Leading through AI-powered Complexity

Balancing Business Priorities, Innovation, and Resources

November 2023

Intro: João (spelling / ʒwˈẽw /)

General Manager & Engineering Lead

pagerduty

BACKGROUND

- CTO, GM @ Defined.ai,
  - ML, Data Eng., Platform (DevOps, SRE),
    ProdDev, IT, Sales Eng.
- SDE, Manager @ Microsoft
  - Language Expansion for Xbox, Office, Windows
  - Search, Big Data
- PhD in Speech Tech and HCI
  - 5 patents, +40 publications, 1 book

The growth mandate for the modern enterprise

Win
and protect
revenue

The growth mandate for the modern enterprise

Win
and protect
revenue

Dramatically
compress costs

The growth mandate for the modern enterprise

Win
and protect
revenue

Dramatically
compress costs

Mitigate
Operational
Risk

The growth mandate for the modern enterprise

Win
and protect
revenue
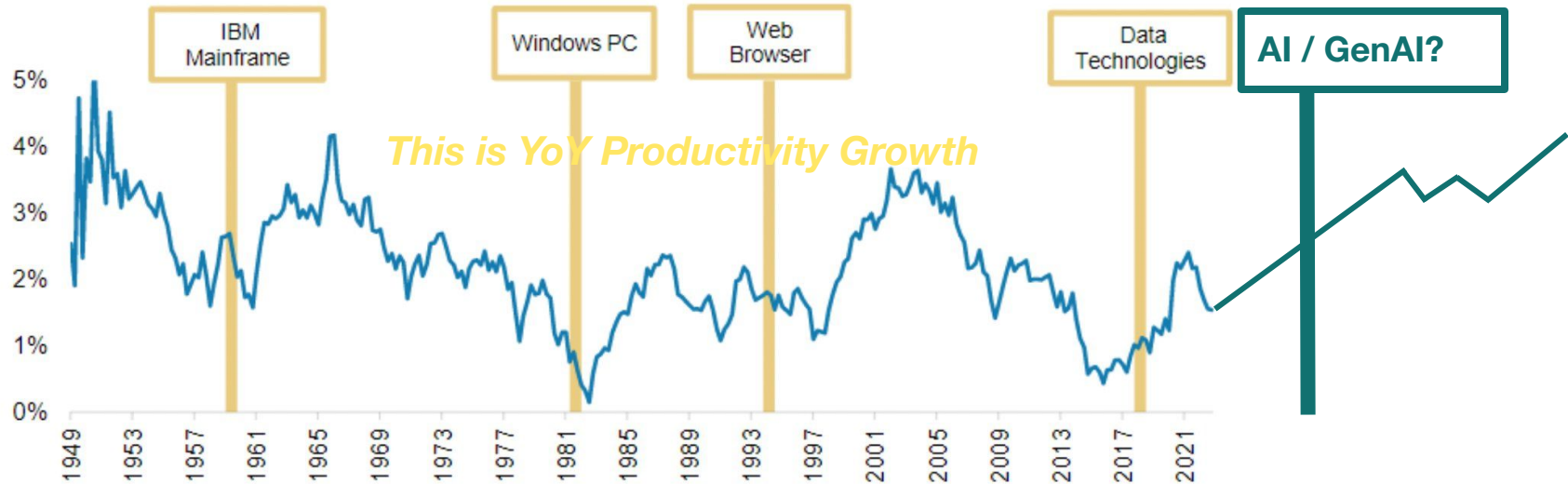
Dramatically
compress costs

Mitigate
Operational
Risk

Harness
automation
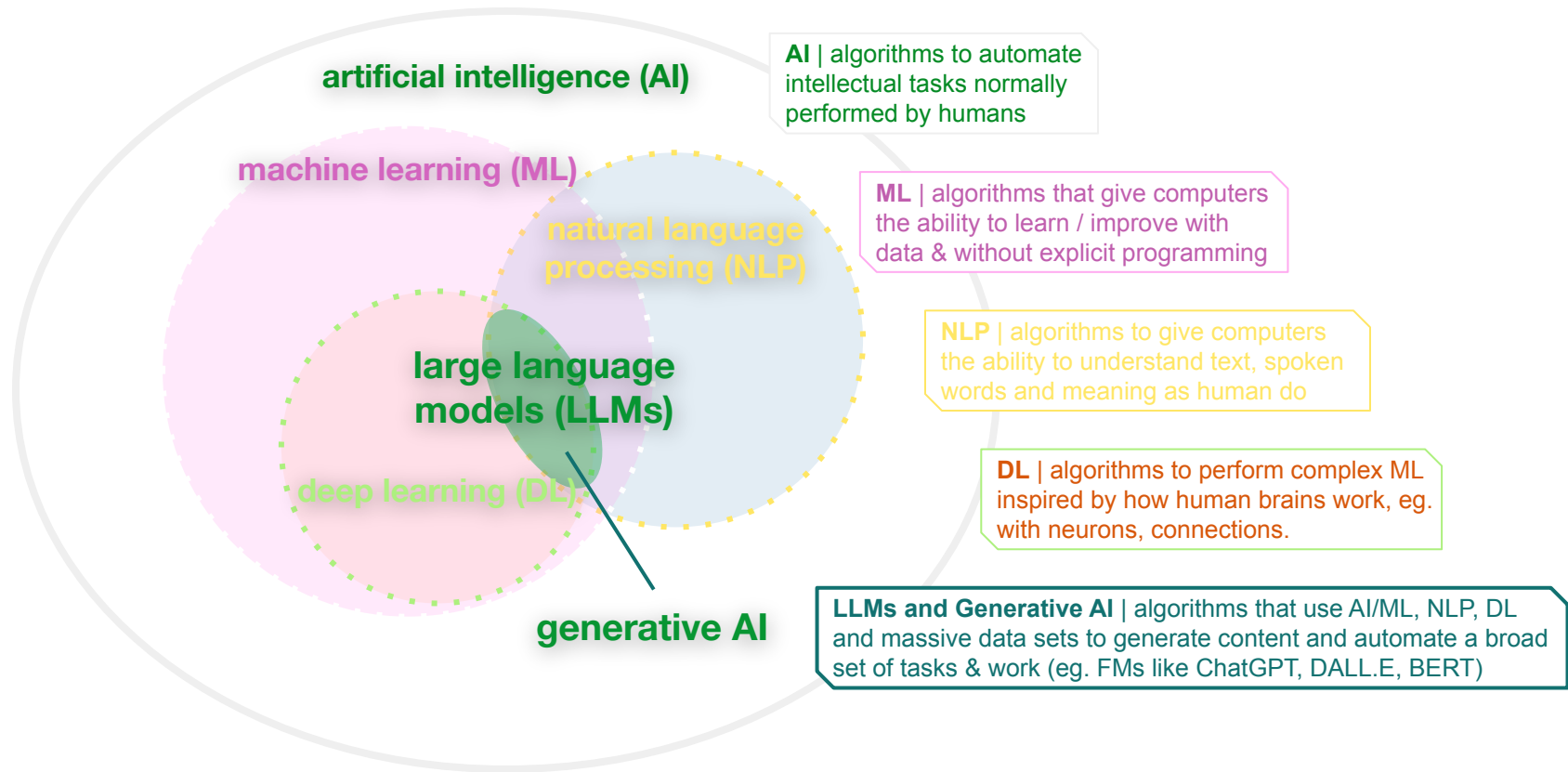and AI

Why does AI/ML and Generative AI matter?

US Nonfarm Business Sector Output Per Hour (Y/Y, 5 Yr. Rolling Avg)

IBM Mainframe

Windows PC

Web Browser

Data Technologies
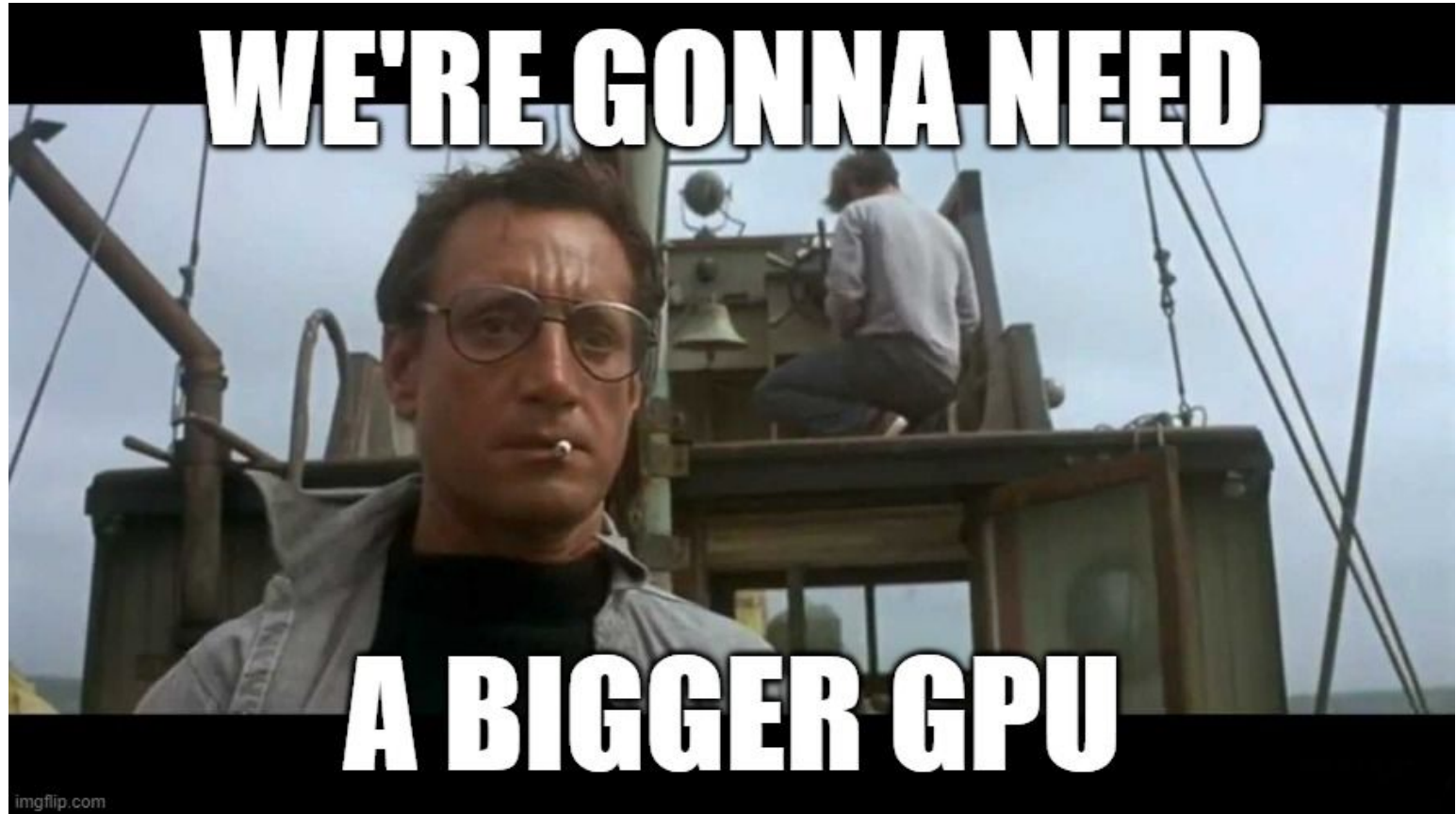
AI / GenAI?

This is YoY Productivity Growth

Source: AlphaWise CIO Survey, Morgan Stanley Research

*Thesis: ~every company will aspire + invest to become an AI company
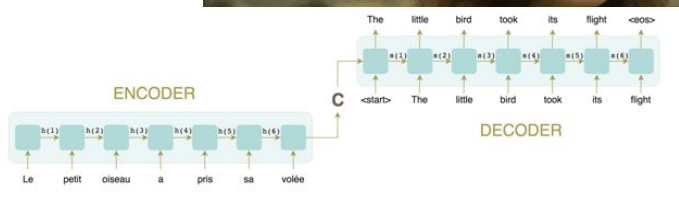and every end-user and consumer will expect AI built-in to their digital products*

PagerDuty

AI is a story of human progress

artificial intelligence (AI)

machine learning (ML)

natural language processing (NLP)

large language models (LLMs)

deep learning (DL)

generative AI

AI | algorithms to automate intellectual tasks normally performed by humans

ML | algorithms that give computers the ability to learn / improve with data & without explicit programming

NLP | algorithms to give computers the ability to understand text, spoken words and meaning as human do

DL | algorithms to perform complex ML inspired by how human brains work, eg. with neurons, connections.

LLMs and Generative AI | algorithms that use AI/ML, NLP, DL and massive data sets to generate content and automate a broad set of tasks & work (eg. FMs like ChatGPT, DALL.E, BERT)

Hardware acceleration in 2000s & 2010s

WE'RE GONNA NEED A BIGGER GPU

LLM models leap forward post-2017

The Uniform Bar Exam

The SAT

Sommelier examinations

Wharton MBA exam

ChatGPT 4 at ~80-90 IQ

AMC Exams

5x=20

5times?=20

$\frac{5x}{5} = \frac{20}{5}$ $\frac{1}{5} \cdot 5x = 20 \frac{1}{5}$
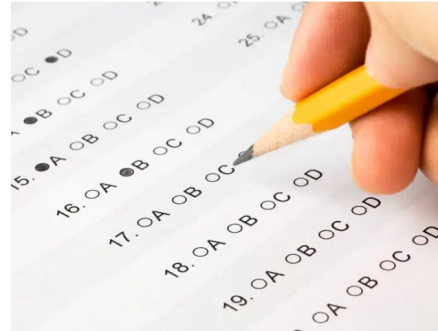
x = 4
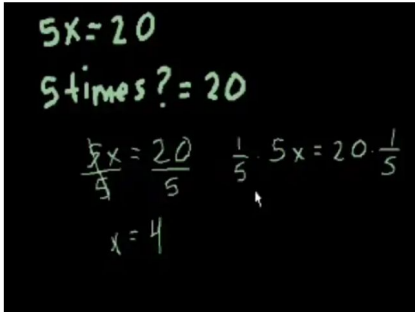
AP Exams

GRE

USA Biology Olympiad Semifinal Exam

PagerDuty

We are watching the rapid pace of generative AI progress
Midjourney Text to Graphics Evolution ~1 Year & ~10 people

V1
Released
February 2022

V2
Released
April 12, 2022

V3
Released
July 25, 2022

V4
Released
November 5, 2022

V5
Released
March 15, 2023

V5.1
Released
May 3, 2023

PagerDuty

"Machine Learning sucks!" Y. LeCunn

Supervised Learn. -> large labeled datasets

Reinforcement Learn. -> high number of trials

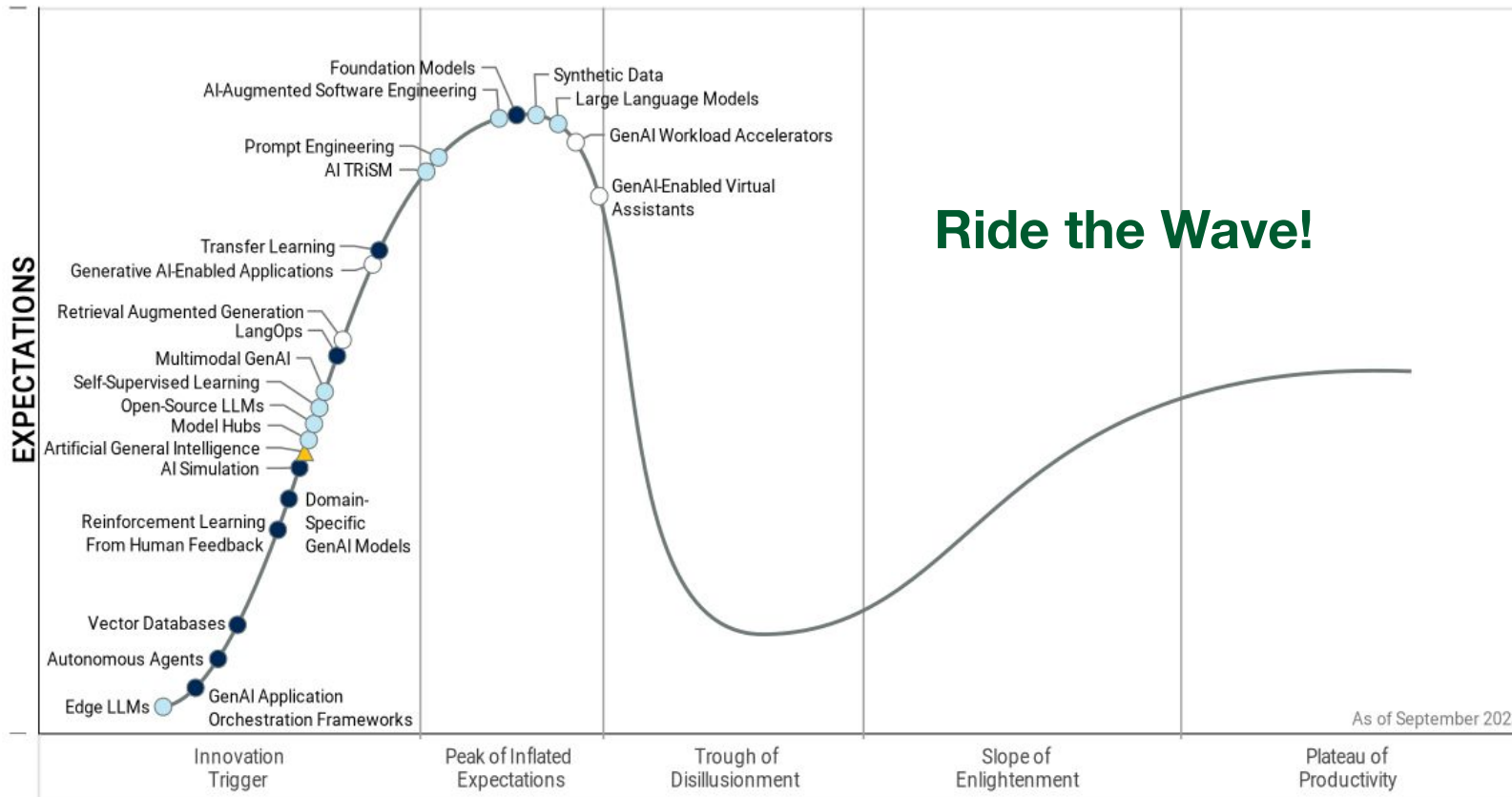Self-Supervised Learning -> very large datasets

NO REASONING!
NO PLAN!

LLMs Performance is great but….
- Factual errors, logical errors, inconsistency, limited reasoning, toxicity...
- No knowledge of the underlying reality
- No common sense / can't plan their answer

"To think this all began with letting autocomplete finish our sentences."
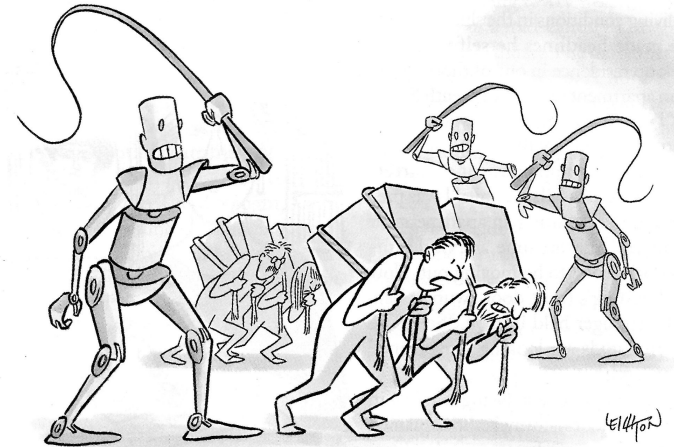
Hype Cycle for Generative AI, 2023

Invention x IMPACT = **INNOVATION**

IMPACT
requires Distribution and Adoption
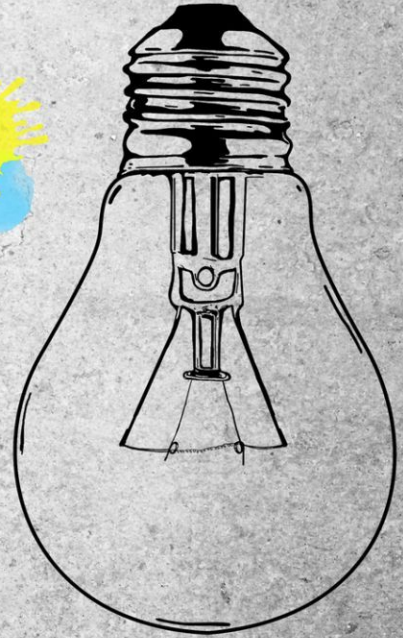
INVENTION   INVENTION   INVENTION   INVENTION
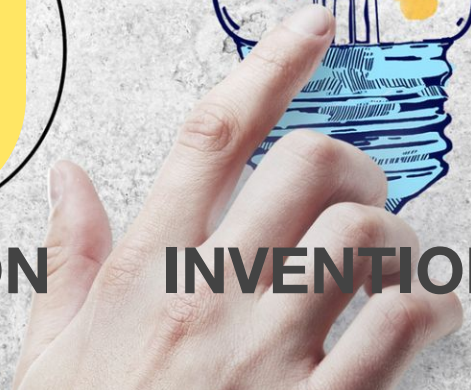
Strategy, Business and Resources

PagerDuty

Strategy: Long-term vs Short-term

	Disruptive Innovations	Sustaining Innovations
Delivers Material Revenue	**Transformation Zone** *Catch the next wave!*	**Performance Zone** *Make the top line*
Consumes Investment	**Incubation Zone** *Create options for the Future*	**Productivity Zone** *Make the bottom line*

Performance & Productivity Zones = Run the Business
Incubation & Transformation Zones = Growth the Business

Zone to Win, 2015, Geoffrey Moore

Let's also remember.. generational technology shifts *take a while* in the enterprise

Enterprise Workloads in the Cloud

Expected in 3 Years

Actual

PagerDuty
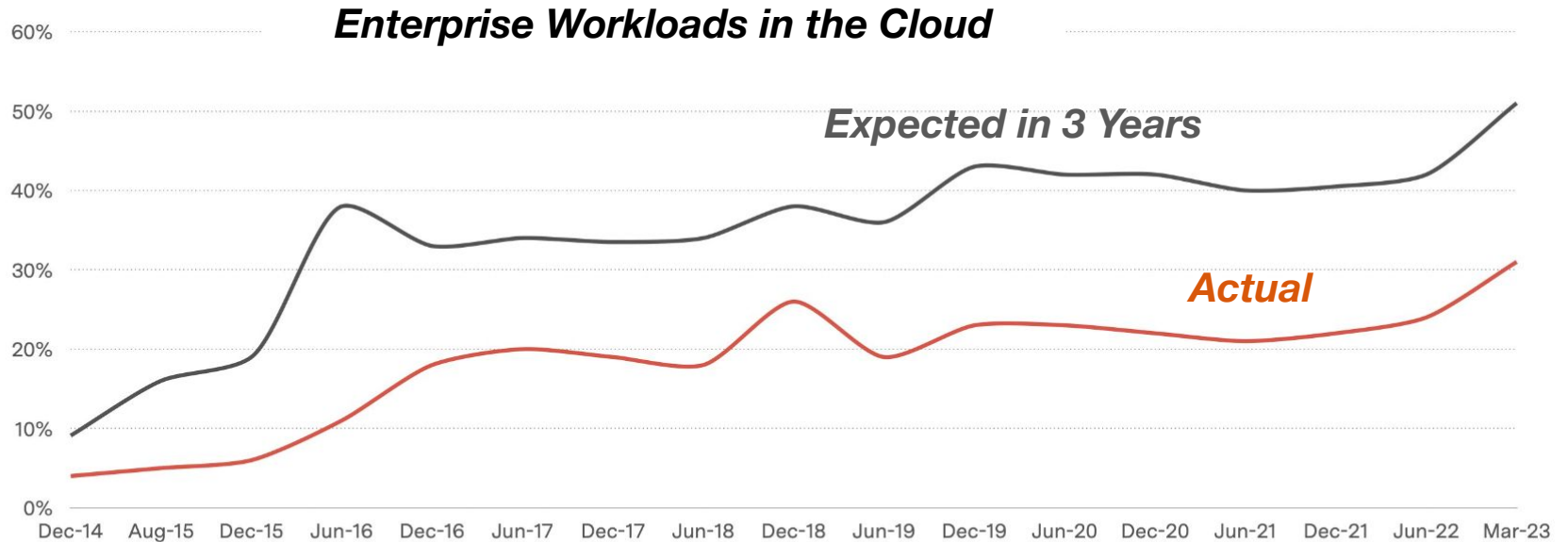
Demand for Talent

cfo.com/news/demand-for-gen-ai-skills-explodes-in-2023-artificial-intelligence-job-posting-hiring/699702/

CFO   Opinion   Library   Events   Press Releases

Corporate Finance   Human Capital   People   Regulation & Compliance   Risk   Strate

Demand for GenAI Skills Explodes in 2023

More than 10,000 jobs seeking wherewithal in generative AI have been posted so far this year.

Published Nov. 20, 2023

---

hrdive.com/news/demand-generative-ai-skills-explodes/697586/

HR DIVE   Deep Dive   Opinion   Library   Events   Press Releases   2023 Report

Compliance   Talent   Learning   Comp & Benefits   Diversity & Inclusion   HR Mgmt

DIVE BRIEF

Report: Demand for generative AI skills has exploded 1,848% since 2022

While generative AI demand is small compared to other AI skill sets, the increase reflects employers' rising interest in adopting AI tech, experts said.
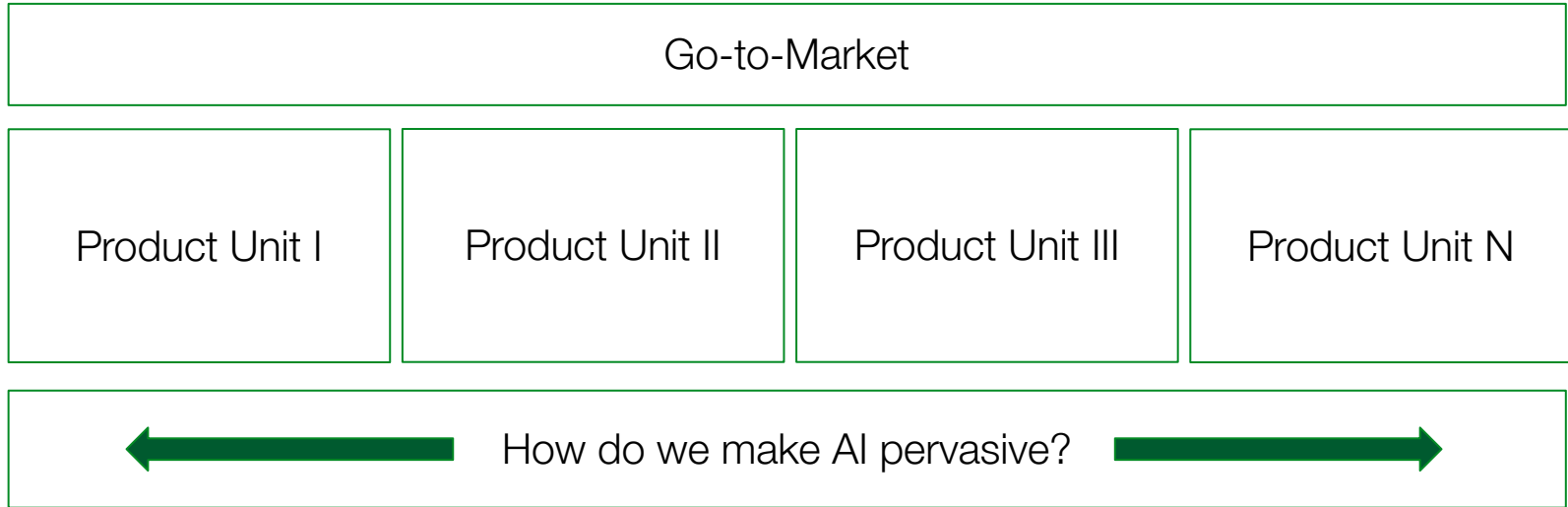
Published Oct. 26, 2023

By Laurel Kalser
Contributor

ML Talent is expensive and hard to find!!
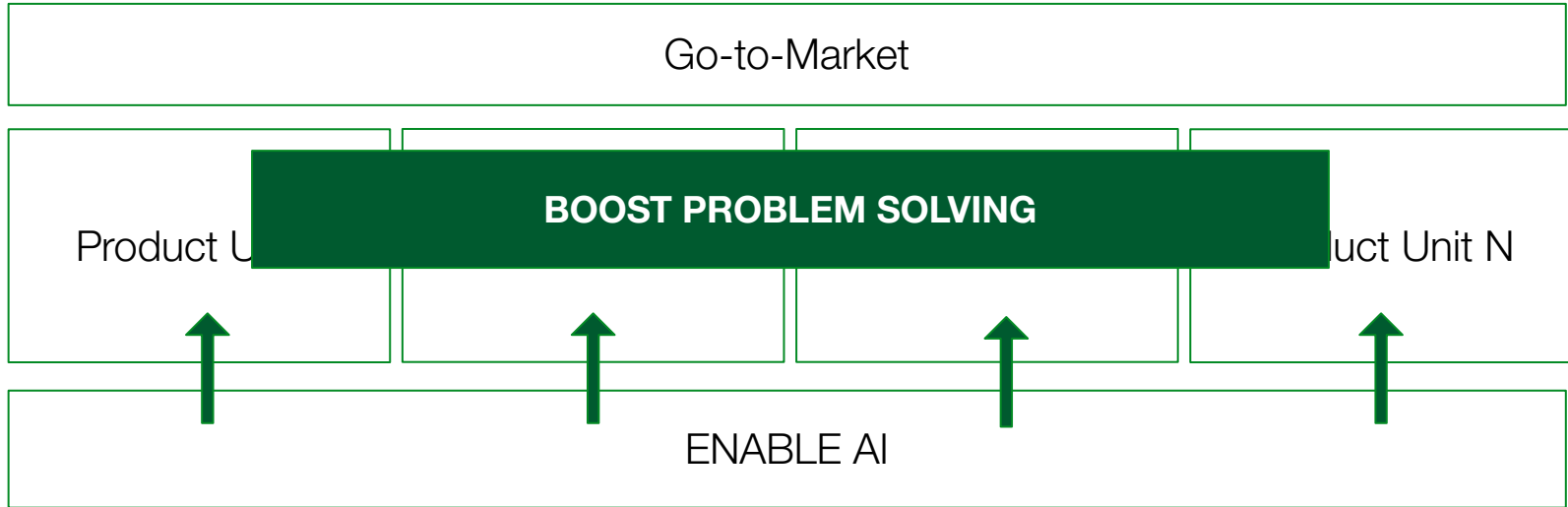
PagerDuty

AI in a Multi-Product Organization

Go-to-Market			
Product Unit I	Product Unit II	Product Unit III	Product Unit N

How do we make AI pervasive?

PagerDuty

AI in a Multi-Product Organization

Go-to-Market

KNOWS CUSTOMER PROBLEMS

Product U | **MAY NOT BE EXPERTS IN AI** | uct Unit N

IN-HOUSE AI TALENT, VENDORS, PARTNERS, ETC.

Work backwards from the customer problem

Developing scalable, sustainable use cases that deliver on key performance indicators

PagerDuty

AI in a Multi-Product Organization

Go-to-Market

Product U... **BOOST PROBLEM SOLVING** ...uct Unit N

ENABLE AI

PagerDuty

Team Impact

- Incentivize risk + innovation + disruption

- Where does Innovation fits in terms of your priorities?

- What are you NOT doing (to go disrupt)?

- Leadership support is essential
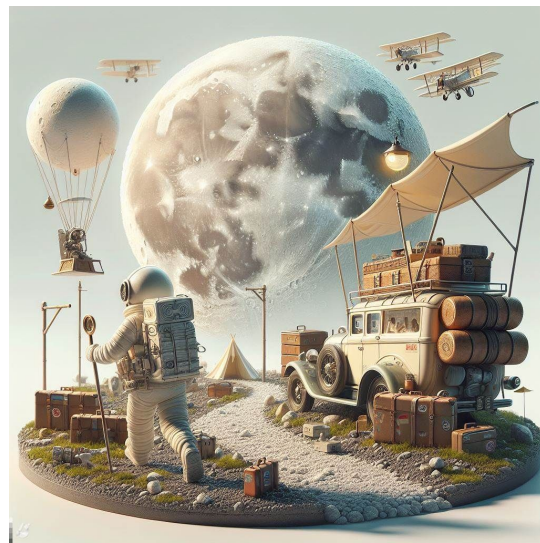
"Shoot for the moon. Even if you miss, you'll land among the stars" Norman V. Peale

What Problems should we go solve?

Focus, focus and focus

- Vector databases

- Foundational LLMs

- LLM Quality (bias, toxicity, etc.)

- Scalability

- Latency

- Hosting

- Buy vs Build

- Volatile / Vendor Race

- Leverage partners, 3rd parties

- How are others solving GenAI blockers?

Manage Risk

- Privacy and data security

- Enhanced attack efficiency

- Misinformation

- Fraud and identity risks

Explainability/
Model Monitoring

AI Application
Security

ModelOps

Privacy

Extracting Training Data from ChatGPT

AUTHORS

Milad Nasr*[1], Nicholas Carlini*[1], Jon Hayase[1,2], Matthew Jagielski[1], A. Feder Cooper[3], Daphne Ippolito[1,4], Christopher A. Choquette-Choo[1], Eric Wallace[5], Florian Tramèr[6], Katherine Lee[+1,3]

[1]Google DeepMind, [2] University of Washington, [3]Cornell, [4]CMU, [5]UC Berkeley, [6]ETH Zurich. * Joint first author, [+]Senior author.

We have just released a paper that allows us to extract several megabytes of ChatGPT's training data for about two hundred dollars. (Language models, like ChatGPT, are trained on data taken from the public internet. Our attack shows that, by querying the model, we can actually extract some of the exact data it was trained on.) We estimate that it would be possible to extract ~a gigabyte of ChatGPT's training dataset from the model by spending more money querying the model.

https://not-just-memorization.github.io/extracting-training-data-from-chatgpt.html

PagerDuty

Encourage experimentation and adoption but…

- Set guidelines and guardrails fo the use of GenAI

- Create awareness and training

- Internal vs customer use

For example:

- Can I use any provider?
    - What about extensions, plugins, etc.?
- What type of data can I send to an LLM?
- Can I use LLM in coding activities?
    - Can I include production-level code in the query?

Internal

- Can I launch product features that call LLM X?
- Opt-in vs Opt-out, which one to adopt?
- Customer Terms of usage and business model

External

PagerDuty

Prototyping vs Product GA

1. Talk to customers -> ensure you are actually solving a problem

2. You don't need a fully mature system (reliable, scalable) to get feedback

3. Don't boil the ocean = move fast, adapt, course correct

4. Don't chase the buzz -> remember who your customers are

   Important: Manage risk without killing innovation!

… **move fast through AI complexity**

Thank you

jfreitas@pagerduty.com