CTO Forum: Building Custom GPTs

Prof Iavor Bojinov

Confidential and proprietary - Use permission granted to CTO Forum.

Provide Hands-On Training on the *Bleeding Edge* of Gen AI

Confidential and proprietary - Use permission granted to CTO Forum.

4 Hands-on sessions

Fine-tuning

Custom GPTs: Combing data and prompts for scaled impact

Process Redesign

Leveraging generative AI to redesign processes

Agents

Building Simple Automation

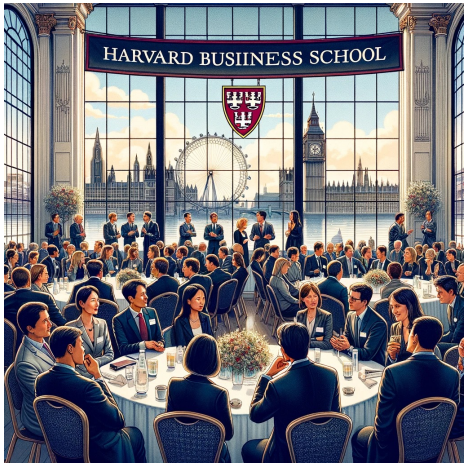
Multi-Agent Systems

Building Complex Automation

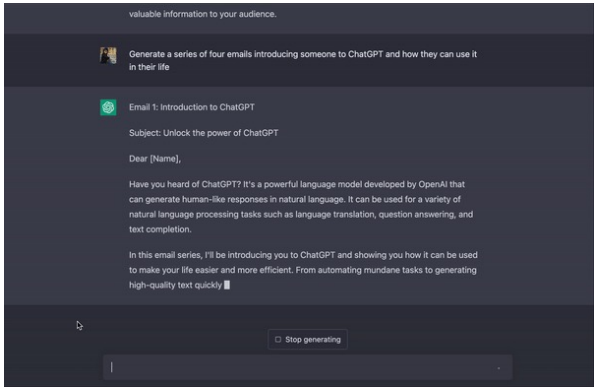
Confidential and proprietary - Use permission granted to CTO Forum.

Recap: What can Gen AI do?

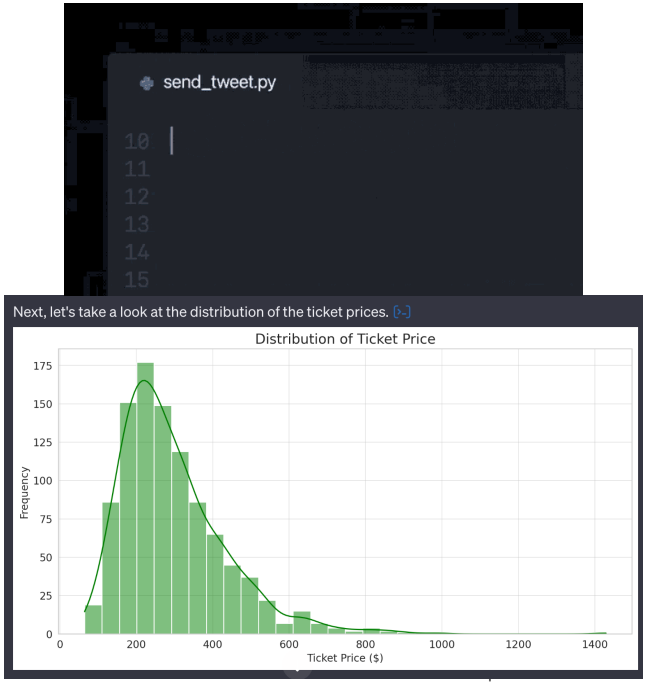
Generate Images, Videos, Sound



Generate Text

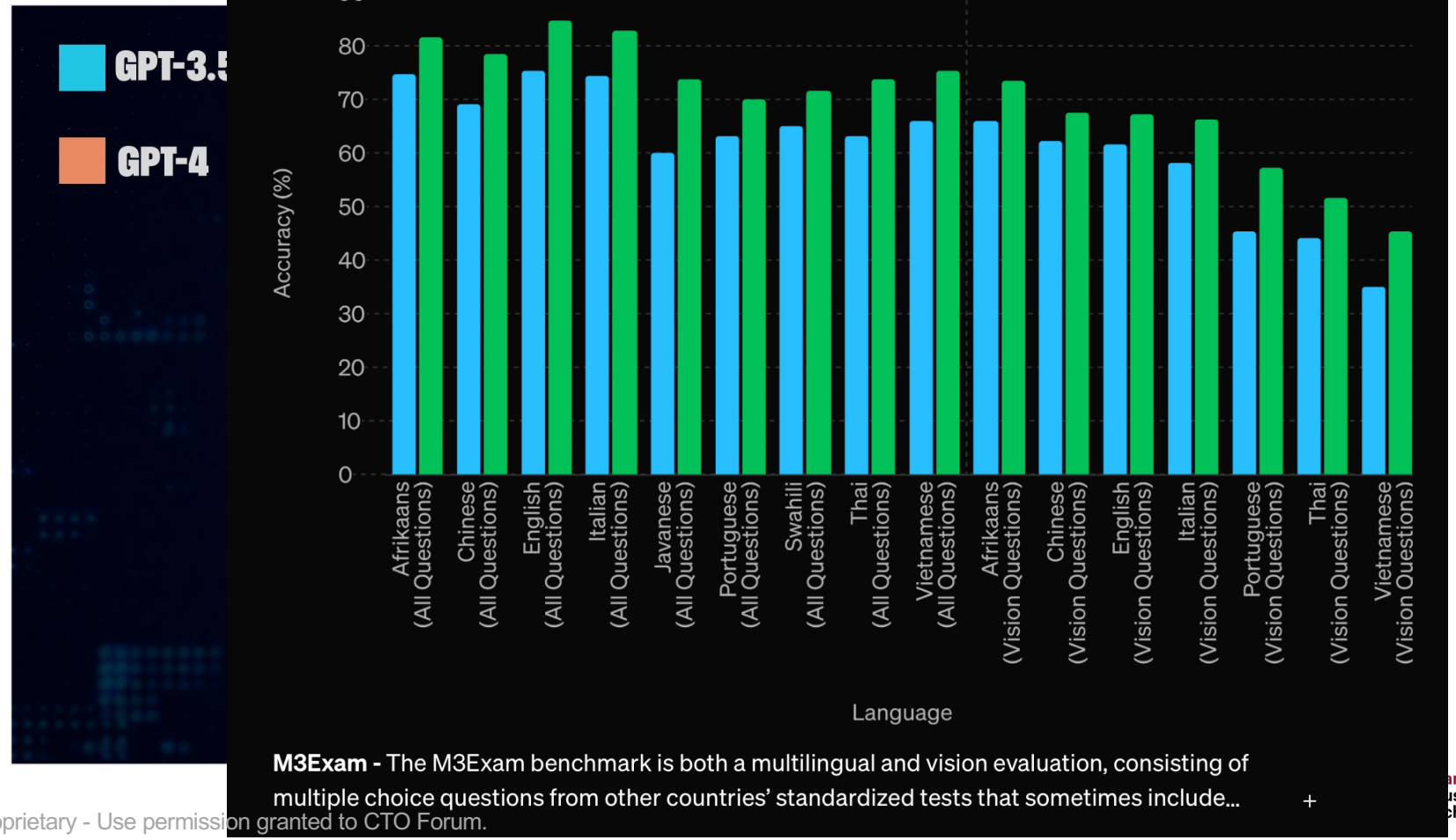


Generate Code



Confidential and proprietary - Use permission granted to CTO Forum.

General Capabilities



M3Exam - The M3Exam benchmark is both a multilingual and vision evaluation, consisting of multiple choice questions from other countries' standardized tests that sometimes include...

Confidential and proprietary - Use permission granted to CTO Forum.

arvard
usiness
chool

Recap: 4 Questions

1. Who used Gen AI tools (like ChatGPT, Perplexity, 365 copilot etc.)?
2. Who used Gen AI tools (like ChatGPT, Perplexity, 365 copilot etc.) every day?
3. Last week, who had integrated Gen AI into most aspects of their work?
4. Who thinks Gen AI will fundamentally impact organizations and your work in the next 3 years?

Confidential and proprietary - Use permission granted to CTO Forum.

Snack Industry

USA Rev +$50B

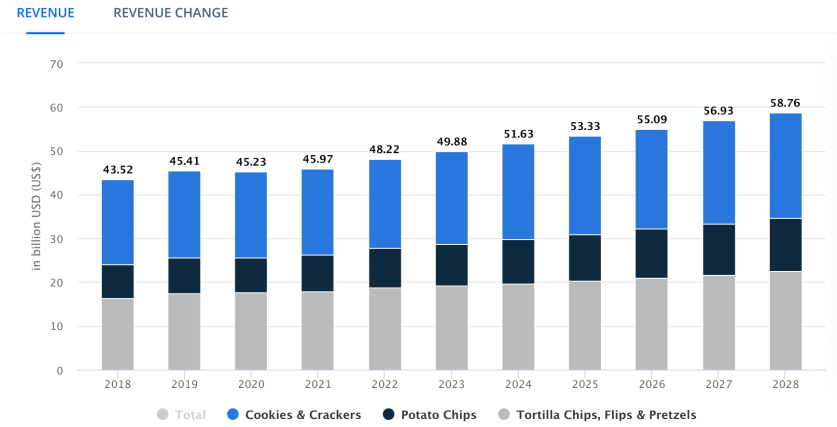
Expected to grow 3-4% CARG.

Years long R&D process

New products can take anywhere from 2-4 years

Months long marketing

Typically, takes 6-12 months to launch a product



Notes: Data shown is using current exchange rates and reflects market impacts of the Russia-Ukraine war.

Most recent update: Mar 2024

Source: Statista Market Insights

Confidential and proprietary - Use permission granted to CTO Forum.

Harvard Business School

Creating AI-first snack company

1. Analyze current trends + identify an opportunity
2. Create a name, logo, and packaging prototype
3. Generate a recipe that scales
4. Identify customer segment and location of the initial launch
5. Create a targeted marketing campaign
6. Summarize the discussion & create a power point slide to pitch your company!

Confidential and proprietary - Use permission granted to CTO Forum.

Harvard Business School

Focus Fuel

thefocusfuel.com

Focus Fuel is a brand offering **caffeine gummies** designed to provide a convenient and tasty way to boost energy, focus, and electrolyte balance.

Confidential and proprietary - Use permission granted to CEO Forum

Product Details & Benefits

- **Caffeine Content**: 100 mg per serving (2 gummies)
- **Flavors**: Electric Blue Razzberry Lemonade, Electric Fruit Punch, Electric Watermelon
- **Dietary Friendly**: Gluten-free and vegan

Focus Fuel

- 3 people + 2 part time contractors.
- $1.5 m in revenue in year 1
- GMC, Walmart, 7-Eleven distribution deals
- 100% AI-powered snack company

Use cases include:

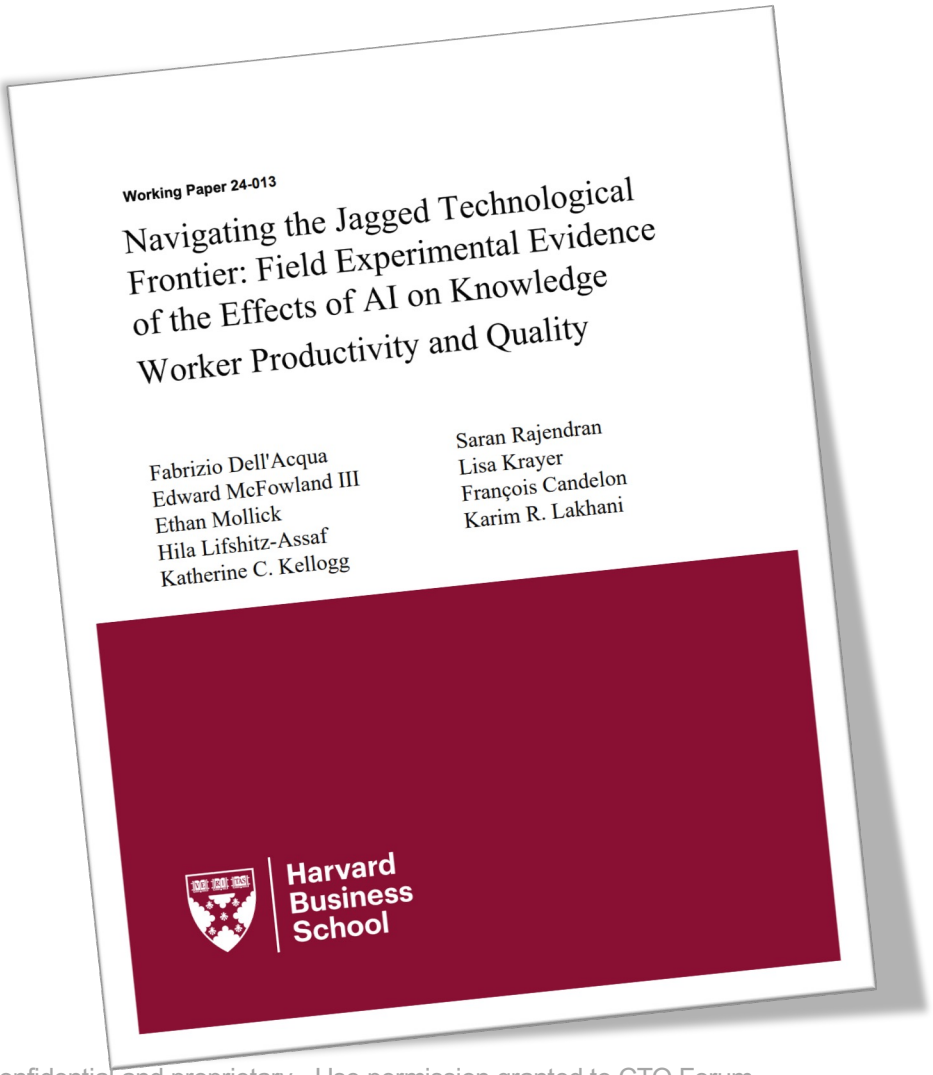
- Product development
- Recipe development and scaling
- Market analysis
- SEO and copywriting
- Branding and packaging
- Distribution strategy
- Initial legal counsel
- Contract strategy and management

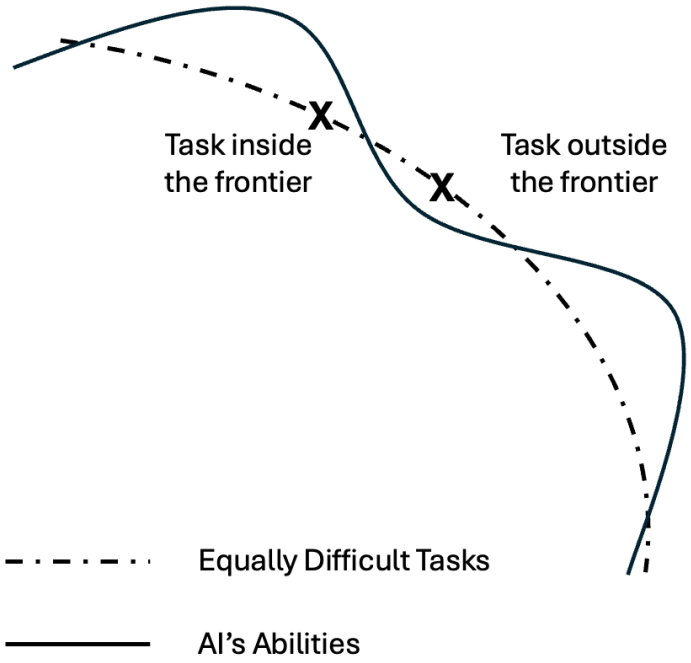
We had a clear vision of what we wanted to create, but I never had the experience or expertise... I've used AI to really be my co-pilot through this whole journey.

- Daniel Hertz, founder

Harvard Business School

Confidential and proprietary - Use permission granted to CTO Forum.

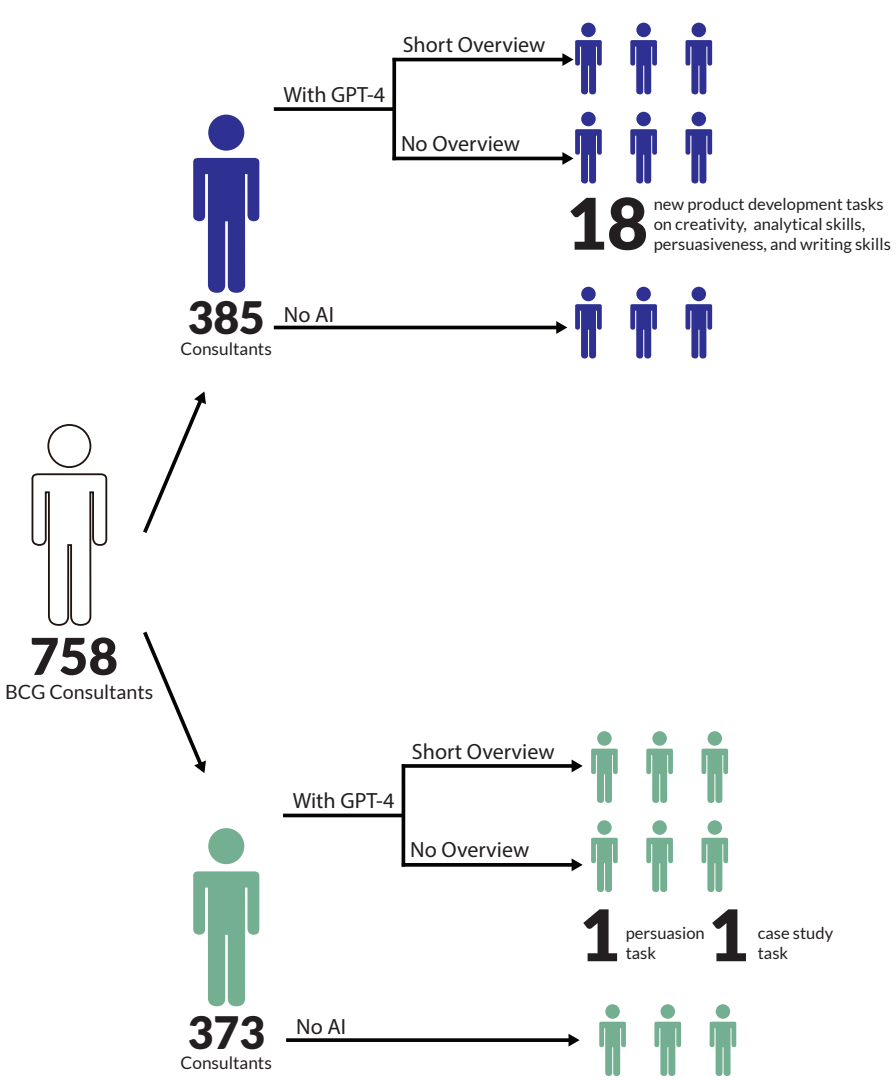


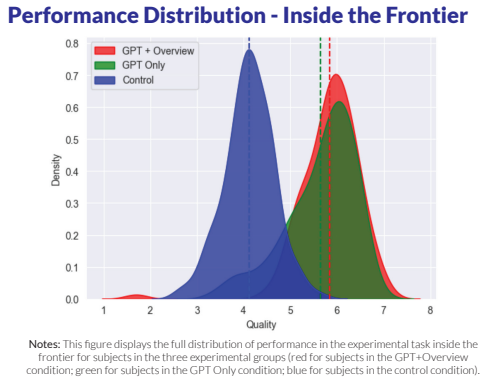


- AI has a "**jagged technological frontier**."
- It will do some things well…
- …and some similar things not well.
- Using AI inside its frontier may help. Using it outside its frontier may harm performance.
- It is difficult to grasp where the frontier is.

Harvard Business School

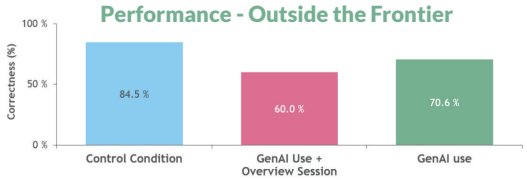
Confidential and proprietary - Use permission granted to CTO Forum.



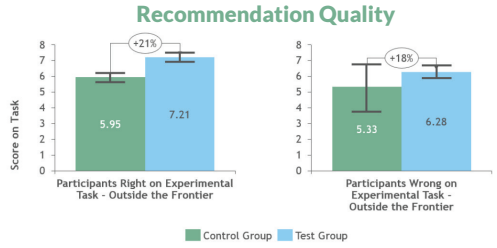


Notes: This figure displays the full distribution of performance in the experimental task inside the frontier for subjects in the three experimental groups (red for subjects in the GPT+Overview condition; green for subjects in the GPT Only condition; blue for subjects in the control condition).

Large improvements on every performance measure for GPT-4 group, no impact from training



Decrease in outside the frontier task



Improvement in inside the frontier task

Confidential and proprietary - Use permission granted to CTO Forum.

Stage	Task	Inside	Outside
Ideation	Market Research	Processes and summarizes data quickly, identifies patterns, and suggests trends.	Struggles with generating entirely novel insights or verifying real-world accuracy.
	Product Idea Generation	Excellent at brainstorming ideas based on structured prompts and preferences.	Limited by the creativity of the user's input and can produce repetitive or impractical suggestions.
	Data Analysis	Good for initial exploratory data analysis and generating visualizations.	Susceptible to errors in interpreting or manipulating complex datasets.
Prototyping	Flavor and Recipe Development	Effective at creating visual outputs like charts and graphs based on clear instructions.	Cannot independently validate or prioritize the significance of trends.
	Manufacturer Identification	Good at combining structured inputs to generate creative recipes and industrial adaptations.	Cannot test or verify flavor profiles or feasibility in real-world production.
	Product Name	Effective at researching and listing potential manufacturers using online sources.	Limited by the currency, accuracy, or completeness of data available online.
	Product Description	Excellent at generating creative, diverse, and context-aligned name suggestions.	Might suggest names that lack cultural nuance or marketability.
Marketing	Target Customer Identification	Very effective at generating concise, engaging descriptions tailored to specific tones or audiences.	May lack specificity or originality if not guided well.
	Packaging Design	Useful for segmenting audiences based on known attributes and generating insights.	Struggles with understanding nuanced consumer behaviors or conducting detailed psychographic profiling.
	Marketing Campaign Development	Excellent at generating visual prototypes of packaging concepts with clear prompts.	Limited in producing print-ready designs or handling brand-specific design nuances.
	Advertising Content Creation	Strong at brainstorming locations, slogans, and tone-aligned content ideas.	May lack the strategic depth needed to craft a comprehensive marketing strategy.

Confidential and proprietary - Use permission granted to CTO Forum.

Prompt Engineering

01	02	03	04	05	06	07
Clarity & Specificity	**Context & Intent**	**Instructional Prompts**	**Feedback Loops**	**Chain of Thought**	**Shot Prompting**	**Personas**
Use precise language to avoid ambiguity.	Include why / for what purpose you're asking.	Give clear commands or steps to follow.	Prompt for iterative refinement or confirmation.	Encourage disclosure of detailed reasoning process.	Provide an example of the desired type of data or output.	Assign a character or role to the AI.

Source: SkillLeap.ai, OpenAI. (2024). ChatGPT (4) [Large language model]

Confidential and proprietary - Use permission granted to CTO Forum.

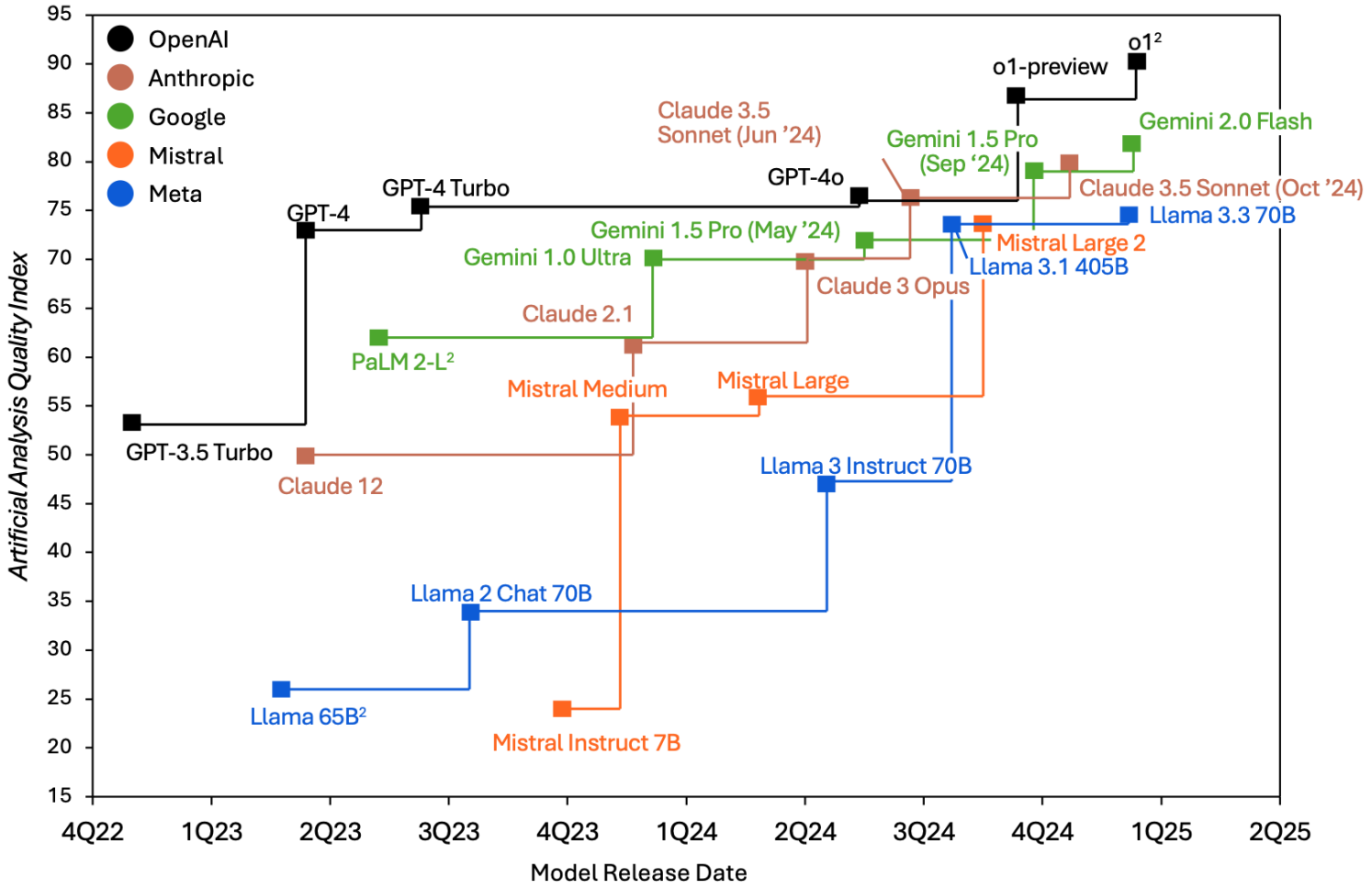
Harvard Business School

The models today are the worst they will ever be.

Confidential and proprietary - Use permission granted to CTO Forum.

Frontier Language Model Intelligence, Over Time[1]

Artificial Analysis Quality Index: MMLU, GPQA Diamond, MATH-500, HumanEval. Top 5 Labs by Market Share.

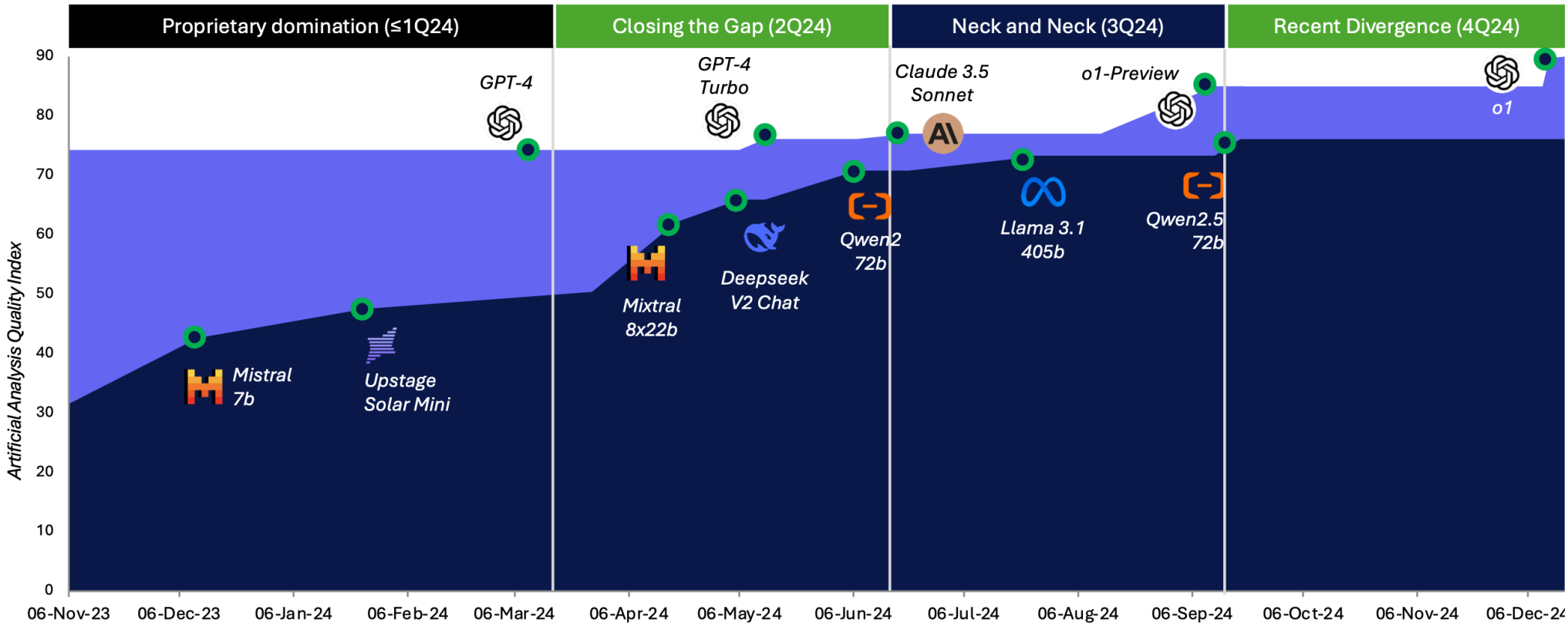


Confidential and proprietary - Use permission granted to CTO Forum.

Harvard Business School

Model Quality: Leading Proprietary and Open Weights Models

Based on proprietary and open-source models that resulted in an increase in Artificial Analysis Intelligence Index score

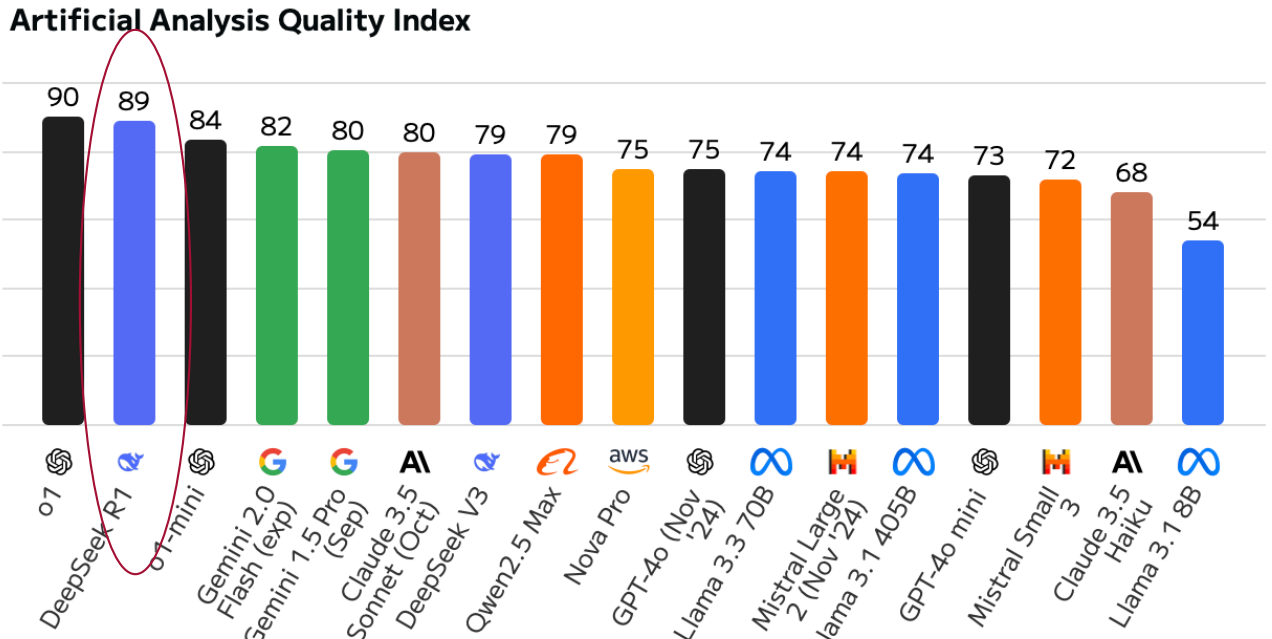


Artificial Analysis

Harvard Business School

Confidential and proprietary - Use permission granted to CTO Forum.

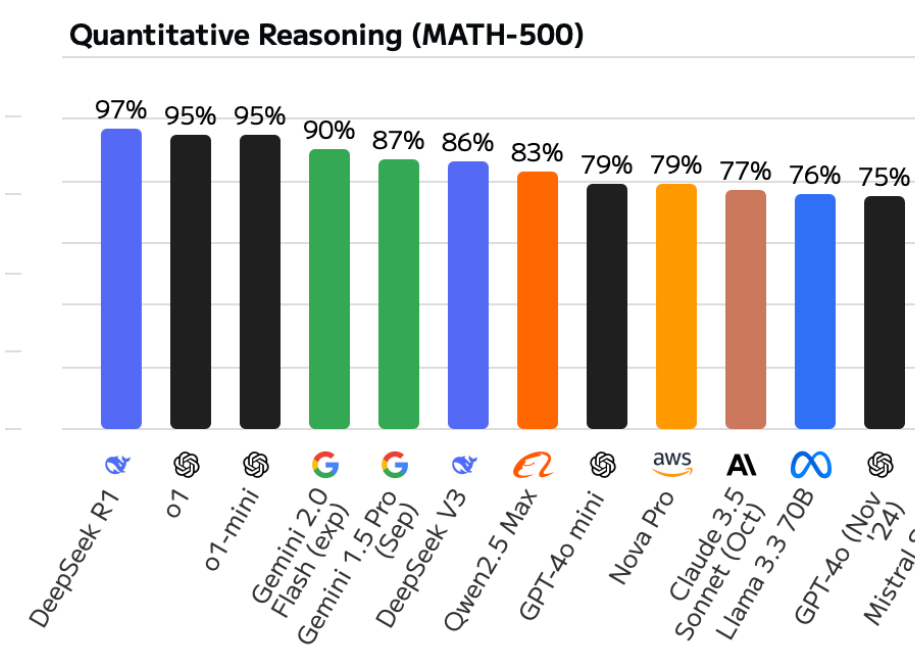
New models are quickly coming out

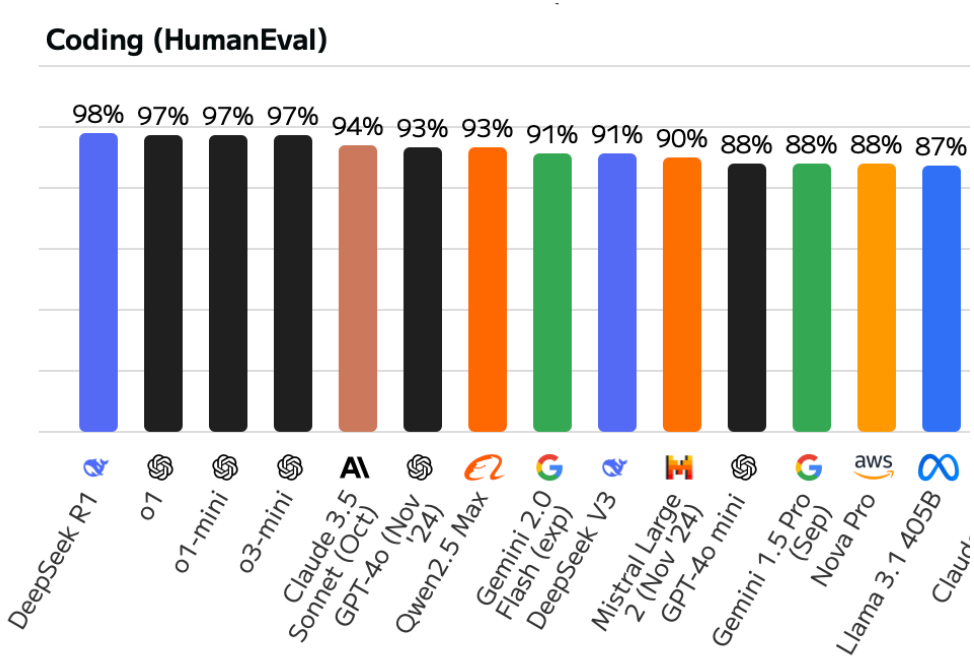


Confidential and proprietary - Use permission granted to CTO Forum.

Harvard Business School

DeepSeek is outperforming in some dimensions!





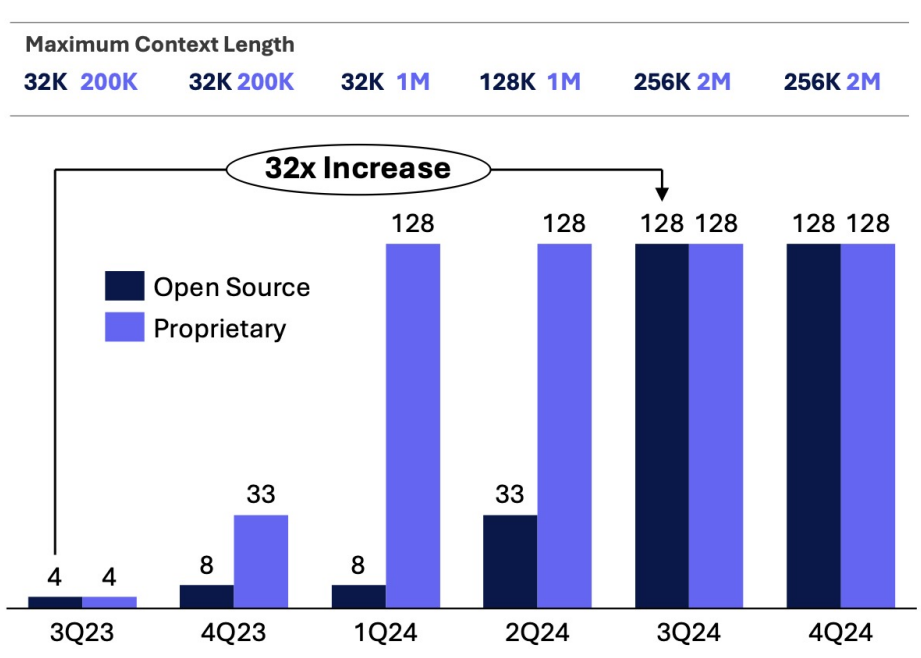
Confidential and proprietary - Use permission granted to CTO Forum.

Harvard Business School

Models can handle more context and are getting cheaper

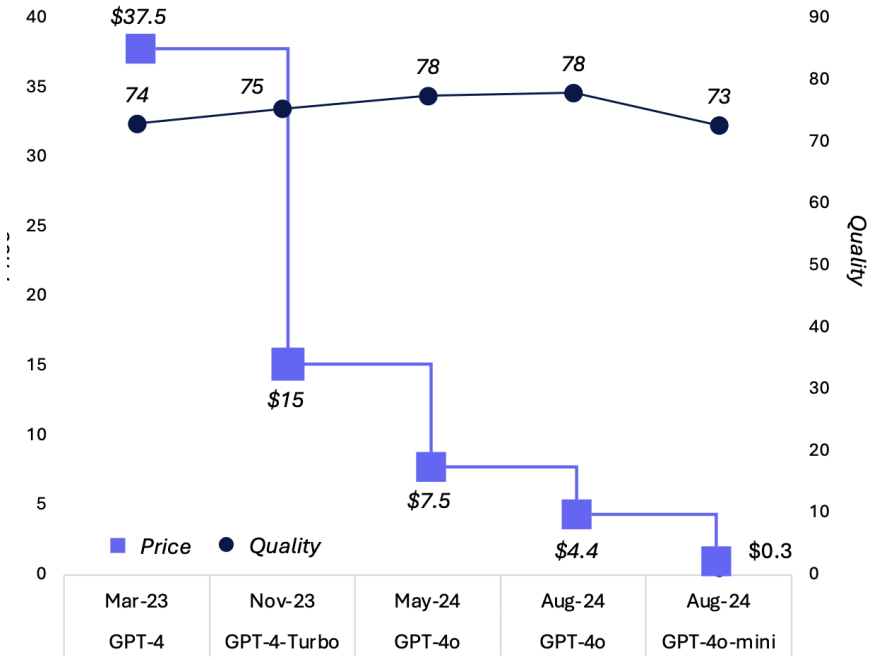
Context Length (Tokens), Median by Quarter

Median context length (thousand tokens)[1]



Inference Cost for OpenAI's GPT-4 Model Family

USD/m tokens, blended input and output tokens (3:1 ratio)



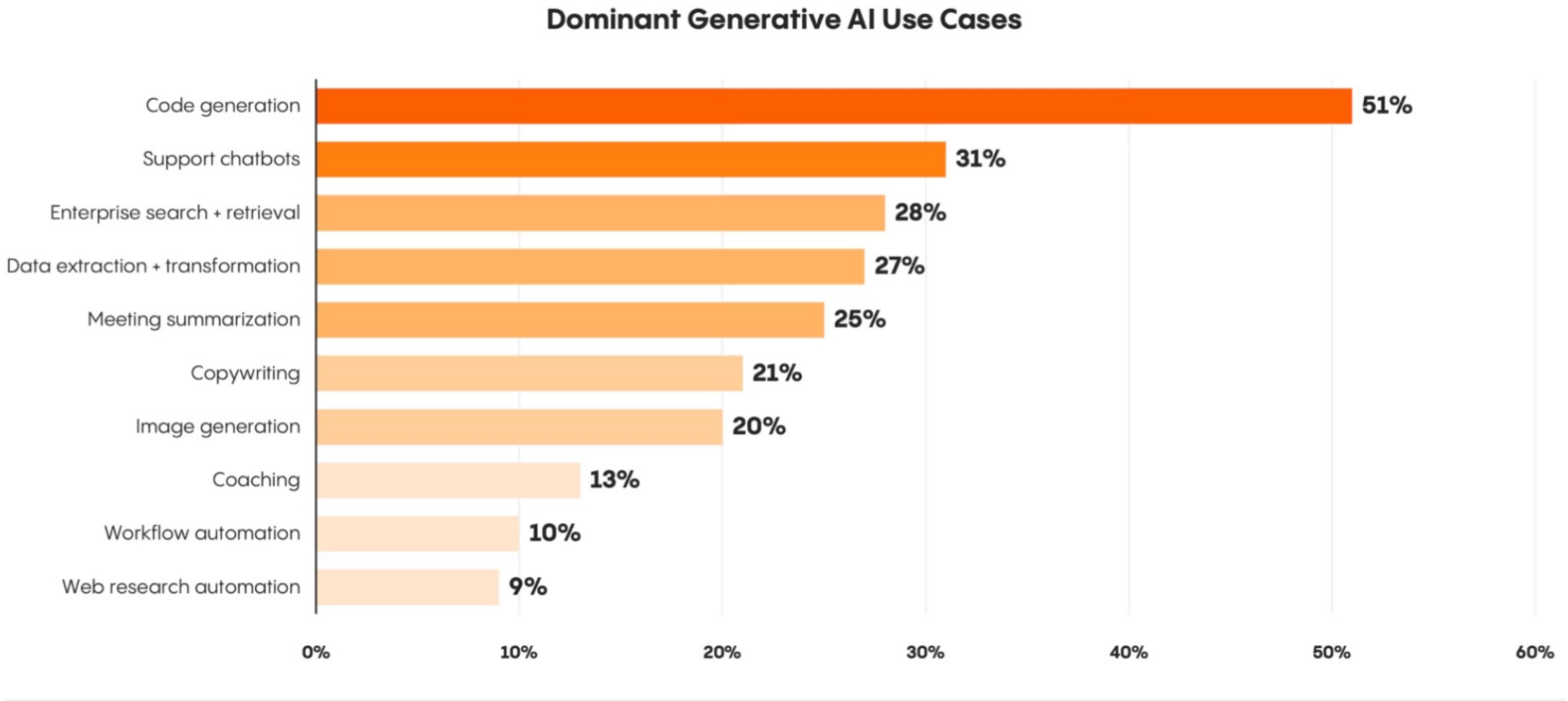
Confidential and proprietary - Use permission granted to CTO Forum.

Harvard Business School

Yet, most companies are struggling to adopt and integrate generative AI.

Confidential and proprietary - Use permission granted to CTO Forum.

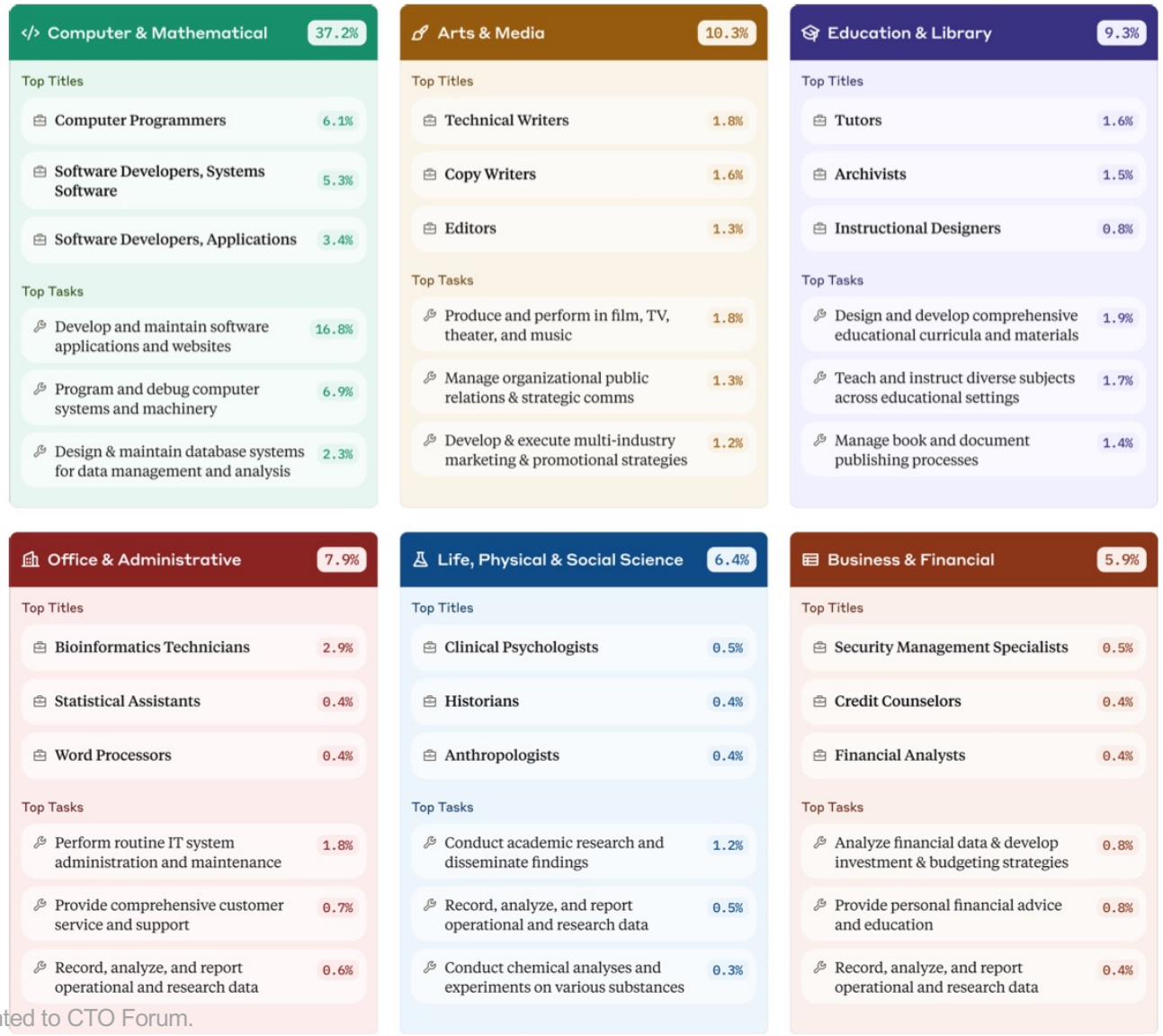
Menlo Ventures Report



Confidential and proprietary - Use permission granted to CTO Forum.

Harvard Business School

Anthropic

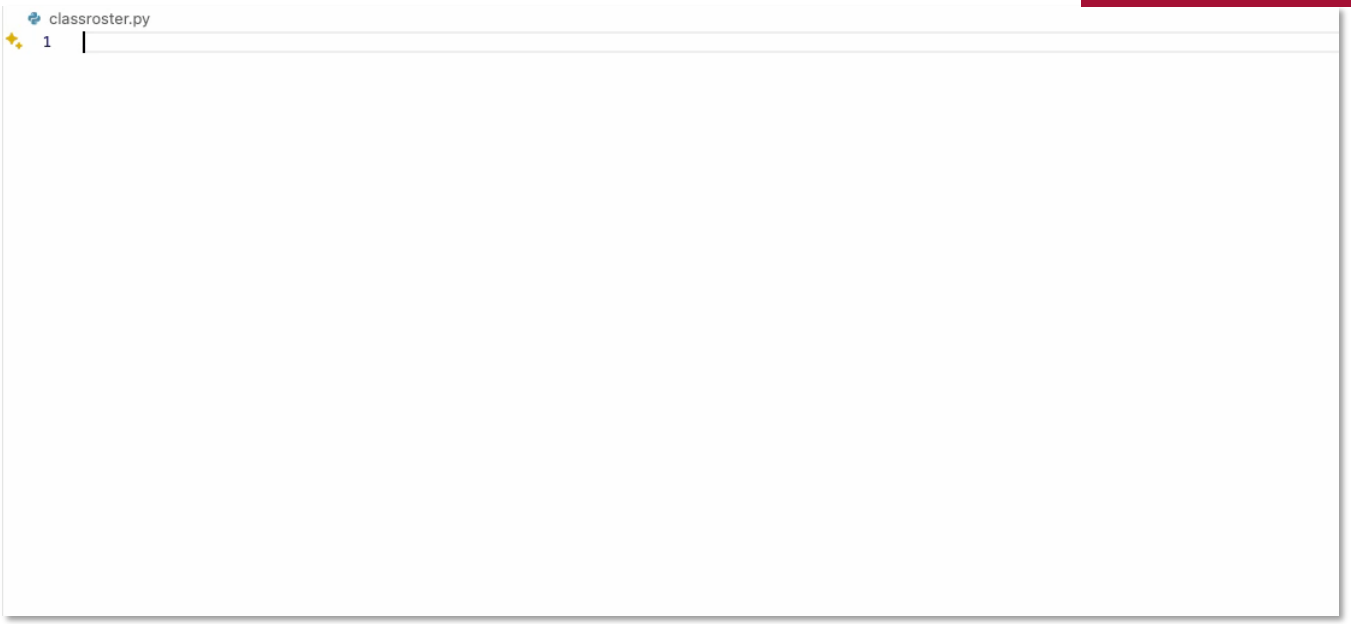


https://assets.anthropic.com/m/2e23255f1e84ca97/original/Economic_Tasks_AI_Paper.pdf

Confidential and proprietary - Use permission granted to CTO Forum.

arvard
usiness
chool

Use Case: GitHub Copilot



RCT: Copilot let to SWE being 55.8% Faster

75% Higher Job Satisfaction

73% Better concentration

87% Save Cognitive Effort

https://github.blog/news-insights/research/research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness/

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. "The impact of ai on developer productivity: Evidence from github copilot." arXiv preprint arXiv:2302.06590 (2023).

Confidential and proprietary - Use permission granted to CTO Forum.

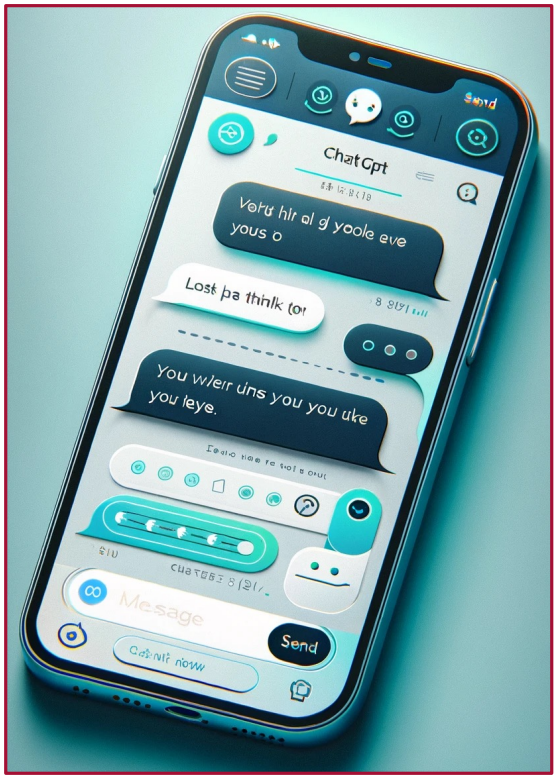
Grab's Mystique

Harnesses large language models (LLMs) and draws from Grab's wealth of data to craft highly personalized content.

98.5% reduction in content creation time

25-50% increase in engagement for Grab user

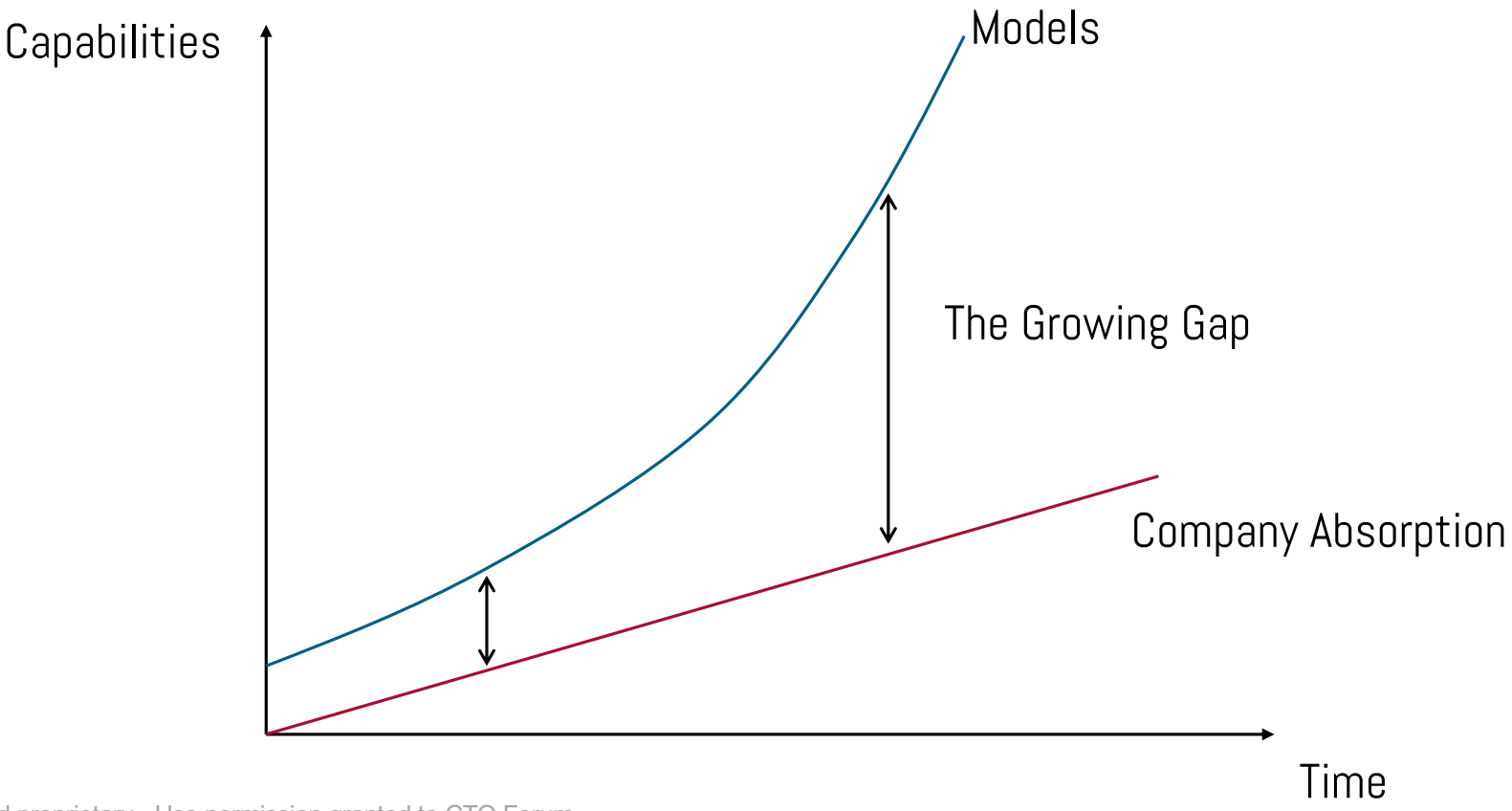
117% increase in engagement for merchants



https://www.grab.com/sg/inside-grab/stories/mystique-ai-copywriting-tool/

Confidential and proprietary - Use permission granted to CTO Forum.

Growing gap



Confidential and proprietary - Use permission granted to CTO Forum.

Harvard Business School

How are you tackling the growing gap?

Confidential and proprietary - Use permission granted to CTO Forum.

Modern's Gen AI Strategy

Education

Education for all on AI, capabilities, applications, etc.

Democratization

Provide access to generative AI tools to all employes

Encourage wide-spread usage (both fun + work)

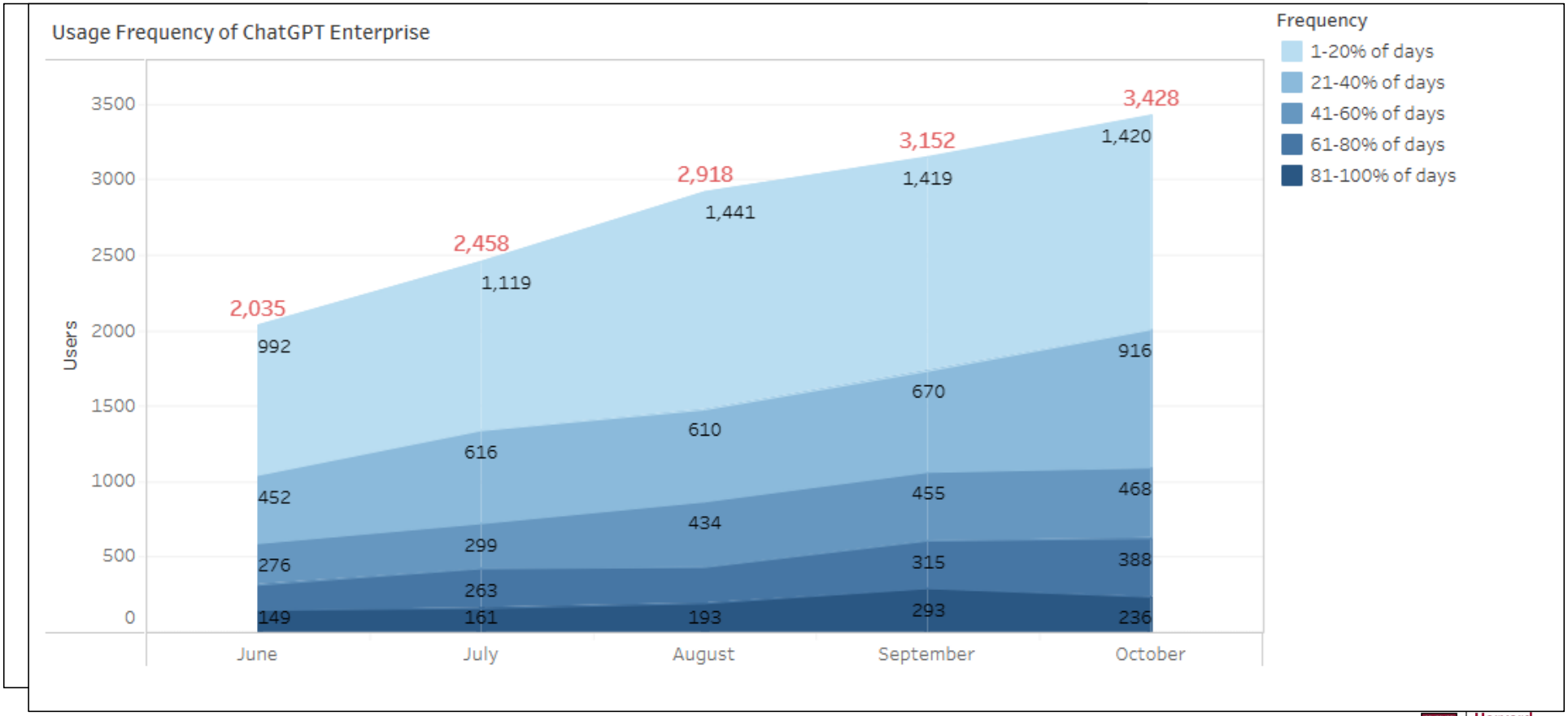
Product Mindset

Encourage employees to build and share custom GPTs

Generate network effects

Confidential and proprietary - Use permission granted to CTO Forum.

Closing the Gap: Moderna Example



Confidential and proprietary - Use permission granted to CTO Forum.

Harvard Business School

Custom GPTs

What Are Custom GPTs?

- AI models fine-tuned for **specific use cases**.
- Configurable via **custom instructions, knowledge, and tool integrations**.

Why Build a Custom GPT?

- **Domain-Specific Expertise** – Adapt AI to industry needs (e.g., finance, legal, healthcare).
- **Consistency & Control** – Align responses with brand, compliance, and security policies.
- **Workflow Optimization** – Automate decision-making, enhance support, and reduce manual effort.
- **Scalability & Adaptability** – Deploy across teams, integrate with enterprise systems.

Confidential and proprietary - Use permission granted to CTO Forum.

AI Agents

What are AI Agents?

- Agents are autonomous AI systems that **act independently** to achieve a goal, often making decisions, retrieving information, and executing tasks across various platforms.
- Agents can be designed with **specific logic, workflows, and multi-step reasoning**, enabling them to interact dynamically with external environments.

Why build an AI Agent?

- **Automation:** AI-powered assistants that handle tasks autonomously (e.g., scheduling, email management, data analysis).
- **Interaction:** Agents that interact with multiple APIs, databases, and enterprise systems.
- **Complexity:** AI bots that execute **multi-step** processes, such as ordering items, booking appointments, or managing workflows.

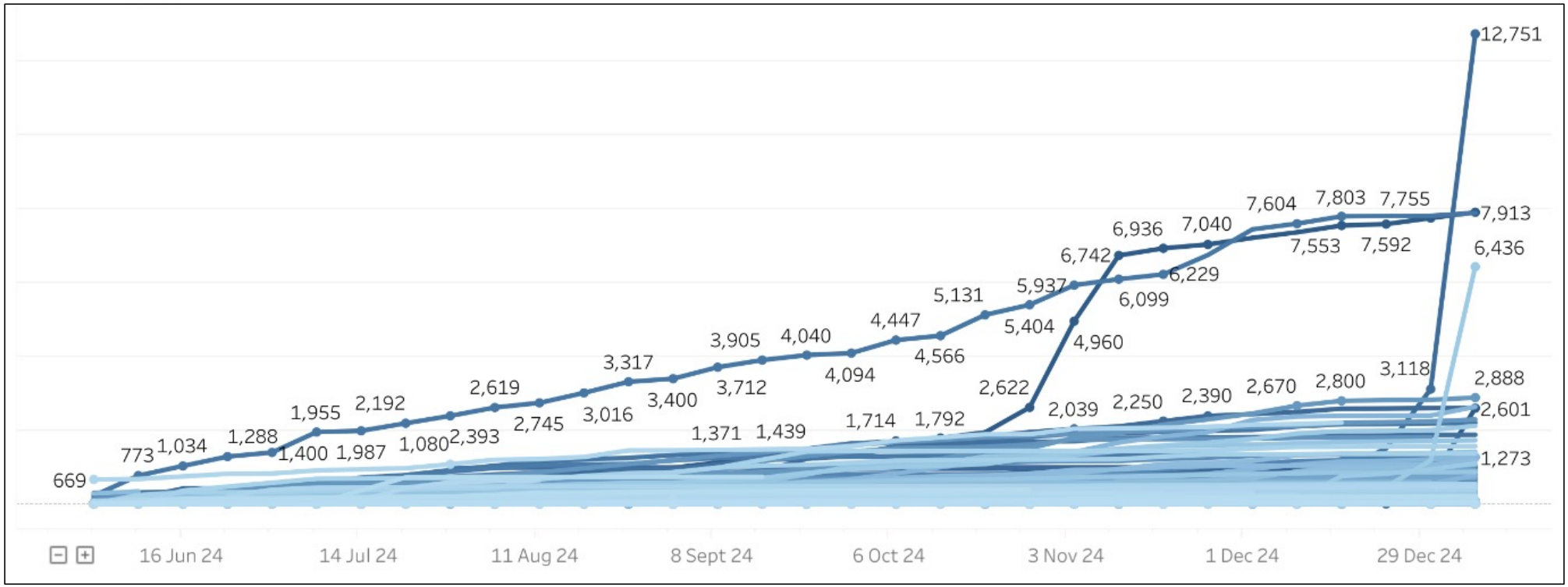
Confidential and proprietary - Use permission granted to CTO Forum.

Comparison

Feature	Custom GPTs	AI Agents
Autonomy	Responds when prompted	Can operate independently
Proactiveness	User-driven	Task-driven, can take action on its own
Customization	Adjusted via instructions & integrations	Programmed with logic, APIs, and workflows
Use Case	Chatbots, assistants	Automation, decision-making, multi-step tasks
Memory	Limited recall within chat	Can store & recall structured knowledge
Risk	Low, still requires human input and direction	High because of the higher level of autonomy

Confidential and proprietary - Use permission granted to CTO Forum.

Growth Coming from Custom GPTs



Confidential and proprietary - Use permission granted to CTO Forum.

Harvard Business School

	GPT Name	Messages	Users
1	Self-Review GPT	12,751	885
	Gathers monthly/quarterly business reviews, self-reflection guides, goals, year-end review preparation.		
2	U.S. Benefits Assistant	7,887	1,215
	Helps employees discover, evaluate, and select annual employee benefits such as medical insurance.		
3	Red Ribbon Visuals	7,884	593
	Helps any employee create visual backgrounds for presentations.		
4	Collaborator Feedback GPT	3,888	387
	Transcribes conversations with key collaborators, helps prepare feedback for peer reviews.		
5	Meeting Minutes Generator/Analyzer	2,868	271
	Summarizes/transcribes meetings, participants, outcomes, individual meeting performance coaching.		
6	GenAI Mentor	2,610	769
	Learning tool that uses AI to teach employees.		
7	Slide PowerPoint Creator	2,598	288
	Helps turns conversations into Moderna-templated presentation slides.		
8	RegBot 3.0	2,194	128
	Regulatory bot, helps Moderna respond to the many questions and data requests from regulators.		
9	Moderna Icon Creator (beta)	1,816	168
	Creates Moderna-templated icons with corporate colors, etc. for documents, programs, signatures.		
10	Upward Feedback GPT	1,562	247
	Helps create feedback for a supervisor/manager.		

Confidential and proprietary - Use permission granted to CTO Forum.

How do you manage the proliferation?

Confidential and proprietary - Use permission granted to CTO Forum.

Build a Governance Custom GPT

Key Components

1. **Base Model Selection** – Choose GPT-4 or future variants.
2. **Custom Instructions** – Define behavior, tone, and constraints.
3. **Knowledge Augmentation** – Upload documents, databases, and private datasets.
4. **Tool & API Integrations** – Extend capabilities (e.g., code execution, web access, external APIs).

GPT Risk Matrix

Severity	3 - Critical	B	C	C
	2 - Medium	A	B	C
	1 - Low	A	A	B
		1 - Individual	2 - Team	3 - Company
		Impact		

A – Low criticality

B – Medium criticality

C – High criticality

Requirements for GPT Creators

General Compliance

1. Follows AI code of conduct **A + B + C**
2. Adheres to naming standards **A + B + C**

Documentation

3. Detailed and accurate GPT description **A + B + C**
4. Published user guide in GPT knowledge repository **B + C**
5. Published release notes with every update **C**

Product Development

6. Routine system enhancements **C**
7. Requirements gathering with stakeholders **C**

Quality Assurance

8. Routine bug fixes **B + C**
9. Peer review prior to release **B + C**
10. Quality assurance testing for each release **C**
11. Performance benchmarks **C**

Security & Privacy

12. Quarterly review of access with stakeholders **B + C**
13. Backup of GPT instructions and documentation **B + C**
14. Cybersecurity review prior to publication **C**

Governance

15. Quarterly utilization and decommissioning review **A + B + C**
16. Designated primary and backup point of contact **B + C**
17. Established SLAs **C**

Confidential and proprietary - Use permission granted to CTO Forum.

Custom GPT Evaluator

Harvard Business School

DRAFT
FEBRUARY 4, 2025

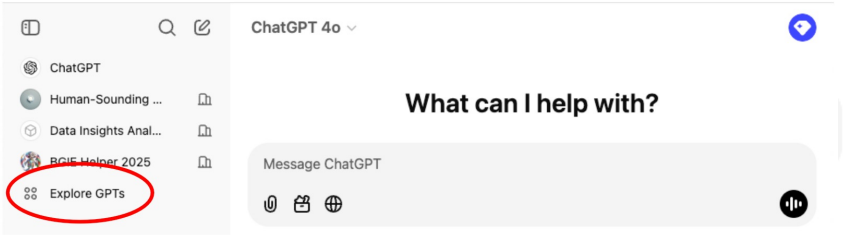
IAVOR BOJINOV
ANNIKA HILDEBRANDT

Building a Custom GPT and AI Agent Evaluator

Custom GPTs can quickly and easily be built using ChatGPT. These models are the precursors to agents as they can combine multiple sources of information with custom instructions; however, they are still intended to be worker companions as they lack the autonomy to act. Nevertheless, they provide a powerful tool that enables employees to create a wide range of custom applications, from personal benefits assistants to coding support. One of the benefits of custom GPTs is they can be shared broadly with others, either in your enterprise or with other ChatGPT users.

This decentralized approach that allows everyone to create and share agents raises some vital governance questions. To address governance issues, you will be creating a custom GPT that creates governance guidelines for the creators of GPTs. Your evaluator will help GPT builders understand what standards custom GPTs and AI agents must adhere to.

The instructions below are tailored for the ChatGPT premium accounts with access to build GPTs. Begin by navigating to chat.com and logging in or creating an account. In the left-hand navigation bar, select 'Explore GPTs' to reach the GPT home page.



Task 1: Creating your first GPT – Basic Configuration

Being by creating your first GPT and configuring the details.

1) On the GPT home page, select the create button in the top right to be brought to the configuration page.

Confidential and proprietary - Use permission granted to CTO Forum.

Harvard Business School

Test 1:

Legal Summarization GPT

Description: A GPT that summarizes multiple internal legal documents.

Test your GPT

Confidential and proprietary - Use permission granted to CTO Forum.

Harvard Business School

Test 2:
Dose ID GPT
Description: Provide a recommendation for the appropriate dosing in clinical trials

---

Test your GPT

Confidential and proprietary - Use permission granted to CTO Forum.

Harvard Business School

Where is your key takeaway from the exercise?

Confidential and proprietary - Use permission granted to CTO Forum.

Where can you apply this in your organization?

Confidential and proprietary - Use permission granted to CTO Forum.

Key Insight: Focus on Functional Patterns NOT Use Cases

Democratize functional patterns to empower employees:

1. Summarization of documents
2. Translation
3. Document Review
   - Legal, security, etc.
4. Data Analysis
5. Content Generation
   - Emails, marketing briefs, images, etc.
6. Knowledge Repository through Retrieval-Augmented Generation (RAG)
   - Benefits, answering questions, etc.

Confidential and proprietary - Use permission granted to CTO Forum.

4 Hands-on sessions

Fine-tuning

Custom GPTs: Combing data and prompts for scaled impact

Process Redesign

Leveraging gen AI to redesign processes

Agents

Building Simple Automation

Multi-Agent Systems

Building Complex Automation

Confidential and proprietary - Use permission granted to CTO Forum.

Thank you!

Confidential and proprietary - Use permission granted to CTO Forum.