IDC Innovators

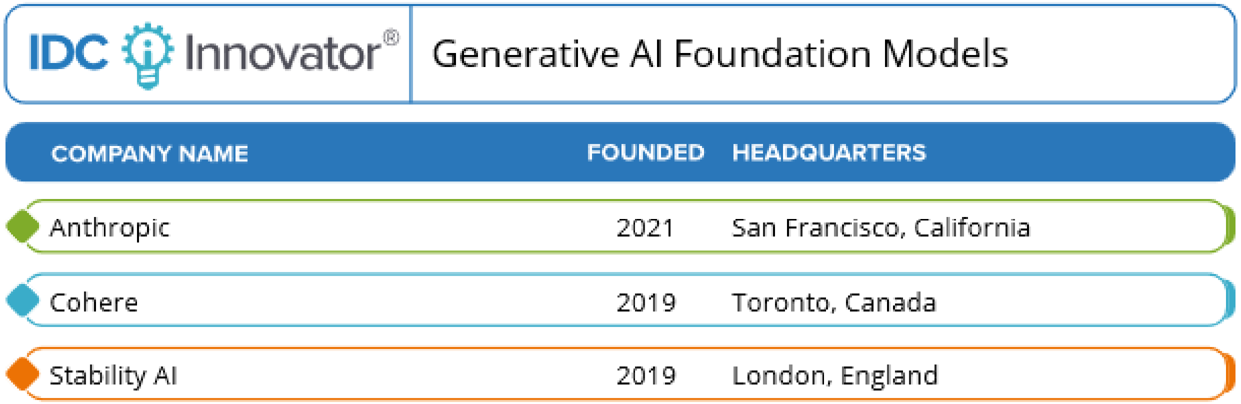
IDC Innovators: Generative AI Foundation Models, 2023 — Part 1

Ritu Jyoti Andrew Gens

IDC INNOVATORS VENDOR LIST

FIGURE 1

IDC Innovator in Generative AI Foundation Models, 2023



IDC Innovator® Generative AI Foundation Models

COMPANY NAME	FOUNDED	HEADQUARTERS
Anthropic	2021	San Francisco, California
Cohere	2019	Toronto, Canada
Stability AI	2019	London, England

Source: IDC, 2023

EXECUTIVE SUMMARY

IDC Innovators are emerging vendors with revenue under $100 million that have demonstrated either a groundbreaking business model or an innovative new technology – or both. This IDC Innovators study profiles three vendors in the generative AI foundation model market: Anthropic, Cohere, and Stability AI.

Generative AI foundation models are a rapidly emerging class of machine learning models that use large volumes of data to generate complete end products in text, image, or other modalities all via simple prompting. As demand for generative AI foundation models has grown rapidly in the past three years, the pool of vendors has expanded with many new entrants. Generative AI foundation models continue to mature quickly, and many new use cases are emerging every year, prompting even more market entrants. Despite this strong overall market growth, generative AI foundation model vendors face various challenges from stiff market competition to strong pressures to differentiate products. Despite these challenges, this market is still new, and this gives vendors an opportunity to adapt their products and business models to increase their reach, relevancy, and reliability. Generative AI

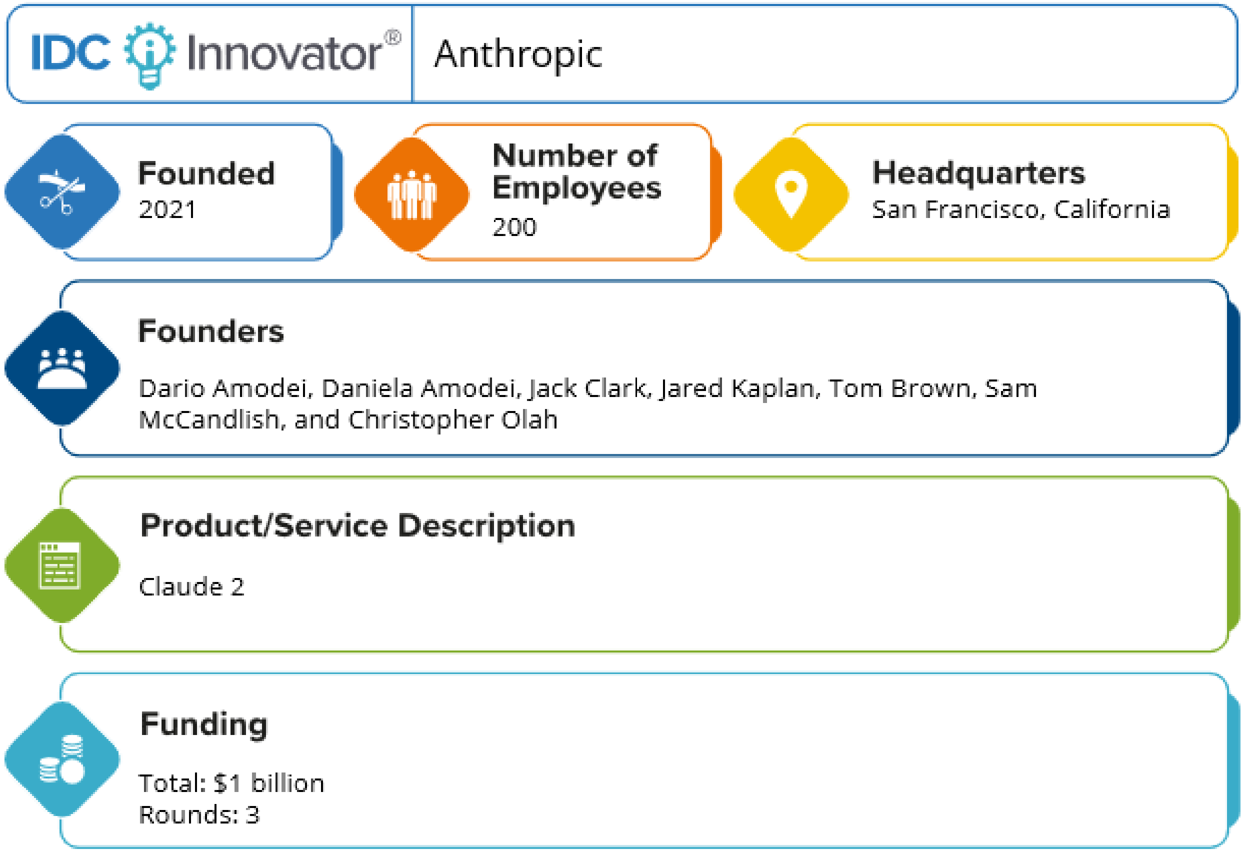
foundation models will only continue to increase in relevancy in the years ahead. As new innovators emerge in this market, potential buyers and industry observers alike should remain vigilant to identify emerging trends and methods.

VENDOR PROFILE: ANTHROPIC

.

FIGURE 2

Anthropic



IDC Innovator®

Anthropic

Founded
2021

Number of Employees
200

Headquarters
San Francisco, California

Founders

Dario Amodei, Daniela Amodei, Jack Clark, Jared Kaplan, Tom Brown, Sam McCandlish, and Christopher Olah

Product/Service Description

Claude 2

Funding

Total: $1 billion
Rounds: 3

Source: IDC, 2023

Why Anthropic Was Chosen as an IDC Innovator

Anthropic was chosen as an IDC innovator due its AI safety and research priorities, focusing on building reliable, interpretable, and steerable AI systems. With a commitment to responsible AI development, Anthropic specializes in developing general AI systems and language models, prioritizing safety and predictability. The company's interdisciplinary team of researchers, engineers, and policy experts explores various areas including natural language, reinforcement learning, interpretability, and human feedback. Through its research-driven approach, Anthropic aims to address the challenges of AI transparency and reliability while creating value for both commercial and public benefit. Its key offerings include Claude and Claude Instant.

IDC Innovator Assessment

- Anthropic's Claude is built using constitutional AI. Constitutional AI is a way to be explicit about the values injected into an LLM. Anthropic gave its AI system a series of ethical and behavioral principles in natural language to evaluate and modify its outputs to be more harmless.
- Claude 2 is a relatively low-cost platform with an approximate cost of $11 per million tokens with ~$32 per million output/completion tokens.
- Anthropic is one of three primary recipients, along with Aleph-Alpha and Cohere, of SAP's strategic investment in generative AI start-ups with SAP emphasizing ease of use, accessibility, and security and data privacy. SAP invested over $1 billion across these three companies in undisclosed totals.

Key Differentiator

A key differentiating factor for Anthropic is Claude 2's ability to help users with a range of tasks. With an industry-leading 100,000 token capacity, roughly equating to 75,000 words of text, Claude 2 can handle far longer prompts for utilization with processing enterprise documents, coding, and automating workflows. With text alone, Claude 2 is capable of editing, rewriting, summarizing, classifying, and so forth within the context of natural conversations. Going off of the token limit alone, Claude 2 has comparatively about three times as large a capacity as many of Anthropic's closest competition, giving the company a substantial lead.

Challenges

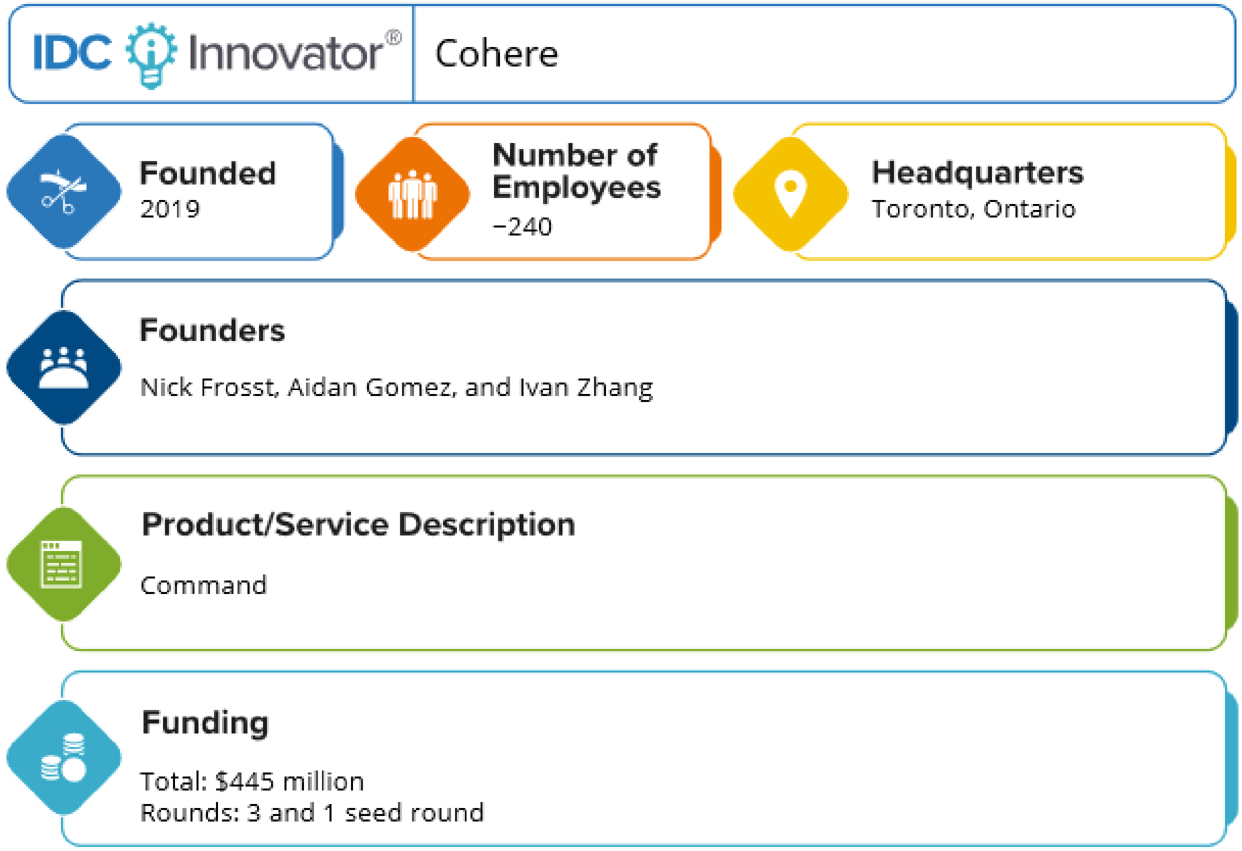
While Anthropic is a leader in prompt token capacity for LLMs, it also faces the challenge of trailblazing new use cases and facing the pitfalls that go along with that. Owing to the 100,000 token capacity of Claude 2, users can potentially use Claude 2 for ingesting and generating content based upon long-form business reports, scientific papers, or other intensive documents. With larger prompts and responses at play, there are also larger stakes. Larger prompts could present a barrier to explainability as well as amplifying the persistent issues of hallucination when applied across a larger output. This could leave Anthropic, and its customers, potentially handling new LLM problems that emerge in real time, necessitating organizational agility.

VENDOR PROFILE: COHERE

.

FIGURE 3

Cohere



IDC Innovator® | Cohere

Founded
2019

Number of Employees
-240

Headquarters
Toronto, Ontario

Founders

Nick Frosst, Aidan Gomez, and Ivan Zhang

Product/Service Description

Command

Funding

Total: $445 million
Rounds: 3 and 1 seed round

Source: IDC, 2023

Why Cohere Was Chosen as an IDC Innovator

Cohere was chosen as an IDC Innovator due to its enterprise-focused generative AI LLMs, which enable companies to incorporate NLP into applications without extensive training or hiring, offering general-purpose language models for generation and representation, enabling the use in chatbots, digital personal assistants, and content moderation. Cohere offers two types of models: text generation and text representation. Command, Cohere's flagship model, is a text-generation model trained for business applications such as summarization, copywriting, dialog, extraction, and question-answering. Cohere's other primary offering is Cohere Embed for text representation.

IDC Innovator Assessment

- Cohere is one of three primary recipients, along with Aleph-Alpha and Anthropic, of SAP's strategic investment in generative AI start-ups, with SAP emphasizing ease of use, accessibility, and security and data privacy. SAP committed over $1 billion across several AI initiatives in undisclosed totals.

- Cohere Command, Cohere's flagship LLM housed within Cohere API, can be operated on a private cloud; a secure cloud through partners — AWS, Oracle, and Google; or Cohere's managed cloud solution. In addition to this, although Cohere is an out-of-the-box product, it can be integrated with organizational databases to increase the volume and relevance of data.
- Cohere Command Nightly was trained with approximately 52.4 billion parameters and can support up to 4,096 tokens in a prompt.

Key Differentiator

A key differentiator for Cohere is its commitment to weekly updates. In the highly competitive LLM industry today, there is a constant push and competition to improve model accuracy, reduce hallucination, and refine semantic understanding. Owing to this highly competitive atmosphere, Cohere's weekly updates sets it apart from other vendors that have slower processes of iteration by ensuring that the company's customers are getting the most advanced product available at almost real time.

Challenges

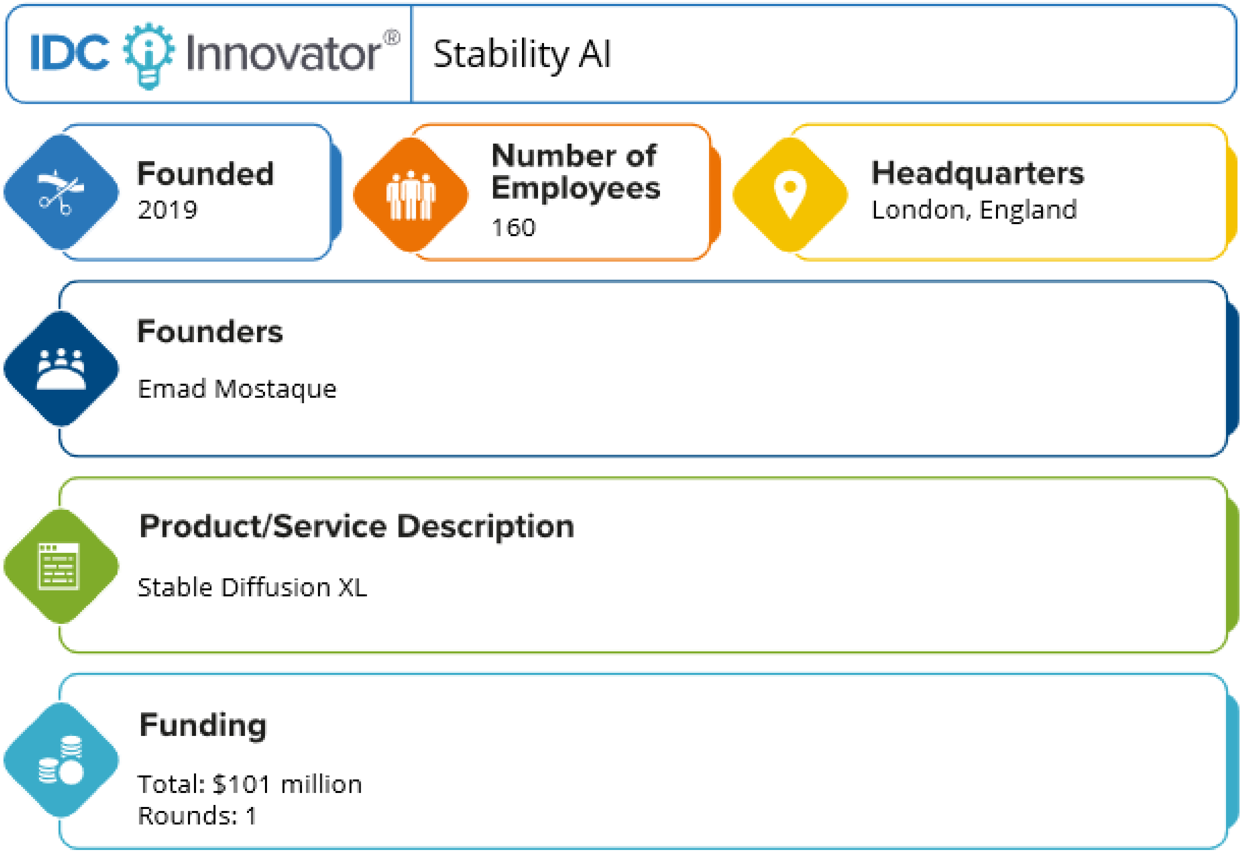
Regular updates present not only a great differentiator for Cohere for enterprises to be getting the latest model advancements but also a challenge for customers to coordinate versions. Cohere provides several model versions for developers to use in production and experimentation. Stable releases are most suitable for production and are released monthly. At the same time, new beta releases for experimentation are available weekly. Customers must plan an upgrade and experimentation strategy to take advantage of continuous model improvements. To try and mitigate this potential challenge, Cohere allows customers to pin the version they're using and can work with customers to make adjustments.

VENDOR PROFILE: STABILITY AI

.

FIGURE 4

Stability AI



IDC Innovator® | Stability AI

Founded
2019

Number of Employees
160

Headquarters
London, England

Founders

Emad Mostaque

Product/Service Description

Stable Diffusion XL

Funding

Total: $101 million
Rounds: 1

Source: IDC, 2023

Why Stability AI Was Chosen as an IDC Innovator

Stability AI was chosen as an IDC Innovator due to its cutting-edge Stable Diffusion Foundation Model, designed specifically for text-to-image generation. It sets itself apart by running efficiently on consumer hardware with standard GPUs, enabling broader accessibility. Stability AI provides users with API access and publicly releases the code, allowing users to leverage this innovative text-to-image technology with greater flexibility and transparency. Stability AI's core models include StableLM, Deepfloyd IF, and Stable Diffusion.

IDC Innovator Assessment

- Stability AI offers its Stable Diffusion XL text-to-image foundation model as an open source solution for configuration through Hugging Face as well as offering it through the company's own managed service, or as an out-of-the-box solution.

- With Stable Diffusion XL, users can use text prompting to create a new image or take a created image and, with no prompting, use digital interaction with image modification through a UI in order to adjust or expand the edges of an image with generated content. In addition, with Stable Diffusion XL, users are given the ability to insert text directly into an image.
- Stable Diffusion XL's base-level model has 3.5 billion parameters, while its second, and largest, model achieves 6.6 billion parameters utilizing an ensemble pipeline.

Key Differentiator

A key differentiator for Stability AI is Stable Diffusion's text to image and text within an image focus. While the majority of mainstream foundation models focus primarily on text-to-text prompting and answering, currently representing the largest business use cases, Stable Diffusion has chosen to go a less traveled path of primarily an image generation focus. While the use cases for text to image are less apparent on the face of it than text to text, many creative endeavors such as advertising through images, art for a number of commercial applications, product mock-ups, and so forth can be generated by Stable Diffusion XL.

Challenge

Owing to the current limited business use cases related to text-to-image generation, as compared with text to text, Stability AI will need to ensure that its product stands far and above broad multimodal LLMs such as GPT-4 and Claude 2. Stability AI will best serve customers that have text-to-image use cases, but customers that have more multimodal needs will need to thoroughly assess whether the capabilities of Stability AI make it worth having on top of a separate text-to-text LLM.

TECHNOLOGY DEFINITION

Generative AI foundation models are a class of machine learning models that are trained on diverse data and can be adapted or fine-tuned for a wide range of downstream tasks. Transformers and LLMs are subsets of foundation models.

IDC INNOVATORS INCLUSION CRITERIA

An "IDC Innovators" document recognizes emerging vendors chosen by an IDC analyst because they offer an innovative new technology or a groundbreaking business model, or both, and were approved by the IDC Innovators Review Panel. It is not an exhaustive evaluation of all companies in a segment or a comparative ranking of the companies.

An IDC Innovators document highlights vendors that meet the following criteria:

- In IDC's opinion, the company exhibits innovative technology or a new business model.
- The company has annual revenue under $100 million at the time of selection.
- Customers are currently using the company's products and services (i.e., the products and services are not conceptual or in the process of being released).
- The product, service, or business model must solve or help to alleviate an IT buyer challenge.

In addition, vendors in the process of being acquired by a larger company may be included provided the acquisition is not finalized at the time of publication of the document. Vendors funded by venture capital firms may also be included even if the venture capital firm has a financial stake in the vendor's company.

LEARN MORE

Related Research

- *Worldwide Artificial Intelligence Applications Market Shares, 2022: Embedded Artificial Intelligence Is Becoming Increasingly Important, Transforming Our Daily Life and Technology Usage* (IDC #US51181823, September 2023)
- *Worldwide Artificial Intelligence Software Forecast, 2023-2027* (IDC #US50027023, September 2023)
- *Generative AI: Preparing for the Long-Term Impact - Practical Applications and Guidance for CIOs/CTOs* (IDC #US51148223, August 2023)
- *Unlocking Business Success with Generative AI* (IDC #US50789223, June 2023)
- *Generative AI Platforms and Applications: Market Trends and Forecast, 2Q23* (IDC #US50700423, May 2023)
- *Generative Artificial Intelligence: A New Chapter for Enterprise Business Applications* (IDC #US50471523, May 2023)
- *Enterprise Automation 2.0: The Connective Tissue of the Digital Business* (IDC #DR2023_GS4_RJ, March 2023)

Synopsis

IDC Innovators are emerging vendors with revenue <$100 million that have demonstrated either a groundbreaking business model or an innovative new technology — or both. This IDC Innovators study profiles software vendors providing innovative generative AI foundation models, be it text, image, or multimodal foundation models. This market is growing rapidly, with many new entrants emerging as well as established vendors establishing new offerings in this space.

"Foundation models represent an important paradigm shift in artificial intelligence (AI) and is integral to the future of AI in the enterprise," says Ritu Jyoti, group vice president, Artificial Intelligence and Automation Research with IDC's Software Market Research and Advisory Practice. "To truly harness the potential of foundation models in the enterprise, there needs to be significant investment in the entire technology stack — not just the models themselves."

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology, IT benchmarking and sourcing, and industry opportunities and trends in over 110 countries. IDC's analysis and insight helps IT professionals, business executives, and the investment community to make fact-based technology decisions and to achieve their key business objectives. Founded in 1964, IDC is a wholly owned subsidiary of International Data Group (IDG, Inc.).

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com

Copyright and Trademark Notice

This IDC research document was published as part of an IDC continuous intelligence service, providing written research, analyst interactions, and web conference and conference event proceedings. Visit www.idc.com to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit www.idc.com/about/worldwideoffices. Please contact IDC report sales at +1.508.988.7988 or www.idc.com/?modal=contact_repsales for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or web rights.