

# 2023年 オープンソースの 生成 AI に関する 調査レポート

オープンソース イノベーションと  
生成 AI の進化が交差する  
企業の視点と調査に基づく洞察

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

序文 Stella Biderman, *EleutherAI*

2023年12月

協力

LF AI & DATA

GENERATIVE AI  
COMMONS



# 2023年 オープンソースの生成 AI に関する調査レポート

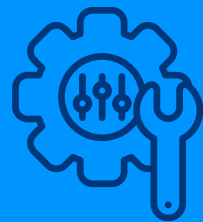
ジェネレーティブAI (生成AI) は企業にとって重要な要素であり、回答者の50%の組織がプロダクションコンテキストで使用しています。



調査対象となった企業の過半数 (60%) が生成AIへの大規模な投資を計画しており、IT予算の大部分をこのテクノロジーに割り当てています。



生成AIは将来計画の重要な要素。調査対象企業の63%が、将来にとって非常に重要、または中程度に重要であると感じています。



調査対象となった企業の大半は、自社のニーズに合わせてカスタマイズし、既存の製品に組み込んだり、生成AIを中心とした新製品を開発する意向です。

一般的に、調査対象となった組織の41%がオープンソースの生成AIテクノロジーを好むのに対し、プロプライエタリなソリューションを好むのは9%でした。



セキュリティは、組織が生成AI関連プロジェクトの導入を計画していない主な理由ですが、プロプライエタリなソリューションがオープンなものよりも安全だとは考えられていません。



回答者によると、オープンソースの生成AIは、プロプライエタリなソリューションよりも、コラボレーション、イノベーション、統合の容易さをサポートする上で優れていると考えられています。



オープンソースの生成AIは、回答者の69%がデータの管理と透明性の向上につながると回答しています。



オープン性は重要。回答者の63%は、自社が使用または開発している生成AIシステムのオープン性に非常に、または中程度の懸念を抱えています。

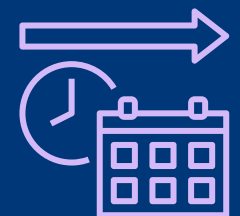


生成AI技術のスケラビリティと精度に関しては、プロプライエタリとオープンソースのソリューションが同じように回答者に好まれています。

回答者のほぼ全員 (95%) によると、中立性は生成AIガバナンスの重要な側面です。



生成AIの長期的な持続可能性については、プロプライエタリなソリューション (32%) よりもオープンソースのソリューション (43%) が好まれます。



# 目次

序文.....	4	パフォーマンスとビジネス ニーズ.....	17
はじめに .....	5	精度と拡張性は、オープンソースとプロプライエタリな生成 AI の 同レベルと判断 .....	17
コンテキスト .....	6	結論.....	18
高い関与と財政的コミットメント .....	6	企業が懸念する生成 AI テクノロジーのオープン性 .....	18
多様なアプリケーション分野と利用戦略 .....	7	調査の回答者は、一般的にオープンソースの方向に傾いている .....	18
生成 AI のオープン性.....	9	中立的なガバナンス アプローチが生成 AI 開発の鍵 .....	18
セキュリティと信頼 .....	12	この調査について .....	19
セキュリティは生成 AI 導入の大きな障壁 .....	12	調査方法 .....	19
セキュリティを考慮した独自の優先順位を示す証拠は 見つからなかった .....	13	人口統計 .....	19
透明性とアクセシビリティ .....	14	Data.World アクセス .....	20
オープンソースの生成 AI がデータ管理と透明性を向上 .....	14	著者について .....	21
生成 AI のアクセシビリティと再現性のためのソリューションとしての オープンソースの評価 .....	15	謝辞.....	21
中立的なガバナンスと責任あるイノベーション .....	16		
生成 AI テクノロジーには中立的なガバナンス アプローチが重要 .....	16		

## 序文

GPT-3 が 2020 年 5 月に登場したとき、人工知能の世界は一変しました。言語モデリング研究の革命として始まった GPT-3 は、画像生成、タンパク質合成、ビデオ編集などにまで広がりました。

残念ながら、これらの革命は世界全体からほとんど知られていませんでした。GPT-3 以降の 2 年間にリリースされた 34 の言語モデルのうち、オープンソースライセンスの下でその成果が公開されたのはわずか 8 つだけで、前世代のクロウズドモデル (GPT-2) よりも強力なモデルの学習に成功した非営利団体や学術機関は、世界でわずか 3 つだけでした。

しかし 2023 年、この傾向にターニングポイントが訪れました。オープンソース AI モデルのリリースがかつてないほど急増し、30 もの新しいベースモデルがオープンソースライセンスの下で利用可能になりました。この変化は、単に量だけでなく、これらのモデルの質と多様性においても顕著であり、15 の言語で学習され、4 大陸にまたがる 13 の異なる国から提供されました。オープンソース AI の爆発的な普及は、最先端技術へのアクセスを民主化し、より幅広い研究者、開発者、組織がこれらの進歩に貢献し、その恩恵を受けることを可能にしました。

オープンソース AI へのコミットメントは、単に寛容にライセンスされたウェイトへのコミットメント以上のものです。オープンソースムーブメントの中核をなす、コンピューターシステムの使用、変更、研究、共有の自由は、とりわけ、大量のコンピューティングリソースへのアクセス、トレーニングを実行するための高度に最適化された HPC ライブラリ、再現可能で透明性の高い評価フレームワーク、および寛容にライセンスされた大規模なトレーニングコーパスを必要とします。GPT-NeoX、OpenCLIP、トレーニングライブラリは、それぞれの作成者を超えて広く使用され、言語モデル評価ハーンズや Open LLM Leaderboards の

ような評価フレームワークは、これらのモデルを作成し研究するための最先端のツールへの前例のないアクセスを提供しています。それでもなお、健全で繁栄するオープンソース AI エコシステムには、生成 AI システムの技術的および物質的な生産手段の両方へのアクセスを向上させる幅広いコミットメントが不可欠です。

世界は、最近の AI 技術の革命から多くのものを得ることができますが、失うものも多々あります。社会、法制度、規制当局がこのテクノロジーに取り組む中で、オープンソースコミュニティが、暗号化などのテクノロジーへのアクセスを広く確保した私たちの歴史的な成功に基づき、AI が一握りの企業によって事実上独占されることのない世界を構築することが不可欠です。経済的・社会的自由が一部のテクノロジー企業の気まぐれで左右されるのではなく、人々が自分の望むものを、望む方法で、自分の価値観に従ってコンピューティングする力を持ち続けることが不可欠なのです。

2024 年、私はこのテクノロジーが引き続き民主化されることを楽しみにしています。これまで生成 AI システムをトレーニングしたことのない国々でトレーニングされた新しいモデルや、作成者の言語を話し、彼らの価値観を反映したモデルを見ることを楽しみにしています。私は、大企業にとって都合の良いものを越えた、責任ある AI の幅広い概念に期待しています。そして、皆さんと一緒にこのすべてを作り上げていくことを楽しみにしています。

**STELLA BIDERMAN**  
**EXECUTIVE DIRECTOR, ELEUTHERAI**

## はじめに

一般的に GenAI と呼ばれる生成 AI (Generative AI) は、技術革命の最前線に立ち、膨大な量のデータを合成して新たなアウトプットを生成することで、様々な分野を大きく変えます。複雑なアートワークの作成や作曲から、新しい医薬化合物の設計やリアルな人間の言語のシミュレーションに至るまで、生成 AI の潜在的なアプリケーションは膨大であり、変革をもたらします。生成 AI は間違いなく、興奮と精査の両方の焦点となっています。

透明性、コラボレーション、共有イノベーションの原則に根ざしたオープンソースアプローチは、生成 AI 技術の進歩のための変革の可能性を秘めています。AI アルゴリズムとデータセットへのアクセスを民主化することで、オープンソースイニシアティブは、広く多様な開発者が生成 AI システムに貢献し、改良し、批評することを可能にします。この集合知は、イノベーションのペースを加速させ、クローズドな開発環境では気づかれなかった可能性のあるバイアスや脆弱性を発見し、修正します。

生成 AI のビジネス業務への統合が勢いを増すにつれ、その複雑さとオープンソースとの関係を理解することが最も重要になります。オープンソースの生成 AI が市場にどのような影響を与えるかを理解するために、LF AI & Data は Linux Foundation Research と提携して世界的な調査を開始しました。本レポートは、企業における生成 AI の現状と生成 AI のオープン性に特に焦点を当て、この調査結果の詳細な調査を提供します。包括的な分析を通じて、私たちは洞察を提供し、ベストプラクティスをハイライトし、このエキサイティングなフロンティアにおける持続可能で倫理的、革新的な開発を確保するための道筋を描くことを目指しています。

この論文で使用する用語を明確にするために、生成 AI を、ある入力に基づいて新しいコンテンツを作成できる AI の種類を表す広範なカテゴリと定義します。生成 AI ツールは、大規模言語モデル (LLM) のような基礎となる AI モデル上に構築されます。LLM は、テキストに特化した生成 AI のサブセットです。この調査では、モデルに限らず、データベース、アプリケーション、フレームワークを含むオープンソースの生成 AI 技術を取り上げました。この論文の執筆時点

では、オープンソースイニシアティブ (OSI) はまだオープンソース AI の定義をリリースしていませんが、[ドラフト 0.0.3 バージョン](#)が利用可能であり、オープンソース AI システムを定義するために 4 つの自由を使用しています。

- システムがどのように機能するかを**学習**し、そのコンポーネントを検査する。
- どのような目的であっても、許可を得ることなくシステムを**使用**する。
- システムを**変更**して、推奨、予測、決定をユーザーのニーズに合わせて変更する。
- いかなる目的であれ、修正あり、または修正なしでシステムを**共有**する。

# コンテキスト

## 高い関与と財政的コミットメント

以下では、調査サンプルの最大の特徴を示します。さまざまなデータと数値から、サンプルは生成 AI に大きく関与している企業で構成されていることがわかります。図 1 に見られるように、調査参加者の 88% が生成 AI は自社の将来にとって重要であると回答しています。このデータは、生成 AI の戦略的重要性を証明しています。図 2 は、調査対象企業が生成 AI テクノロジーに高い関与 (80%) を示し、生成 AI 戦略に多額の投資 (60%) を行うことを示しています。この投資分布は、エンドユーザーとベンダーの両方の組織でほぼ同じであり、サンプルのすべての組織が多額の投資を見込んでいることを示唆しています。この多額の投資は大きな取り組みを反映しており、これらの企業内のいくつかのプロジェクトやインフラ変更に大きな影響を与えることを示しています。

図 1

## 企業の将来計画における生成 AI の重要性

2023 年 生成 AI に関する調査 , Q11, サンプルサイズ = 280

あなたの勤務先企業の将来にとって、生成 AI はどの程度重要ですか? (1 つ選択してください)

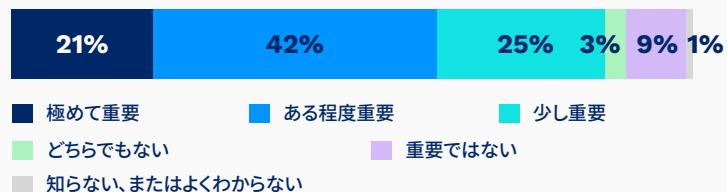
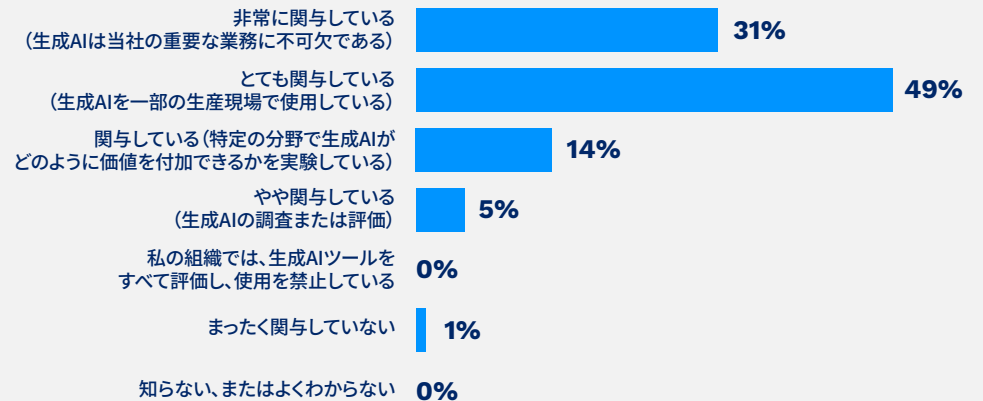


図 2

## 調査対象企業における生成 AI 技術への高い関与と多額の投資

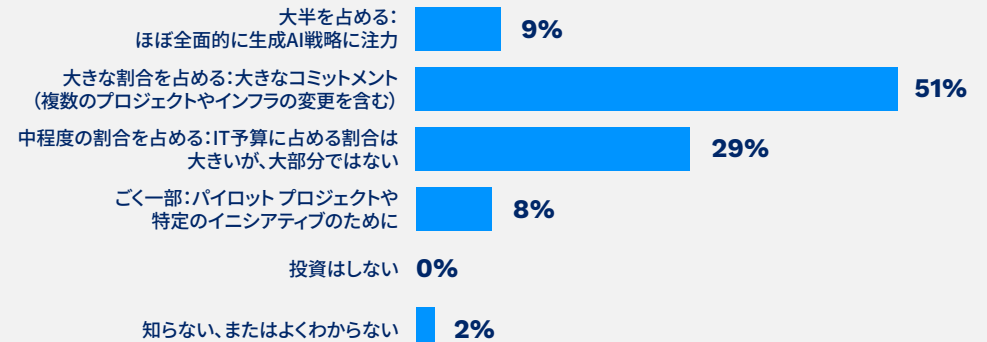
貴社は生成 AI にどの程度関わっていますか? (1 つ選択してください)

2023 年 生成 AI に関する調査 , Q2, サンプルサイズ = 284



貴社では、今後 12 カ月間に生成 AI 戦略にいくら投資する予定ですか? IT 予算全体に占める割合でお答えください。(1 つ選択してください)

2023 年 生成 AI に関する調査 , Q16, サンプルサイズ = 249



## 多様なアプリケーション分野と利用戦略

生成 AI は、図 3 に示すように、特に製品開発と機能強化において業務に大きな影響を与えます。主な分野には、ソフトウェア品質保証 (35%)、ソフトウェアテスト (34%)、サイバーセキュリティ (31%) が含まれ、リスク軽減や製品・サービスの品質確保における可能性を示しています。

さらに、ソフトウェア開発 (29%) とドキュメンテーション (34%) は注目すべきアプリケーションであり、組織は生成 AI を使用してコード生成を自動化し、アプリケーションやソースコードの的ドキュメントを作成しています。

図 3

### 生成 AI の多様な応用分野

2023 年 生成 AI に関する調査, Q12, サンプルサイズ=280

あなたの組織が生成 AI の開発または利用を期待している分野を特定してください。(該当するものをすべて選択してください)

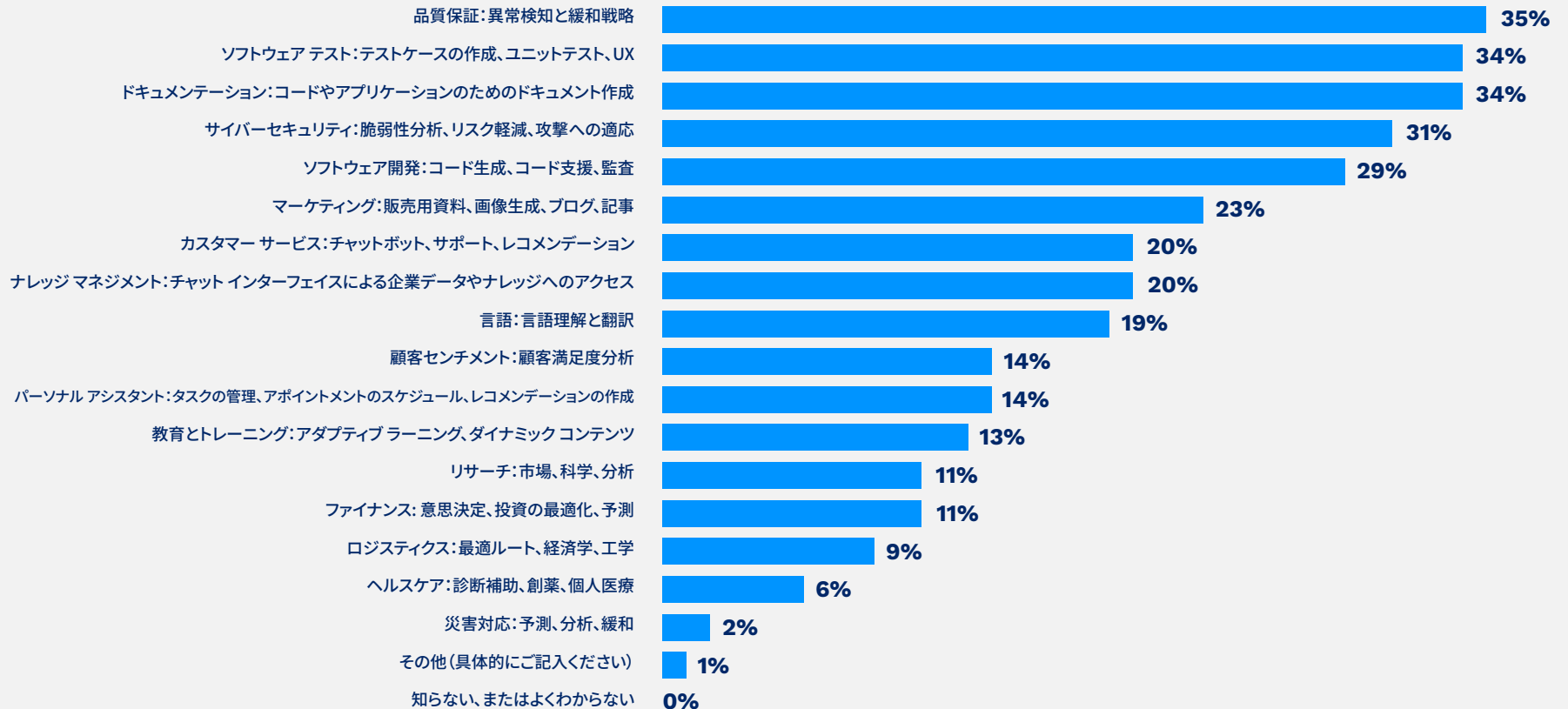


図 4

## 大半の企業は、生成 AI テクノロジーを自社のニーズに合わせてカスタマイズすることを意図している

2023 年 生成 AI に関する調査, Q10 and Q14, サンプルサイズ=245, セルの合計は 100%

貴社では生成 AI をどのように使用していますか、またはどのように使用する予定ですか？（該当するものをすべて選択してください）

	既成の生成AIテクノロジーをほとんどカスタマイズせずに使用	生成AIテクノロジーを使い、ニーズに合わせて幅広くカスタマイズ	自社独自の生成AIテクノロジーを開発	行 合計
社内のプロセス、ワークフロー、タスクなどを強化するために生成AIを使用する	4%	7%	5%	16%
提供する製品やサービスに生成AIを組み込む	6%	18%	6%	29%
生成AIをベースにして新商品や新サービスを創出する	3%	33%	19%	55%
列 合計	13%	57%	30%	

私たちの調査では、図4に示すように、企業が生成AIをどのように利用する予定かを調査することで、生成AI活用における企業のステージを評価しました。私たちは、2つの次元で最も進んだ生成AIの利用によって企業をグループ分けしました。行の合計から、サンプルの組織が生成AIを使用して内部プロセスを強化するだけでなく、製品やサービスに生成AIを組み込む（29%）、または新しい生成AIベースの製品やソリューションを作成する（55%）を目指していることを明らかにしました。カスタマイズのレベルに関しては、多くの組織が生成AIの基盤モデルをカスタマイズし強化することを計画していることがわかります（57%）。また、かなりの数が社内で生成AI技術を開発する予定です（30%）。社内でソリューションを開発することは、必ずしもLLMや他の大規模な基礎モデルをゼロ

から構築することを意味するわけではありません。高価で希少なリソースを必要とし、特定のユースケースをうまく使うことができない可能性があるからです。企業には、データサイエンスの専門知識を活用して、自社のデータセットから小規模でドメイン固有の生成AIモデルを構築する機会もあります。カスタマイズと社内ソリューションの開発の両方は、おそらくオープンソースコミュニティに依存することになるでしょう。オープンソースコミュニティは、計算コストを抑えた微調整アプローチであるLongLoRAのような技術で、生成AIのカスタマイズの課題に対する解決策を生み出してきました。



## 生成 AI のオープン性

オープンソース ソフトウェアは、ソフトウェアがオープンに開発されることを保証することで、大きな利益をもたらします。この属性は、ソフトウェアの学習、使用、共有、改善に対する障壁を取り除きます。これはまた、より自律性、透明性、およびコラボレーションをもたらす可能性があり、生成 AI に適用された場合、ユーザーは信頼性と透明性の高い AI システムを開発する自由を確保することができます。次のセクションでは、調査の生成 AI のオープン性に関する質問の結果を掘り下げます。

現在利用可能な様々な生成 AI モデルによって、オープン性のレベルは大きく異なる可能性があります。基礎となるコード、データ、モデル、ドキュメントへの可用性とアクセスはまれであるため、それらのほとんどはオープンソースの称号を得ることはないでしょう。<sup>1</sup>しかし、生成 AI のエコシステムはモデルに限定されず、ベクトルやグラフ データベースからエージェント フレームワークまでのアプリケーションを含みます。例を挙げると、企業はクローズド モデルの上にオープンソースのアプリケーション開発フレームワーク（例えば、LangChain）を

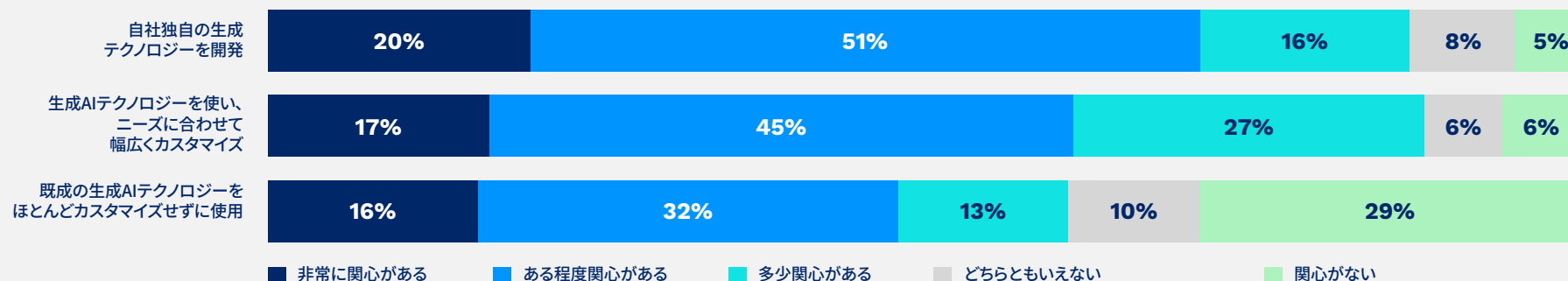
1 Andreas Liesenfeld, Alianda Lopez, and Mark Dingemans. 2023. Opening up ChatGPT: Tracking openness, transparency, and accountability in instruction-tuned text generators. In Proceedings of the 5<sup>th</sup> International Conference on Conversational User Interfaces (CUI '23). Association for Computing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

図 5

### 既存の生成 AI システムのオープン性に対する懸念、特に自社ソリューションをカスタマイズまたは開発する企業における懸念

2023 年 生成 AI に関する調査, Q14 by Q13, サンプルサイズ = 247

貴社は生成 AI 技術をどのように利用していますか、またはどのように利用する予定ですか。(該当するものをすべて選択してください) segmented by あなたの組織は、開発または使用している生成 AI システムのオープン性にどの程度関心がありますか? (1つ選択してください)



活用し、アプリケーションを統合し、オフィスシステムをバックアップし、新しいプラットフォームでイノベーションを起こす機会があります。したがって、オープン性は生成 AI の広い範囲にわたって活用することができます。オープンなアプローチは、私たちの調査回答者が使用している、または開発している生成 AI 技術のオープン性についての懸念によって確認されたように不可欠です。

図 5 は、組織が生成 AI テクノロジーを採用しようとする 3 つの方法（自社開発、あなたのニーズに合わせたカスタマイズ、ほとんどカスタマイズせずに使用）において、生成 AI のオープン性に対する懸念が組織の関与レベルと相関していることを示しています。図 5 では、71% の組織が、これから開発する生成 AI のオープン性について、中程度または非常に懸念しています。これは、今日のオープン性には大きなばらつきがあり、市場が成熟するにつれて業界が否定するようなアプローチに賭けるリスクがあるためと考えられます。あなたのニーズに合うように生成 AI システムをカスタマイズしようとする組織にも同様の状況が存在し、62% の組織がオープン性の必要性を中程度または非常に懸念しています。これとは対照的に、カスタマイズをほとんど行わずに使用される既製の AI テクノロジーのオープン性について、中程度または非常に懸念している組織は 48% に過ぎません。おそらくこれは、あなたの組織はその選択プロセスにおいてすでにデューデリジェンスを行っており、ベンダー / サプライヤーもまた、製品やサービスの品質と信頼性に対して最終的な責任を負っているためだと考えられます。

生成 AI のオープン性についての懸念は、オープンソースとプロプライエタリな生成 AI テクノロジーの間の組織の好みに反映されます。図 6 は、組織の 41% がオープンソースの生成 AI テクノロジーに傾いているのに対し、プロプライエタリなものを好むのは 9% であることを示しています。22% の組織は両方のタイプのソリューションを使用する傾向があり、28% は無関心です、テクノロジーの選択は、最終的にはこれらの嗜好を超えた要因に影響されることを示しています。

図 6

## 一般的に、オープンソースの生成 AI テクノロジーは、プロプライエタリなソリューションよりも好まれると、調査回答者は述べている

2023 年 生成 AI に関する調査, Q17, サンプルサイズ=249

あなたの組織は生成 AI について、プロプライエタリとオープンソースのどちらの配布モデルを好みますか？（1 つ選択してください）

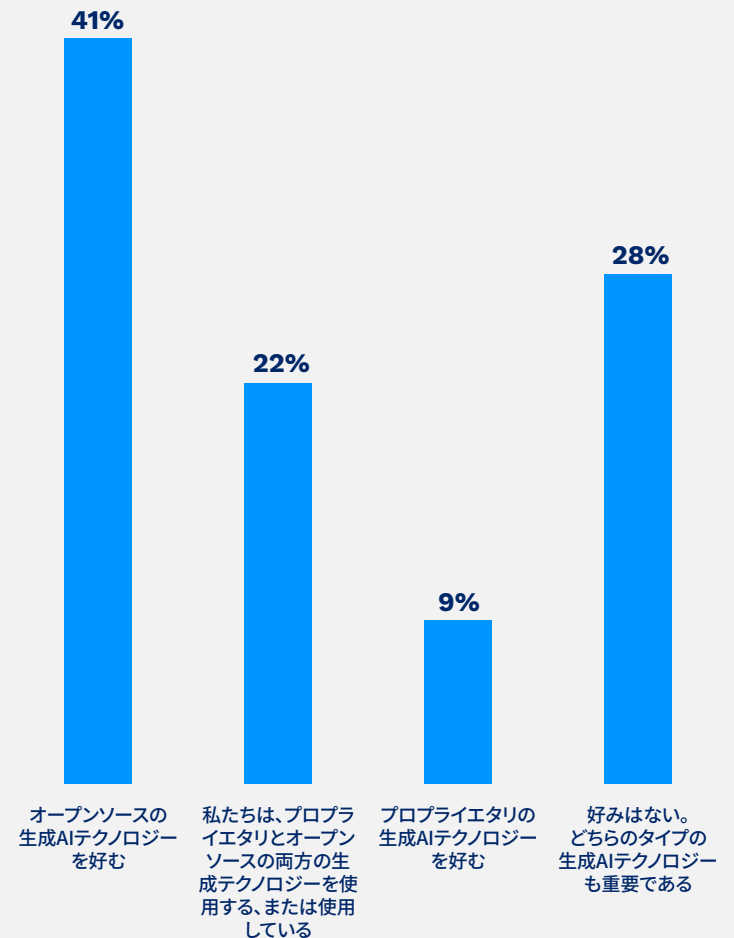
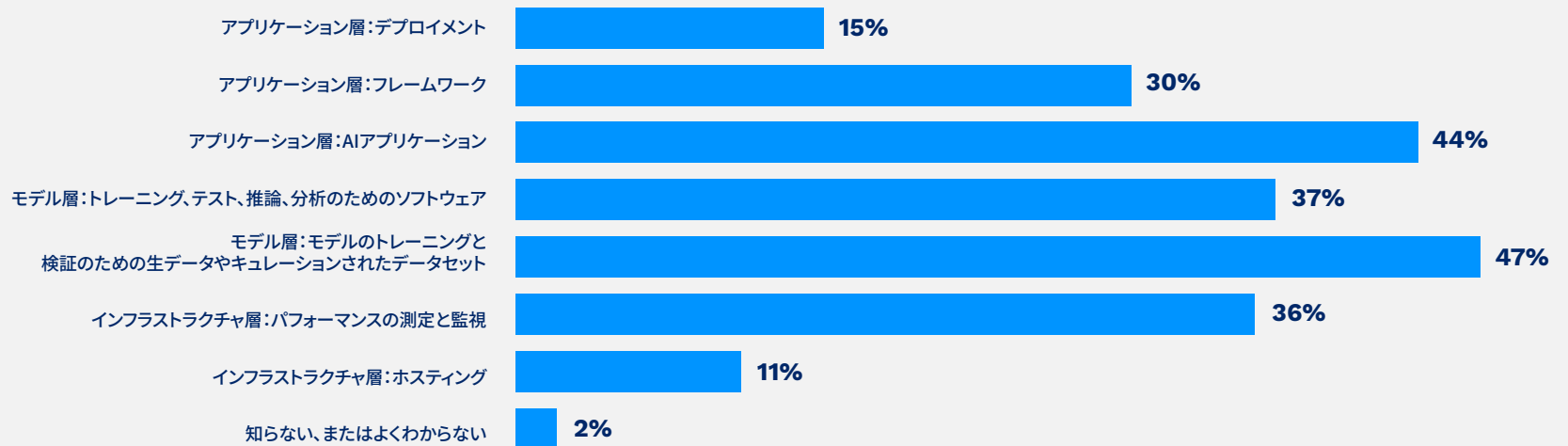


図 7

## 生成 AI スタックの 3 つの主要レイヤーにおいて、モデルレイヤーのオープンデータセットは、最も好まれるオープンソーステクノロジー

2023 年 生成 AI に関する調査, Q22, サンプルサイズ = 249

生成 AI スタックは一般的に、アプリケーション層、モデル層、インフラストラクチャ層の 3 つの主要レイヤーに分けられます。これらのレイヤーのうち、オープンソース技術に基づくべきコンポーネントはどれだと思いますか？（該当するものをすべて選択してください）



オープンソース ソフトウェアの定義がソースコードを中心に展開するのに対し<sup>2</sup>、オープンソース AI システムの定義は、生成 AI スタックを構成する様々なレイヤーを考慮する必要があります。私たちの調査では、アプリケーション層、モデル層、インフラ層の 3 つの主要な層を概説しました。図 7 は、回答者がオープンデータセットを最も評価していることを示しています (47%)。生成 AI のためのオープンデータセットは、イノベーションを加速し、コラボレーションを促進し、データの可用性を通じてバイアスを軽減することができます。調査回答者はさらに、

オープンソース テクノロジーが生成 AI モデルに基づくアプリケーション (44%)、トレーニングとテストのためのソフトウェア (37%)、インフラストラクチャ層におけるパフォーマンスの測定と監視のためのツール (36%) を改善できると述べています。生成 AI システムを分解し、そのオープン性を評価する方法は他にもあります：研究者は、様々な LLM の [オープントラッカー](#) を開発しました。

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

## セキュリティと信頼

### セキュリティは生成 AI 導入の大きな障壁

図 8 に見られるように、セキュリティ (49%) が生成 AI を採用する上で最も大きな障壁となっています。生成 AI に関するセキュリティ上の懸念の例としては、プライバシー、信頼、意図しない結果、データ侵害、悪用などが挙げられます。生成 AI システムは、設計上、学習し最適に動作させるために膨大な量のデータを取り込みます。このデータには、機密情報、安全でない情報、不正確な情報、または偏った情報が含まれている可能性があります。

また、モデルやインフラストラクチャの他の部分が、トレーニング、テスト、ま

たは検証プロセス中に導入された情報を不注意に開示しないようにするという課題もあり、機密情報の漏洩につながる可能性があります。生成 AI テクノロジーのセキュリティを確保することは、技術的な必要性だけでなく、信頼と規制遵守を維持するために極めて重要です。このことは、脆弱性を不明瞭にする可能性のあるこれらのブラックボックスモデルの複雑さによってさらに複雑になり、組織が潜在的なセキュリティリスクを完全に理解し、軽減することを困難にしています。回答者の一人は、生成 AI の課題に関する自由形式の質問で次のように答えています。”生成 AI は、消費者データを保護し、プライバシー規制の遵守を保証するために、より効果的にセキュリティ対策を講じる必要があるため、データセキュリティの懸念は、当社における最大の問題です。”

図 8

### 企業が生成 AI 関連プロジェクトを開始しない主な理由は、セキュリティ上の懸念

2023 年 生成 AI に関する調査, Q15, サンプルサイズ = 249

貴社が今後 1 年間に生成 AI 関連のプロジェクトを展開または開始する予定がないとしたら、そのおもな理由は何ですか? (該当するものをすべて選択してください)

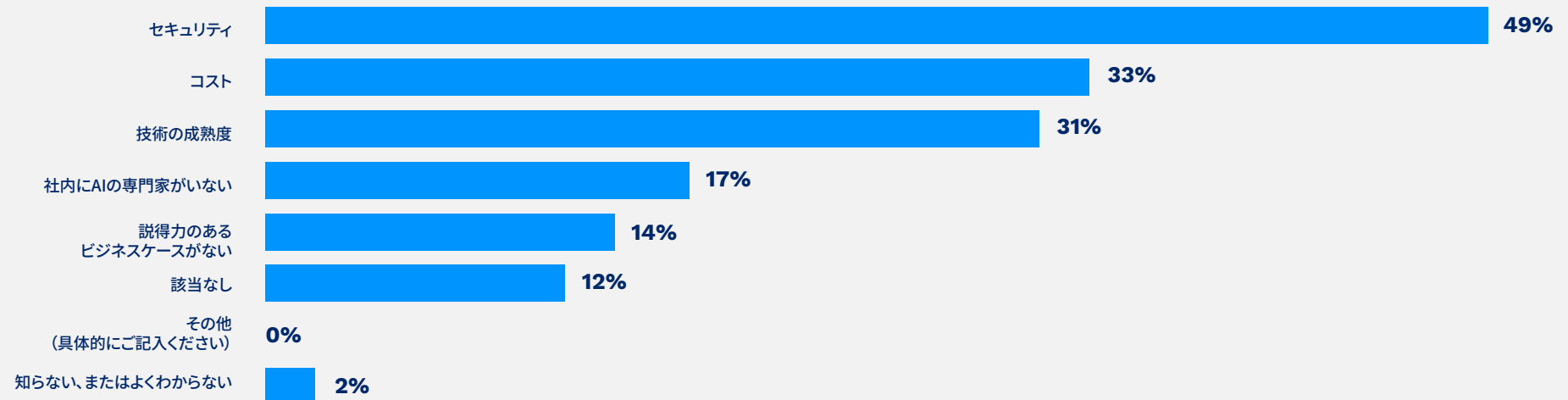
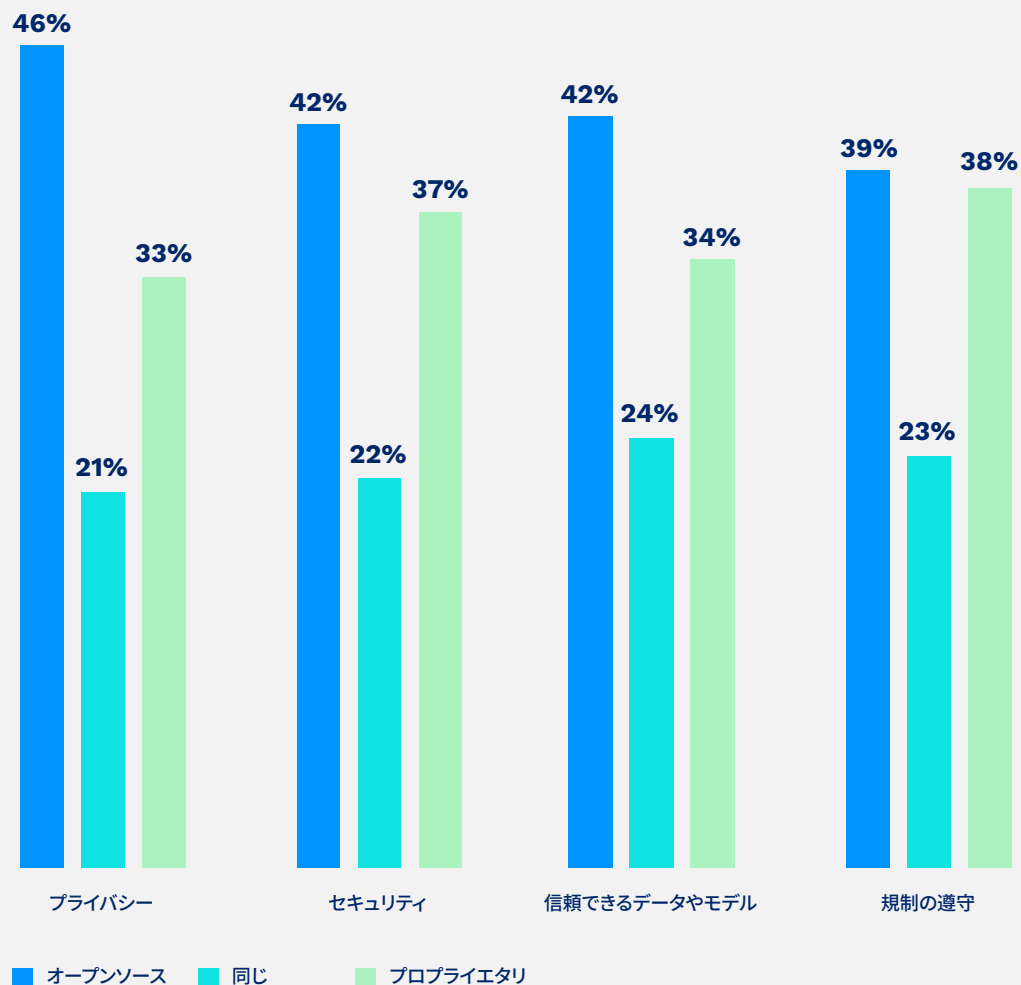


図 9

## セキュリティ問題を考慮した場合、企業がオープンソースのソリューションよりもプロプライエタリな生成 AI テクノロジーを好むという証拠は見つからなかった

2023 年 生成 AI に関する調査, Q18 and Q19, サンプルサイズ = 249

以下の各項目について、どのタイプの生成 AI ソリューションを選びますか？ (1 行につき 1 つ 選択してください)



## セキュリティを考慮した独自の優先順位を示す証拠は見つからなかった

図 8 を議論する際に言及されたセキュリティ上の懸念に対処することは極めて重要ですが、プロプライエタリなソリューションがこれらの問題を効果的に解決することは保証されていません。私たちは、アンケート回答者に、4 つの個別のセキュリティ懸念について、オープンソースとプロプライエタリな生成 AI ソリューションのどちらを好むかを検討してもらいました。

図 9 は、生成 AI テクノロジーにおけるセキュリティ、プライバシー、および規制コンプライアンスという重要な検討事項に関しては、企業の間でオープンソースの選択肢よりもプロプライエタリなソリューションを好む傾向が強いという実質的な証拠はないことを明らかにしています。図 9 は、回答者がプロプライエタリよりもオープンソースを好む傾向にあることを示していますが、調査の誤差範囲では、2 つの選択肢の間に大きな差は見られません。この結果は、プロプライエタリのソリューションの方が規制に準拠しており、生成 AI 開発にとって安全であると主張する論拠に疑問を投げかけるものです。

# 透明性とアクセシビリティ

## オープンソースの生成 AI がデータ管理と透明性を向上

生成 AI モデルのオープン性は、一般市民や学者が AI モデルを精査する機会を提供します。生成 AI モデルがどのように意思決定を行うかについての理解や透明性の欠如は、自分のデータがどのように使用されているかを知る個人の権利を妨げます。説明責任を果たすための適切なメカニズムがなければ、プライバシーが一貫して守られることを保証するのは困難です。あるアンケートの回答者は、透明性に関する自由形式の質問に対して、“私たちの会社は、AI テクノロジーの使用について透明性を保ち、AI システムがどのように意思決定を行うかについて明確な説明を提供することによって、顧客との信頼関係を構築することを

優先しなければなりません（中略）一部の顧客は、AI システムとの対話をためらい、人間との対話を好むかもしれません”と回答しています。図 10 に見られるように、69% もの企業が、オープンソースの生成 AI テクノロジーを利用することで、あなたの組織のデータ管理と透明性が向上すると考えています。

図 10

## オープンソース生成 AI テクノロジーによるデータ管理と透明性の向上に関する合意

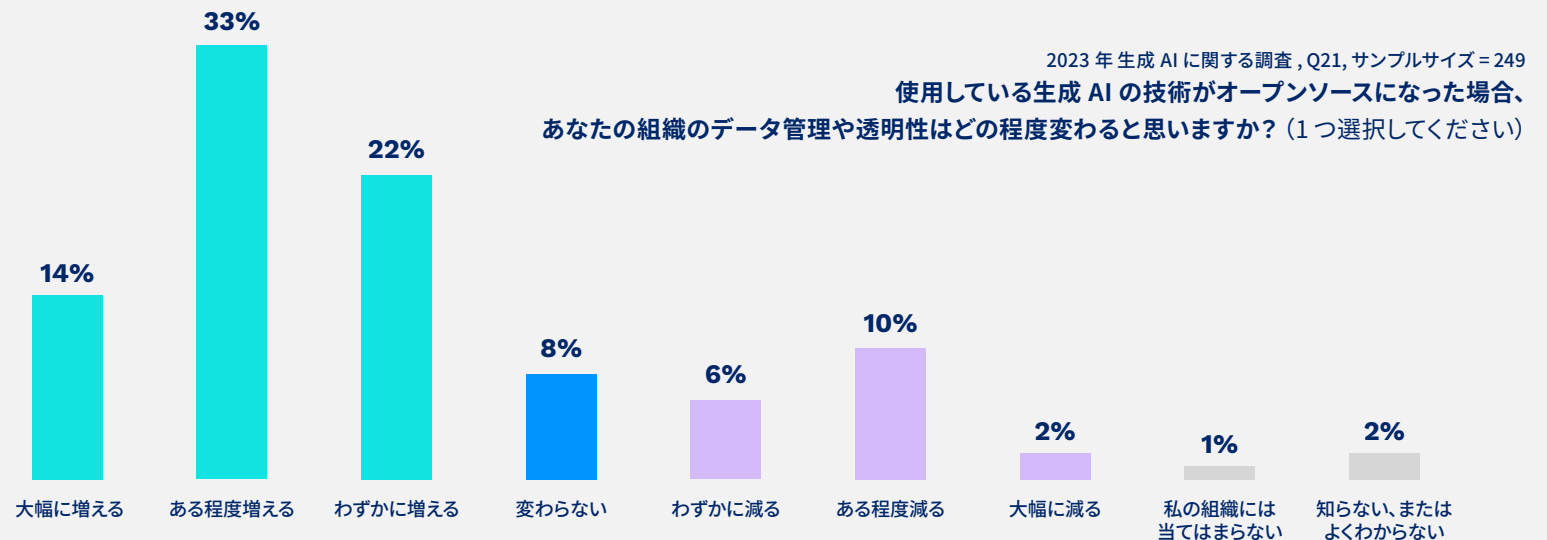
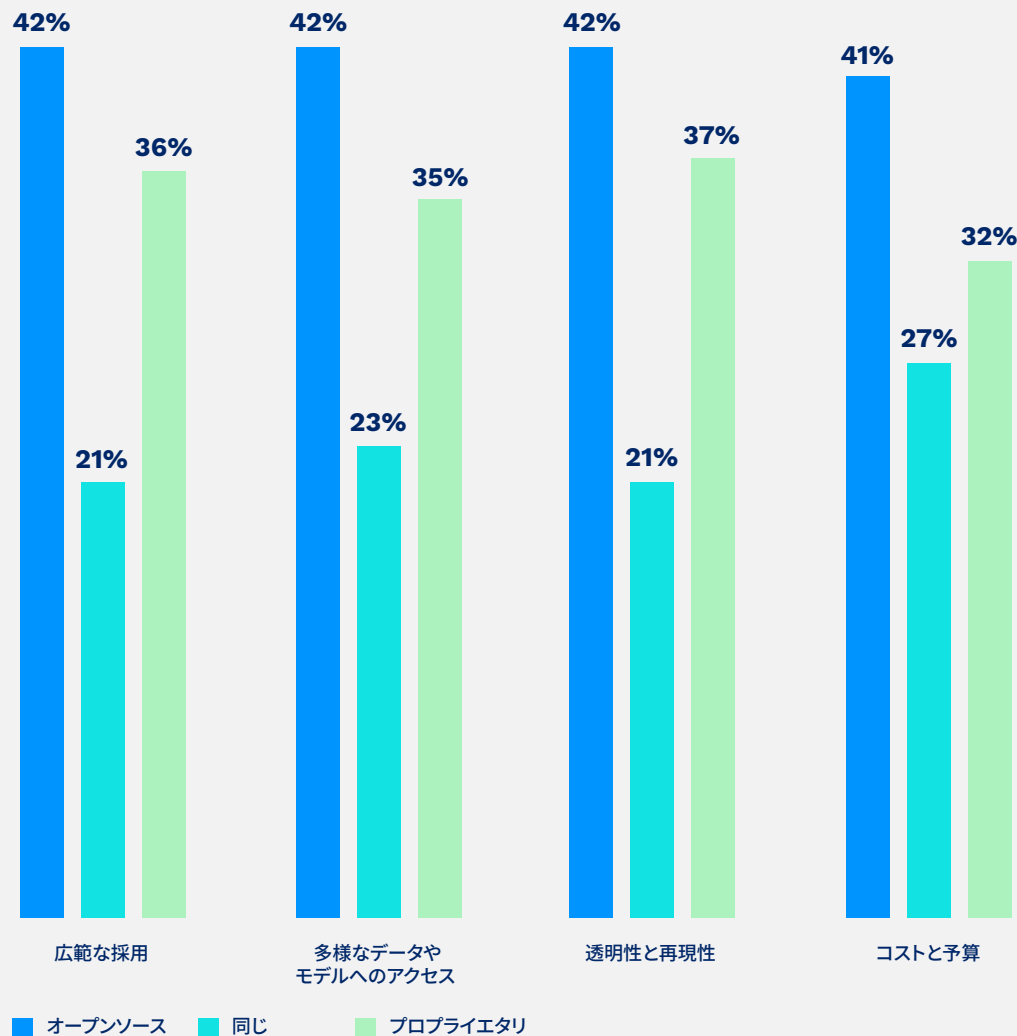


図 11

## オープンソースを生成 AI の採用、アクセシビリティ、再現性のためのソリューションとして評価

2023 年 生成 AI に関する調査 , Q18 and Q19, サンプルサイズ = 249

以下の各項目について、どのタイプの生成 AI ソリューションを選びますか？  
(1 行につき 1 つ選択してください)



## 生成 AI のアクセシビリティと再現性のためのソリューションとしてのオープンソースの評価

図 11 は、生成 AI の採用に関連する 4 つの考慮事項に対する好みを示しています。オープンソース モデルは、生成 AI の普及に有利であると考えられています。その理由は、アクセスしやすいことと、共同作業の機会が提供され、幅広いユーザー層への迅速な普及と反復が可能になるからです。

多様なデータやモデルへのアクセスについては、プロプライエタリなもの 35% であったのに対し、オープンソースが 42% であったことから、オープンソースはデータやモデリングの選択肢の豊富さにつながることを示唆されます。これは、データセットの多様性により、より堅牢で偏りの少ない AI システムを実現できる AI 開発において非常に重要です。

42% が透明性と再現性のためにオープンソースを好むという結果は、AI コミュニティにおいてオープン性が重視されていることを強調しています。透明性は信頼を築き、AI システムの独立した検証を可能にする鍵であり、再現性は科学の進歩と結果の検証に不可欠です。

コストと予算の観点から、プロプライエタリ (32%) よりもオープンソース (41%) が好まれることは、オープンソースのソリューションがよりコスト効率に優れていると認識されていることを示しています。これは、特に予算の制約が要因となっている場合に、組織が AI テクノロジーへの投資の効率を最大化しようとしている状況において、特に関連性があります。

図 12

## 生成 AI テクノロジーの開発に注力している企業は、中立的なガバナンス アプローチの採用が重要であると考えている

2023 年 生成 AI に関する調査, Q24, サンプルサイズ=72

生成 AI テクノロジーを開発する上で、中立的なガバナンスのオープンソース アプローチを持つことはどの程度重要ですか？ (1 つ選択してください)

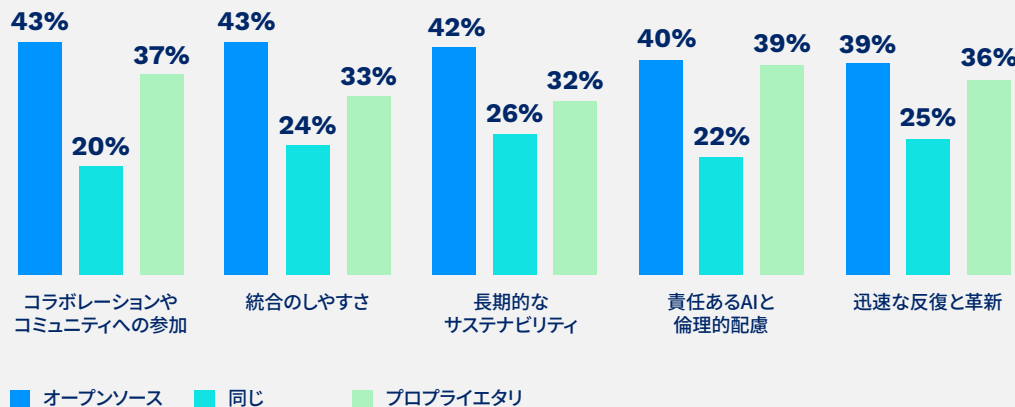


図 13

## 中立的なガバナンスの下、オープンソースの生成 AI テクノロジーは、責任あるイノベーションを実現する可能性を秘めている

2023 年 生成 AI に関する調査, Q18 and Q19, サンプルサイズ=249

以下の各項目について、どのタイプの生成 AI ソリューションを選びますか？ (1 行につき 1 つ選択してください)



## 中立的なガバナンスと責任あるイノベーション

### 生成 AI テクノロジーには中立的なガバナンスアプローチが重要

生成 AI テクノロジーにとって透明性とアクセシビリティが重要であるのと同様に、オープンソースは生成 AI から関連付けられるリスクを軽減するのに十分ではないかもしれません。図 12 は、中立的なガバナンスアプローチが調査回答者にとって重要であることを示しており、88% が生成 AI テクノロジーを開発する際に非常に重要または非常に重要であると回答しています。中立的なガバナンスは、真のオープンソース モデルのもう一つの側面であり、さまざまな方法で生成 AI テクノロジーに利益をもたらします。中立的なガバナンスは、イノベーションが一部の企業だけに左右されないようにするために重要です。さらに、中立的なガバナンスは、テクノロジーの悪用を防ぐための倫理的なスタンダードやガイドラインの設定にも役立ちます。中立的なガバナンスは、我々の調査で調査された様々な考慮事項と結びついています。

図 13 では、コラボレーションとコミュニティへの参加 (43%)、長期的な持続可能性 (42%)、責任ある AI と倫理的配慮 (40%) の各分野で、オープンソースソリューションへの傾倒を示しています。このようなガバナンスは、プロプライエタリなシステムの利益に縛られないため、開発プロセスにおける多様性と包括性を促進する公平なフレームワークを提供します。中立的なガバナンスは、イノベーションと反復が迅速であるだけでなく、倫理的に整合し、長期にわたって持続可能であることを保証し、テクノロジーをよりアクセスしやすくし、生成 AI 分野でより公平な成果をもたらす可能性があります。



# パフォーマンスとビジネス ニーズ

## 精度と拡張性は、オープンソースとプロプライエタリな生成 AI の同レベルと判断

生成 AI の有効性は、多くの場合、精度やスピードなどのパフォーマンス指標に基づいて企業によって評価されます。図 14 は、主要なビジネスニーズに関連して、オープンソースとプロプライエタリの生成 AI テクノロジーの比較嗜好をハイライトしたものです。このデータから明らかのように、オープンソースとプロプライエタリなソリューションの好みは、さまざまな技術的な検討事項にわたって密接に一致しています。例えば、オープンソースとプロプライエタリのソリューションは、その精度の点でほぼ同じように好まれており、プロプライエタリが 36%、オープンソースが 35% となっています。同様のパターンは、サポートと

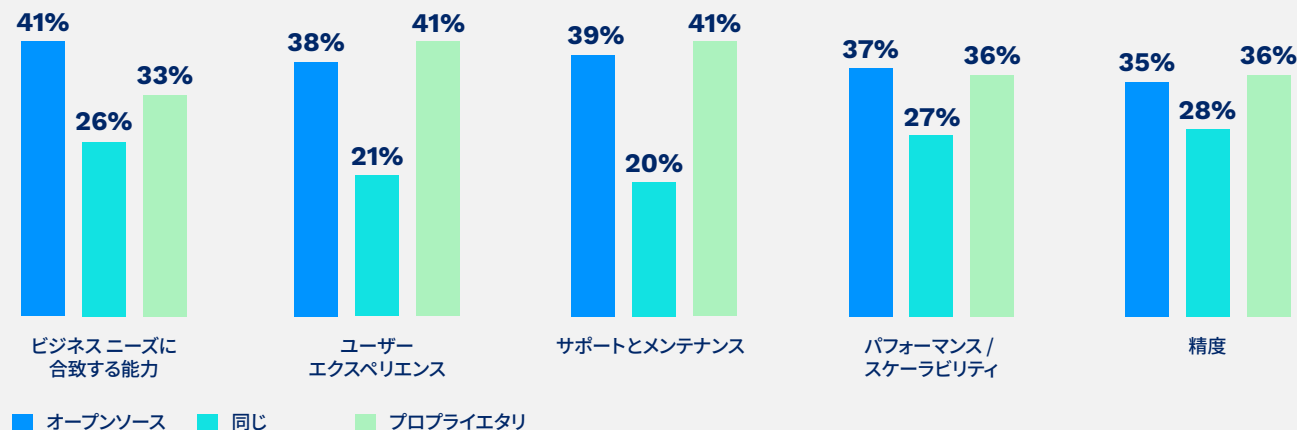
メンテナンス、パフォーマンス / スケーラビリティなどの他のカテゴリーでも観察されます。ユーザー エクスペリエンスに関しては、プロプライエタリなソリューション (41%) とオープンソースのソリューション (38%) を好む回答者がわずかに多くなっています。このように好みバランスよく分散していることから、重要な技術的ニーズを満たす上で、オープンソースのソリューションがプロプライエタリなものと同様に有利であると考えられている競争環境があることがわかります。

図 14

## スケーラビリティや正確さといったビジネス ニーズに関して、オープンソースの生成 AI テクノロジーとプロプライエタリなソリューションが同程度に支持されている

2023 年 生成 AI に関する調査, Q18 and Q19, サンプルサイズ = 249

以下の各項目について、どのタイプの生成 AI ソリューションを選びますか? (1 行につき 1 つの回答)



## 結論

### 企業が懸念する生成 AI テクノロジーのオープン性

この調査では、生成 AI システムのオープン性に関する回答者の強い懸念が明らかになりました。回答者の約 3 分の 2 がこの点を非常に懸念しているか、中程度に懸念しており、テクノロジー展開における透明性と管理の重要性を反映しています。回答者の 69% によると、オープンソースの生成 AI は、倫理的で責任ある AI 開発にとって重要なデータ管理と透明性の向上につながると回答しています。

### 調査の回答者は、一般的にオープンソースの方向に傾いている

私たちの調査結果は、生成 AI に対する組織の現在の態度や好みについて説得力のある洞察を提供し、特にオープンソース ソリューションへの顕著な傾向をハイライトしています。この発見は、透明性、再現性、多様なデータやモデルへのアクセス、統合の容易さなど、オープンソース技術に関連する利点の認識をハイライトしたものです。どのようなテクノロジー導入においても重要な懸念事項であるセキュリティは、オープンソースの生成 AI 採用の抑止力にはなっていないようです。実際、ほとんどの回答者は、プロプライエタリなソリューションがオープンなソリューションよりもセキュリティに適しているとは考えていません。

### 中立的なガバナンス アプローチが生成 AI 開発の鍵

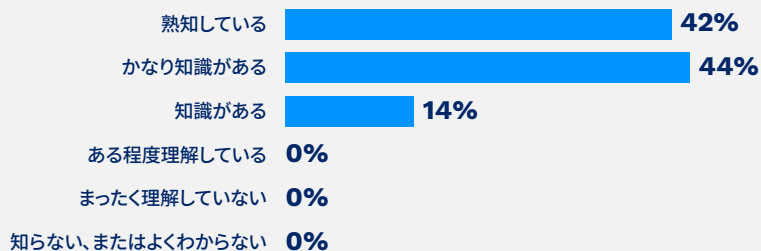
生成 AI における中立的なガバナンスの重要性は、アンケートの回答者の 95% が支持しています。このガバナンスの枠組みは、コミュニティの参加と協力を通じて、より倫理的で公平な生成 AI 技術の発展を保証します。中立的なガバナンスは、生成 AI の責任ある成長を促進するだけでなく、その恩恵が広く行き渡り、社会的価値と一致することを保証するためにも極めて重要です。このアプローチは、生成 AI の進歩の完全性と持続可能性を維持し、コミュニティと利害関係者の両方に役立つことを保証するために不可欠です。

図 15

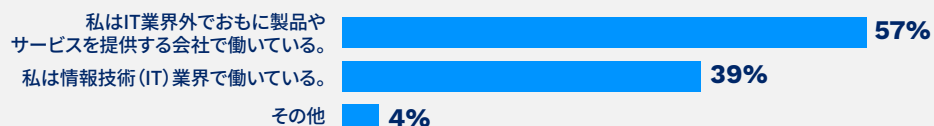
## 選択された人口統計データ

2023 年 生成 AI に関する調査, Q1, Q3 and Q4, サンプルサイズ = 284

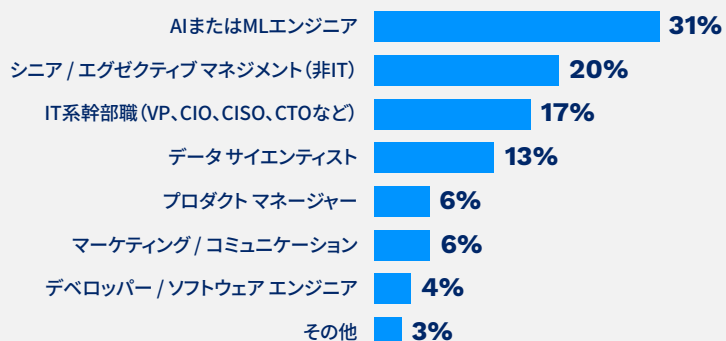
あなたの組織におけるジェネレーティブ AI (生成 AI) の導入について、どの程度ご存じですか? (1つ選択してください)



あなたの勤務先企業について、最もあてはまるものをお答えください。(1つ選択してください)



あなたの職業上の役割について、最も適切なものはどれですか? (1つ選択してください)



## この調査について

2023 年 9 月から 10 月にかけて、LF AI & Data と Linux Foundation Research は、生成 AI に関連するさまざまな質問について、組織の個人を対象にオンライン調査を実施しました。この調査は、LF のソーシャル メディアや LF のイベントで宣伝されました。また、より多様なサンプルを作成するために、サードパーティのパネル プロバイダーから適格な回答者を集めました。

### 調査方法

284 名の有効な調査票を入手し、249 名の回答者がすべての質問に回答しました。サンプル数 249 の誤差は、90% 信頼水準で 5.2% です。このサンプル数は、さまざまなスクリーニングおよびフィルタリング基準を満たした回答者を反映しています。主なスクリーニング基準には、雇用状況 (学生、無職、または退職している回答者は失格)、組織の生成 AI 採用への精通度 (まったく知らない、少し知っている、「知らない、わからない」と回答した回答者も失格) が含まれていました。

四捨五入の関係上、本レポートのパーセンテージの合計が 100% にならない場合があります。

### 人口統計

図 15 と図 16 は、調査サンプルの人口統計です。

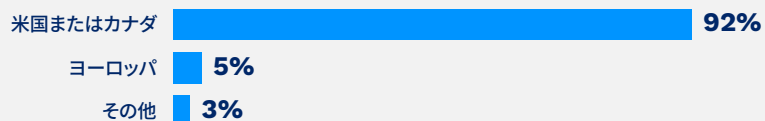
図 15 の左側のパネルでは、回答者の 42% が生成 AI を「熟知している」で、44% が「かなり知識がある」です。「ある程度理解している」、「まったく理解していない」、「知らない、またはよくわからない」という回答者が少ないのは意図的なものです。というのも、この質問は、回答者が信頼できる視点と洞察を提供できるように、私たちのスクリーニング プロセスの一部だったからです。回答者の 86% が生成 AI を「熟知している」または「かなり知識がある」と回答していることから、このような高いレベルの専門知識が、本調査で提供される品質と洞察力を向上させると考えています。図 15 の中央のパネルは、回答者が業種間でうまく分散していることを示しており、57% がエンドユーザー組織 (IT を使用、あるいは組み込んでいるが、主に IT 以外の業界に焦点を当てた製品やサービスを提供している企業) に勤務し、39% が IT ベンダーやサービスプロバイダーに勤務しています。図 15 の右側のパネルでは、回答者が AI または ML エンジ

図 16

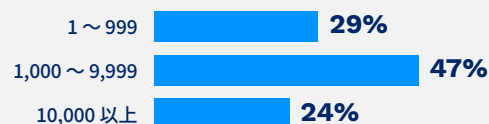
## 人口統計データ

2023 年 生成 AI に関する調査, Q5, Q6 and Q7, サンプルサイズ = 284

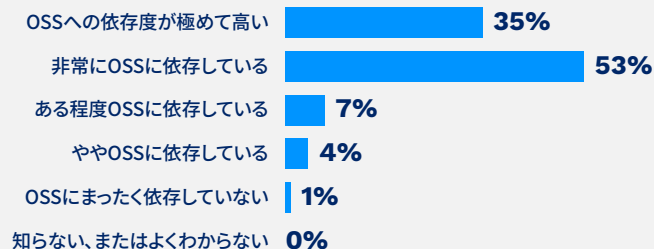
貴社の本社はどの地域にありますか? (1つ選択してください)



貴社の総従業員数をお答えください。(1つ選択してください)



貴社はオープンソースソフトウェア(OSS) にどの程度依存していますか? (1つ選択してください)



ニア (31%)、非 IT シニア / エグゼクティブ マネージャー (20%)、IT マネージャー (17%)、データサイエンティスト (13%) など、さまざまな役割に分散していることが示されています。

図 16 は、この人口統計データの続きです。図 16 の左側のパネルは、地域別の分布を示しています。地域別に層別化する努力はしておらず、その結果、ほとんどの回答が米国またはカナダからのものです (92%)。図 16 の中央のパネルは、回答組織の従業員規模別分布を示しています。回答者は、1 ~ 999 人 (29%)、1,000 ~ 9,000 人 (47%)、10,000 人以上 (24%) の 3 つのグループに適度に分散しています。図 16 の右側のパネルでは、ほとんどの組織がオープンソースソフトウェアに依存しており、53% が「非常に依存している」、35% が「非常に依存している」と回答しています。

## Data.World アクセス

LF Research では、実証プロジェクトの各データセットを Data.World で公開しています。このデータセットには、調査票、生の調査データ、スクリーニングとフィルタリングの基準、調査の各質問の度数表が含まれています。このプロジェクトを含む LF リサーチのデータセットは、[data.world/thelinuxfoundation](https://data.world/thelinuxfoundation) で見ることができます。

## 著者について

**ADRIENN LAWSON** は LF のデータ アナリストです。オックスフォード大学でソーシャル データ サイエンスの修士号を取得しました。調査開発、分析、レポート作成で LF Research をサポートしています。これまでにオックスフォード大学、ブダペスト政策分析研究所、英国国家統計局で調査を実施しました。

**Dr. MARCO GEROSA** は、ノーザンアリゾナ大学のコンピュータ サイエンスの正教授であり、LF Research のリサーチ アナリストです。主な研究分野はソフトウェア工学とオープンソース ソフトウェア。彼は現在、コンピュータ サイエンス教育と OSS コミュニティにおける新しい開発者のオンボーディングをサポートするツールとしての生成 AI の使用を調査しています。これらのプロジェクトは、全米科学財団の支援を受けており、一流の出版物で紹介されています。200 以上の論文を発表し、ICSE、FSE、MSR などの重要な会議のプログラム委員や、いくつかのジャーナルの査読者を務めています。現在、一流研究機関で研究者として活躍する博士号取得者や修士号取得者を輩出し、20 年以上の指導経験あり。詳細は <http://www.marcoagerosa.com>。

**STEPHEN HENDRICK** は、Linux Foundation の研究担当副社長であり、OSS が IT の生産者と利用者にとってイノベーションの原動力となることを Linux Foundation が理解する上で中核となるさまざまな研究プロジェクトの主任研究員を務めています。Steve は、ソフトウェア業界のアナリストとして 30 年以上にわたって培ってきた主要な調査手法を専門としています。Steve は、DevOps、アプリケーション管理、意思決定分析など、アプリケーション開発とデプロイメントに関するトピックの専門家です。市場ダイナミクスを深く洞察する

さまざまな定量・定性調査手法の経験を生かし、多くのアプリケーション開発・導入領域で先駆的な調査を行ってきました。1,000 以上の出版物を執筆し、シンジケート リサーチやカスタム コンサルティングを通じて、世界有数のソフトウェア ベンダーや著名な新興企業に市場ガイダンスを提供しています。

**Matt White** は、Linux Foundation's AI & Data Foundation の Generative AI Commons ディレクターです。彼はまた、Amdocs の AI & データ部門責任者であり、AI 研究グループ Berkeley Synthetic の創設者でもあります。カリフォルニア大学バークレー校で大学院生にデータサイエンスを指導しています。また、Linux Foundation の一部である Open Metaverse Foundation の共同設立者兼議長であり、Metaverse Standards Forum の理事でもあります。AI、データ、オープンソースに 25 年以上の経験あり。カリフォルニア大学バークレー校でデータサイエンス修士号、デンバー大学で MBA、ヨーク大学で IT 学士号を取得しています。詳細は [www.matt-white.com](http://www.matt-white.com)。

**Lucy Hyde** は、人工知能とデータのオープンソース イノベーションをサポートする機械学習を専門とするシニア プログラム マネージャーです。熟練したプロフェッショナルとして、データサイエンス、ソフトウェア エンジニアリング、技術開発に重点を置いた職務に従事し、技術的な専門知識で高く評価されています。彼女は現役軍人および政府民間人として国防総省でキャリアをスタートし、民間部門のデジタル フォレンジックに携わり続けました。心理学、アラビア語、諜報活動の学士号、分析学の修士号を取得し、ジョージ・メイソン大学で計算科学/情報学の博士号を、ジョンズ・ホプキンス大学で人工知能の修士号を取得中です。

## 謝辞

2023 年の生成 AI のあり方について洞察と経験を快く共有して下さった調査参加者の皆様に感謝いたします。特に調査プロセスの様々な段階に関わってくれた査読者と LF の同僚、Hilary Carter、Michael Dolan、Ibrahim Haddad、Anna Hermansen に感謝します。

この日本語レポートは、以下の文書の参考訳です。  
[2023 Open Source Generative AI Survey Report](#)  
翻訳協力：吉田行男

2021年に設立されたLinux Foundation Researchは、オープンソース コラボレーションの規模の拡大を調査し、新たな技術動向、ベストプラクティス、オープンソース プロジェクトの世界的な影響に関する洞察を提供しています。プロジェクトのデータベースやネットワークを活用し、定量的・定性的手法のベストプラクティスに取り組むことで、Linux Foundation Researchは、世界中の組織のためにオープンソースの知見を提供するライブラリを構築しています。



## LF AI & DATA

[lfaidata.foundation](https://lfaidata.foundation) | [genacommons.org](https://genacommons.org)



Copyright © 2023 [The Linux Foundation](https://www.linuxfoundation.org/)

本レポートは [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/) の下でライセンスされています。

本著作物を引用する場合は、以下のように記載してください。  
Adrienn Lawson, Marco Gerosa, and Stephen Hendrick,  
"2023 Open Source Generative AI Survey Report:  
Enterprise perspectives and survey-based insights at the  
intersection of open source innovation and generative AI  
advancements", foreword by Stella Biderman,  
The Linux Foundation, December 2023.

