2023 Open Source Generative AI Survey Report

Enterprise perspectives and survey-based insights at the intersection of open source innovation and generative AI advancements

Adrienn Lawson, *Linux Foundation*
Marco Gerosa, *Linux Foundation*
Stephen Hendrick, *Linux Foundation*
Matt White, *Linux Foundation*
Lucy Hyde, *Linux Foundation*

Foreword by Stella Biderman, *EleutherAI*

December 2023

In partnership with

LF AI & DATA   GENERATIVE AI COMMONS

2023 Open Source Generative AI Survey Report

Generative AI (GenAI) is a key component for businesses, with 50% of respondents' organizations using it in a production context.

A majority (60%) of companies surveyed plan to significantly invest in GenAI, allocating a large percentage of their IT budgets to the technology.

GenAI is a key factor in future planning. 63% of companies surveyed feel it is extremely or moderately important to the future.
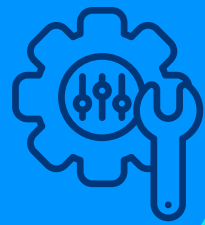
The majority of businesses surveyed intend to tailor GenAI technologies to their needs, embedding them in existing products or creating new products around it.

In general, 41% of organizations surveyed would prefer open source GenAI technologies, compared with 9% who would prefer proprietary solutions.

Security is the primary reason why organizations do not plan to deploy GenAI-related projects, but proprietary solutions are not considered more secure than open ones.

Open source GenAI is considered better at supporting collaboration, innovation, and ease of integration over proprietary solutions, according to our respondents.

Open source GenAI leads to increased data control and transparency, according to 69% of respondents.

Openness is important. 63% of respondents are extremely or moderately concerned by the openness of GenAI systems their companies are using or developing.
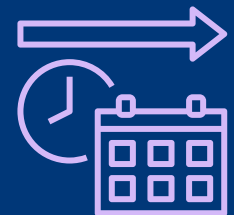
Proprietary and open source solutions are equally preferred by respondents when it comes to the scalability and accuracy of GenAI technologies.

Neutrality is a key aspect of GenAI governance, according to almost all of our respondents (95%).

For the long-term sustainability of GenAI, open source solutions (43%) are preferred over proprietary (32%) solutions.

Contents

Foreword

When GPT-3 came out in May 2020, the world of artificial intelligence was forever changed. What began as a revolution in language modeling research has expanded to image generation, protein synthesis, video editing, and more.

Unfortunately these revolutions were largely kept from the world writ large: only eight of the thirty-four language models released in the two years since GPT-3 had their weights released under an open source license, and only three non-profit or academic institutions in the world successfully trained models more powerful than the previous generation of closed models (GPT-2).

2023, however, marked a turning point in this trend. We witnessed an unprecedented surge in the release of open source AI models, with thirty new base models being made available under open source licenses. This shift was not just in quantity but also in the quality and diversity of these models, trained on 15 languages and coming from 13 different countries across four continents. Moreover, these base models served as the foundation for thousands of fine-tuned models, each tailored for specific applications. This explosion in open source AI has democratized access to cutting-edge technology, enabling a broader range of researchers, developers, and organizations to contribute to and benefit from these advancements.

A commitment to open source AI is more than just a commitment to permissively licensed weights however. The core tenants of the open source movement — the freedom to use, modify, study, and share computer systems — requires access to large amounts of computing resources, highly optimized HPC libraries to carry out the training, reproducible and transparent evaluation frameworks, and large permissively licensed training corpora

among other things. Some of these barriers are beginning to fall, with GPT-NeoX, OpenCLIP, training libraries seeing widespread use beyond their respective creators and evaluation frameworks such as the Language Model Evaluation Harness and Open LLM Leaderboards providing unprecedented access to state-of-the-art tools for creating and studying these models. Still, a broad commitment to increased access to both the technological and the material means of production of generative AI systems is essential to a healthy and thriving open source AI ecosystem

The world has a lot to gain from the recent revolution in AI technology, but it also has a lot to lose. As society, legal systems, and regulators to grapple with this technology it's essential that the open source community builds on our historical successes of securing widespread access to technology such as encryption to build a world where AI is not held in a de facto monopoly by a handful of companies. It is essential that people continue to be empowered to compute what they want, how they want, and according to their own values, rather than having their economic and social freedoms be at the whims of a few technology companies.

In 2024 I look forward to seeing the continued democratization of this technology. I look forward to seeing new models trained in countries that have never trained generative AI systems before, models that speak their creators' languages and reflect their values. I look forward to broader notions of responsible AI that go beyond what is expedient for large corporations. And I look forward to building all this alongside you.

STELLA BIDERMAN
EXECUTIVE DIRECTOR, ELEUTHERAI

Introduction

Generative AI, commonly referred to as GenAI, stands at the forefront of a technological revolution, profoundly altering diverse sectors by synthesizing vast amounts of data and generating new outputs. From creating intricate artworks and composing music to designing novel pharmaceutical compounds and simulating realistic human language, the potential applications of GenAI are vast and transformative. GenAI has undoubtedly become a focal point of both excitement and scrutiny.

The open source approach, rooted in principles of transparency, collaboration, and shared innovation, holds transformative potential for the advancement of GenAI technologies. By democratizing access to AI algorithms and datasets, open source initiatives allow a broad and diverse pool of developers to contribute to, refine, and critique GenAI systems. This collective intelligence accelerates the pace of innovation and uncovers and rectifies biases or vulnerabilities that might otherwise go unnoticed in closed development environments.

As the integration of GenAI into business operations gains momentum, understanding its intricacies and its relation to open source becomes paramount. To understand how open source GenAI can impact the market, LF AI & Data, in partnership with Linux Foundation Research, launched a worldwide survey. This report provides an in-depth exploration of this survey's results, with a special focus on the current state of GenAI in enterprise and GenAI openness. Through comprehensive analysis, we aim to offer insights, highlight best practices, and chart a path forward that ensures sustainable, ethical, and innovative development in this exciting frontier.

To clarify the terminology present in this paper, we refer to GenAI as a broad category for a type of AI that can create new content based on some input. GenAI tools are built on underlying AI models, such as a large language model (LLM). LLMs are a subset of GenAI with a specialized focus on text. In this survey, we have covered open source GenAI technologies not limited to models but including databases, applications, and frameworks. Although at the time of the writing of this paper, the Open Source Initiative (OSI) had not yet released an open source AI definition, a **draft 0.0.3 version** is available and uses four freedoms to define an open source AI system:

- **Study** how the system works and inspect its components.

- **Use** the system for any purpose and without having to ask for permission.

- **Modify** the system to change its recommendations, predictions, or decisions to adapt to the users' needs.

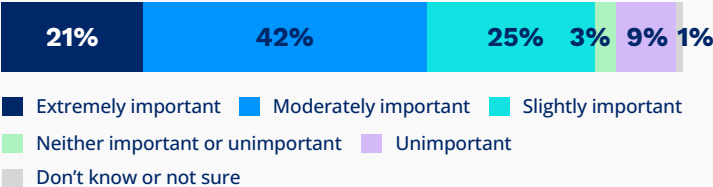- **Share** the system with or without modifications, for any purpose.

Context

High involvement and financial commitment

In the following section, the report lays out the most important features of the survey sample. The varied data and figures reveal that the sample comprises companies that are highly involved in GenAI. As observed in Figure 1, 88% of survey participants indicate that GenAI is important to the future of their companies. This data evidences the strategic importance of GenAI. Figure 2 shows that the surveyed companies show high involvement in GenAI technologies (80%) and will invest heavily in their GenAI strategies (60%). This investment distribution is nearly identical for both end-user and vendor organizations, suggesting that all organizations in our sample are anticipating heavy investment. This considerable investment reflects a major commitment, indicating a significant impact on several projects and infrastructure changes within these companies.

FIGURE 1

HIGH IMPORTANCE FOR GENAI FOR THE FUTURE PLANS OF COMPANIES

2023 GenAI Survey, Q11, Sample Size = 280

How important is GenAI to the future of the company you work for? (select one)

| 21% | 42% | 25% | 3% | 9% | 1% |

- ■ Extremely important
- ■ Moderately important
- ■ Slightly important
- ■ Neither important or unimportant
- ■ Unimportant
- Don't know or not sure

FIGURE 2

HIGH INVOLVEMENT AND LARGE INVESTMENT IN GENAI TECHNOLOGIES WITHIN THE SURVEYED COMPANIES

To what extent is your company involved with GenAI? (select one)

2023 GenAI Survey, Q2, Sample Size = 284

Extremely involved (GenAI is business critical to key aspects of what our company does)	31%
Very involved (GenAI is being used in production in selected areas)	49%
Involved (experimenting with how GenAI can add value in selected areas)	14%
Slightly involved (researching or evaluating GenAI)	5%
My organization has evaluated and banned all use of GenAI tools	0%
Not involved at all	1%
Don't know or not sure	0%

How much is your company planning to invest in its GenAI strategies in the next 12 months as a percentage of its overall IT budget? (select one)

2023 GenAI Survey, Q16, Sample Size = 249

A majority: almost entirely focused on GenAI strategies	9%
A large percentage: a major commitment encompassing several projects or infrastructure changes	51%
A moderate percentage: significant but not a major portion of the IT budget	29%
A small percentage: for pilot projects or specific initiatives	8%
No investment	0%
Don't know or not sure	2%
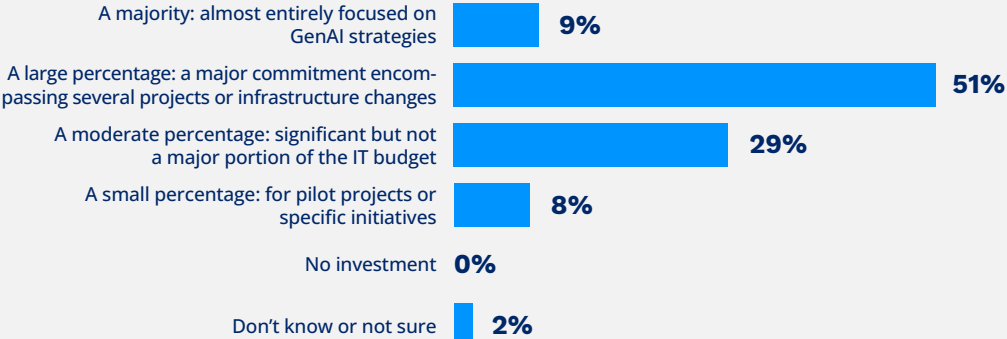
Diverse application areas and usage strategies

GenAI significantly impacts operations, as shown in Figure 3, particularly in product development and enhancement. Key areas include software quality assurance (35%), software testing (34%), and cybersecurity (31%), demonstrating its potential in risk mitigation and ensuring product and service quality. Additionally, software development (29%) and documentation

DIVERSE APPLICATION AREAS FOR GENAI UTILIZATION

2023 GenAI Survey, Q12, Sample Size = 280

Please identify those areas where your organization expects to develop or use GenAI. (select all that apply)

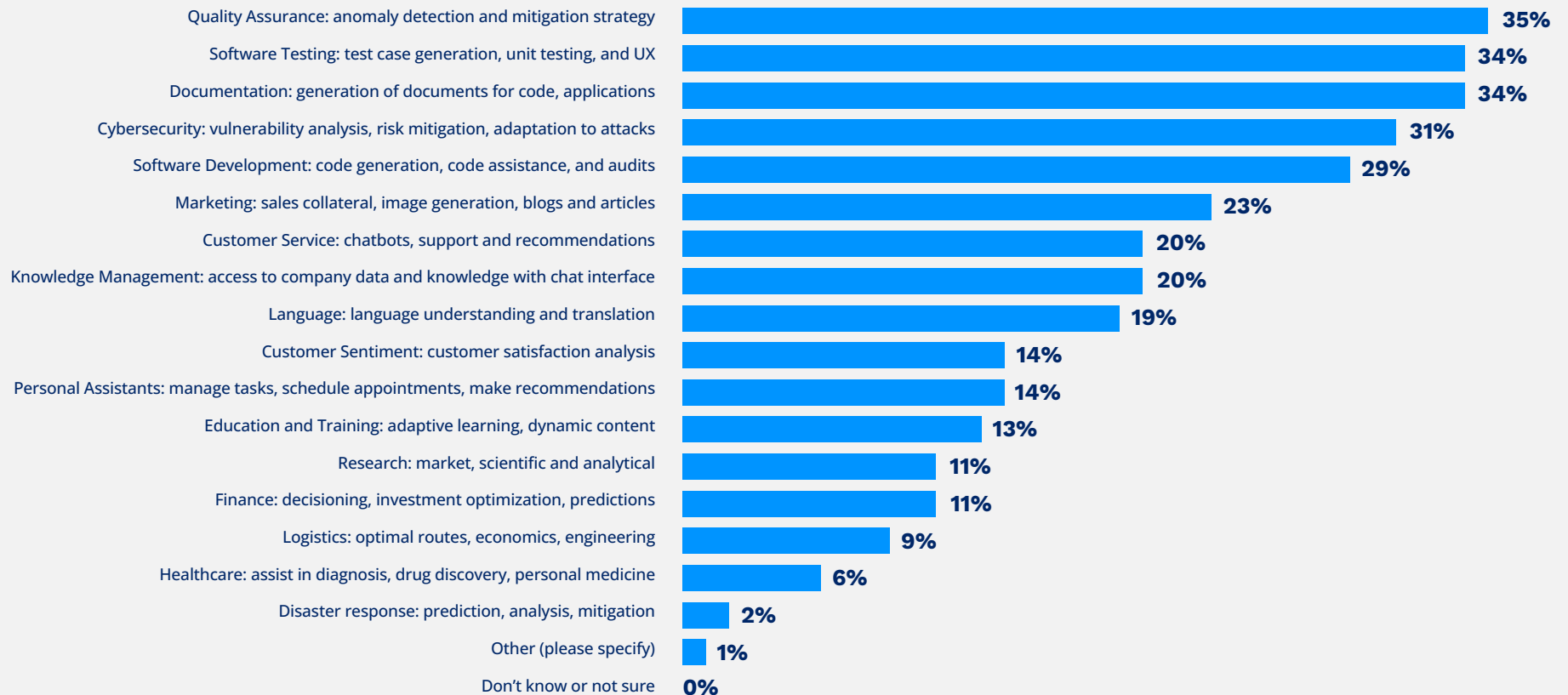| Area	Percentage
Quality Assurance: anomaly detection and mitigation strategy	35%
Software Testing: test case generation, unit testing, and UX	34%
Documentation: generation of documents for code, applications	34%
Cybersecurity: vulnerability analysis, risk mitigation, adaptation to attacks	31%
Software Development: code generation, code assistance, and audits	29%
Marketing: sales collateral, image generation, blogs and articles	23%
Customer Service: chatbots, support and recommendations	20%
Knowledge Management: access to company data and knowledge with chat interface	20%
Language: language understanding and translation	19%
Customer Sentiment: customer satisfaction analysis	14%
Personal Assistants: manage tasks, schedule appointments, make recommendations	14%
Education and Training: adaptive learning, dynamic content	13%
Research: market, scientific and analytical	11%
Finance: decisioning, investment optimization, predictions	11%
Logistics: optimal routes, economics, engineering	9%
Healthcare: assist in diagnosis, drug discovery, personal medicine	6%
Disaster response: prediction, analysis, mitigation	2%
Other (please specify)	1%
Don't know or not sure	0%

FIGURE 4

THE MAJORITY OF BUSINESSES INTEND TO TAILOR GENAI TECHNOLOGIES TO THEIR SPECIFIC NEEDS

2023 GenAI Survey, Q10 and Q14, Sample Size = 245, cells add up to 100%

How does your company use or plan to use GenAI technologies? (select all that apply)

	Use out-of-the-box GenAI technologies with little to no customization	Use GenAI technologies and customize them extensively to fit our needs	Develop our own in-house GenAI technologies	Row totals
Employ GenAI to enhance internal processes	4%	7%	5%	16%
Embed GenAI into products and services	6%	18%	6%	29%
Create new products based on GenAI or create GenAI solutions for third parties	3%	33%	19%	55%
Column totals	13%	57%	30%	

(34%) are notable applications, with organizations using GenAI to automate code generation and create dynamic documentation for applications and source code.

Our survey assessed companies' stages in their GenAI journey by examining how they plan to use GenAI, as shown in Figure 4. We grouped companies by their most advanced GenAI usage on two dimensions. The row totals reveal that organizations in our sample aim not only to enhance their internal processes with GenAI but also to embed GenAI into products and services (29%) or create new GenAI-based products or solutions (55%). In terms of customization level, the column totals show that many organizations plan to customize and enhance GenAI foundation models (57%),

potentially through methods such as fine-tuning or RAG (retrieval-augmented generation). A significant number also intend to develop in-house GenAI technologies (30%). Developing in-house solutions does not necessarily mean building LLMs or other large foundation models from scratch, as it can require expensive and scarce resources and might not serve specific use cases well. Companies also have an opportunity to build small, domain-specific GenAI models from their own datasets by leveraging expertise in data science. Both customization and the development of in-house solutions will likely rely on the open-source community, which has been creating solutions to the challenges of GenAI customization with techniques such as **LongLoRA**, a fine-tuning approach with limited computation cost.

Generative AI openness

Open source software provides significant benefits by ensuring that software is developed in the open. This attribute removes barriers to learning, using, sharing, and improving software. This can also result in more autonomy, transparency, and collaboration, which, if applied to GenAI, could ensure that users have the freedom to develop reliable and transparent AI systems. The following section delves into the results of the survey's GenAI openness questions.
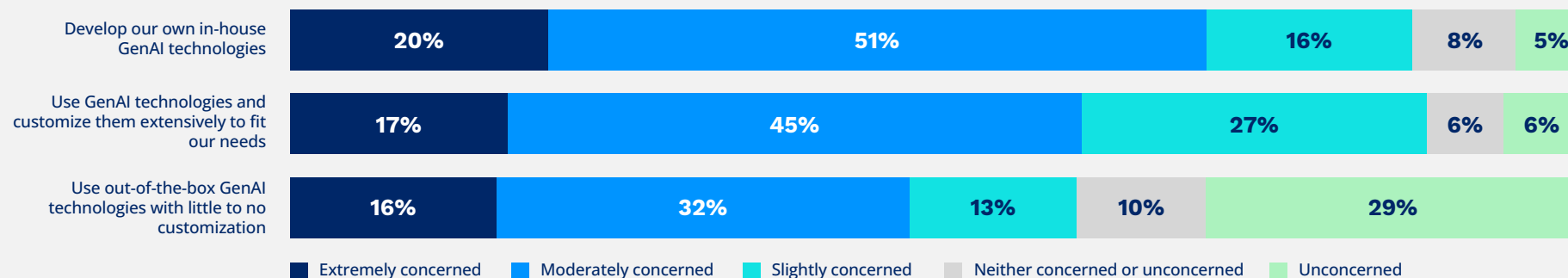
The level of openness can vary greatly between the different GenAI models currently available, but most of them would likely not earn the open source title, since availability and access to the underlying code, data, model, and documentation are rare.[1] However, the GenAI ecosystem is not limited to models but includes applications from vector and graph databases to agent frameworks. To illustrate, companies have the opportunity to leverage open source application development frameworks

1   Andreas Liesenfeld, Alianda Lopez, and Mark Dingemanse. 2023. Opening up ChatGPT: Tracking openness, transparency, and accountability in instruction-tuned text generators. In Proceedings of the 5th International Conference on Conversational User Interfaces (CUI '23). Association for Computing Machinery, New York, NY, USA, Article 47, 1–6. https://doi.org/10.1145/3571884.3604316

FIGURE 5

CONCERNS OVER OPENNESS IN EXISTING GENAI SYSTEMS ESPECIALLY AMONG COMPANIES CUSTOMIZING OR DEVELOPING IN-HOUSE SOLUTIONS

2023 GenAI Survey,Q14 by Q13, Sample Size = 247

How does your company employ or plan to employ GenAI technologies? (select all that apply) segmented by
How concerned is your organization about the openness of the GenAI systems you are developing or using? (select one)

	Extremely concerned	Moderately concerned	Slightly concerned	Neither concerned or unconcerned	Unconcerned
Develop our own in-house GenAI technologies	20%	51%	16%	8%	5%
Use GenAI technologies and customize them extensively to fit our needs	17%	45%	27%	6%	6%
Use out-of-the-box GenAI technologies with little to no customization	16%	32%	13%	10%	29%

(e.g., LangChain) on top of closed models to integrate their applications, back their office systems and innovate with new platforms. Therefore, openness can be leveraged across a wide range of GenAI. The open approach is vital, as confirmed by our survey respondents' concerns about the openness of the GenAI technologies they are using or developing.

Figure 5 shows that, across the three ways in which organizations intend to employ GenAI technologies (develop in-house, customize to their needs, and use with little or no customization), concern over the openness of the GenAI is correlated with the level of organizational involvement. In Figure 5, 71% of organizations are moderately or extremely concerned about the openness of the GenAI they will be developing. This may be due to the wide variations in openness today and the risk of betting on an approach that the industry dismisses as the market matures. A similar situation exists for organizations that intend to customize GenAI systems to better fit their needs, where 62% of organizations are moderately or extremely concerned about the need for openness. By contrast, only 48% of organizations are moderately or extremely concerned about the openness of out-of-the-box AI technologies that are used with little or no customization. Presumably, this is because organizations have already done due diligence in their selection process and the vendor/supplier is also ultimately responsible for the quality and reliability of the product or service.

Concern about the openness of GenAI translates into organizational preferences between open source and proprietary GenAI technologies. Figure 6 shows that 41% of organizations lean toward open source GenAI technologies, compared with 9% favoring proprietary ones. Twenty-two percent of organizations are inclined to use both types of solutions while 28% are indifferent, indicating that their choice of technology will ultimately be influenced by factors beyond these preferences.

Which distribution model does your organization prefer for GenAI, proprietary or open source? (select one)

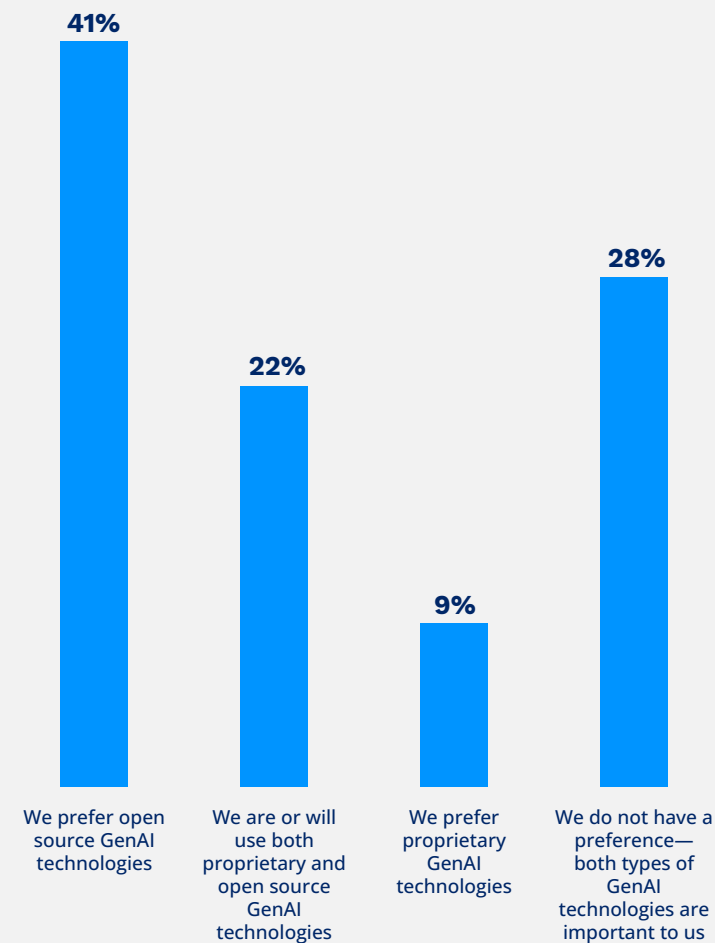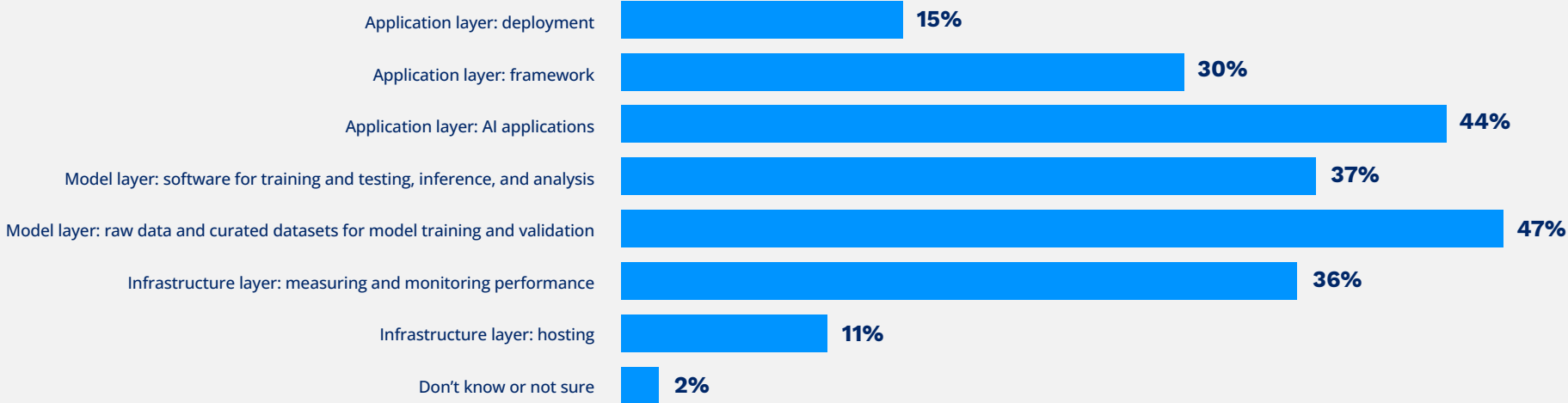|	
41%	We prefer open source GenAI technologies
22%	We are or will use both proprietary and open source GenAI technologies
9%	We prefer proprietary GenAI technologies
28%	We do not have a preference— both types of GenAI technologies are important to us

FIGURE 7

ACROSS THE THREE PRIMARY LAYERS OF THE GENAI STACK, OPEN DATASETS AT THE MODEL LAYER WOULD BE THE MOST FAVORED OPEN SOURCE TECHNOLOGY

2023 GenAI Survey, Q22, Sample Size = 249

The GenAI stack can generally be divided into three primary layers: the Application Layer, the Model Layer, and the Infrastructure Layer.
Which components of these layers, if any, do you believe should be based on open source technologies? (select all that apply)

Application layer: deployment	15%
Application layer: framework	30%
Application layer: AI applications	44%
Model layer: software for training and testing, inference, and analysis	37%
Model layer: raw data and curated datasets for model training and validation	47%
Infrastructure layer: measuring and monitoring performance	36%
Infrastructure layer: hosting	11%
Don't know or not sure	2%

While the open source *software* definition revolves around the source code[2], an open source *AI system* definition will have to consider the various layers that make up the GenAI stack. In our survey, we outlined three primary layers: the application layer, the model layer, and the infrastructure layer. Figure 7 shows that respondents appreciate open datasets most (47%). Open datasets for GenAI can accelerate innovation, promote collaboration, and mitigate bias through data availability. Survey respondents further mentioned that open source technologies could improve the applications based on GenAI models (44%), the software for training and testing (37%), and the tools for measuring and monitoring performance in the infrastructure layer (36%). Other ways exist to deconstruct GenAI systems and assess their openness: Researchers have developed an **openness tracker** for various LLMs.

---

2   Open Source Initiative: The Open Source Definition, available at https://opensource.org/osd/

Security and trust

Security is a major barrier to deploying GenAI

Security (49%) is by far the most relevant obstacle to employing GenAI, as observed in Figure 8. Some examples of security concerns regarding GenAI are privacy, trust, unintended consequences, data breaches, and misuse. GenAI systems, by design, ingest vast amounts of data to train and operate optimally. This data may include sensitive, insecure, incorrect, or biased information.

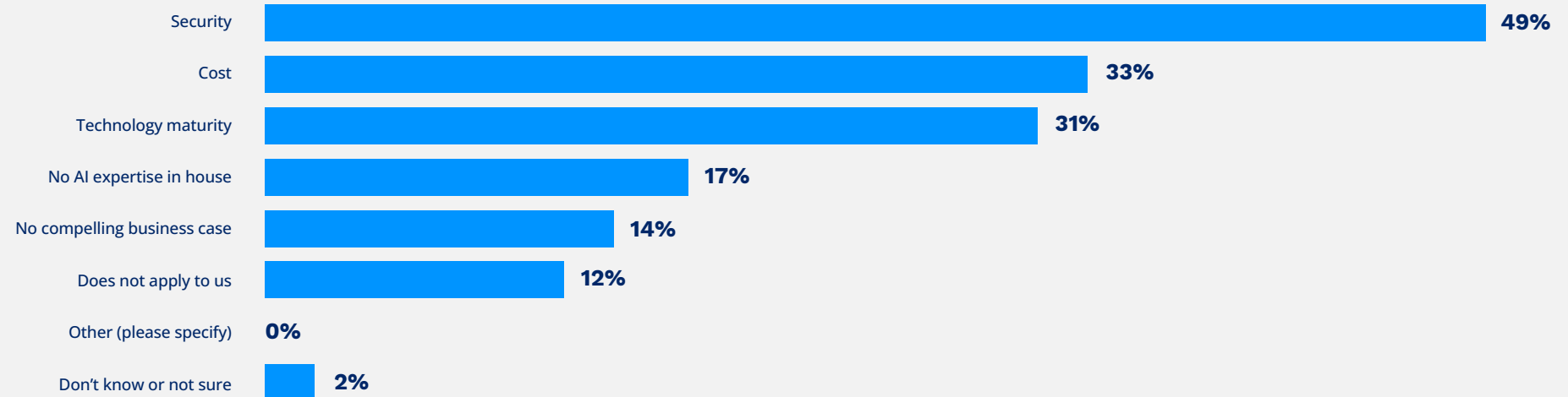There is also the challenge of ensuring that the model or other parts of the infrastructure do not inadvertently disclose information introduced during the training, testing, or validation process, which could lead to leakage of confidential information. Ensuring the security of GenAI technologies is not just a technical necessity but crucial for maintaining trust and regulatory compliance. This is further complicated by the complexity of these black-box models that can obscure vulnerabilities, making it challenging for organizations to fully understand and mitigate potential security risks. As one of the respondents answered in an open-

FIGURE 8

SECURITY CONCERNS ARE THE PRIMARY REASON WHY COMPANIES DO NOT INITIATE GENAI RELATED PROJECTS

2023 GenAI Survey, Q15, Sample Size = 249

If your company does not plan to deploy or initiate any GenAI-related projects in the next 12 months, what are the primary reasons? (select all that apply)

Reason	Percentage
Security	49%
Cost	33%
Technology maturity	31%
No AI expertise in house	17%
No compelling business case	14%
Does not apply to us	12%
Other (please specify)	0%
Don't know or not sure	2%

FIGURE 9

WHEN CONSIDERING SECURITY MATTERS, NO EVIDENCE FOUND THAT COMPANIES PREFER PROPRIETARY GENAI TECHNOLOGIES TO OPEN SOURCE SOLUTIONS

2023 GenAI Survey, Q18 and Q19, Sample Size = 249

For each of the following considerations, which type of GenAI solution would you prefer? (one response per row)

Legend: ■ Open source ■ The same ■ Proprietary
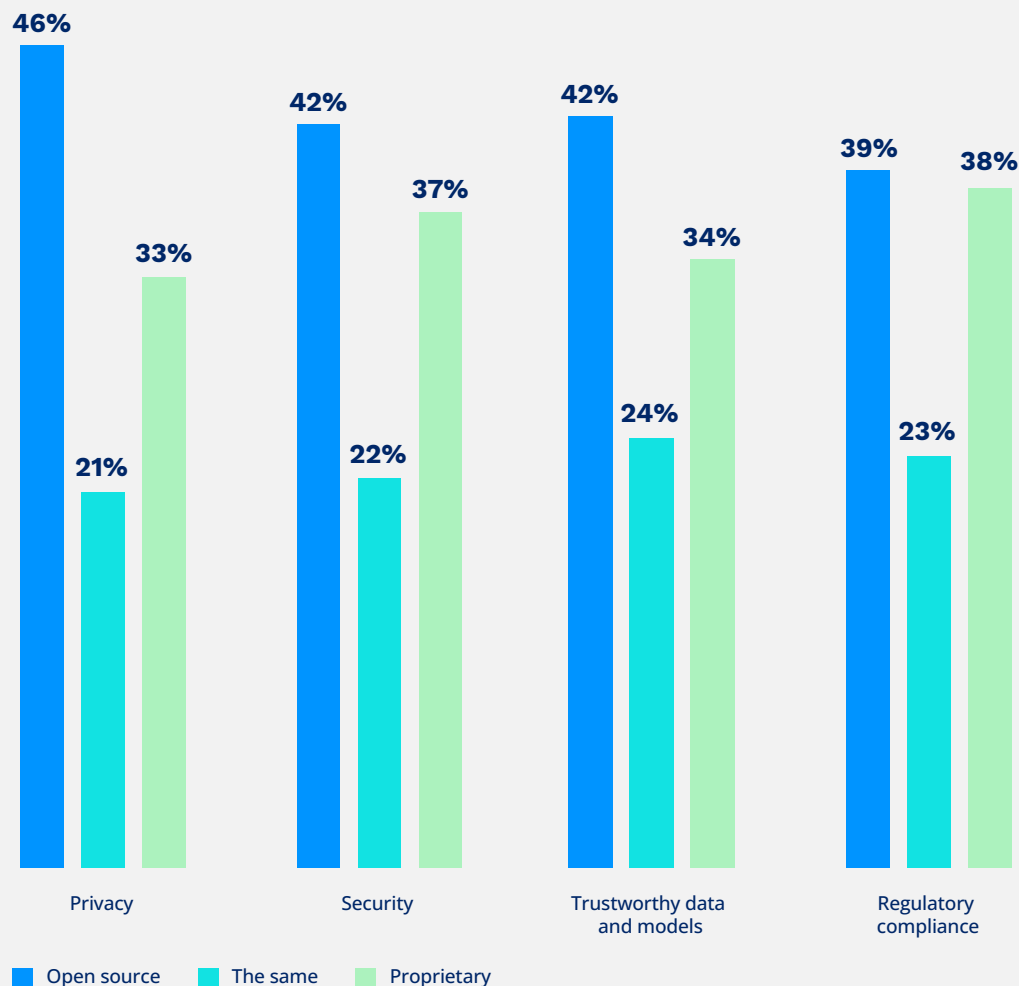
ended question on the challenges of GenAI: *"Data security concerns are the biggest problem at our company since GenAI needs to adopt security measures more effectively to protect consumer data and guarantee compliance with privacy regulations."*

No evidence found for proprietary preference in security considerations

While it is crucial to address the security concerns mentioned when discussing Figure 8, it is not guaranteed that proprietary solutions will effectively resolve these issues. We asked our survey respondents to consider whether they would prefer open source or proprietary GenAI solutions across four discrete security concerns.

Figure 9 reveals that when it comes to the vital considerations of security, privacy, and regulatory compliance in GenAI technologies, there is no substantial evidence of a prevailing preference for proprietary solutions over open source options among companies. Figure 9 shows that respondents lean toward preferring open source over proprietary, but the survey's margin of error does not show a significant difference between the two alternatives. This finding challenges the arguments that claim that proprietary solutions are more compliant with regulation and safer for GenAI development.
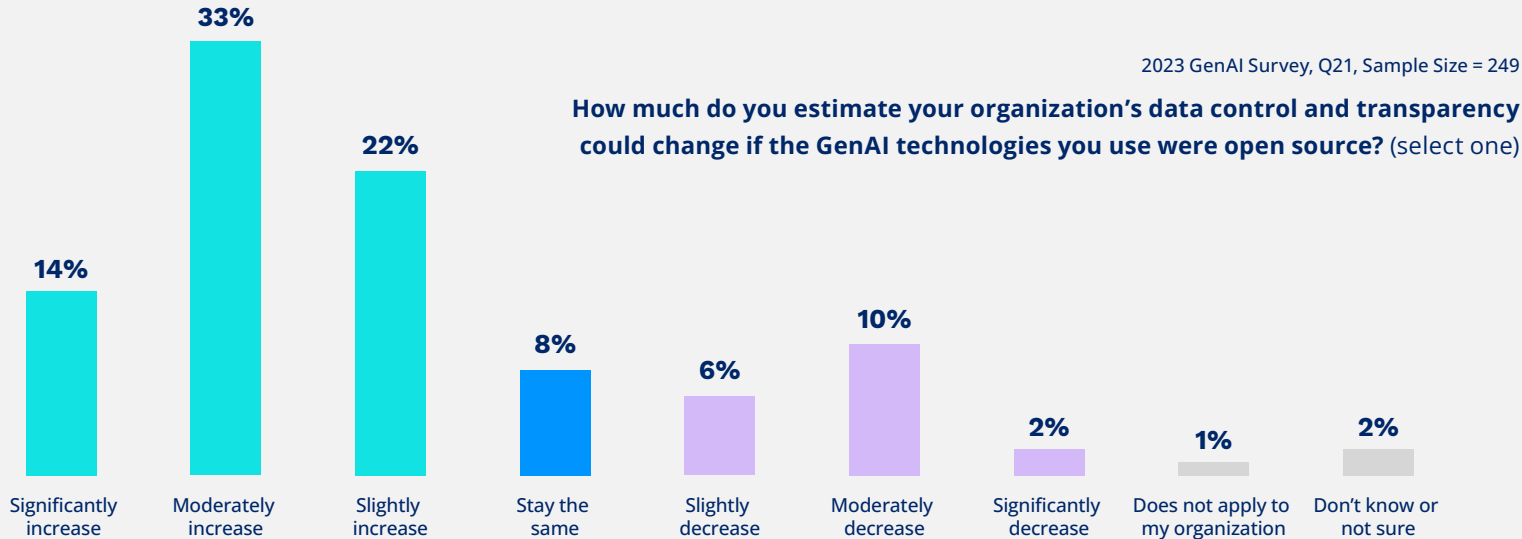
Transparency and accessibility

Open source GenAI increases data control and transparency

The openness of GenAI models provides opportunities for the public and academics to scrutinize AI models. A lack of understanding or transparency about how GenAI models make decisions can hinder individuals' rights to know how their data is being used. Without proper mechanisms for accountability, it is challenging to ensure that privacy is consistently upheld. A survey respondent answered an open-ended question on transparency by saying, *"Our company must prioritize building trust with their customers by*

being transparent about their use of AI technology and providing clear explanations of how AI systems make decisions (...) some customers may be hesitant to interact with AI systems, preferring human interaction." As observed in Figure 10, a significant 69% of companies believe that their organization's data control and transparency would see an improvement if they were to use open source GenAI technologies.
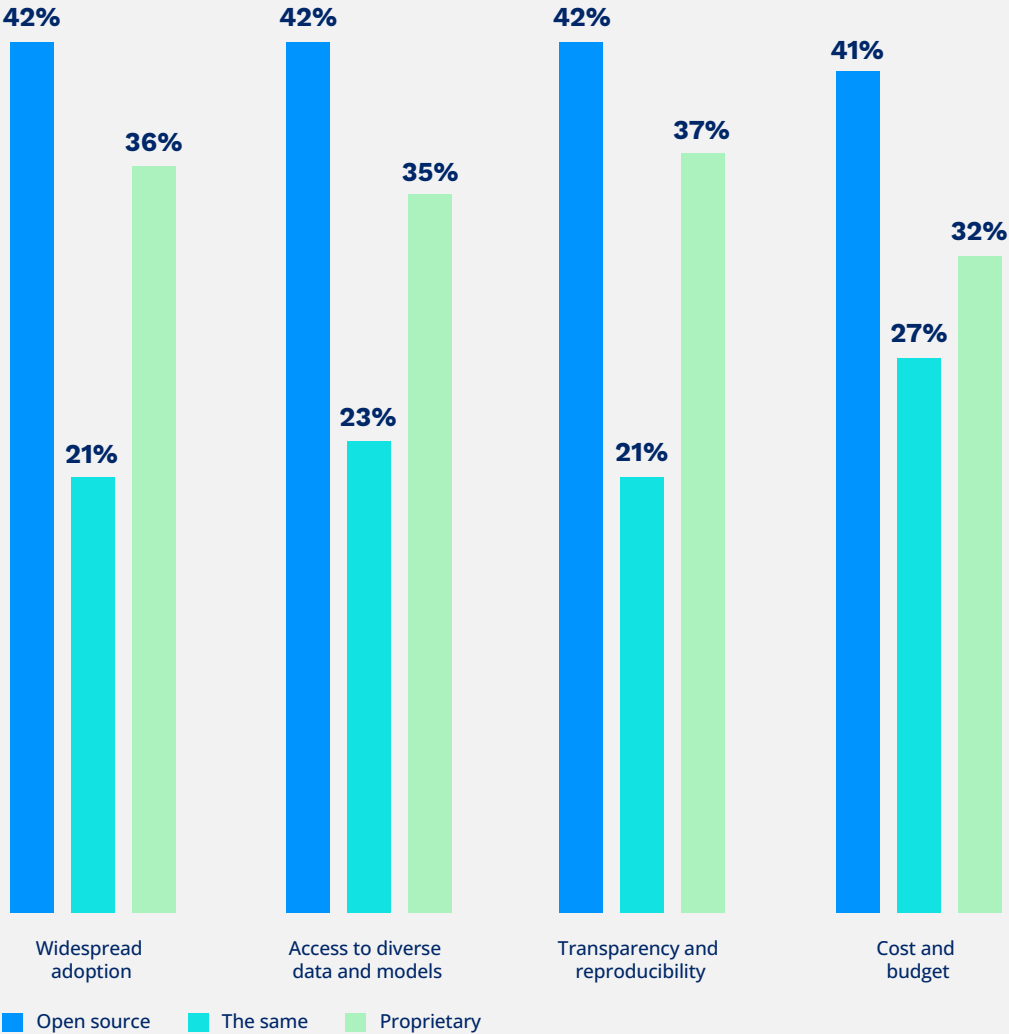
FIGURE 10

AGREEMENT ON THE INCREASE IN DATA CONTROL AND TRANSPARENCY BY OPEN SOURCE GENAI TECHNOLOGIES

2023 GenAI Survey, Q21, Sample Size = 249

How much do you estimate your organization's data control and transparency could change if the GenAI technologies you use were open source? (select one)

Significantly increase	Moderately increase	Slightly increase	Stay the same	Slightly decrease	Moderately decrease	Significantly decrease	Does not apply to my organization	Don't know or not sure
14%	33%	22%	8%	6%	10%	2%	1%	2%

FIGURE 11

EVALUATING OPEN SOURCE AS A SOLUTION FOR ADOPTION, ACCESSIBILITY, AND REPRODUCIBILITY OF GENAI

2023 GenAI Survey, Q18 and Q19, Sample Size = 249

For each of the following considerations, which type of GenAI solution would you prefer? (one response per row)

	Widespread adoption	Access to diverse data and models	Transparency and reproducibility	Cost and budget
Open source	42%	42%	42%	41%
The same	21%	23%	21%	27%
Proprietary	36%	35%	37%	32%

Legend: ■ Open source ■ The same ■ Proprietary

Evaluating open source as a solution for accessibility and reproducibility of GenAI

Figure 11 shows preferences for four considerations related to GenAI adoption. Open source models may be perceived as more favorable for widespread adoption of GenAI due to their accessibility and the collaborative opportunities they offer, allowing for rapid dissemination and iteration across a broad user base, as shown.

With 42% favoring open source for access to diverse data and models compared with 35% for proprietary, there is an implication that open source is associated with a richer variety of data and modeling options. This is critical in AI development, where diversity in datasets can lead to more robust and less biased AI systems.

The result that 42% prefer open source for transparency and reproducibility underscores the value placed on openness in the AI community. Transparency is key to building trust and allowing for independent verification of AI systems, while reproducibility is essential for scientific progress and validation of results.

The preference for open source (41%) over proprietary (32%) in terms of cost and budget considerations indicates that open source solutions are perceived as more cost-effective. This is particularly relevant in a context where organizations are seeking to maximize the efficiency of their investments in AI technologies, especially when budget constraints are a factor.

FIGURE 12

COMPANIES FOCUSING ON THE DEVELOPMENT OF GENAI TECHNOLOGIES CONSIDER THE ADOPTION OF A NEUTRAL GOVERNANCE APPROACH TO BE OF SIGNIFICANT IMPORTANCE
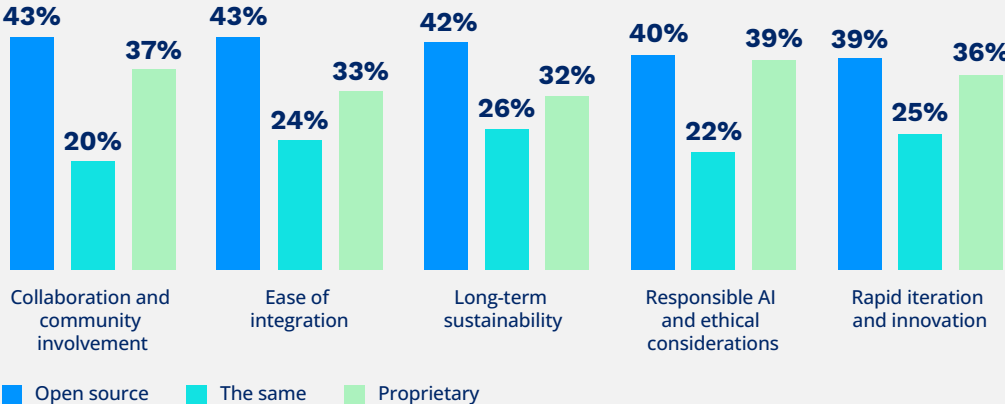
2023 GenAI Survey, Q24, Sample Size = 72

How important is having a neutral governance open source approach to developing GenAI technologies? (select one)

46%	42%	7%	3%

- ■ Extremely important
- ■ Very important
- ■ Important
- ■ Slightly important
- ■ Not important at all
- ■ Don't know or not sure

FIGURE 13

OPEN SOURCE GENAI TECHNOLOGIES, UNDER NEUTRAL GOVERNANCE, HAS THE POTENTIAL TO ACHIEVE RESPONSIBLE INNOVATION

2023 GenAI Survey, Q18 and Q19, Sample Size = 249

For each of the following considerations, which type of GenAI solution would you prefer? (one response per row)

Consideration	Open source	The same	Proprietary
Collaboration and community involvement	43%	20%	37%
Ease of integration	43%	24%	33%
Long-term sustainability	42%	26%	32%
Responsible AI and ethical considerations	40%	22%	39%
Rapid iteration and innovation	39%	25%	36%

- ■ Open source
- ■ The same
- ■ Proprietary

Neutral governance and responsible innovation

A neutral governance approach is important for GenAI technologies

As important as transparency and accessibility are for GenAI technologies, open source might not be enough to mitigate the risks that we associate with GenAI. Figure 12 shows that a neutral governance approach is important for our survey respondents, with 88% indicating that it is extremely or very important when developing GenAI technologies. Neutral governance is another aspect of true open source models and can benefit GenAI technologies in multiple ways. Neutral governance is important to ensure innovation is not subject to only a few companies' futures. In addition, neutral governance can help set ethical standards and guidelines to prevent misuse of the technology. Neutral governance is tied to various considerations explored in our survey.

Figure 13 shows a lean toward open source solutions in the realms of collaboration and community involvement (43%), long-term sustainability (42%), and responsible AI and ethical considerations (40%). Such governance provides an impartial framework that likely encourages diversity and inclusion in the development process as it is not tied to the interests of proprietary systems. Neutral governance can ensure that innovation and iteration are not only rapid but also ethically aligned and sustainable over time, making the technology more accessible and potentially leading to more equitable outcomes in the GenAI space.

Performance and business needs

Accuracy and scalability are deemed to be at similar levels of open source and proprietary GenAI
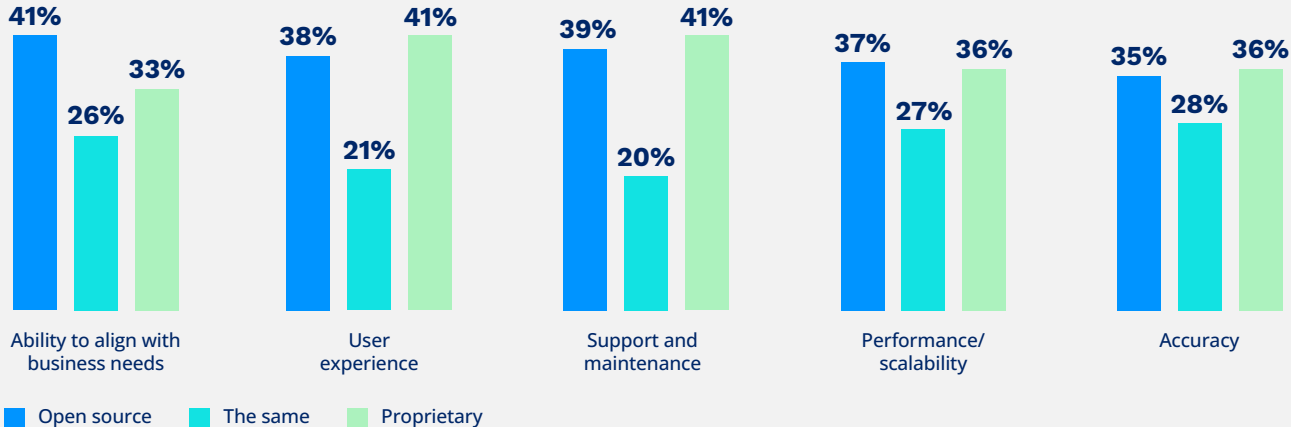
The effectiveness of GenAI is often evaluated by companies based on performance indicators such as accuracy and speed. Figure 14 highlights the comparative preferences for open source versus proprietary GenAI technologies in relation to key business needs. It is evident from the data that the preference for open source and proprietary solutions is closely matched across various technical considerations. For example, open source and proprietary solutions are almost equally preferred in terms of

their accuracy, with 36% for proprietary and 35% for open source. Similar patterns are observed in other categories, such as support and maintenance and performance/scalability. In terms of user experience, slightly more respondents prefer proprietary solutions (41%) to open source ones (38%). This balanced distribution of preferences acknowledges a competitive landscape where open source solutions are considered nearly as favorable as proprietary ones in meeting critical technical needs.

FIGURE 14

SIMILAR LEVELS OF PREFERENCE OF OPEN SOURCE GENAI TECHNOLOGIES AND PROPRIETARY SOLUTIONS REGARDING BUSINESS NEEDS, SUCH AS SCALABILITY AND ACCURACY

2023 GenAI Survey, Q18 and Q19, Sample Size = 249

For each of the following considerations, which type of GenAI solution would you prefer? (one response per row)

Ability to align with business needs: Open source 41%, The same 26%, Proprietary 33%
User experience: Open source 38%, The same 21%, Proprietary 41%
Support and maintenance: Open source 39%, The same 20%, Proprietary 41%
Performance/scalability: Open source 37%, The same 27%, Proprietary 36%
Accuracy: Open source 35%, The same 28%, Proprietary 36%

Legend: ■ Open source ■ The same ■ Proprietary

Conclusions

Businesses are concerned by the openness of the GenAI technologies they are using

The survey reveals a strong concern among respondents regarding the openness of GenAI systems. Around two-thirds of respondents are either extremely or moderately concerned about this aspect, reflecting the importance of transparency and control in technology deployments. Open source GenAI, according to 69% of respondents, leads to increased data control and transparency, which are critical for ethical and responsible AI development.

Survey respondents generally lean in the direction of open source

The findings from our survey provide compelling insights into the current attitudes and preferences of organizations toward GenAI, particularly highlighting a notable inclination toward open source solutions. This finding highlights a recognition of the benefits associated with open source technologies, including transparency, reproducibility, access to diverse data and models, and ease of integration. Security, an important concern for any technology deployment, does not appear to be a deterrent for open source GenAI adoption. In fact, most respondents do not view proprietary solutions as more suitable for security considerations than open ones.

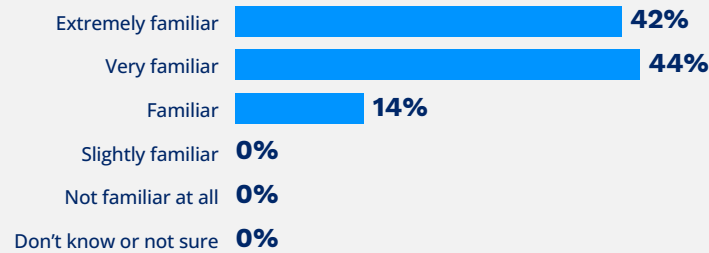## A neutral governance approach is key to GenAI development

The importance of neutral governance in GenAI was supported by 95% of respondents in the survey. This governance framework ensures a more ethical and equitable development of GenAI technologies through community involvement and collaboration. Neutral governance is not only crucial for fostering responsible growth of GenAI but also for ensuring that its benefits are widespread and aligned with societal values. This approach is vital in maintaining the integrity and sustainability of GenAI advancements, ensuring that they serve both communities and stakeholders.
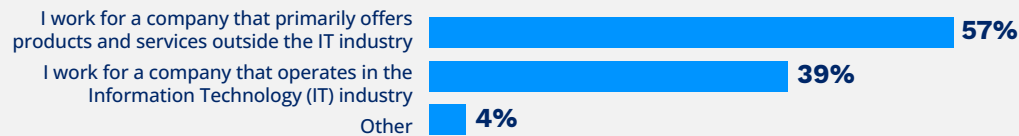
FIGURE 15

SELECTED DEMOGRAPHIC DATA

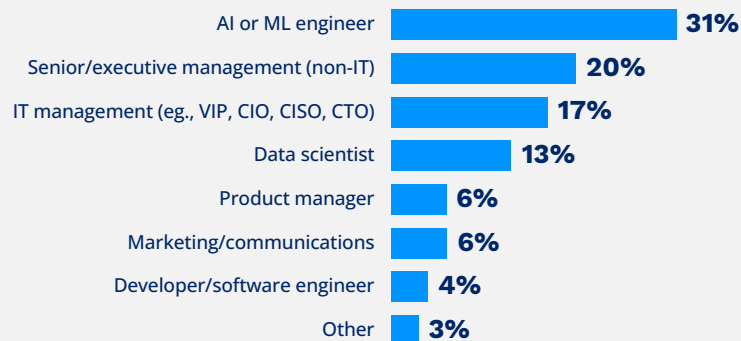2023 GenAI Survey, Q1, Q3 and Q4, Sample Size = 284

How familiar are you with your organization's adoption of Generative AI (GenAI)? (select one)

Extremely familiar	**42%**
Very familiar	**44%**
Familiar	**14%**
Slightly familiar	**0%**
Not familiar at all	**0%**
Don't know or not sure	**0%**

What best describes the company you work for? (select one)

I work for a company that primarily offers products and services outside the IT industry	**57%**
I work for a company that operates in the Information Technology (IT) industry	**39%**
Other	**4%**

Which of the following best describes your professional role? (select one)

AI or ML engineer	**31%**
Senior/executive management (non-IT)	**20%**
IT management (eg., VIP, CIO, CISO, CTO)	**17%**
Data scientist	**13%**
Product manager	**6%**
Marketing/communications	**6%**
Developer/software engineer	**4%**
Other	**3%**

About this study

During September and October 2023, LF AI & Data and Linux Foundation Research fielded an online survey of individuals at organizations on a range of questions related to GenAI. The survey was promoted via LF social media and at LF events. We also sourced qualified respondents from a third-party panel provider to craft a more diverse sample.

Methodology

We received 284 valid survey starts, and 249 respondents completed all relevant questions. The margin of error for the sample size of 249 is 5.2% at the 90% confidence level. This sample size reflects those respondents who met a variety of screening and filtering criteria. The primary screening criteria included employment (respondents who were students, unemployed or retired were disqualified) and familiarity with the organization's adoption of GenAI (not familiar at all, slightly familiar, and those responding "Don't know or not sure" were also disqualified).

The percentage values in this report may not total exactly 100% due to rounding.

Demographics

Figures 15 and 16 provide selected demographics of the survey sample.

In the left-hand panel of Figure 15, we see that 42% of our respondents were extremely familiar with GenAI, 44% were very familiar, and just 14% were familiar. The lack of respondents who were either slightly familiar or not familiar or didn't know or were not sure was intentional. This is because this question was part of our screening process so that respondents would be capable of

FIGURE 16

SELECTED DEMOGRAPHIC DATA
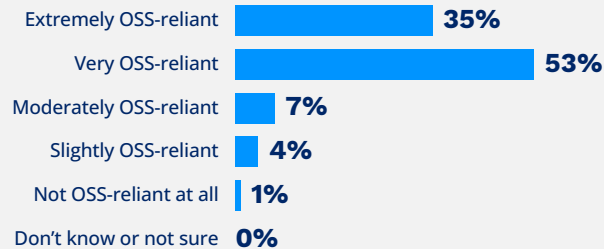
2023 GenAI Survey, Q5, Q6 and Q7, Sample Size = 284

In which region does your company have its headquarters? (select one)

United States or Canada	92%
Europe	5%
Other	3%
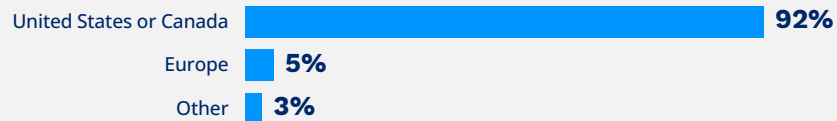
Please estimate how many total employees are in your company. (select one)

1 to 999	29%
1,000 to 9,999	47%
10,000 or more	24%

How reliant is your company on open source software (OSS)? (select one)

Extremely OSS-reliant	35%
Very OSS-reliant	53%
Moderately OSS-reliant	7%
Slightly OSS-reliant	4%
Not OSS-reliant at all	1%
Don't know or not sure	0%

providing us with reliable perspectives and insights. Because 86% of respondents were either very or extremely familiar with GenAI, we believe this higher level of expertise will improve the quality and insight provided by this survey. The central panel in Figure 15 shows that the respondents are well distributed across industries, with 57% working in end-user organizations (those companies that use or even embed IT but primarily offer products and services focused on industries outside of IT itself) and 39% working for IT vendors or service providers. The right-hand panel of Figure 15 shows that respondents are distributed across a variety of roles, including AI or ML engineer (31%), non-IT senior/executive manager (20%), IT managers (17%), and data scientists (13%).

Figure 16 is a continuation of this demographic data. The left-hand panel of Figure 16 shows a distribution by region. We did not make an effort to stratify by region, and, as a result, most responses come from the U.S. or Canada (92%). The center panel in Figure 16 shows the distribution of respondent organizations by company size in employees. Respondents are reasonably well distributed across three groups: 1 to 999 (29%), 1,000 to 9,000 (47%), and 10,000 or more (24%). The right-hand panel in Figure 16 shows that most organizations are reliant on open source software, with 53% reporting being very reliant and 35% being extremely reliant.

Data.World access

LF Research makes each of its empirical project datasets available on Data.World. Included in this dataset are the survey instrument, raw survey data, screening and filtering criteria, and frequency charts for each question in the survey. LF Research datasets, including this project, can be found at **data.world/ thelinuxfoundation**.

About the authors

ADRIENN LAWSON is a data analyst at the LF. Adrienn obtained a master's degree from the University of Oxford in social data science. She supports LF Research with survey development, analysis, and report writing. Adrienn has previously conducted research at the University of Oxford, the Budapest Institute for Policy Analysis, and the U.K.'s Office for National Statistics.

Dr. MARCO GEROSA is a full professor in computer science at Northern Arizona University and a research analyst at LF Research. His main areas of investigation are software engineering and open source software. He is currently investigating the use of GenAI as a tool to support computer science education and the onboarding of new developers in OSS communities. These projects are supported by the National Science Foundation and have been featured in publications in top-tier venues. He has published over 200 papers and serves on the program committee of important conferences, such as ICSE, FSE, and MSR, and as a reviewer for several journals. He graduated several Ph.D. and M.Sc. students who are now researchers in top institutions and have more than 20 years of teaching experience. For more information, visit http://www.marcoagerosa.com.

STEPHEN HENDRICK is vice president of research at the Linux Foundation, where he is the principal investigator on a variety of research projects core to the Linux Foundation's understanding of how OSS is an engine of innovation for producers and consumers of IT. Steve specializes in primary research techniques developed over 30 years as a software industry analyst. Steve is a subject matter expert in application development and deployment topics, including DevOps, application management, and decision analytics. Steve brings experience in a variety of quantitative and qualitative research techniques that enable deep insight into market dynamics and has pioneered research across many application development and deployment domains. Steve has authored over 1,000 publications and provided market guidance through syndicated research and custom consulting to the world's leading software vendors and high-profile start-ups.

MATT WHITE is the Director of the Generative AI Commons at the Linux Foundation's AI & Data Foundation. He is Head of AI & Data at Amdocs, as well as the Founder of the AI research group Berkeley Synthetic. He teaches graduate students Data Science at UC Berkeley. He is also the Co-Founder and Chair of the Open Metaverse Foundation a part of the Linux Foundation and a Board Director at the Metaverse Standards Forum. He has over 25 years of experience in AI and data and open source. He holds a Master of Data Science from UC Berkeley, an MBA from the University of Denver and a BSc IT from York University. For more information, visit www.matt-white.com.

LUCY HYDE is a Senior Program Manager specializing in Machine Learning supporting open source innovation in artificial intelligence and data. As an accomplished professional, she is highly regarded for technical expertise working in roles focusing on data science, software engineering, and technical exploitation. She started her career with the Department of Defense as an active duty service member and as a government civilian, and continued in private sector digital forensics. She graduated with undergraduate degrees in Psychology, the Arabic language, and Intelligence Operations; an MS in Analytics; and is pursuing a Ph.D. in Computational Science/Informatics from George Mason University and a second MS degree in Artificial Intelligence from Johns Hopkins University.

Acknowledgments

THE LINUX FOUNDATION | Research

Founded in 2021, **Linux Foundation Research** explores the growing scale of open source collaboration, providing insight into emerging technology trends, best practices, and the global impact of open source projects. Through leveraging project databases and networks, and a commitment to best practices in quantitative and qualitative methodologies, Linux Foundation Research is creating the go-to library for open source insights for the benefit of organizations the world over.

□LF AI & DATA

lfaidata.foundation | **genaicommons.org**