Cool Vendors in AI Core Technologies

Published 23 December 2022 - ID G00776792 - 11 min read

By Analyst(s): Arun Chandrasekaran, Erick Brethenoux, Anthony Mullen, Pieter den Hamer

Initiatives: Artificial Intelligence

AI core technologies seek to expand the applications of AI with novel techniques, which can boost AI performance, scaling and accuracy. D&A leaders can consider these Cool Vendors to explore how causal AI, foundation models and decision intelligence can help them propel their initiatives.

Overview

Key Findings

- Causal artificial intelligence (AI) goes beyond correlations, capturing causal relationships and providing new insights. While the market is quite nascent currently, there are promising startups working on a broad set of enterprise applications for causal AI.
- Foundation models represent a huge change in the field of artificial intelligence, due to their massive pretraining and significant improvements in accuracy across a variety of tasks. Identifying the right use cases and creating safe deployment processes are key challenges that organizations need to navigate.
- There is growing organizational interest in decision intelligence (DI) to optimize decision flow. DI can identify inefficiencies, risks and areas for improvement by modeling the processes and workflows that inform a decision.

Recommendations

Data and analytics (D&A) leaders responsible for AI strategy and planning should:

- Augment and automate decision intelligence processes by educating their data science teams on causal AI, then using it to generate predictions and learning how to affect the outcomes that it predicts.

- Select the foundation model best suited to their use case by assessing the models' benefits, risks and use cases, using cloud-based APIs to consume them and prioritizing vendors that promote responsible deployment of models.
- Develop skills in decision augmentation and decision automation techniques by including descriptive, diagnostic, predictive and prescriptive capabilities to fully leverage the value of the DI platform.

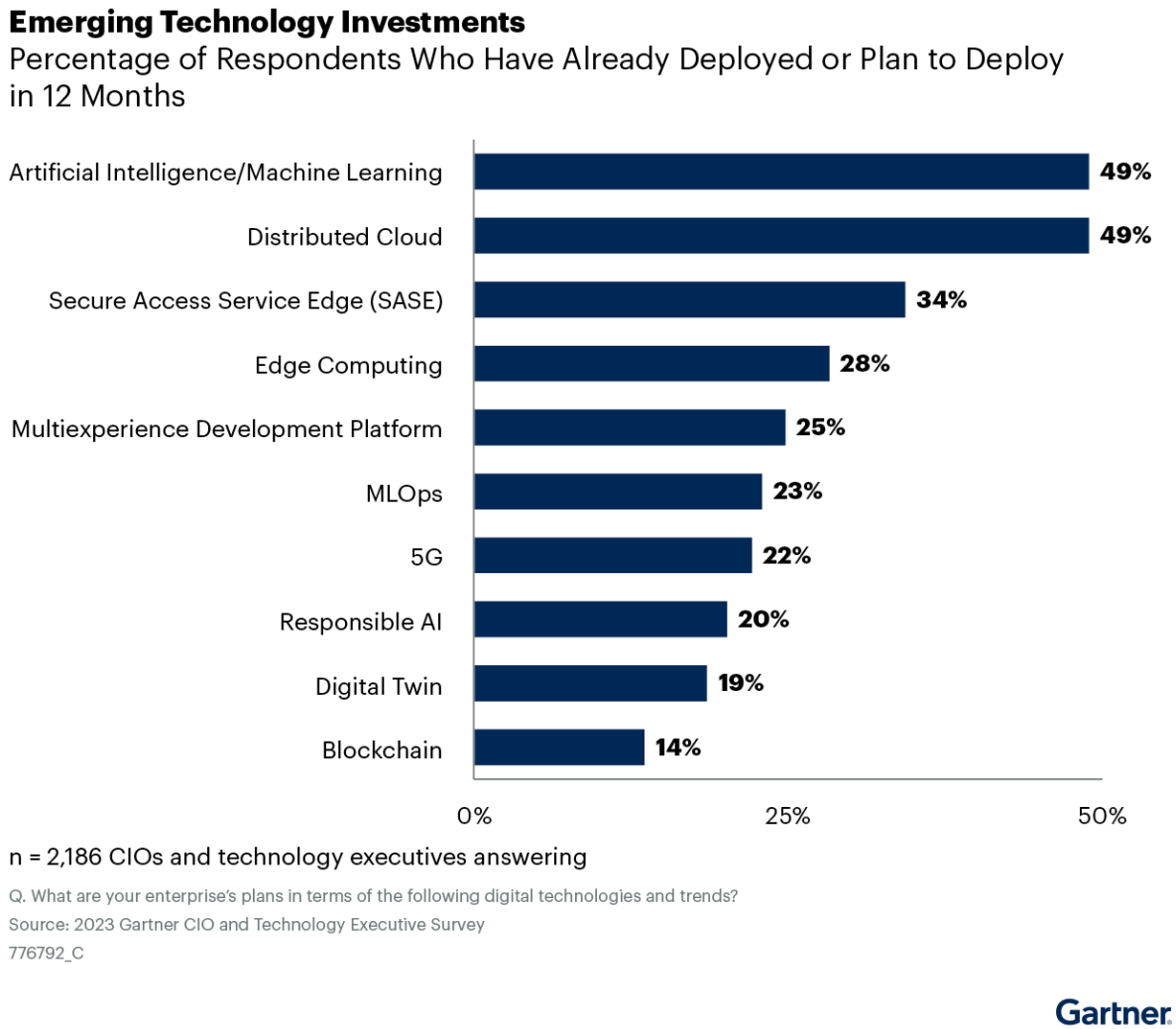
Analysis

This research does not constitute an exhaustive list of vendors in any given technology area, but rather is designed to highlight interesting, new and innovative vendors, products and services. Gartner disclaims all warranties, express or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

What You Need to Know

Gartner's 2023 CIO survey found that 49% of surveyed leaders indicated existing or imminent adoption of AI. [1] This is the highest rate of adoption among emerging digital technologies, as seen below in Figure 1.

Figure 1. Emerging Technology Investments



Enterprises are looking to innovate with AI, which continues to be viewed as a key differentiator for many organizations. The Cool Vendors in this research have demonstrated novel capabilities ranging from causal AI to foundation models and decision intelligence that enable continuous intelligence design patterns.

To get a more holistic view of the Cool Vendors and their market categories, D&A leaders can also consider reviewing the following:

Innovation Insight: Causal AI

Innovation Insight for Artificial Intelligence Foundation Models

Innovation Insight for Decision Intelligence Platforms

causaLens

London, U.K. (https://www.causalens.com)

Analysis by Pieter den Hamer

Why Cool: causaLens is one of the trailblazers in causal AI (see Innovation Insight: Causal AI); its products represent a new approach to AI that goes beyond correlation-based machine learning (ML) by providing tools for cause-and-effect analysis. Causal AI improves decision making and action taking by identifying cause-and-effect relations between variables and an outcome. Other benefits include better transparency and explainability of AI, and AI that is more resilient to change, as it is based on more profound causality, rather than more volatile correlation.

causaLens offers a causal AI platform, which includes:

- **causaLake**: Enables the identification of data sources that are likely most relevant to the AI project at hand.
- **causaLab**: Provides guided discovery of causal relations in data, represented in editable causal graphs, which then can be used to create causal models that make inferences and generate predictions.
- **decisionOS**: Allows the development of decision models that leverage a causal model to generate prescriptions and that incorporate constraints for decision optimization.

In addition, the company offers a number of "decision apps," which are prebuilt solutions, based on their causal AI platform, for common business functions and decisions. The resulting portfolio reflects causaLens' belief that it is necessary to bring AI and domain expertise together, rather than focusing solely on technology. This positions causaLens as a potential vendor in the emerging market of DI platforms (see Innovation Insight for Decision Intelligence Platforms).

Challenges:

- With causal AI being a relatively new and innovative approach to AI, the awareness among AI practitioners and causaLens' potential clients about the relevancy and benefits of causal AI is still limited.

- The discovery of causal relationships in data is a major addition to the palette of AI design and analysis tools. However, additional steps are required to connect them to other AI methods (e.g., deep neural networks). Moreover, resulting causal graphs and models still need to be extended, or at least validated, by human domain experts.
- Causal AI brings AI closer to decision makers across all industries and business units. Realizing this enormous potential requires causaLens to develop a scalable business and delivery model that combines technical know-how with adequate domain expertise delivered by a network of partners.

Who Should Care: Heads of AI and data science and their fusion teams. Organizations that want to create more impact and value by adopting causal AI and taking it beyond predictions, toward better transparency, robustness, decision intelligence and autonomous systems.

Hugging Face

New York City, New York, U.S. (https://huggingface.co)

Analysis by Arun Chandrasekaran

Why Cool: Hugging Face wants to be the largest open-source community for machine learning, bringing to life large ML models to enable new use cases and to improve the performance and accuracy of a huge variety of enterprise applications. Its core product is a model library called Transformers, which uses cutting-edge foundation models such as BERT and GPT2, as well as fine-tuned models such as Stable Diffusion to enable a variety of use cases. These use cases include text generation, summarization, question answering, sentiment analysis, language translation, image classification and generation, text to speech, speech recognition and even multimodal use cases such as text to image generation.

The key capabilities offered by Hugging Face are:

- **Hugging Face Hub —** Hugging Face Hub is a cloud platform with over 100,000 models and 40,000 datasets, which enables developers and data scientists to rapidly explore and experiment with various models and datasets, accelerating their business use cases. To promote responsible model usage and development, model repos are equipped with model cards to inform users of each model's limitations and biases.

- **Governance —** This cloud platform offers versioning, commit history, diffs, branches and over a dozen library integrations, all out of the box. In addition, security capabilities such as access control, user tokens and malware scanning are offered in its cloud platform.
- **Open source —** The models and core library are all available on Github under the Apache License 2.0, a permissive license that allows others to build on and deploy the models in a flexible way.
- **Cloud-hosted AutoML and inferencing —** AutoTrain is their automated machine learning (AutoML) tool for discovering the most effective models and accelerating their training. In addition, Hugging Face offers an inference API, which simplifies operationalization of models and delivers quick time to value for clients. Both these capabilities are cloud-hosted and delivered as a fully managed service.
- **Freemium product —** While the models themselves are free, the company's AutoTrain capability for easier model training and its inference API are only free up to a limited scale. The PRO plan (charged at less than $10 per month) increases those limits, while their enterprise plan adds more features such as a private hub, dedicated hosts, advanced security controls and 24/7 support.

While Hugging Face started as an open-source community for natural language processing (NLP) use cases, its growth mirrors the growth of foundation models. The use cases it addresses is expanding to other domains, such as computer vision and software engineering (language to code). In the past, Hugging Face had left the core research of the models to other technology giants. However, it is adapting quickly to the market need for better transparency and the openness of the large language models. It recently released BLOOM, a large multilingual open-source model, in coordination with academia and research labs, with greater degree of transparency into model performance and behavior.

Challenges:

- Large technology companies such as Google and Microsoft, as well as research labs such as OpenAI, are making significant investments in foundation models and view both the training and inferencing aspects as competitive moats. Competing against these deep-pocketed entities, on a sustainable basis, will prove challenging for Hugging Face.
- There are numerous startups, such as Cohere and AI21 Labs, that are operationalizing foundation models for NLP with a greater degree of fine-tuning and use case alignment. This competes with the DIY approach of Hugging Face.

- Newer startups, such as Stability AI with its Stable Diffusion, its text-to-image model, and MidJourney have been able to achieve viral adoption of their products. The advent of these viral startups, with similar paths to monetization, creates more competitive tension for Hugging Face.

Who Should Care: Data scientists, D&A leaders and CTOs should consider Hugging Face's suite of models to improve accuracy and accelerate the time to market for their greenfield and brownfield NLP and computer vision use cases.

Accrete AI

New York City, New York, U.S. (https://www.accrete.ai)

Analysis by Erick Brethenoux and Anthony Mullen

Why Cool: Accrete's Nebula platform extracts and automates insights from high-volume multistructured content, both inside and outside the enterprise. The platform differentiates itself through its insights-spotting capability, which is designed to discover and identify bias and contradictory insights, whether by accident or design (misinformation). The knowledge engineering platform is centered around a knowledge graph and a variety of modular natural language and other AI techniques, and importantly makes low demands of users in labeling data and model development.

Key capabilities offered by Accrete include:

- **Data hooks** — Able to ingest structured and unstructured content from inside the organization, as well as from external sources like social media, corporate filings and multilingual content.
- **Knowledge functions** — Modular AI building blocks (and models) to help build a knowledge engineering pipeline, including capabilities such as data extraction, normalization, linking, influence modeling, topic classification, semantic search, anomaly detection and semantic translation.
- **Explainable AI** — The platform provides intuitive interfaces called "glass boxes" that enable users to visualize, compare and drill down into model performance and inference metrics.
- **Flexible consumption** — The platform offers delivery to humans in the form of configurable dashboards and automated report generation, or as APIs for downstream automation.

- **Progressive networks —** Accrete's platform learning capabilities rest on the concept of progressive neural networks. Progressive neural networks are designed to effectively transfer knowledge across a series of tasks. In contrast to transfer learning models that incorporate prior knowledge only at initialization, progressive networks retain a pool of pretrained models throughout training, and learn lateral connections from these models to extract useful features for new tasks.

Challenges:

- Competing with deep vertical and end-to-end solution expertise. Well-defined industry-tailored solutions and offerings, such as Palantir and the Insight Engines, will be seen as a friendlier way to kickstart products than the blank page of "the platform that can do anything." This perception will still hold even if Accrete's approach can help bridge the gap between AI teams and software engineering functions.
- Providing a semantic framework for managing data and knowledge. While there are many tools available from Accrete to support the engineering of knowledge, the information architecture and data used to represent it is not as evolved.
- Overcoming the design paradigm for the many. Despite its intuitive "Lego-like" knowledge functions that can be incrementally assembled, the Nebula platform still represents a novel, and therefore unusual, way to build AI-powered solutions.

Who Should Care:

- Organizations that need a low staffing approach to analytical insight development and are happy to use automation-led approaches.
- Organizations that need to detect complex signals from a complex landscape of structured and unstructured data.
- Organizations looking to detect misinformation and concealing behavior, such as government, military and cybersecurity use cases, as well as finance with fraud, anti-money-laundering and insider trading use cases.
- Business functions that do not benefit from deep expertise in composite AI techniques and are looking for solutions that leverage those techniques for content analysis and management.

Evidence

Detailed interviews were conducted with all vendors profiled in this research.

[1] **2023 Gartner CIO and Technology Executive Survey:** This survey was conducted to help CIOs and technology executives overcome digital execution gaps by empowering and enabling an ecosystem of internal and external digital technology producers. It was conducted online from 2 May through 25 June 2022 among Gartner Executive Programs members and other CIOs. Qualified respondents are each the most senior IT leader (for example, CIO) for their overall organization or some part of their organization (for example, a business unit or region). The total sample is 2,203 respondents, with representation from all geographies and industry sectors (public and private). *Disclaimer: Results of this survey do not represent global findings or the market as a whole, but reflect the sentiments of the respondents and companies surveyed.*

Document Revision History

Cool Vendors in AI Core Technologies - 11 October 2021

Cool Vendors in AI Core Technologies - 1 October 2020

Cool Vendors in AI Core Technologies - 28 August 2019

Recommended by the Authors

Some documents may not be available as part of your current Gartner subscription.

Innovation Insight for Artificial Intelligence Foundation Models

Hype Cycle for Natural Language Technologies, 2022

Hype Cycle for Emerging Technologies, 2022

Hype Cycle for Data Science and Machine Learning, 2022

Top Strategic Technology Trends for 2023

© 2022 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by Gartner's Usage Policy. Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "Guiding Principles on Independence and Objectivity."