AI Transformation Brief

Featuring AI best practices and insights to enable our members to strategize, plan, develop, deploy, manage, and govern AI-based technologies and solutions.

In This Issue → AI in the News | AI Research Highlights | Vendor Spotlight | Upcoming Events & Resources

AI IN THE NEWS

DeepSeek R1 disrupts the AI industry

Read the DeepSeek paper - DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning

On January 20, 2025, DeepSeek launched its R1 LLM and claimed performance that rivaled the best AI models at a mere fraction of the training cost and infrastructure typically required. The DeepSeek R1 models were also announced under an open-source license, allowing organizations to use the models at no cost. As a result, DeepSeek suddenly became the most popular downloaded free app in the US from the Apple App Store, surpassing ChatGPT, and the world stock markets, particularly in the technology sector, experienced massive fluctuations.

ANALYST ANALYSIS

The hype behind DeepSeek, one of the latest chain-of-thought LLMs, has captured the AI industry's attention, with organizations questioning and scrutinizing whether they should continue accumulating massive compute infrastructure to train and build the next generation of AI models.

Novel approaches to training that were introduced by DeepSeek that dramatically reduced compute requirements included:

- Using reinforcement learning without requiring human-supervised fine-tuning to train the model.
- Mixed number of experts introduces a technique that divides an LLM into specialized sub-models (or experts), each dedicated to process a component of the LLM.

Shortly after, cloud vendors around the world announced the availability of DeepSeek on their platforms, including AWS, Google, and Azure. At the same time, numerous countries have either banned (Italy) or placed restrictions on the use of DeepSeek on government devices/agencies, including, but not exclusively: the US, India, Canada, Australia, South Korea, and Taiwan.

European AI organizations announce the OpenEuroLLM Project

Visit the OpenEuroLLM Project website

On February 3, 2025, Europe's leading AI companies, research institutions, and high-performance computing (HPC) centers announced that they will join forces and expertise to develop the next-generation of open-source foundation models that will support all official and future European Union (EU) languages. Funded largely from EU programs, initial funding is starting at €52m.

ANALYST ANALYSIS

The consortium of over 20 organizations will focus on building a family of performant, multilingual, LLMs for commercial, industry, and public use. The models will be open source, sharing weights and data, and will democratize access to high-quality AI technology.

While it's aspirational to develop an AI ecosystem so that European organizations can become independent from the Big Tech vendors based largely in Silicon Valley, the consortium will be challenged and needs to address their lack of experience in developing these foundation models and delivering them to market.

AI IN THE NEWS

OpenAI announces Deep Research [Read the OpenAI Deep Research announcement](#)

On February 2, 2025, OpenAI released its latest agentic offering, Deep Research, positioned as a chain-of-thought capable model that can provide PhD-level research capabilities.

OpenAI CEO, Sam Altman, claims Deep Research could do a "single-digit percentage of all economically valuable tasks in the world."

ANALYST ANALYSIS

Deep Research leverages OpenAI's o3's advanced reasoning capabilities to produce research reports. The capabilities include:
- Ability to perform in-depth, multistep queries across the internet by retrieving and synthesizing data.
- Ability to autonomously perform searches, scan, and assess online information from a variety of sources to produce comprehensive reports.

However, despite these capabilities, reports that have been produced clearly show that this application can hallucinate facts in its responses and make incorrect inferences. Sam Altman's statements on Deep Research's capabilities are "over-hyped," and he should have qualified that the single-digit percentage of valuable tasks are tasks that are non-mission critical and that can tolerate the risk of inaccuracies.

OpenAI announces Operator [Read the OpenAI Operator announcement](#)

On January 23, 2025, OpenAI announced one of its first agentic offerings, Operator. This new offering is OpenAI's strategic tool designed to simply the creation and management of AI agents. Operator helps autonomous agents to assess and interact with the digital environment they operate in. Users will be able to describe a task (e.g. Make a dinner reservation, buy concert tickets) and Operator will execute the necessary steps to complete the task, just as a human would interacting with the website.

ANALYST ANALYSIS

OpenAI Operator is powered by a model called Computer-Using Agent (CUA), leveraging GPT-4o vision and reasoning capabilities to interpret screenshots and interact with sites using typical browser controls like mouse and keyboard actions. Its design enables digital task completion, bypassing the need for OS- or web-specific APIs.

The CUA operates using an interactive process that integrates perception, reasoning, and action.
- Perception: Takes visual screenshots to establish the computer's current state.
- Reasoning: Using chain-of-thought reasoning, it assesses the next steps taking into consideration its current state and past screenshots and actions.
- Action: It performs the actions – clicking, scrolling, or typing – until it decides that the task has been completed or user input is still required.

AI Energy Score Unveiled [Salesforce Unveils its AI Energy Score Measuring AI Model](#)

On February 10, 2025, at the Artificial Intelligence Action Summit in Paris, Salesforce, Hugging Face, Cohere, and Carnegie Mellon University launched the AI Energy Score – a new benchmarking tool to measure and compare AI model energy consumption. Salesforce becomes the first AI model developer to disclose its energy efficiency data using this framework. The mission of this coalition is to drive awareness, encourage transparency, and promote AI sustainability.

ANALYST ANALYSIS

The goal of the AI Energy Score is to establish a standardized approach for assessing the energy efficiency of AI model inference. The AI Energy Score features:
- Standardized energy ratings: A standardized framework for measuring and comparing AI model energy efficiency.
- Public leaderboard: A comprehensive ranking of 166 models across 10 common AI tasks like text generation, image generation, and summarization.
- Benchmarking portal: A platform for AI developers to submit their open or proprietary models for evaluation and leaderboard inclusion.
- Recognizable energy use label: A new 1- to 5-star label that rates AI model energy use, with five stars indicating the highest efficiency.

INFO~TECH
RESEARCH GROUP

Assessing DeepSeek: Disruption in the AI Industry

AI customers and companies are questioning if they are overpaying for AI capabilities

DeepSeek model highlights

According to DeepSeek researchers, it cost US$6 million to train its chain of thought model and took only two months to build, using older and slower NVIDIA H800 chip technology.[1] In contrast, in 2024, OpenAI secured US$6.6 billion to pursue artificial general intelligence with its chain of thought model.

DeepSeek costs vs. OpenAI costs[2]

DeepSeek R1 API (1 million tokens)	OpenAI o1 API (1 million tokens)
55 cents for input	$15 for input (over 27x higher costs than DeepSeek)
$2.19 for output	$60 for output (over 27x higher costs than DeepSeek)

Resource utilization comparison

DeepSeek achieved results with **2.78 million GPU hours**,[1] significantly lower than Meta's **30.8 million GPU hours** for similar-scale models.

1. DeepSeek API Docs
2. Liu et al., 2024

Source: Assessing DeepSeek: Disruption in the AI Industry
Info-Tech Research Group

CIO Priorities 2025

01 **DISTRIBUTE DATA & AI ACCESS**

AI may be new, but the business demands it

Most organizations only started investing in AI in the past two years but must quickly move from experimentation to transformation.

AI is different from many technology investments made because it holds transformative potential. Rather than touching only one part of an organization, or making a broader but incremental impact, AI's promise is both broad and significant, and moving from investing in AI to squeezing the best value out of it is an important journey.

It's one that many organizations will struggle with. In the same way that digital transformation efforts were slowed by technical debt and different data silos in traditional organizations, now AI efforts are hindered. IT leaders say they feel that their teams lack the data management skills needed to set the foundation for AI. They also worry about how to govern its use and recognize their data platform isn't optimized for AI.

The more knowledge workers are provided with regarding AI tools, the more they can tailor it to their workflows and improve productivity. But before AI can be democratized, it must be governed. And before it's governed, it must be useful.

To avoid errors in generative AI output and

provide useful context, large language models (LLMs) benefit from retrieval augmented generation (RAG) in which additional data is pre-loaded into the memory of the model. This is a common method to customize AI to an organization's context, although additional pre-training and fine turning are also options.

Any of these approaches requires good data quality. Feed a LLM a biased data set, get biased answers. Provide conflicting answers from many different data sets, get different answers. The bad news is that most organizations struggle with data quality.

THE DATA QUALITY GAP
Info-Tech's CIO Business Vision diagnostic benchmark for the past year shows that data quality suffers the largest gap between importance and stakeholder satisfaction out of all core services. The related analytical capability is the third-largest gap.

The business wants access to new AI tools that will help them work. Yet many organizations feel they aren't ready to fully realize the value of AI because of challenges with their data. So there's work to be done.
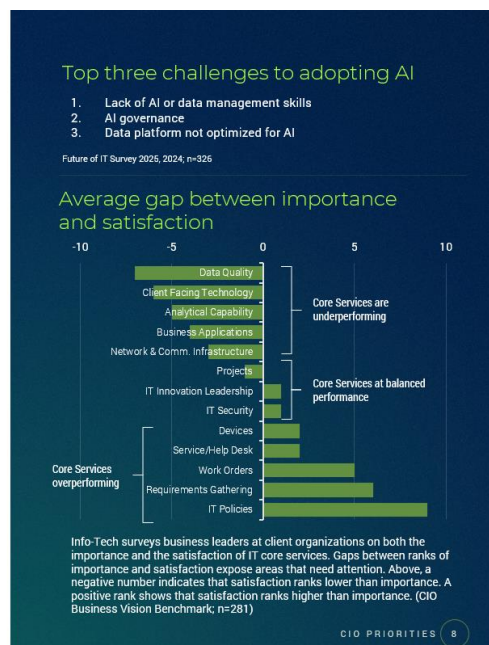
Source: CIO Priorities
Info-Tech Research Group

VENDOR SPOTLIGHT:

AI Marketplace

Unlock the potential of AI tailored to your needs and transform possibilities into reality with our dedicated support.

[Analytics Model: How AI Transforms Data Into Actionable Insights](#)

[DataChat: How AI Is Used for Conversational and Data Analytics](#)

[Neteera: How AI Can Deliver Personalized and Proactive Care](#)

[BusinessGPT: How AI Can Remain Accountable With Full Visibility and Control](#)

UPCOMING AND RECENT EVENTS

[LIVE Australia](#) Brisbane, Queensland – March 18-19, 2025 →

[Info-Tech LIVE 2025](#) Las Vegas, NV – June 10-12, 2025 →

[LIVE Europe](#) Barcelona, Spain October 21-22, 2025 →

AI AND DATA ANALYTICS SOLUTIONS – RESOURCES

[AI Marketplace](#)

[Artificial Intelligence Research Center](#)

[AI Workforce Development Program](#)

[Workshops](#)

AI Strategy Workshop

AI Proof of Value Workshop

AI MLOps Workshop

AI Governance Workshop

AI EDITOR-IN-CHIEF

[Bill Wong](#) – Info-Tech AI Research Fellow

INFO~TECH RESEARCH GROUP