

CTO Forum: Building an AI Agent for Cybersecurity Incident Reports

Prof Iavor Bojinov



Plan

1. Recap (what we discussed last time)
2. How to keep your AI projects on track
3. Building an AI Agent
4. Discussion, learnings, and next steps.



4 Hands-on sessions

1. Fine-tuning & (baby) RAG

Custom GPTs: Combining data and prompts for scaled impact

2. Agents

Building Simple Automation

3. AI & ML with Gen AI

Leveraging generative AI to for data analytics, ML, and AI

Preparing for the future

Ensuring your organization is planning for the next 5 years

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




Recap

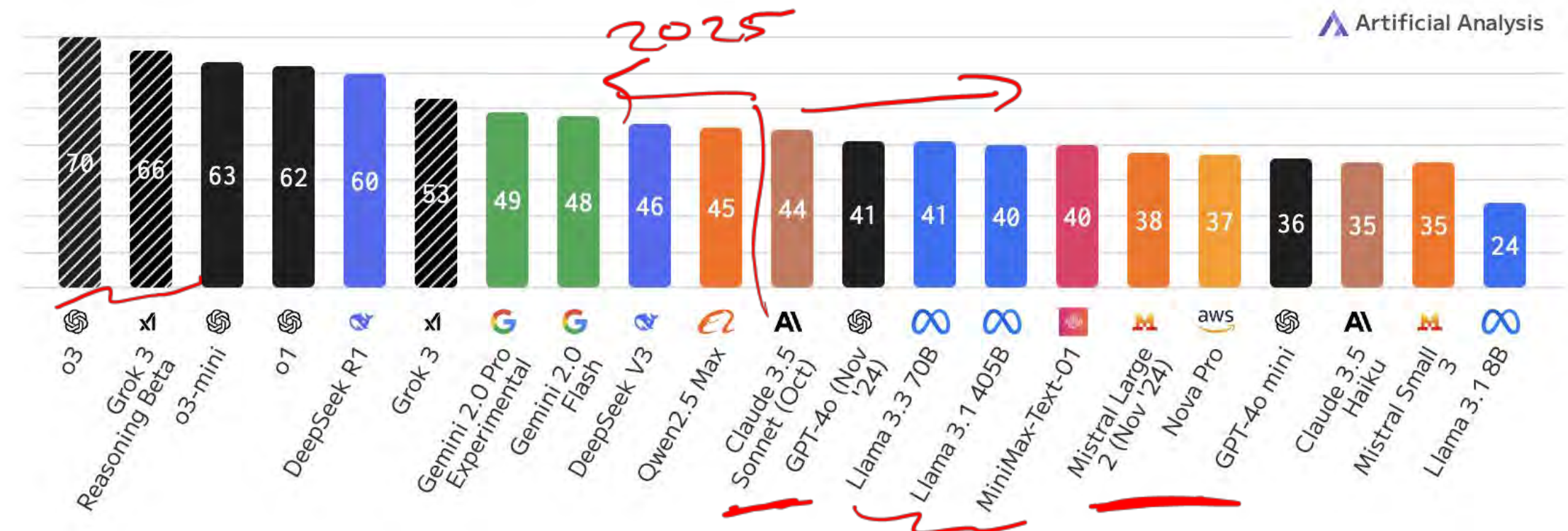
What has changed over the last few months?

What did we do last time?

State of the art when we last met

Artificial Analysis Intelligence Index (Version 2, released Feb '25), includes 7 evaluations spanning reasoning, knowledge, math, coding and more

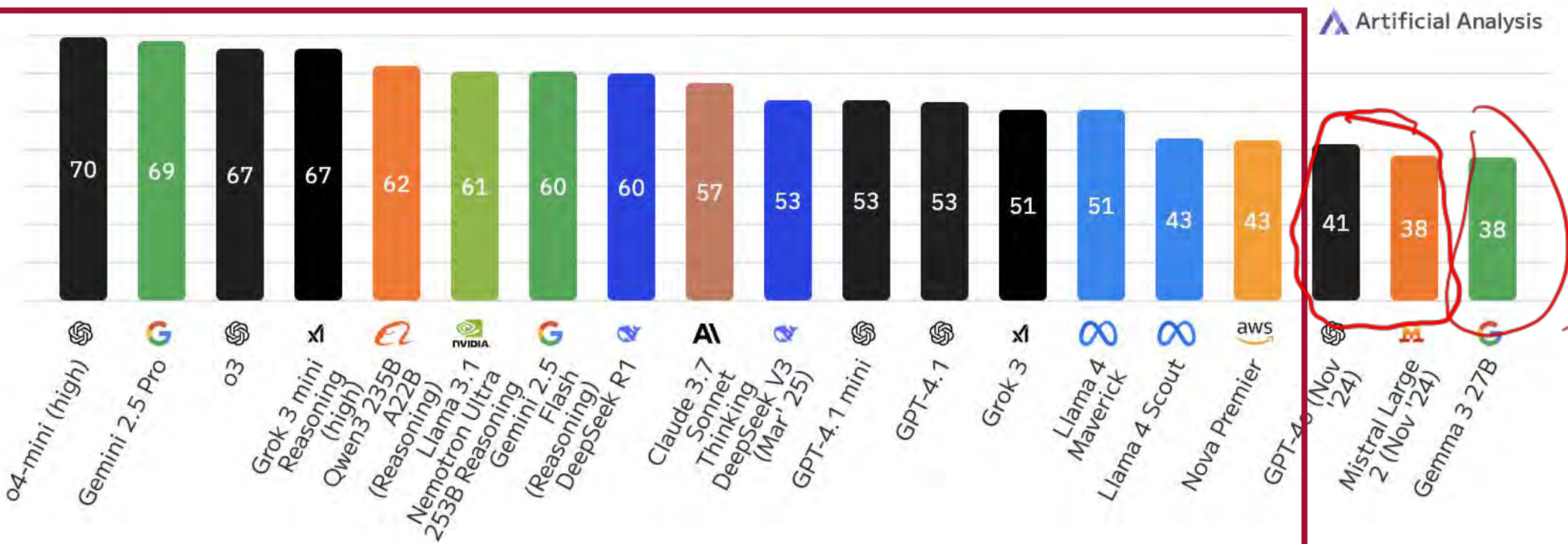
 Estimate (independent evaluation forthcoming)



State of the art today

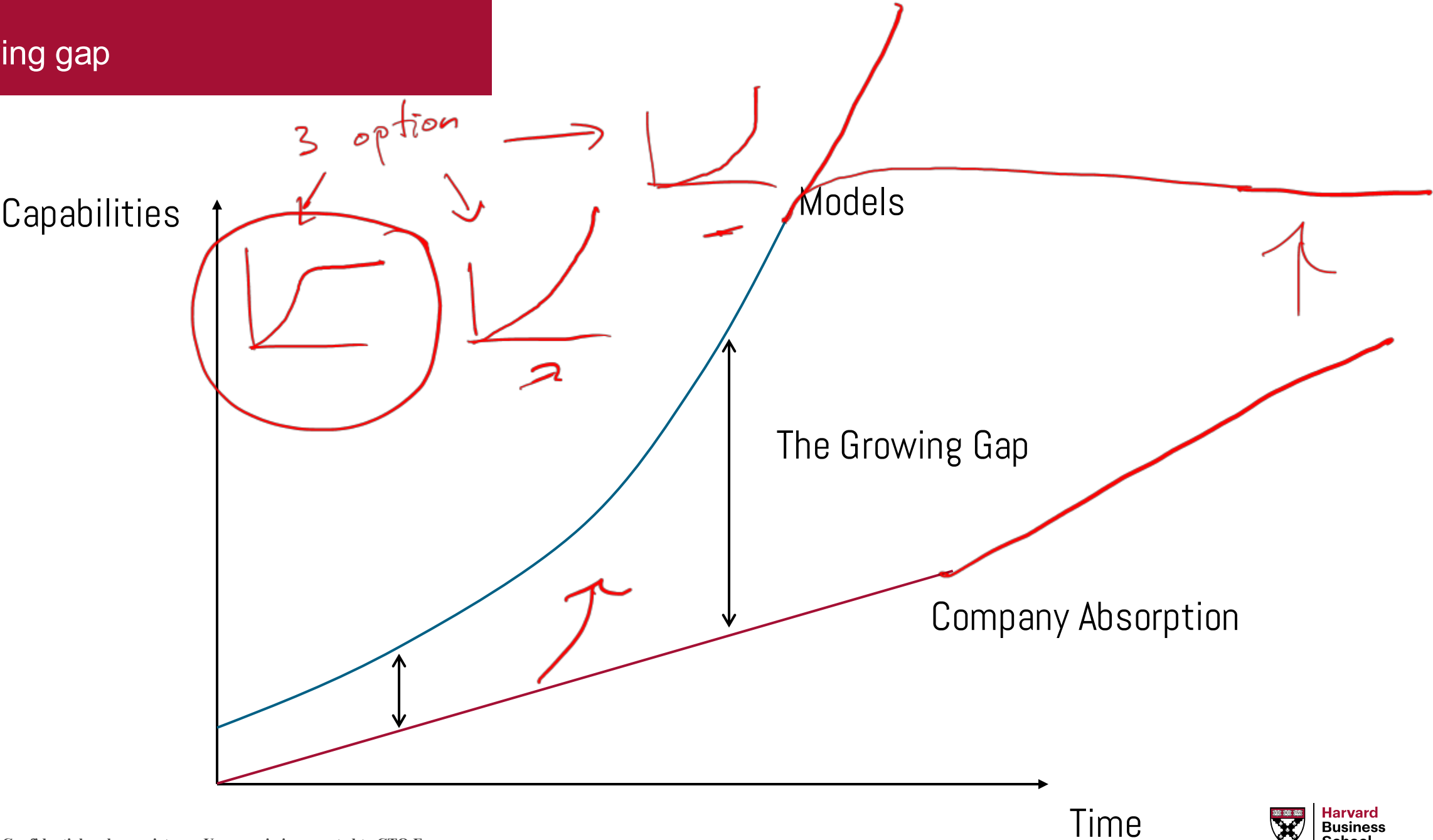
Artificial Analysis Intelligence Index

Intelligence Index incorporates 7 evaluations: MMLU-Pro, GPQA Diamond, Humanity's Last Exam, LiveCodeBench, SciCode, AIME, MATH-500



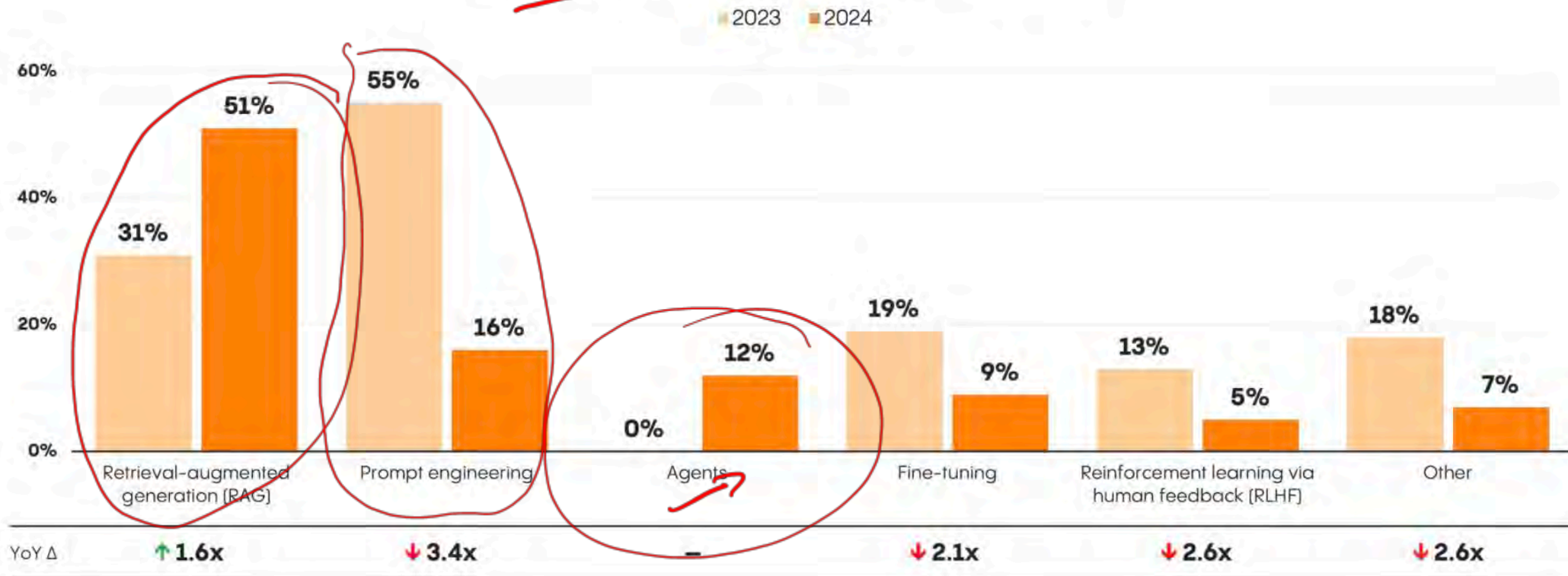
Models from 2025

Growing gap



Current trends

Primary Architectural Approach: 2023 vs. 2024

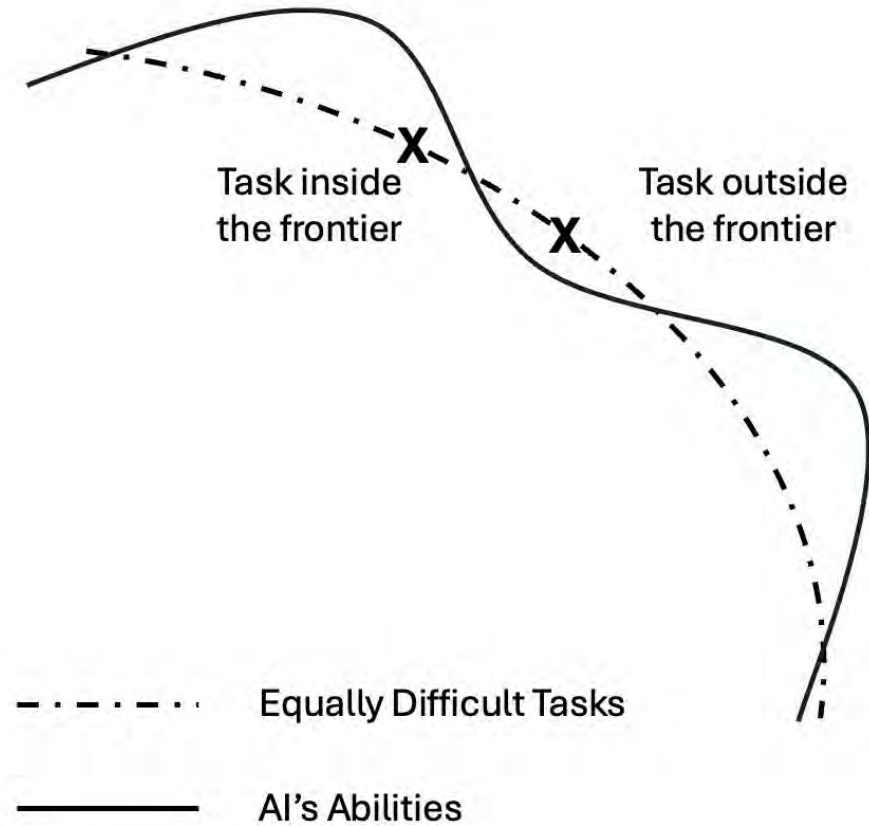
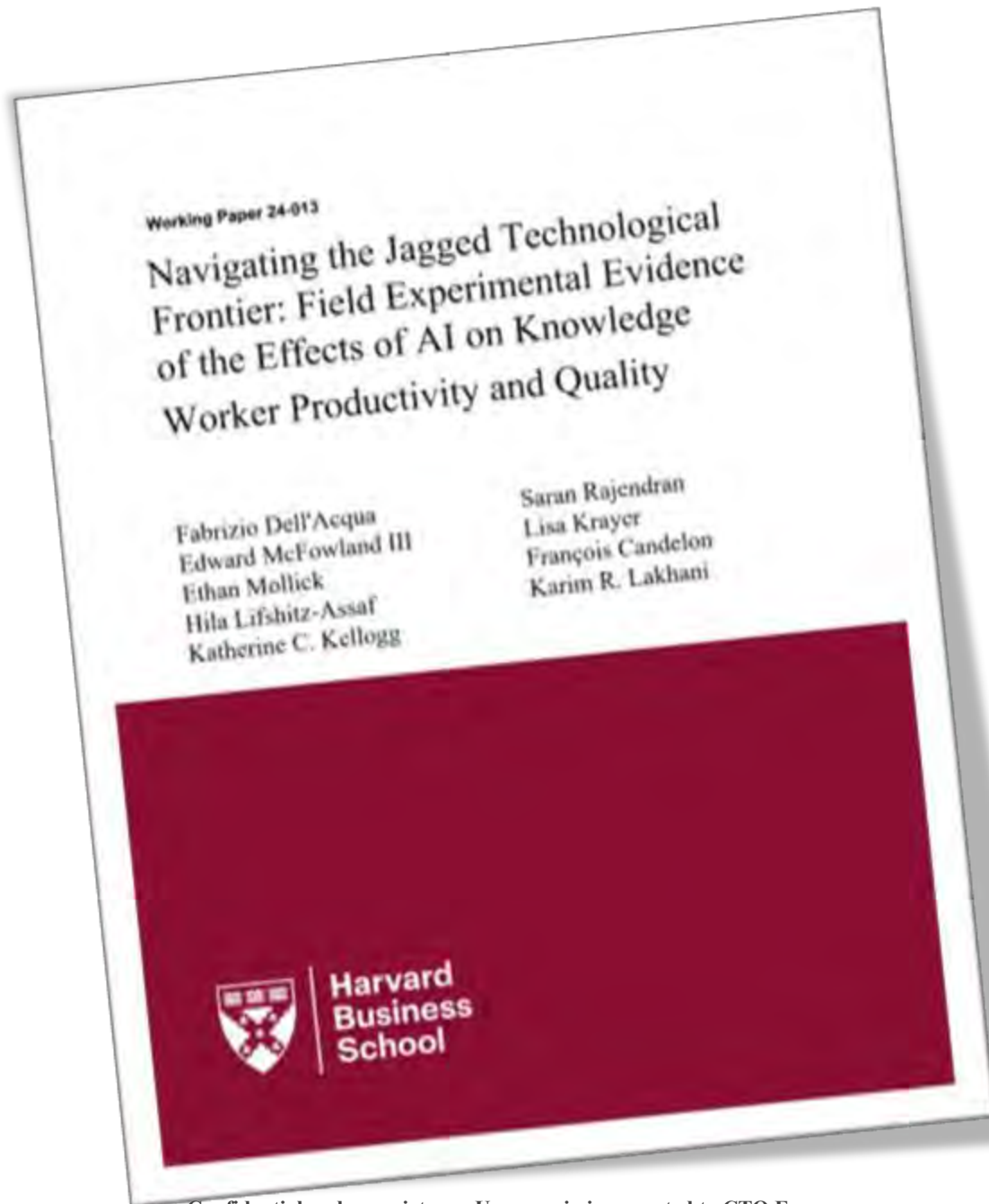


© 2024 Menlo Ventures

Session 0 (October): 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!





- 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.

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.

Session 1:

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

Confidential and proprietary - Use permission granted to CTO Forum.

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**



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

Begin 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.

Motivation for the GPT Evaluator: Modern's Gen AI Strategy



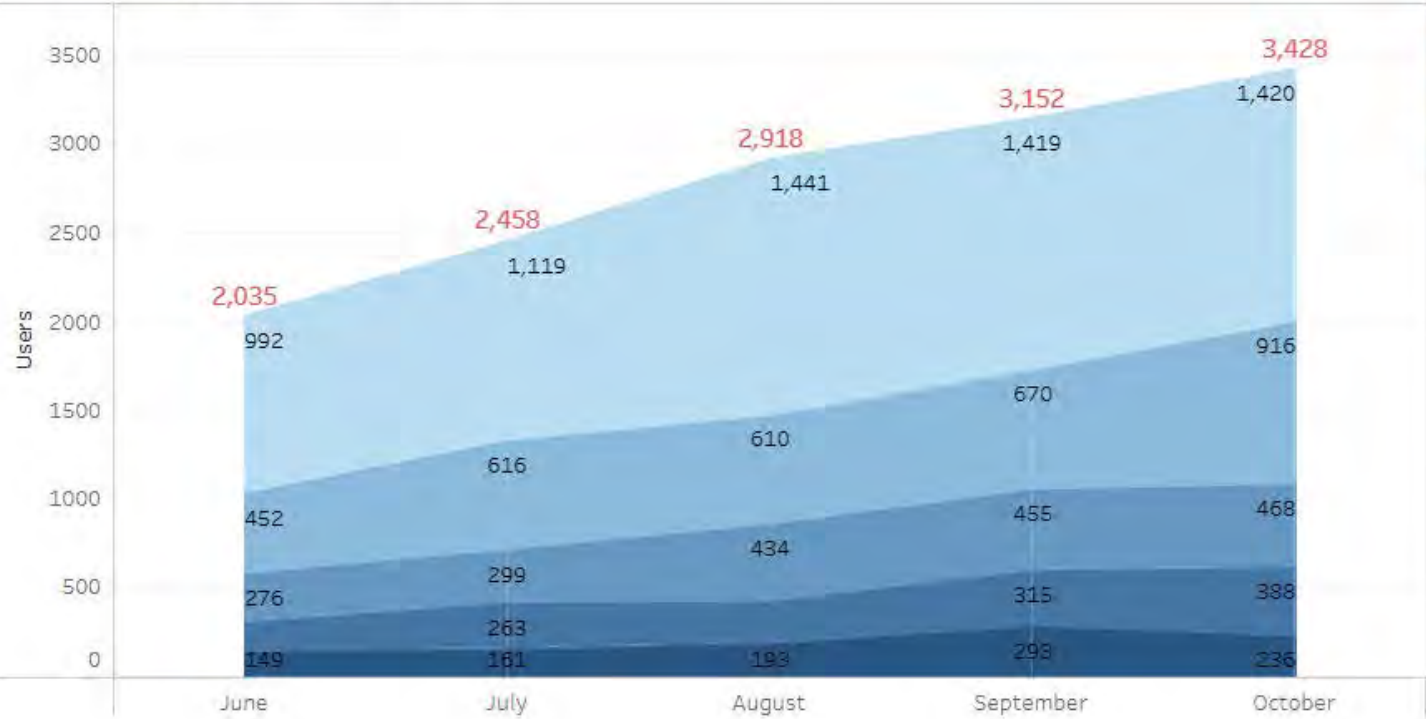
Harvard
Business
School

DRAFT

JANUARY 17, 2024

IAVOR BOJINOV
KARIM LAKHANI
ANNIKA HILDEBRANDT
JAMES WEBER

Usage Frequency of ChatGPT Enterprise



Moderna: Democratizing Artificial Intelligence

In late December 2024, Vice President of AI Products and Innovation Brice Challamel met with CEO Stéphane Bancel, Chief People and Digital Technology Officer Tracey Franklin, and Chief Information Officer Brad Miller to review the adoption of generative AI at Moderna. Over the past year, the biotechnology company had provided all employees with access to OpenAI's ChatGPT Enterprise and encouraged them to incorporate the tool into their daily work. GPTs—Generative Pre-trained Transformers—were a powerful form of artificial intelligence that could reshape a variety of standard business processes. One of ChatGPT's key functionalities allowed users to create and share custom GPTs, each fine-tuned with specific instructions and data to deliver more accurate, relevant responses for particular use cases. (See **Exhibit 1** for Moderna's top 20 GPTs and **Exhibit 2** for GPT usage activity.)

From the outset, Moderna had been a digital-first, AI-focused company. Bancel famously described it as "a technology company that happens to do biology." By 2024, Moderna aimed to obtain 10 new drug approvals within three years. Bancel believed that sustained AI-driven innovation would enable the company's nearly 6,000 employees to keep pace with rival pharmaceutical firms employing more than 100,000 people. To foster this innovation, the company encouraged employees to develop, publish, and maintain custom GPTs, embracing a model akin to the Apple App Store or Google Play Store, where employees could share their creations with each other. Yet AI was not without its flaws. Employees were still learning to wield these emerging tools, and GPTs sometimes produced inaccuracies—so-called "hallucinations." Challamel recognized that as a publicly traded and highly regulated pharmaceutical company, GPT errors in critical processes could have serious consequences for Moderna. To balance risk management with speed and innovation, he implemented governance practices for AI use.

As the Moderna leadership team discussed generative AI adoption, concerns about the use and governance of custom GPTs began to resurface. During the meeting, Challamel highlighted the recent spike in usage of the Self-Review GPT, a tool assisting employees with quarterly and annual performance reviews. Franklin expressed concern: "I'm worried that the Self-Review GPT is potentially problematic, as it is augmenting—and to some extent replacing—a critical process in developing employees. Maybe there are some processes and work that should be kept off-limits? How can Moderna lead the way in pioneering human-AI augmentation in all the work that gets done here?" Seizing on the point, Challamel turned the group's attention to a new GPT called DoseID, created by physician and medical writer Lee Quist, which provided drug dosing recommendations for clinical

Professors Iavor Bojinov and Karim Lakhani, Research Associate Annika Hildebrandt, and Case Researcher James Weber (Case Research & Writing Group) prepared this case. It was reviewed and approved before publication by a company designate. Funding for the development of this case was provided by Harvard Business School and not by the company. HBS cases are developed solely as the basis for class discussion. Cases are not intended to serve as endorsements, sources of primary data, or illustrations of effective or ineffective management.

Copyright © 2025 President and Fellows of Harvard College. To order copies or request permission to reproduce materials, call 1-800-545-7685, write Harvard Business School Publishing, Boston, MA 02163, or go to www.hbsp.harvard.edu. This publication may not be digitized, photocopied, or otherwise reproduced, posted, or transmitted, without the permission of 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.



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.



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

Agent terminology

Feature	Level 1 Agent: Custom GPTs	Level 3 Agent
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



J3016 Automation Levels (cars)

Level	Name	Narrative		Responsibility for:			Mode coverage
				Vehicle direction & speed	Monitoring environment	Fallback	
0	No Automation	Full-time performance by the driver of all aspects of driving, even when "enhanced by warning or intervention systems"		Driver	Driver	Driver	n/a
1	Driver Assistance	Driving mode-specific control by an ADAS of either steering or speed	ADAS uses information about the driving environment; driver is expected to perform all other driving tasks.	Driver and system			Some
2	Partial Automation	Driving mode-specific execution by one or more ADAS for both steering and speed		System			
3	Conditional Automation	Driving mode-specific control by an ADAS of all aspects of driving	Driver must appropriately respond to a request to intervene.				
4	High Automation		If a driver does not respond appropriately to a request to intervene, the car can stop safely.				
5	Full Automation		System controls the vehicle under all conditions and circumstances.				

Range vs Agency trade-off

		Agency	
		Low (prompted)	High (Autonomous)
Task Range (Specialization)	Narrow	RAG/ custom GPTs + Guardrails	Agentic systems
	Broad	Generic Chat Bots	RISK!

Klarna (2024)

<https://www.klarna.com/international/press/klarna-ai-assistant-handles-two-thirds-of-customer-service-chats-in-its-first-month/>

66% of customer service

RETAIL · A.I.

As Klarna flips from AI-first to hiring people again, a new landmark survey reveals most AI projects fail to deliver

BY IRINA IVANOVA

May 9, 2025 at 7:07 AM EDT

Updated May 9, 2025 at 1:01 PM EDT



After leaning hard on AI for customer service, fintech Klarna says it's hiring more humans.

GETTY IMAGES



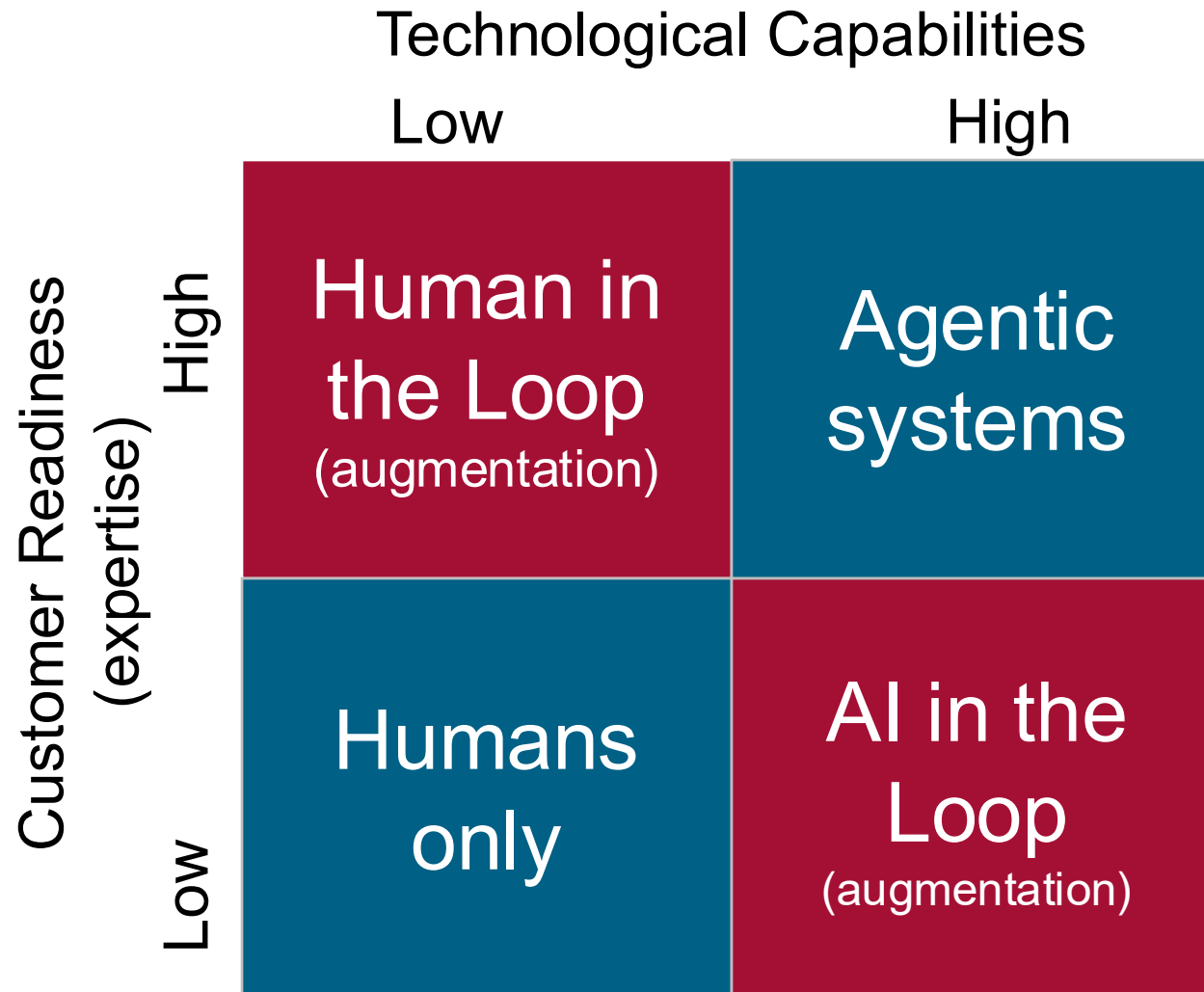
the Klarna app

Track your delivery, handle returns and manage your payments in the Klarna app. Get 24/7 help in our chat, come and go, you'll never miss a message.

\$40 million
Profit Improvement



Customer were not ready...



But... Are your employees ready
for Gen AI?

Three modes of democratizing AI

CONSUMER MODEL



CONTROLLED DEMOCRATIZATION



PRODUCER MINDSET



Consumer Model

Employees have access to centrally managed tools (internal/external).

Tools: GitHub Copilot, LLMs through sandbox, etc.

When: Low level of employee readiness

Examples: Most Companies (P&G, Pernod Ricard, etc.)

Pros:

- Helps employees get accustomed to new tools
- Captures basic efficiencies in operations
- Centralized process redesign work

Cons:

- No AI network effect
- Limited focus on self-improvement
- Challenges in adoption as limited value



Controlled Democratization



Employees have access to centrally managed tools (internal/external) and are empowered to build their own automations for specific functional patterns (such as RAGs or translators).

Tools: Either custom-built or generic tools like Microsoft Copilot Studio and ChatGPT Enterprise with extensive control over development and sharing.

Functional Patterns: Summarization, internal Q&A through RAGs, etc.

When: Medium level of employee readiness

Examples: A few Companies (JP Morgan, etc.)

Pros:

- Enable employees to build and share some customization
- Basic network effects
- Begin to transform individual work

Cons:

- Challenges in adoption as limited value
- Requires changing mindset: consumer → producer.

The 6 most common use democratized functional patterns

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.

Producer Mindset



Employees have access to centrally managed tools (internal/external) and are empowered to build their own automations for **ANY** use cases (such as RAGs or translators) with a centralized effort to transform all working processes.

Tools: Either custom-built or generic tools like Microsoft Copilot Studio and ChatGPT Enterprise, with limited control over development and sharing.

When: High level of employee readiness

Examples: A few companies (Moderna, etc.)

Pros:

- Widespread innovation through a producer mindset
- Significant network effects
- Fundamental rewiring of the company's operating model

Cons:

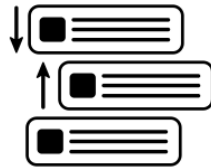
- Managing complexity (e.g., 1,000s of custom GPTs at Moderna)

Process Redesign is Key



70 – 80%
of AI projects fail

Keep your AI projects on track (HBR):



SELECTION

Prioritizing & sequencing effectively



DEVELOPMENT

Accelerate through the AI Factory



EVALUATION

Does it really



ADOPTION

Release and drive growth



MANAGEMENT

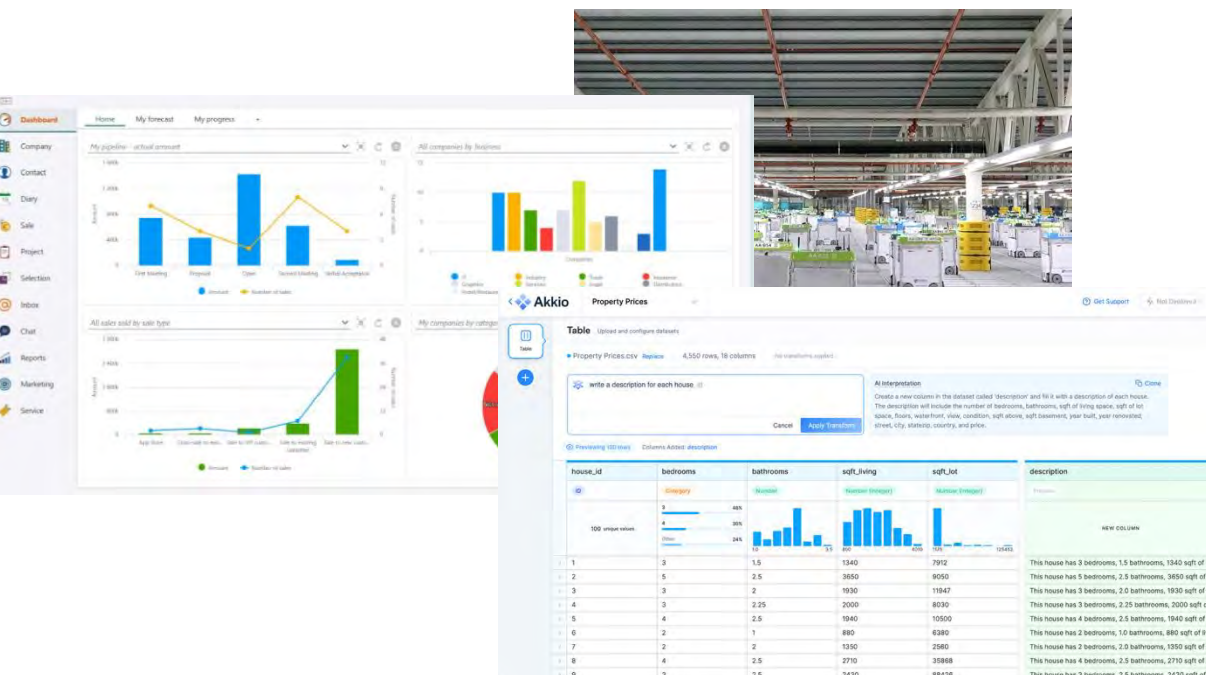
Monitor, manage, improve



Where can you use AI?

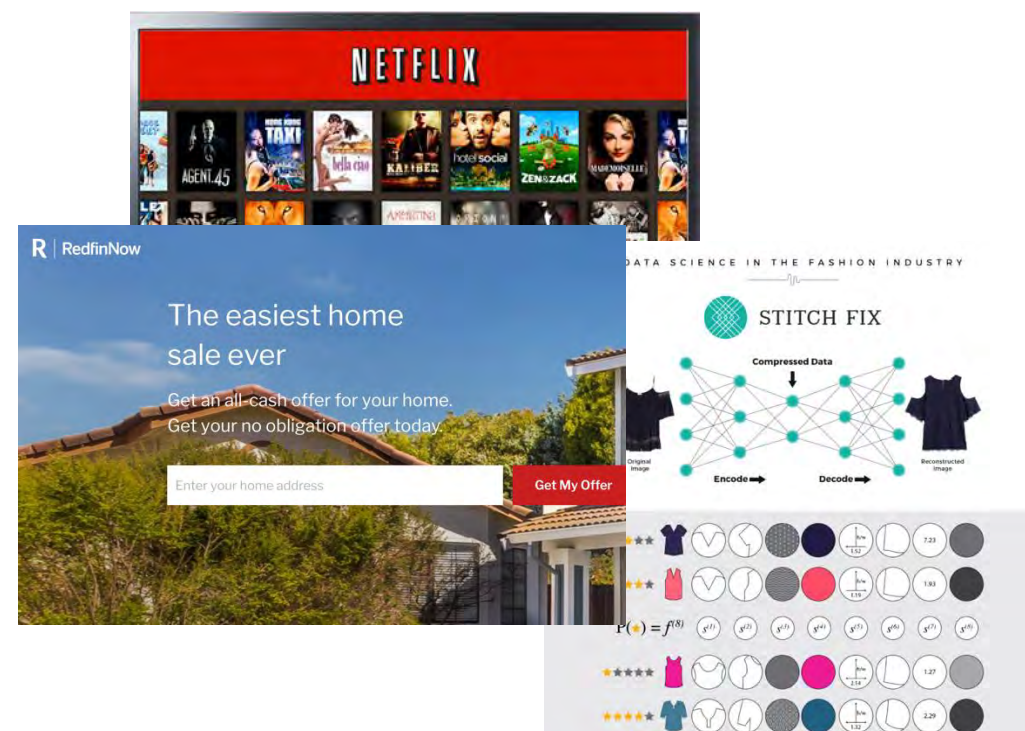
INTERNAL (Operating Model)

Employee facing tool



EXTERNAL (Business Model)

Customer facing product



IMPACT CHECKLIST

- ✓ Clear strategic alignment?
- ✓ Directly measurable return on investment?
- ✓ Augment v Replace

FEASIBILITY CHECKLIST

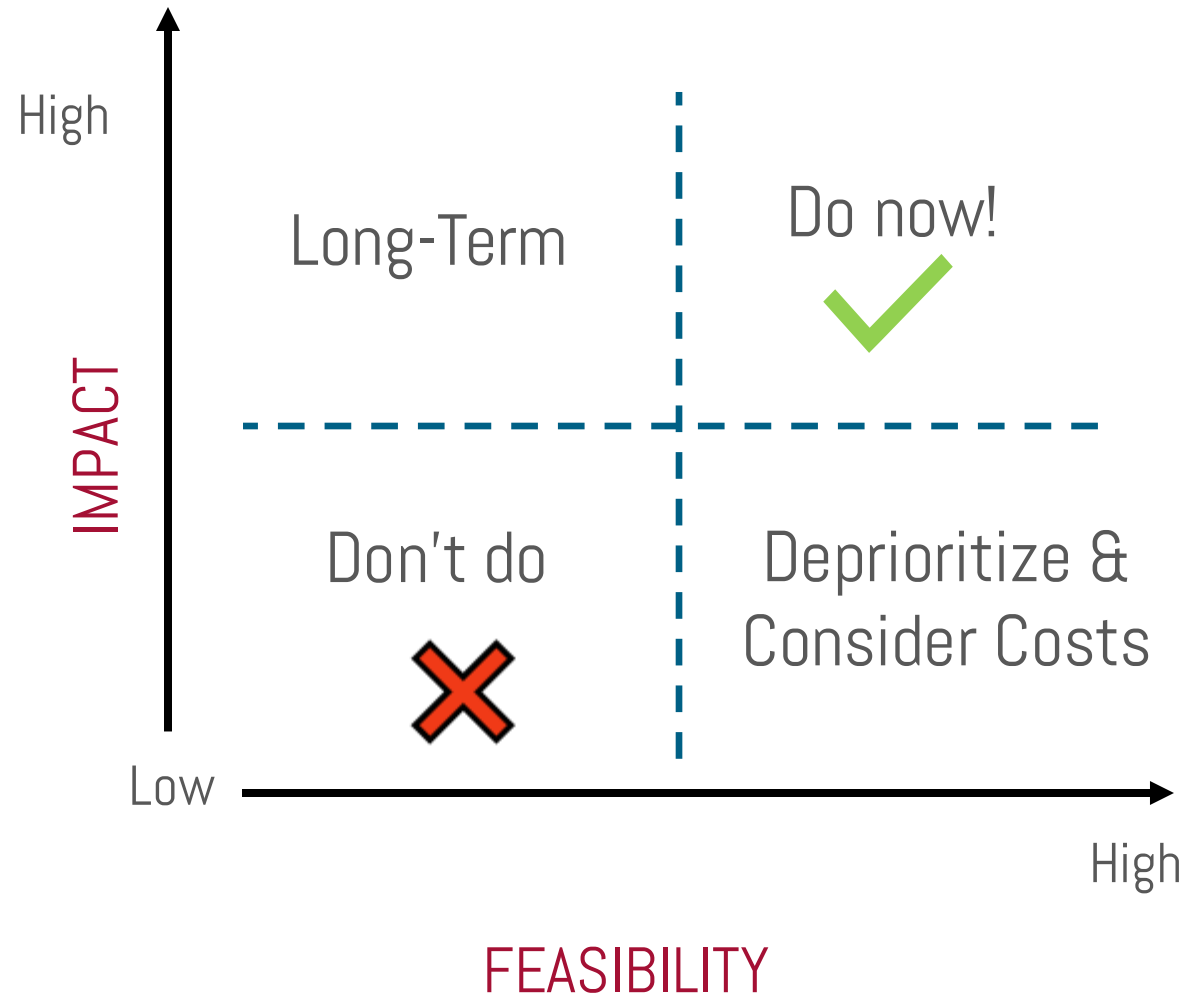
- ✓ Nature of the problem?
- ✓ Do we have the necessary data?
- ✓ Do we have the necessary technology & skills?
- ✓ What are the ethical considerations?
 - ✓ Privacy, Fairness, Bias,...

SCIENTIFIC METHOD

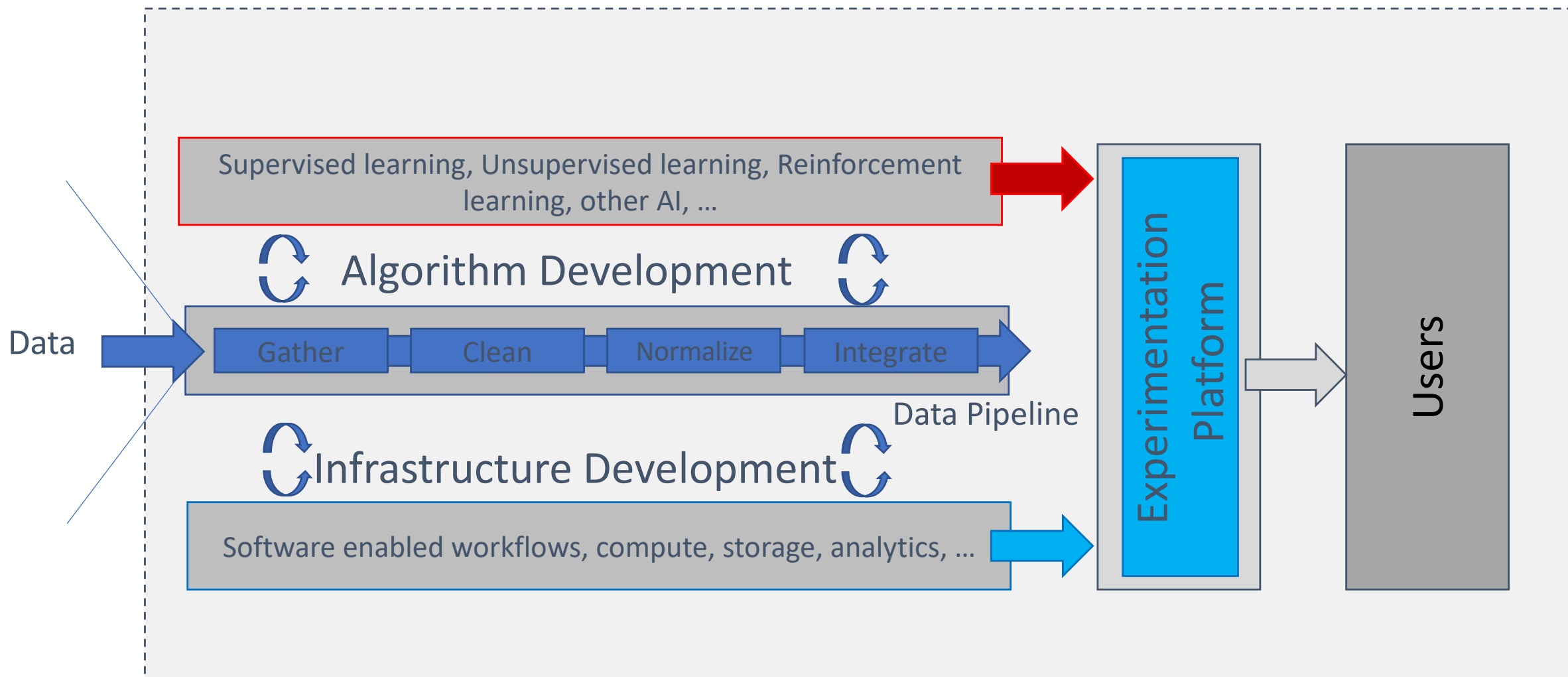
Hypothesis driven

[If____then____by____because____]

Impact – Feasibility 2x2



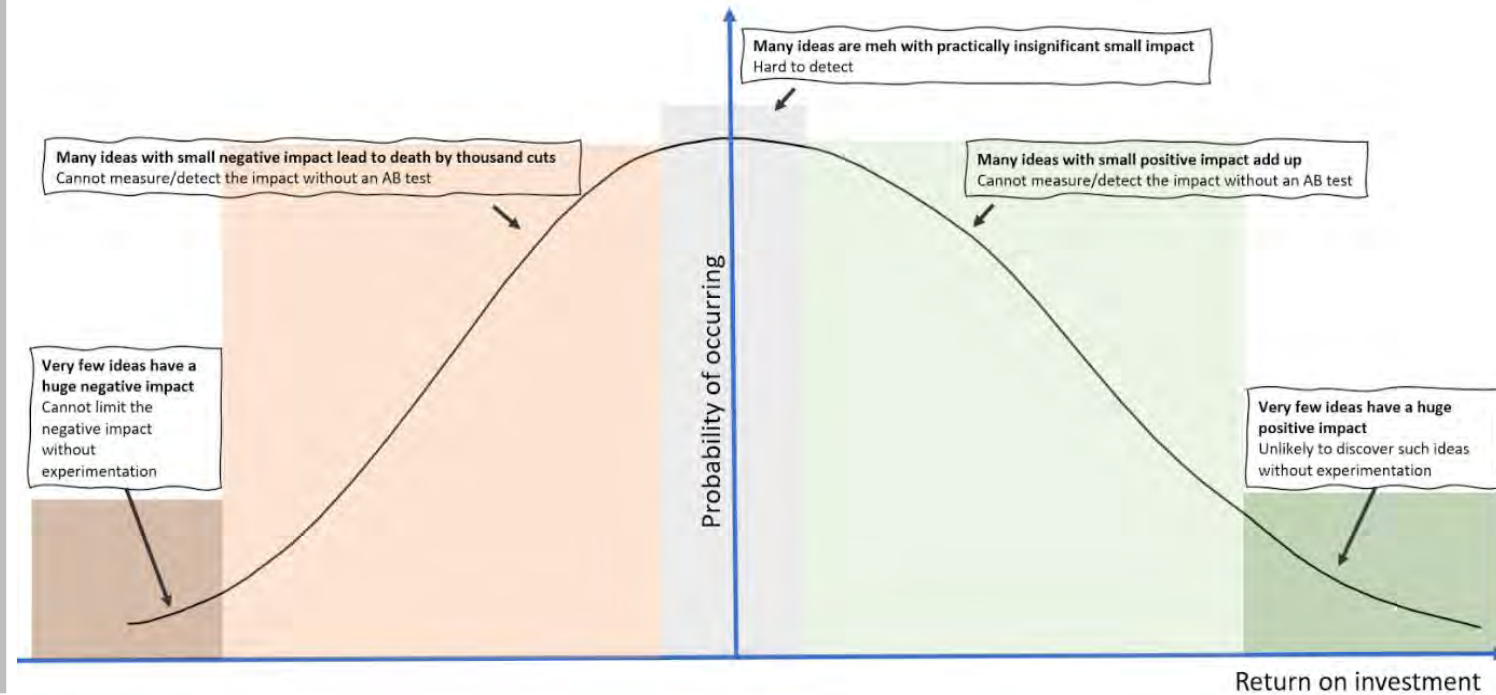
Development: AI Factory



Evaluation

80 – 90%

of initial product changes shipped by Bing & Google have negative or neutral impacts on metrics they were designed to improve



Adoption: Three Pillars of Trust in the Age of AI

ALGORITHM

Quality?

Hallucination?

Fair/Transparent?

DEVELOPER

Involvement?

Replacement?

Hidden Intentions?

PROCESS

Integration?

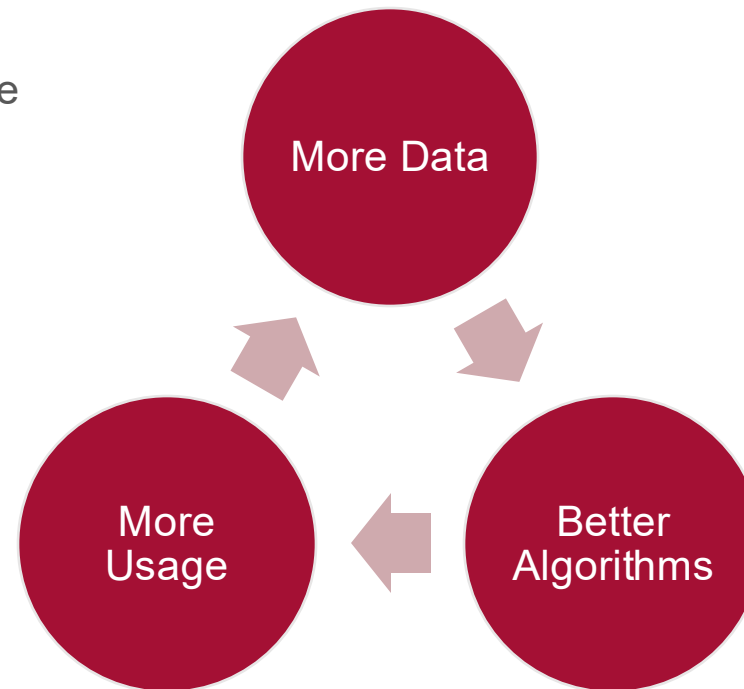
Mistakes?

Safeguards?

OWN

- Every AI needs an owner
- Depending on risk, owners might be teams

IMPROVE



MONITOR

- Performance changes over time
- Bias, fairness & privacy
- AI Audits



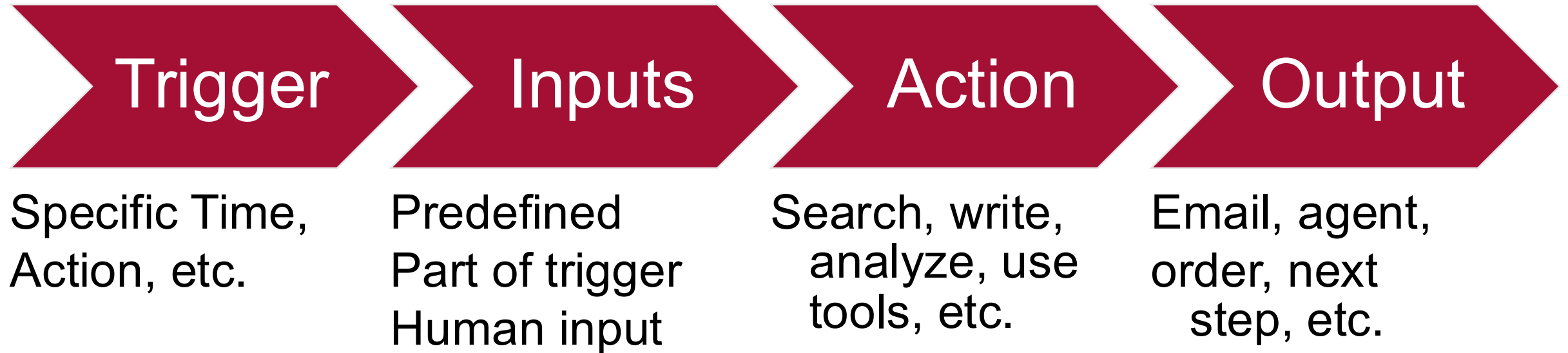


Building an agent

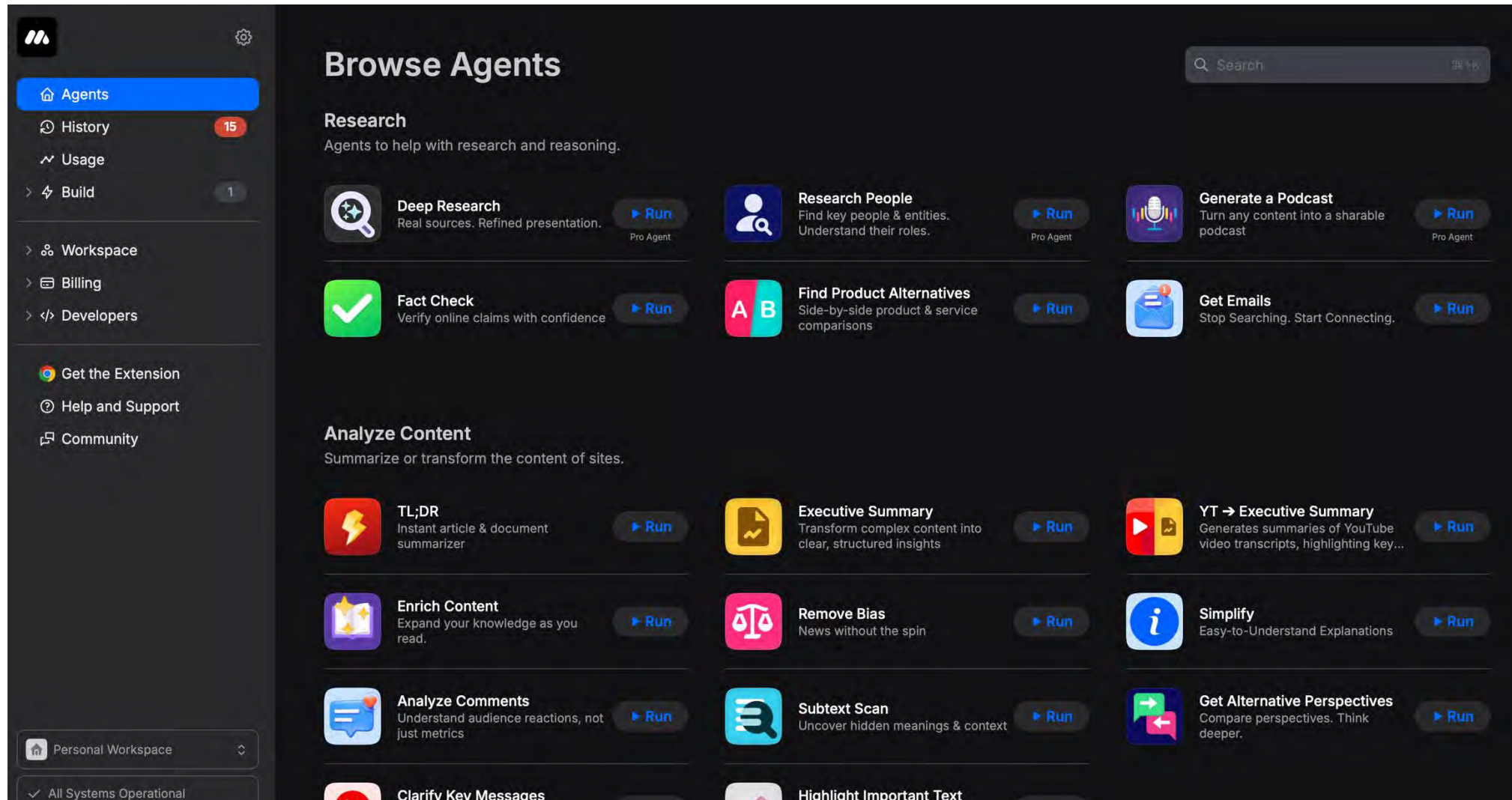
Cyber security agent

Experience with building AI Agents?

4 Key Ingredients for any level 2-3 agent



Mindstudio: Leading platform for AI Agent dev



Today's exercise



Build an agent that searches Google News for company-specific data breaches and sends a daily report.

1. Work as a breakout group by designating 1 person to be the primary “writer.” Have that person screen share.
2. Everyone else should work in parallel using similar prompts and steps.
3. The final output requires showcasing a powerful and useful AI Agent.
4. You'll have about 45-1 hour to work on this.
5. When you finish the basic agent, build on more complicated features and tailor it to your setting.



IAVOR I. BOJINOV
ANNIKA HILDEBRANDT

Creating a Custom AI Agent for Cybersecurity Incident Updates

Introduction

Custom AI agents can leverage generative AI to automate different workflows. These agents can enable employees to create a wide range of custom applications, from asset creation to A/B testing.

To test the power of AI agents, you will be creating an AI that alerts you to possible cybersecurity incidents at your organization. The agent will search for news articles relating to cybersecurity incidents at your company, identify the most recent results, and create and send an email with the findings. While this is a relatively simple example, it should demonstrate the power of AI agents and inspire you to create AI agents for your organizational use cases.

The instructions below are tailored for the MindStudio AI platform. Begin by navigating to mindstudio.ai and creating a free account. With a free account, you are able to build up to 3 custom agents a month and run workflows 1000 times (although please note that usage limits are subject to change).

Preliminaries: Familiarizing yourself with Workflows and Variables

Most agentic AI platforms leverage what are known as workflows. Workflows define a flow of different tasks that should be completed in a specific sequence. These workflows may consist of both AI and non-AI components, often called blocks. For example, a block may create text based on some inputs using generative AI. Other blocks may not use AI, completing tasks such as sending an email or even running some traditional Python code. These workflows combine different technologies to create a powerful automated process.

Where is your key takeaway from the exercise?

How are you tackling the growing gap?

Where can you apply this in your organization?

Thank you!

