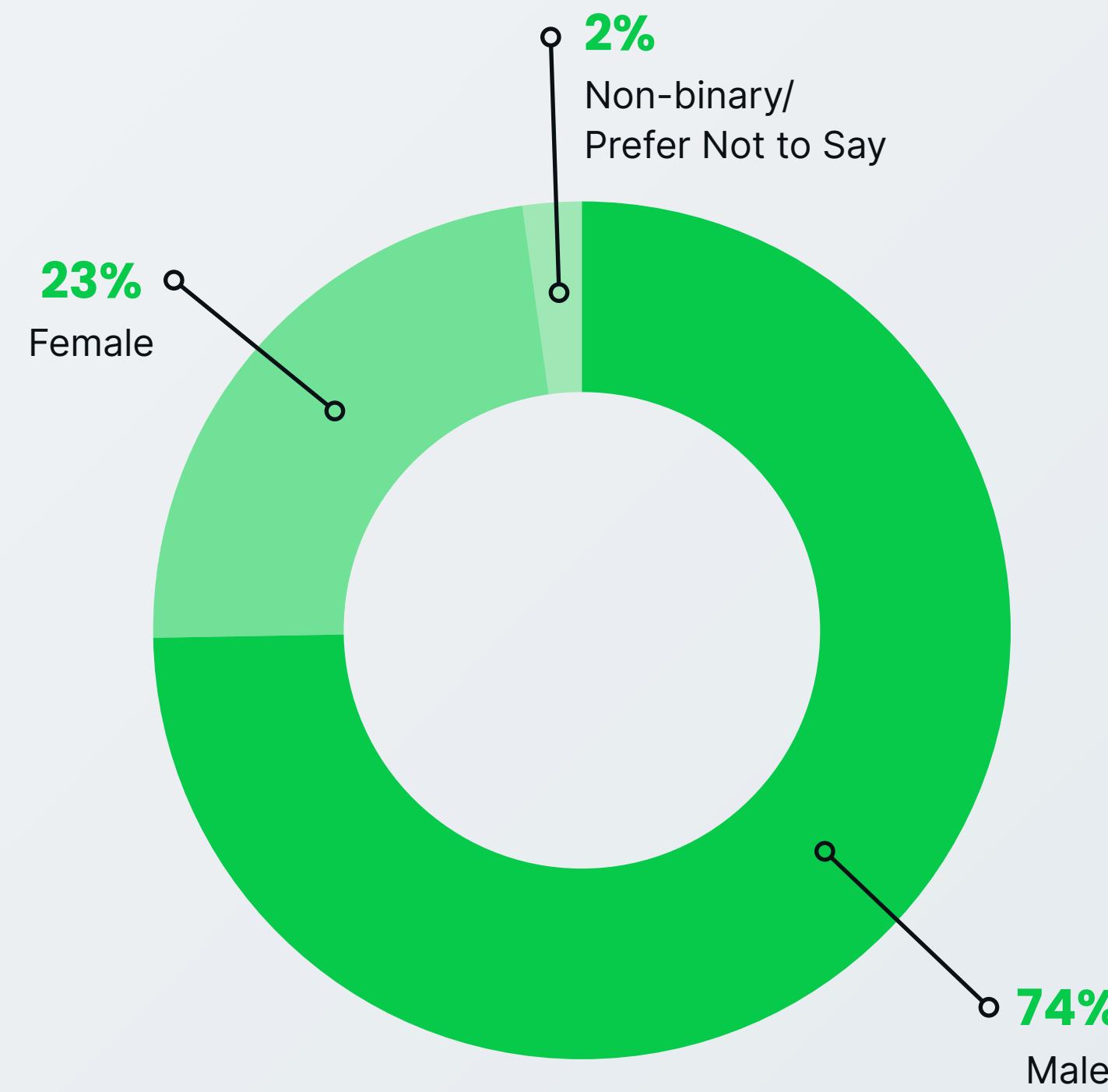
### DEMOGRAPHIC

#### Age



### DEMOGRAPHIC

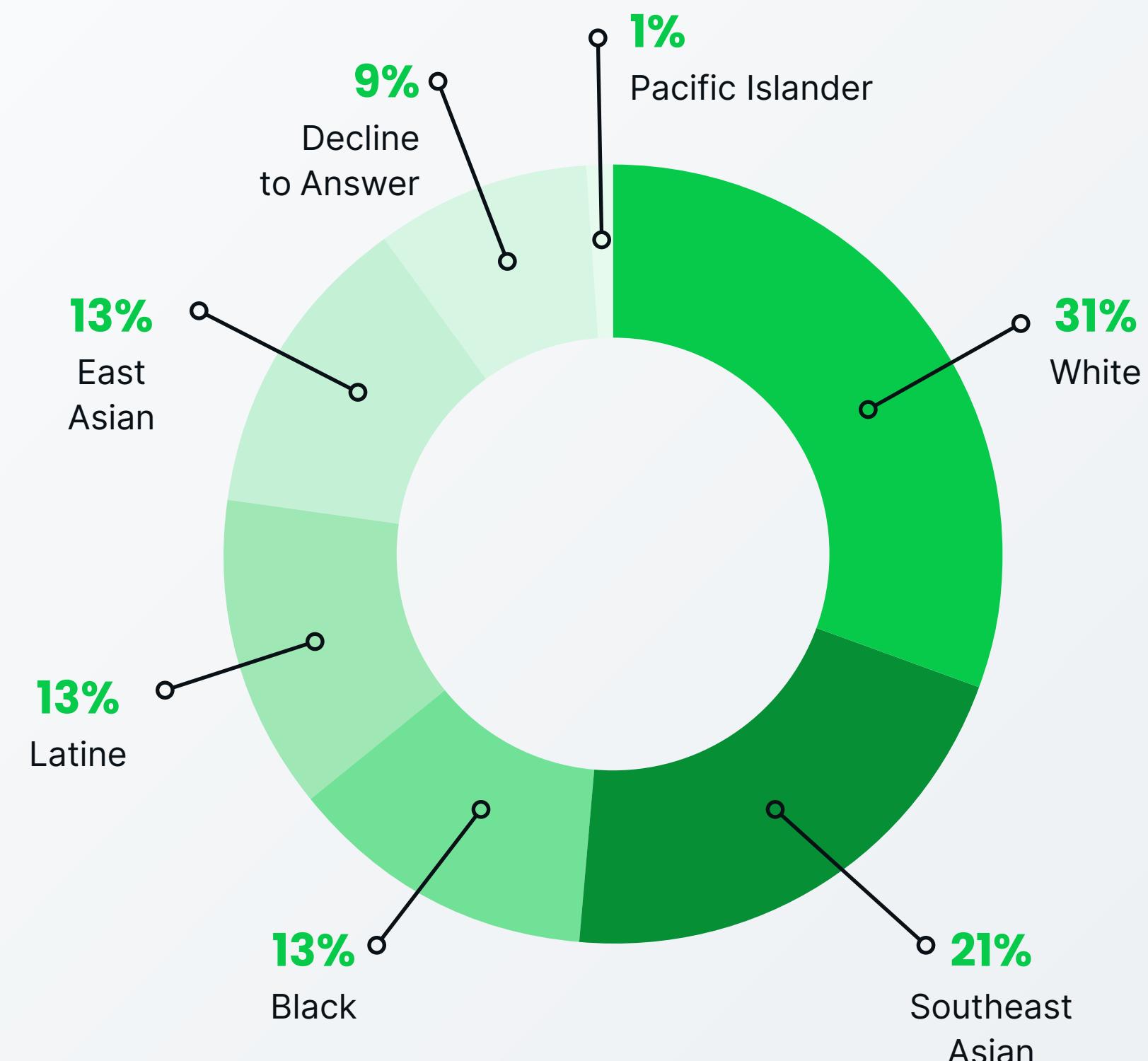
#### Gender



## DEMOGRAPHIC OVERVIEW

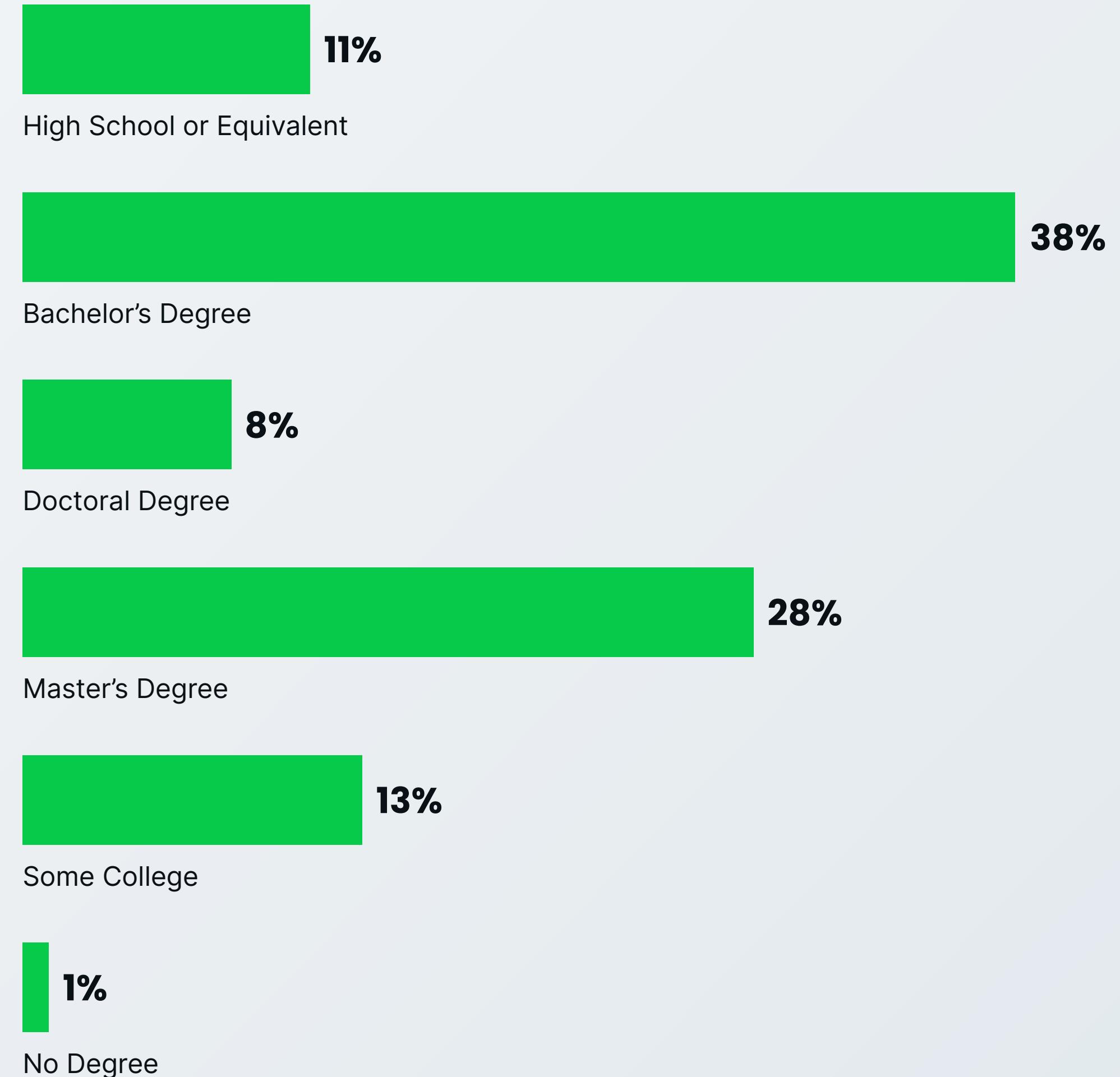
### DEMOGRAPHIC

#### Race



### DEMOGRAPHIC

#### Education



# Open Source and AI Bring Value – And Challenges

“

When sushi first got introduced here, people said, ‘Ew, raw oysters, raw fish?’ Now, you can get it down at the grocery store. Yet you can’t just eat any random raw fish – you’re going to have a bad time. Open source is good for consumption, but you’ve got to care about the supply chain differently than when you buy something like canned beans.”



**Peter Wang**

Chief AI and Innovation Officer and Co-Founder, Anaconda



# Open Source and AI Bring Value – And Challenges

**Python continues to serve as the foundation of data science, with 93% of respondents using it in some capacity.** The simplicity and flexibility of Python — plus its expansive libraries and collaborative tools — make it a key pillar in AI and machine learning. Data science practitioners find value in open source, particularly its agility and flexibility. We asked respondents to rank their top values of open-source software; “Most economical option (31%),” “Speed of innovation (30%),” and “Most useful tools for my needs (19%)” were the top selections.

Concerns remain around open-source software security, which was cited as the biggest technical challenge for AI adoption and usage today (42%). **About one in four respondents (26%) said their companies have an interest in AI but don't have the support or budget to drive business value.** Open source and AI have never been more accessible,

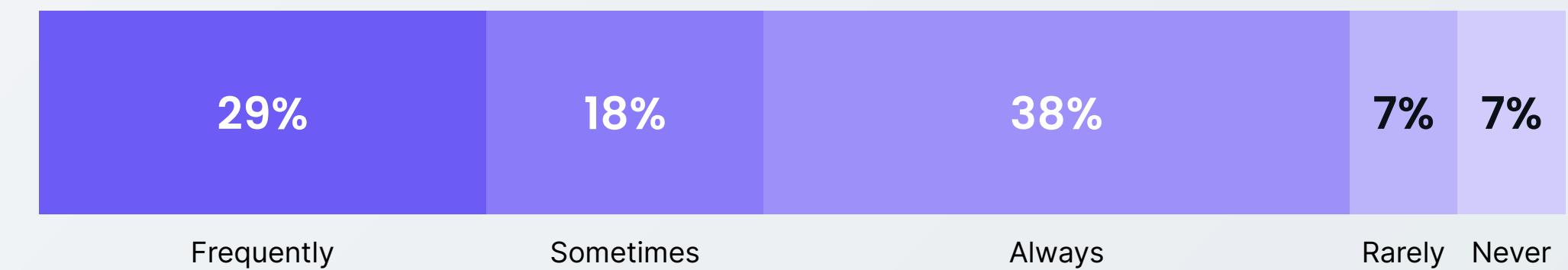
which is exciting but also a cause for concern. Peter Wang, Chief AI and Innovation Officer and Co-founder of Anaconda, believes companies can find plenty of value in open source, though they must be diligent and realistic in defining their goals and how they go about setting them.

**43% of respondents feel unprepared or unsure about taking on rising challenges in the data science space,** such as government regulations, technological tools, and the increase in AI usage across roles. Additionally, 38% are concerned with ensuring transparency and explainability of AI models, while 36% are concerned with addressing bias and fairness in AI algorithms.

As data practitioners and companies increase their use of third-party tools, there's a critical need for a distribution platform provider like Anaconda that delivers well-vetted, secure, and trusted packages for all users.

## GENERAL SENTIMENT

How often do you use Python?



## GENERAL SENTIMENT

What do you value most of open-source software?



70%

of respondents ranked “most useful, most economical, and speed” as a top 3 value of open-source software. This clearly states the value that users see.

## OPEN SOURCE AND AI BRING VALUE — AND CHALLENGES

### GENERAL SENTIMENT

Where do you see the biggest technical challenges for AI adoption/usage today?

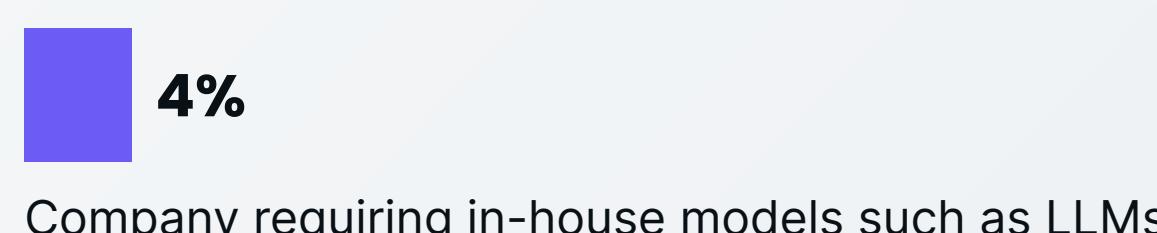
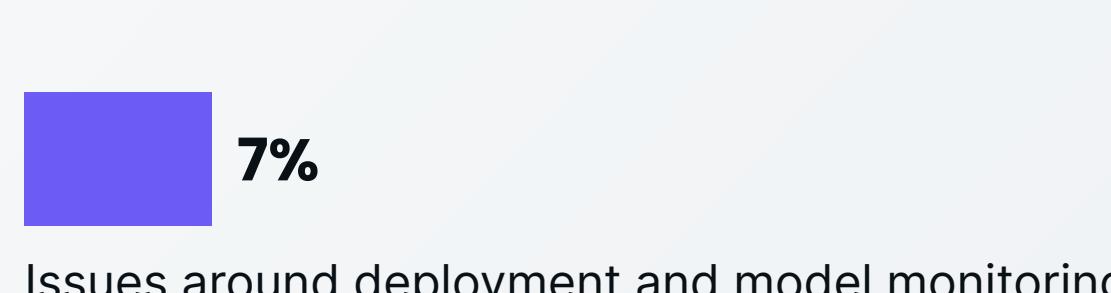
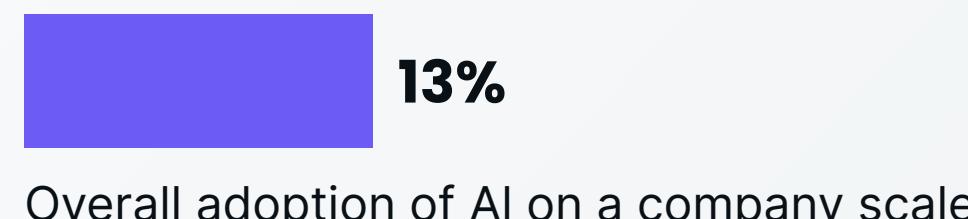
Chart shows % of top concern



Security concerns around open-source software

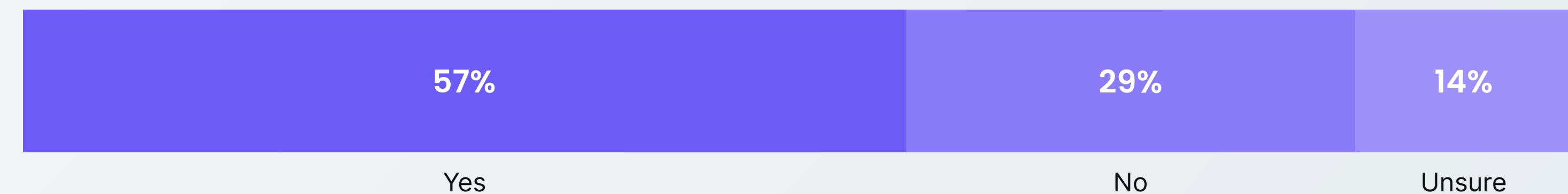


Higher interest but less support/budget for driving business value



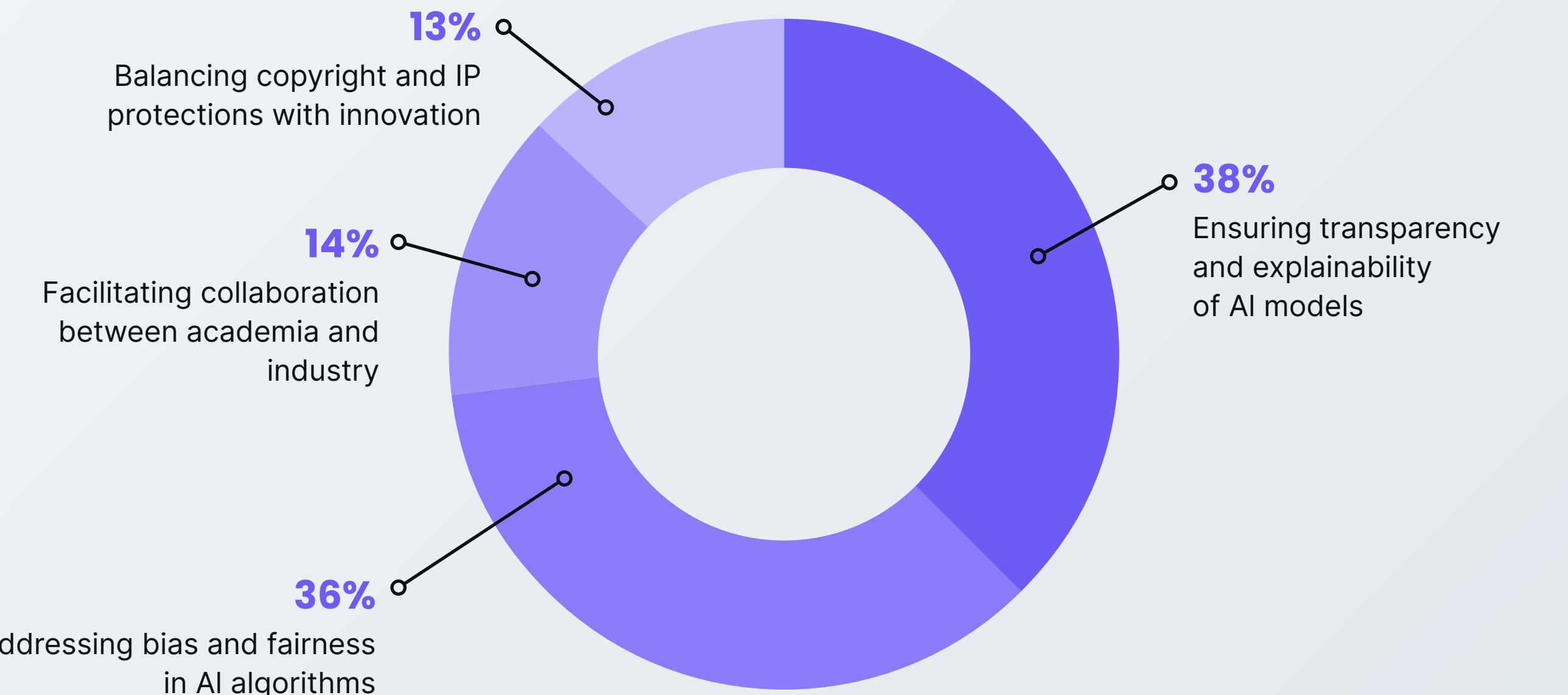
### GENERAL SENTIMENT

Do you feel prepared to take on the rising challenges in the data science space related to government regulations, new technological tools, and the rise in AI?



### GENERAL SENTIMENT

What are your biggest concerns regarding regulations for AI?



# The Community Must Collaborate for Innovation and Security

“

Having established processes internally with a really strong sense of what ‘good’ looks like is very important. If you don’t have an internal way to evaluate the quality of the response, it’s going to be difficult for you to apply AI to it effectively. So much about applying AI to any problem is understanding how you iterate the system to get an increasingly better-quality answer.”



**Greg Jennings**

VP of Engineering for AI, Anaconda



# The Community Must Collaborate for Innovation and Security

Another value of AI and open-source software lies in the power of community. **Half of data science practitioners said they're using open source in their work pipeline**, while 66% of IT administrators say their companies are using open-source software.

Among data science practitioners, 58% report their companies are building new tools with AI, with 56% building AI models for internal use. **AI is becoming a more significant part of work, as 87% of data science practitioners are spending as much or more time on AI techniques this year compared to last year.** Almost one-third of respondents (31%) said they spent "significantly more time" on these AI techniques. PwC calls generative AI the "missing link" for data, so companies are poised for impressive results if they embrace AI with a concerted focus on what they're trying to accomplish.

That additional focus on implementing AI has led to less concern about AI overtaking humans. Only 22% of IT administrators and 21% of data science practitioners admitted to being fearful that AI would take their jobs, which are

steep decreases from last year. **Instead, most respondents are finding ways AI can help support their daily tasks while they focus on more complex, innovative, and interesting projects.** Here's where AI is fitting in:

## DATA SCIENCE PRACTITIONERS

### What are you using AI for?

Select all that apply.

**67%**

Data cleaning, visualization, analysis

**52%**

Automating tasks

**49%**

Prediction or detection models (e.g., churn prediction)

**44%**

LLMs (chatbots)

**38%**

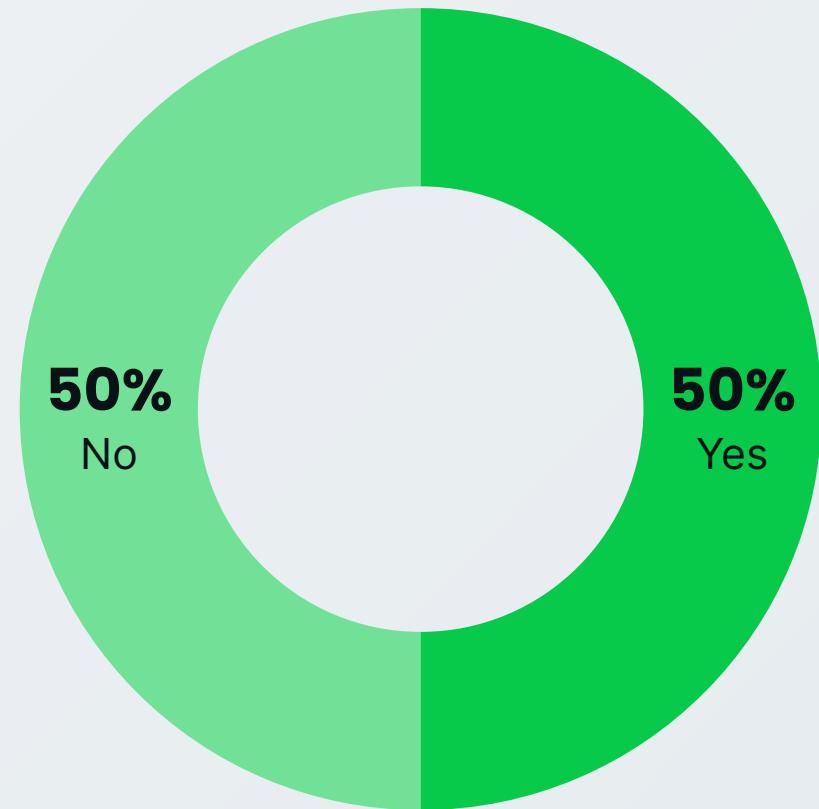
Content creation (e.g., images, video, written word)

AI handling the above helps save time and energy — **88% and 83% of data practitioners ranked data preparation and data cleaning, respectively, as their two most time-consuming tasks.** These AI use cases are being incorporated across teams and functions, showcasing value throughout an organization.

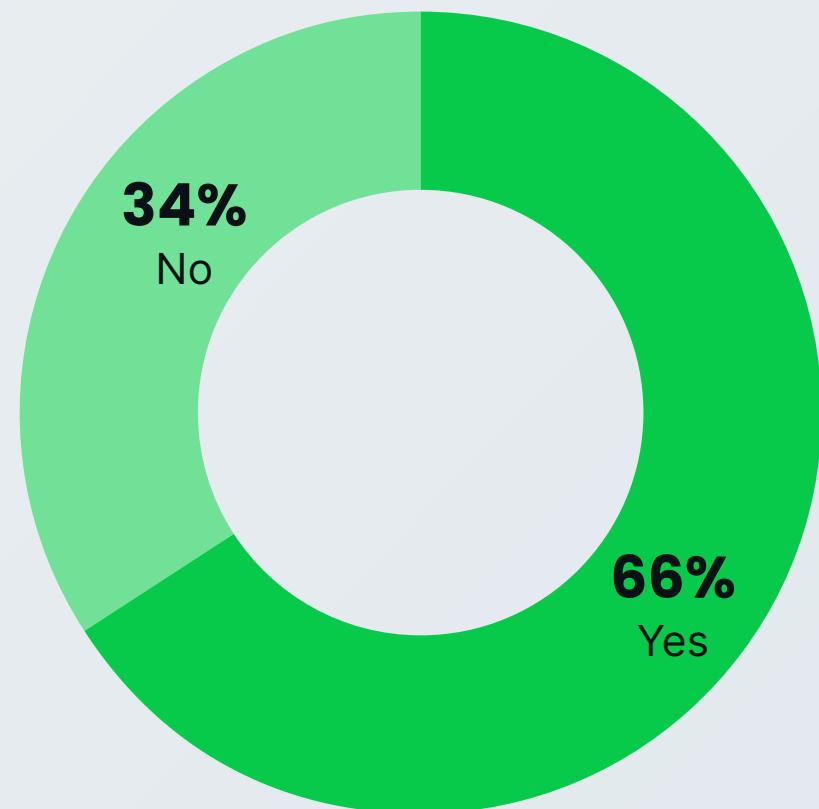
AI and open source function best when collaboration is involved. However, there's room for improvement: 34% of IT administrators don't feel empowered to voice their concerns about the potential security risks of using AI and open-source software. **As these tools and platforms proliferate, employees at all levels must be willing to speak about their thoughts and experiences — and companies must create a culture that welcomes open feedback.**

We see knowledge sharing and collaboration every day at Anaconda through our community of 45 million users and more than one billion package downloads. Users, partners, and suppliers are exchanging expertise and putting others in positions to succeed and innovate. These collaborative ecosystems are what will drive AI and open source forward.

**DATA SCIENCE PRACTITIONERS**  
Do you use open-source software in your current work pipeline?



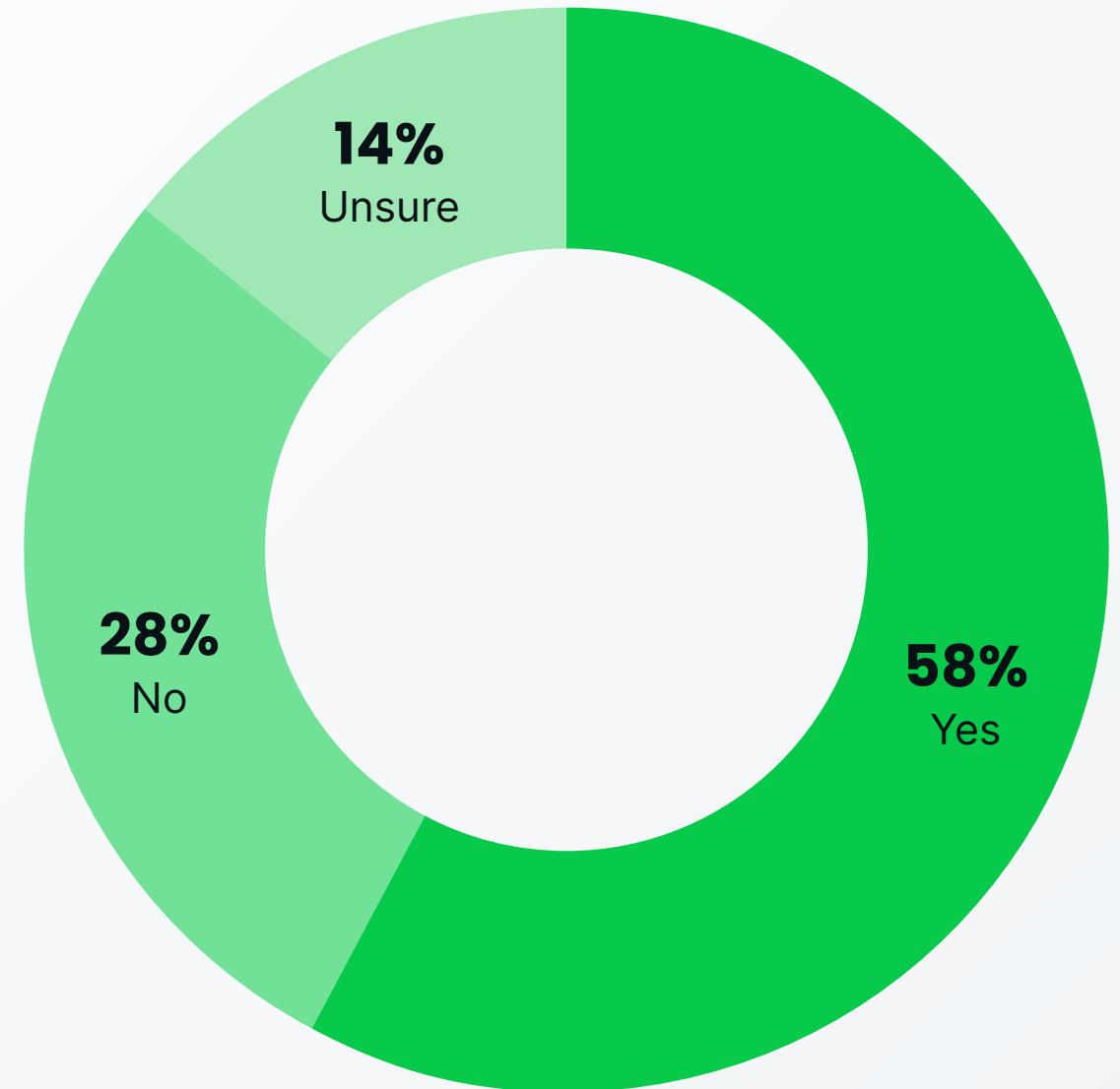
**IT ADMINISTRATORS**  
Does your company use open-source software?



## THE COMMUNITY MUST COLLABORATE FOR INNOVATION AND SECURITY

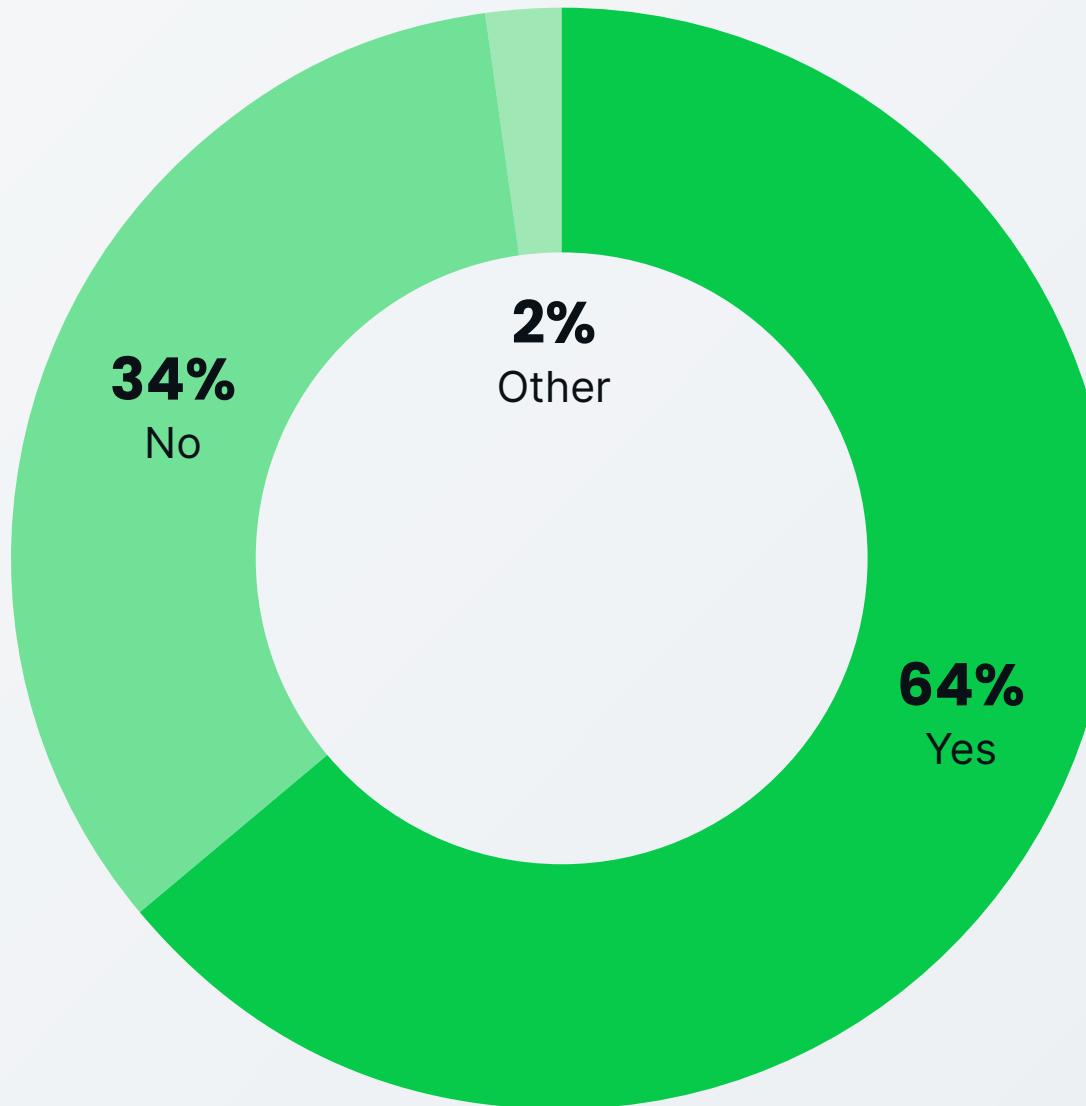
### DATA SCIENCE PRACTITIONERS

Is your company building new tools with AI?



### IT ADMINISTRATORS

At your current company, do you feel empowered to voice concerns?



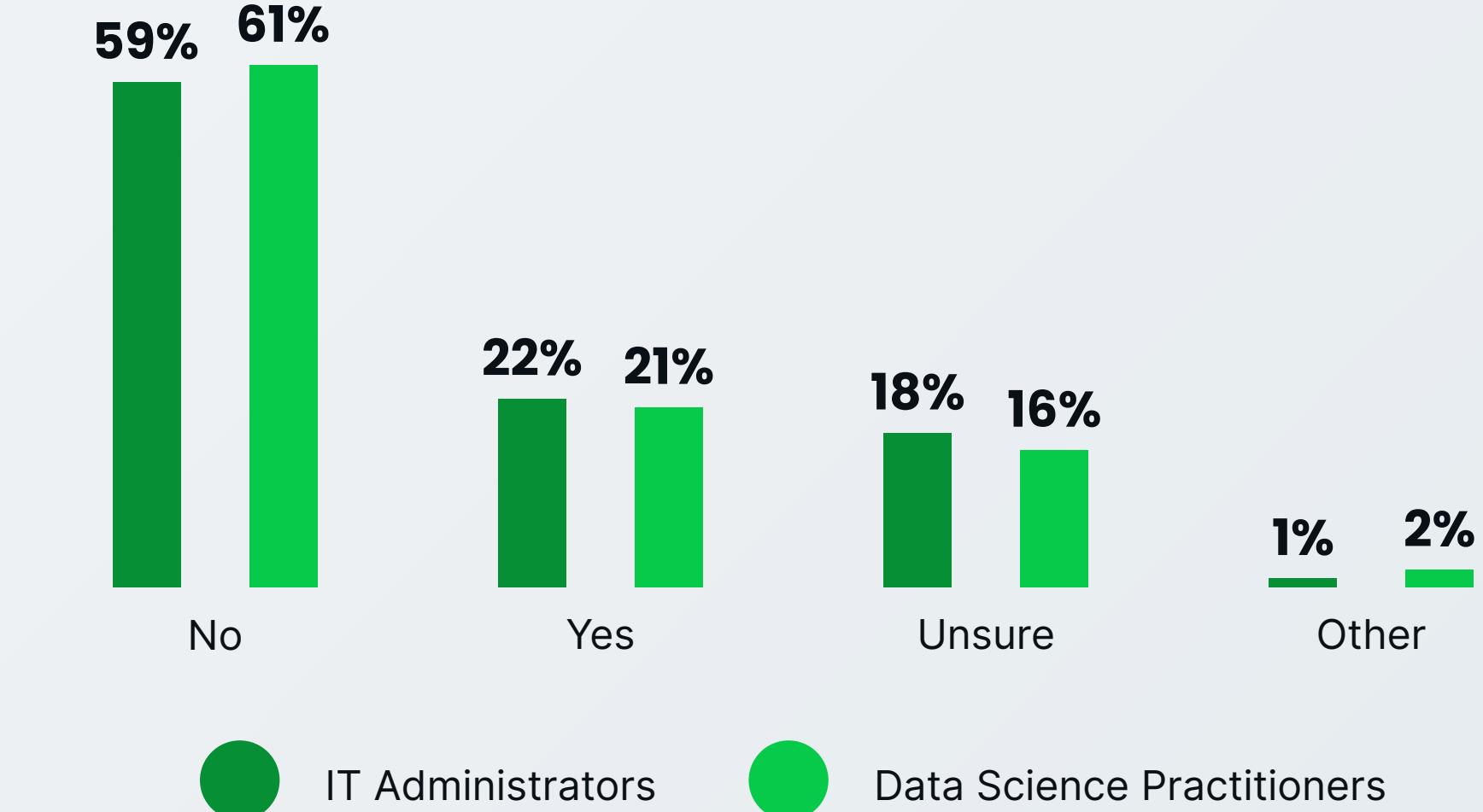
### DATA SCIENCE PRACTITIONERS

Compared to last year, how much time have you spent this year on AI techniques such as generative adversarial networks (GANs), deep learning (DL), transformer models, etc.?



### IT ADMINISTRATORS / DATA SCIENCE PRACTITIONERS

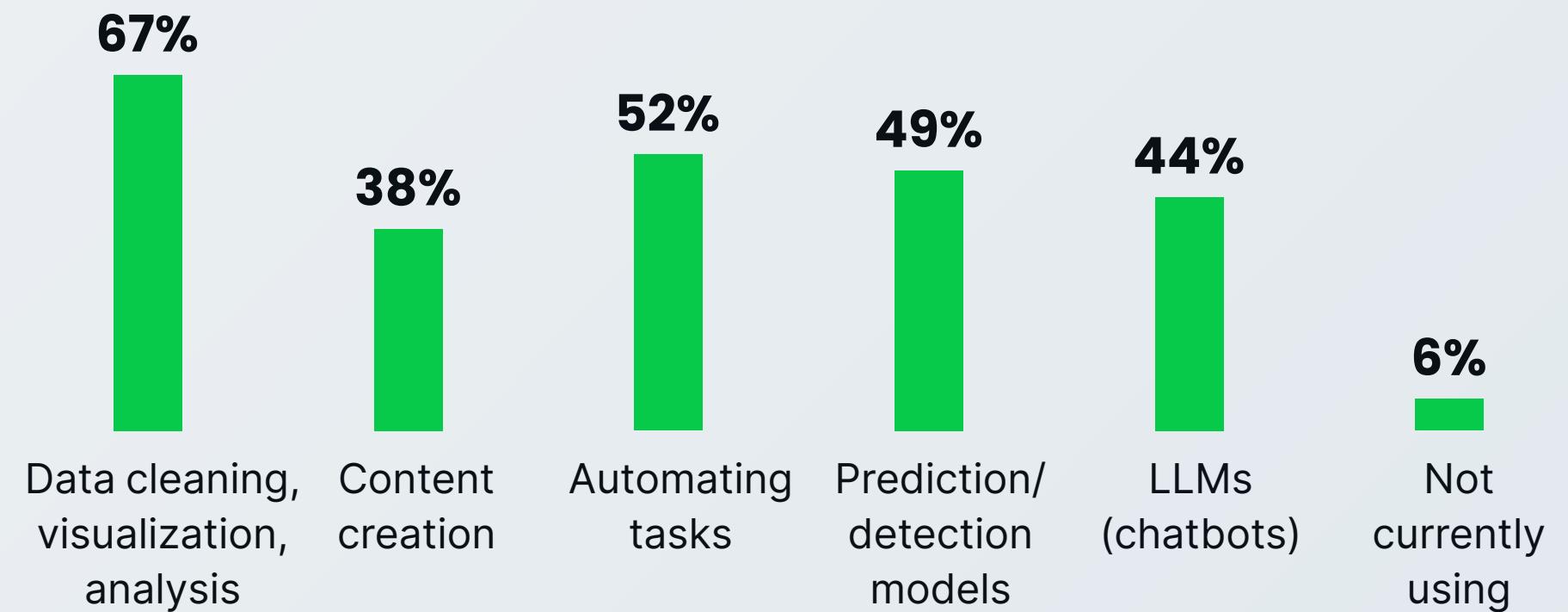
Do you feel your job is threatened by the rise of generative AI tools?



### DATA SCIENCE PRACTITIONERS

What are you using AI for?

Select all that apply.



# How Data Science Practitioners and IT Administrators are Working

“

Small models are getting really good, really fast. Organizations are figuring out how to run things efficiently. That means a few different things. If I have a big data center and a lot of compute, I can run that model cheaper. It also means I can get the inference out faster, and I can run those models on local hardware more effectively.”



**Greg Jennings**

VP of Engineering for AI, Anaconda



# How Data Science Practitioners and IT Administrators are Working

Data science practitioners and IT administrators have their individual preferences when it comes to developing new products and services, with local laptops leading the way. **Only 19% of data science model development and training is happening in the cloud; local laptops (38%) and local desktops and workstations (21%) are more common working environments.**

We're supporting this desire through AI Navigator and Anaconda Code, which both enable developers, engineers, and IT staff to work locally.

However, **deploying models into production is more evenly split, with 31% using the cloud compared to 27% on-prem.** Companies that move data science or AI models from development into production may experience certain roadblocks. Here are the most common:

## DATA SCIENCE PRACTITIONERS

**28%**

A skills gap within the organization

**27%**

Securing data connectivity

**26%**

Meeting IT/InfoSec standards

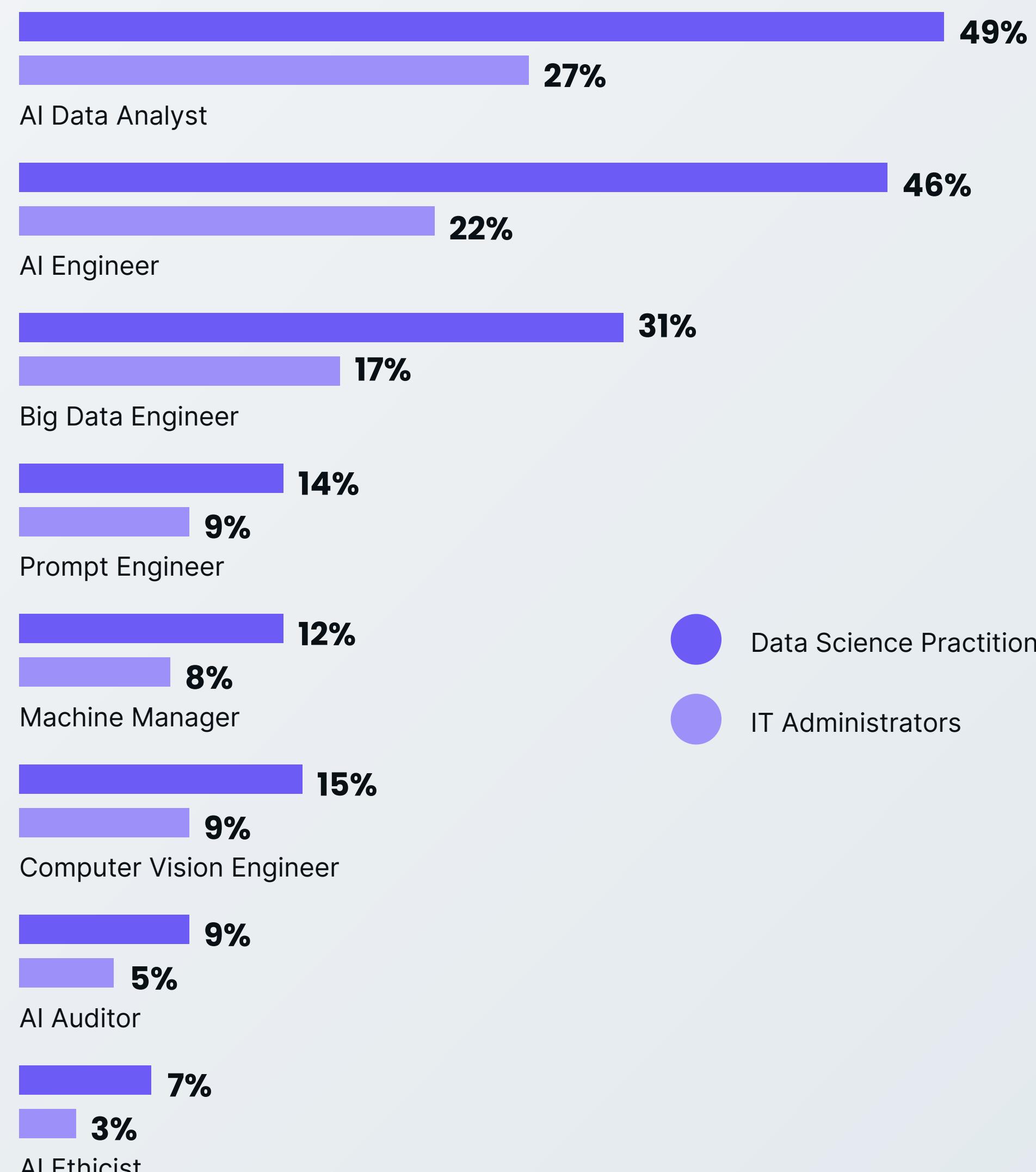
Some companies are addressing the increase in AI usage by hiring new, tailored roles, preparing for the future in an AI-driven world.

**The most frequent of these targeted hires are AI data analysts and AI engineers.** Among data science practitioner respondents, AI data analysts (49%) and AI engineers (46%) are the most common, followed by big data engineers (32%), computer vision engineers (15%), prompt engineers (14%), and machine managers (12%). IT administrator respondents reported AI data analysts (27%) are the most common roles, followed by AI engineers (22%), big data engineers (17%), and computer vision engineers and prompt engineers (9% for both roles).

Roles like AI auditors and AI ethicists are less common, but we expect to see them grow in the coming years, since responsible AI and the ethical use of data and models are a rising area of focus. The [US Bureau of Labor Statistics](#) projects 36% growth in data scientist careers by 2033. Companies that invest in these areas and then support those growing teams — or collaborate with communities and networks who are doing so — will be poised for long-term success across the AI landscape.

## DATA SCIENCE PRACTITIONERS / IT ADMINISTRATORS

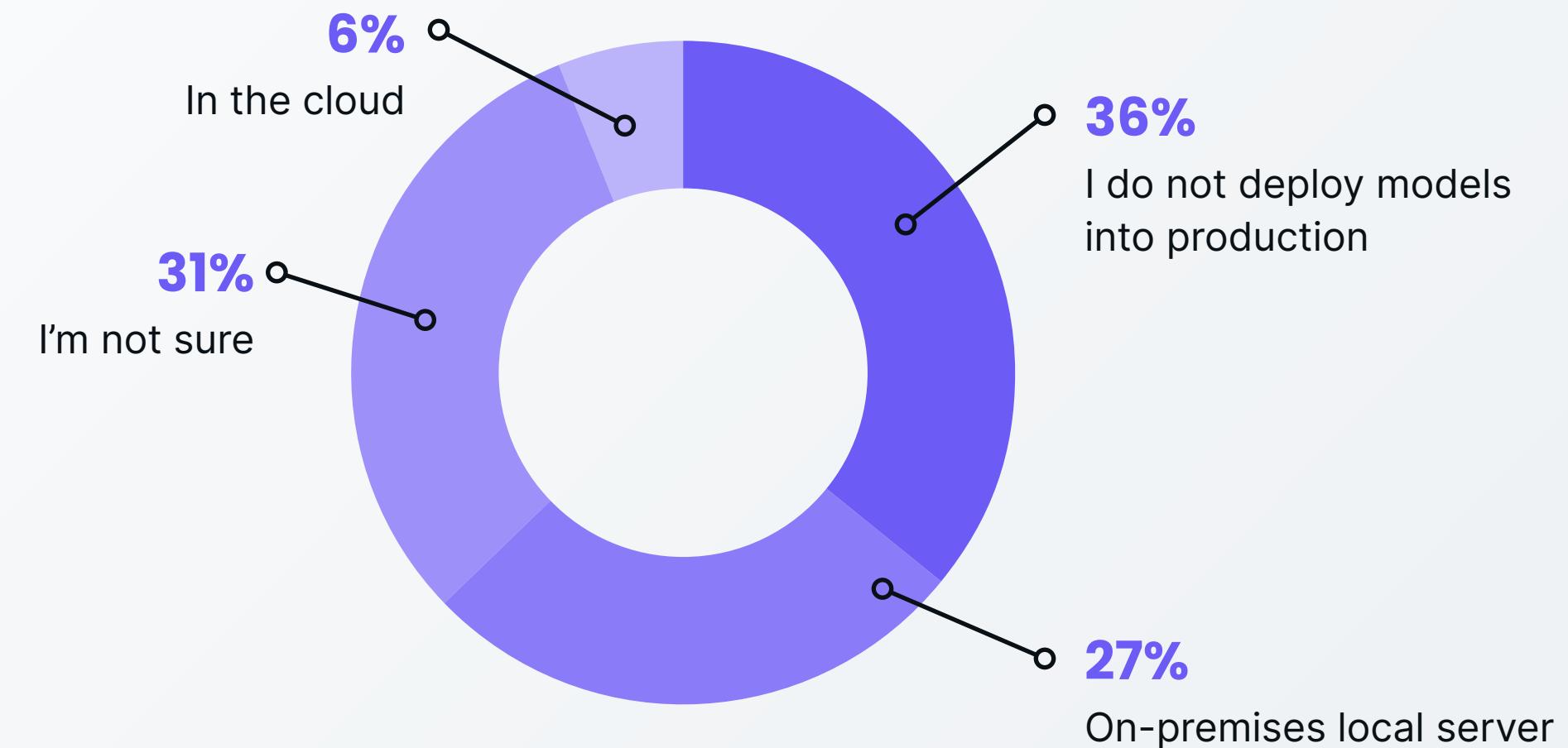
Has your company hired/do they plan on hiring the following roles?



## HOW DATA SCIENCE PRACTITIONERS AND IT ADMINISTRATORS ARE WORKING

### DATA SCIENCE PRACTITIONERS

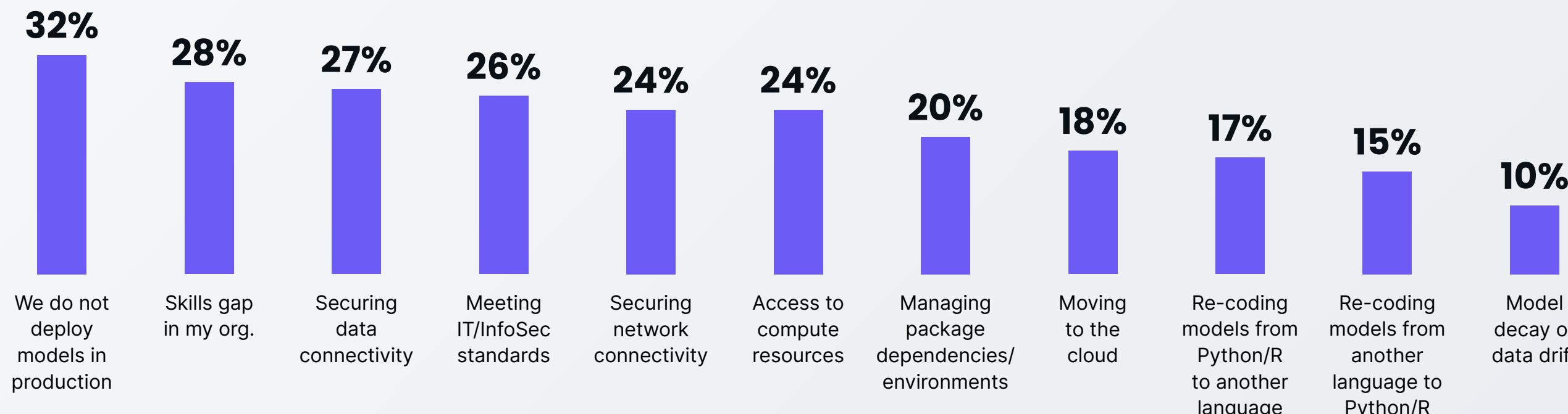
Where do you most often deploy your models into production?



### DATA SCIENCE PRACTITIONERS

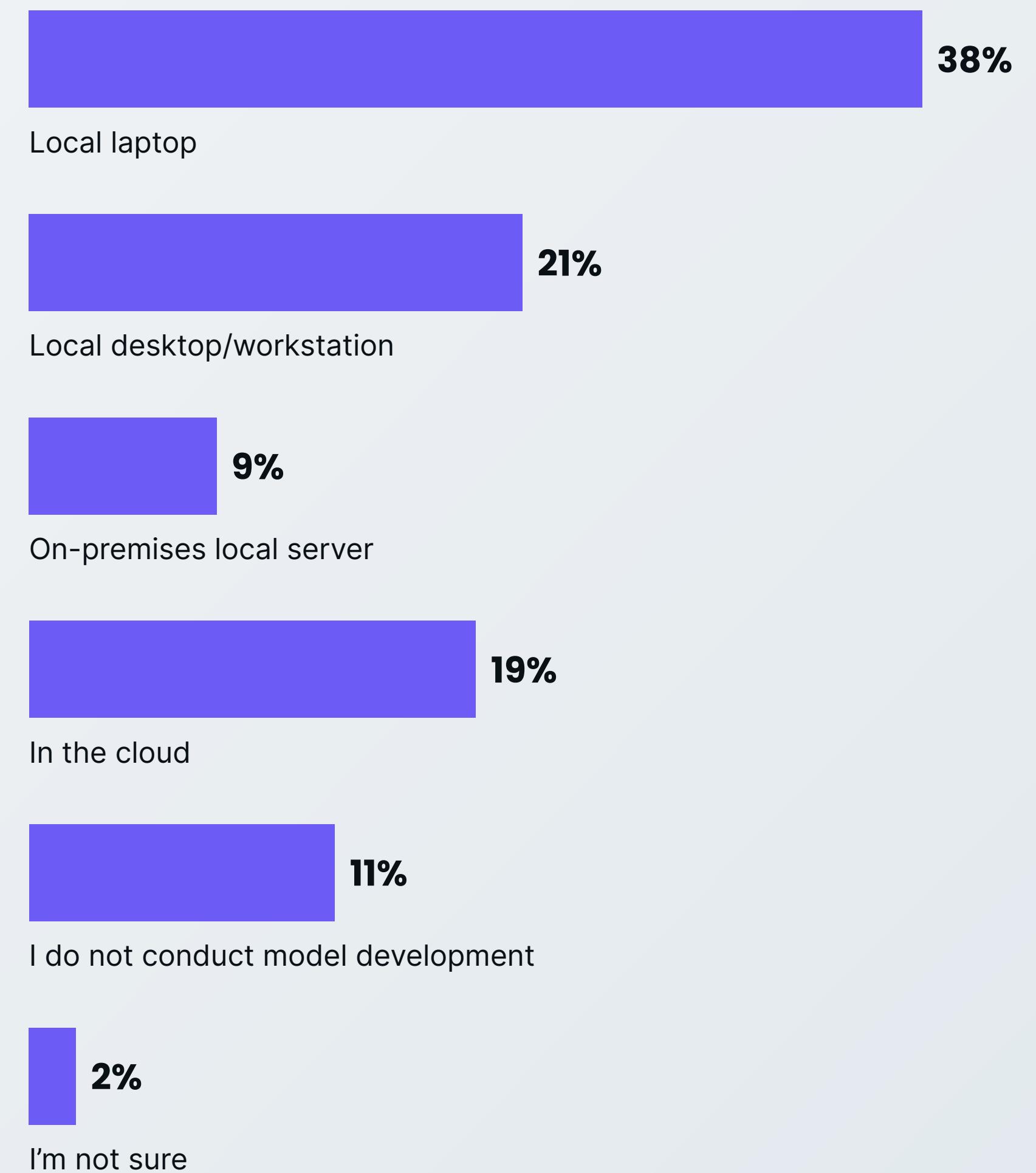
What roadblocks do you/your company face when moving data science or AI models to a production environment?

Select all that apply.



### DATA SCIENCE PRACTITIONERS

Where do you conduct most of your AI or data science model development and training?



# Supporting Education and Teaching

“

Companies need to encourage people to learn more about how to both do their jobs and expand their horizons professionally. What is the point of learning? It's actually to get people to achieve something in a better way.”



**Peter Wang**

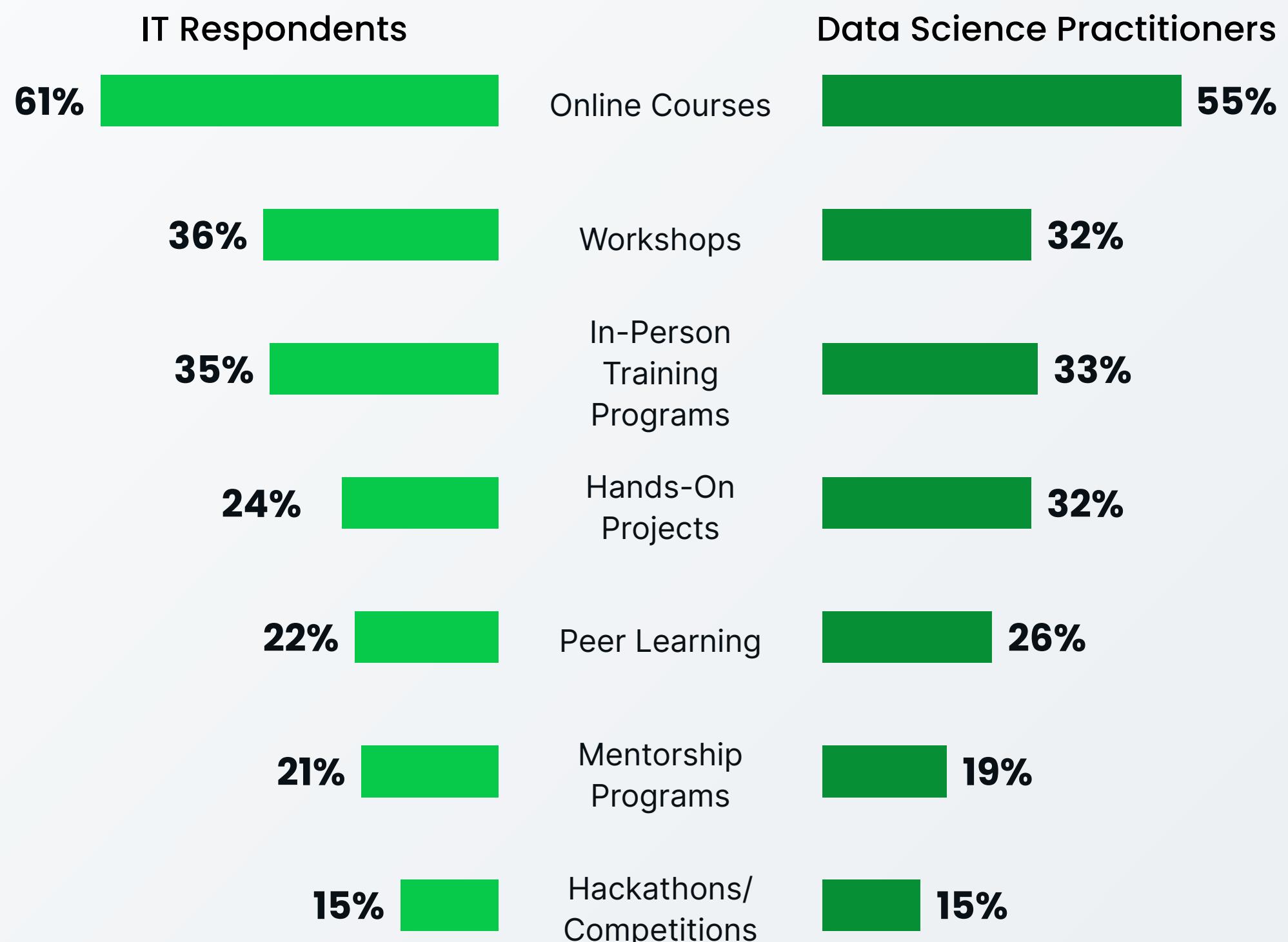
Chief AI and Innovation Officer and Co-Founder, Anaconda



# Supporting Education and Teaching

Even with tremendous AI advances, we're still in the earlier phases of this technological shift. As new technologies emerge and AI and open source become even more commonplace across workflows, it's important to continue learning — and for companies to allow employees to do so. Here's how respondents are gaining knowledge:

## IT ADMINISTRATORS / DATA SCIENCE PRACTITIONERS

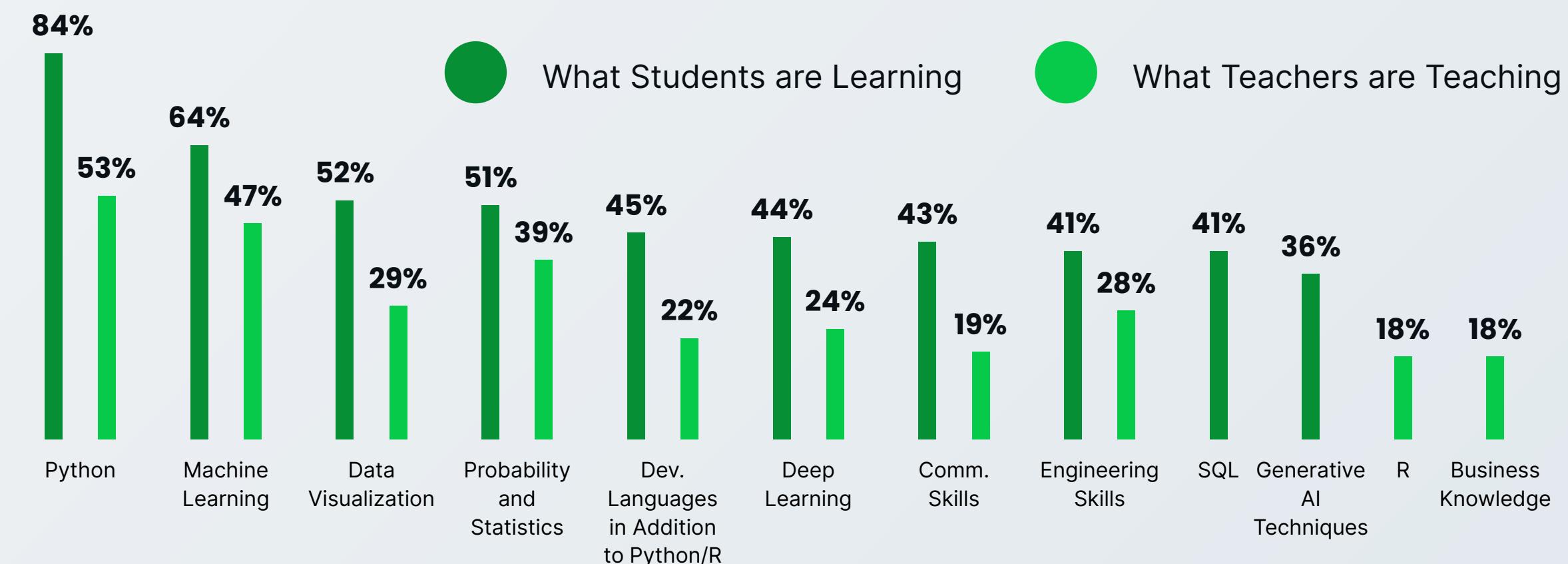


With courses and workshops readily available online, companies have an opportunity to empower data literacy among practitioners, partners, and students with educational resources. Anaconda focuses on [live and on-demand courses](#). More than [90% of Fortune 500 companies](#) and [95% of organizations overall](#) use open-source software in the development of their products and services. For example, IBM embeds Anaconda's curated repository of open-source Python packages for [IBM watsonx enterprise AI](#), streamlining collaboration, development, and deployment of enterprise AI apps.

We're encouraged to see that topics related to open-source software, AI, and data analysis are being taught in schools, highlighting the importance of a well-rounded skill set. Our student respondents shared their top 10 focus areas as they prepare to enter the industry. Students' input generally aligns with the topics professors are teaching. An interesting callout is data visualization, which has a fairly steep drop-off. [AI can help alleviate time and money spent on visualizing data, so future generations would be well-served to learn how to incorporate AI into their data workflows to better interpret it.](#)

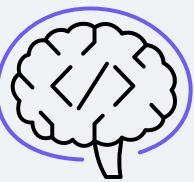
## STUDENTS AND ACADEMICS / STUDENTS AND ACADEMICS

What topics, tools, or skills are you (as a student) learning in preparation for entering the data science/ML/IT field?/What top three topics, tools, or skills is your institution teaching students of data science, machine learning, and AI?



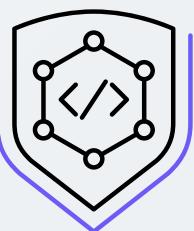


# Key Takeaways



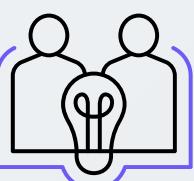
## Greater Focus on AI

We're seeing an increase in the usage of AI techniques, such as generative adversarial networks (GANs), deep learning, and transformer models. 87% of data science practitioners are spending as much or more time on these techniques this year compared to last year.



## Addressing Security

Security will always be a consideration for open source. Organizational leaders must implement security protection processes, encourage collaboration among employees to ensure the secure use of data and models, and offer safety for employees who speak up to highlight security risks. Technology distribution hubs like Anaconda can provide peace of mind as trusted providers of secure packages and repositories.



## Educate, Experiment, Innovate

Development and learning are mainly happening on laptops and personal workstations, while production is more varied, with a slight lean toward the cloud. With work from anywhere fitting into the concept of open-source software, businesses should be encouraging continued education among employees, and a willingness to test different iterations and processes when developing new products and services. Additionally, companies hiring additional AI and data scientist roles suggest a trend toward a concerted focus in these areas, which should allow more innovation.



With 45 million users, Anaconda is the world's most popular data science platform and the foundation of modern AI development. To learn more, visit our [website](#), create your [free Anaconda account](#), attend an [event](#), and follow us [online](#).

[Create your free account >](#)

