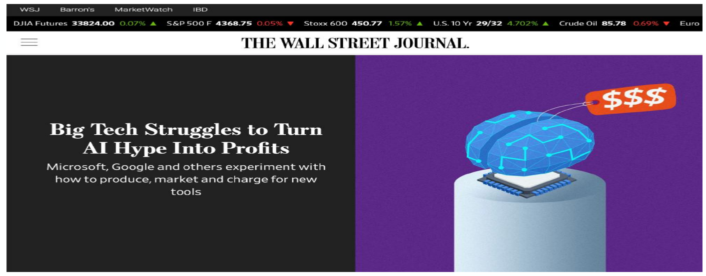
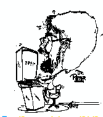


AI–Handle Me with Care



peter.aiken@anythingawesome.com +1.804.382.5957

Peter Aiken, Ph.D.

- I've been doing this a long time
- My work is recognized as useful
- Associate Professor of IS (vcu.edu)
- Institute for Defense Analyses (ida.org)
- DAMA International (dama.org)
- MIT CDO Society (iscdo.org)
- Anything Awesome (anythingawesome.com)
- Experienced w/ 500+ data management practices worldwide
- 13 books and dozens of articles
- Multi-year immersions
  - US DoD (DISA/Army/Marines/DLA)
  - Nokia
  - Deutsche Bank
  - Wells Fargo
  - Walmart
  - HUD …

$1,500,000,000.00 USD

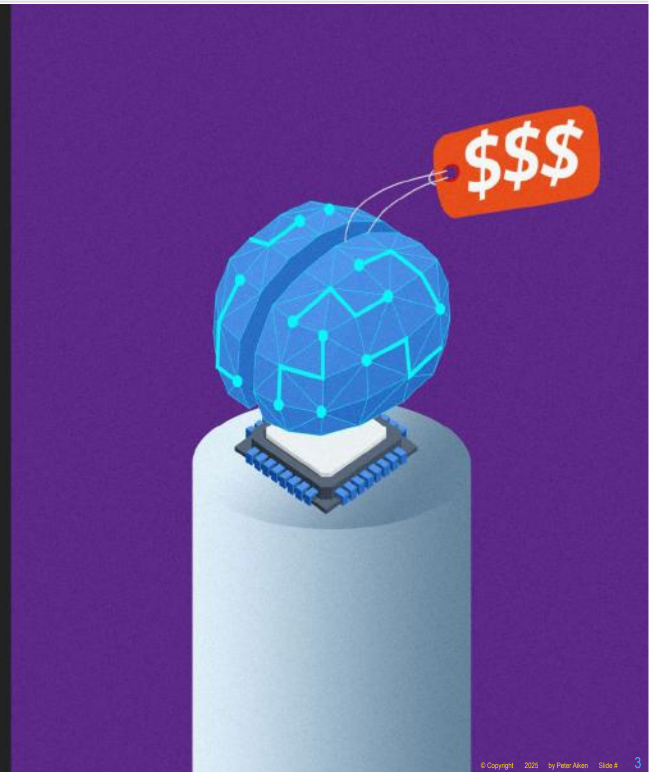
WSJ Barron's MarketWatch IBD

DJIA Futures **33824.00** 0.07% ▲ S&P 500 F **4368.75** 0.05% ▼ Stoxx 600 **450.77** 1.57% ▲ U.S. 10 Yr **29/32** 4.702% ▲ Crude Oil **85.78** 0.69% ▼ Euro

THE WALL STREET JOURNAL.

Big Tech Struggles to Turn AI Hype Into Profits

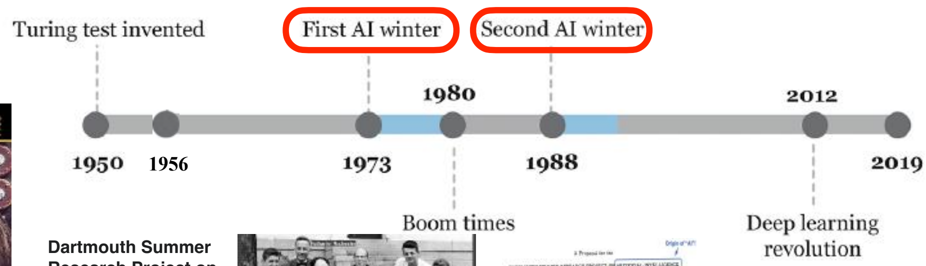
Microsoft, Google and others experiment with how to produce, market and charge for new tools

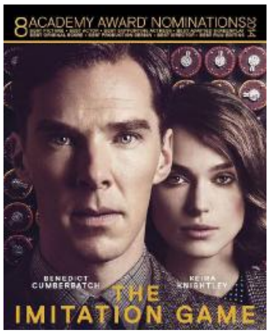


AI’s version of the Traveling Wilburys

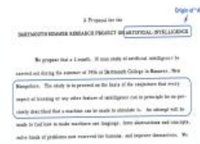
Been beat up and battered ‘round
Been sent up, and I've been shot down

Dartmouth Summer Research Project on Artificial Intelligence

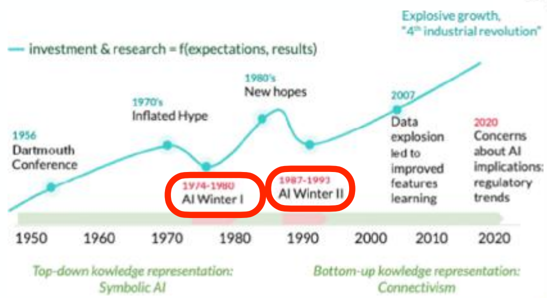


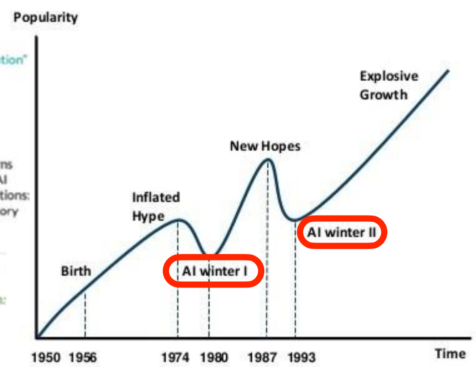


Dartmouth Summer Research Project on Artificial Intelligence
a group to clarify and develop ideas about thinking machines



AI HAS A LONG HISTORY OF BEING "THE NEXT BIG THING"...





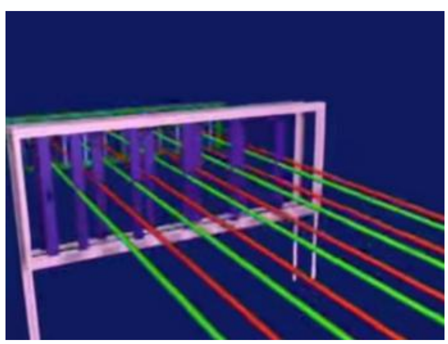
Timeline of AI Development

- **1950s-1960s**: First AI boom - the age of reasoning, prototype AI developed
- **1970s**: AI winter I
- **1980s-1990s**: Second AI boom: the age of Knowledge representation (appearance of expert systems capable of reproducing human decision-making)
- **1990s**: AI winter II
- **1997**: Deep Blue beats Gary Kasparov
- **2006**: University of Toronto develops Deep Learning
- **2011**: IBM's Watson won Jeopardy
- **2016**: Go software based on Deep Learning beats world's champions

https://anythingawesome.com

Augusta Ada King

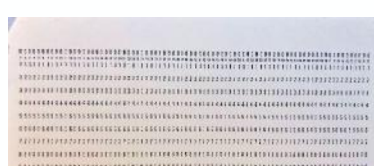
(aka Lady Ada, Countess of Lovelace)



Jacquard machine 1804

≈

https://people.well.com/user/adatoole/bio.htm



https://anythingawesome.com

Credit for the Hype Cycle should go to Lady Ada!

VISIBILITY

Publisher of the first computing program

Scientific Memoirs, Selections from The Transactions of Foreign Academies and Learned Societies and from Foreign Journals, edited by Richard Taylor, F.S.A.,Vol III London: 1843

Peak of Inflated Expectations: Early publicity produces a number of success stories—often accompanied by scores of failures. Some companies take action; many do not.

In considering any new subject,

There is frequently a tendency first to overrate what we find to be already interesting or remarkable, and

Secondly - by a sort of natural reaction - to undervalue the true state of the case.

Plateau of Productivity: Mainstream adoption starts to take off. Criteria for assessing provider viability are more clearly defined. The technology's broad market applicability and relevance are clearly paying off.

Slope of Enlightenment: More instances of how the technology can benefit the enterprise start to crystallize and become more widely understood. Second- and third-generation products appear from technology providers. More enterprises fund pilots; conservative companies remain cautious.

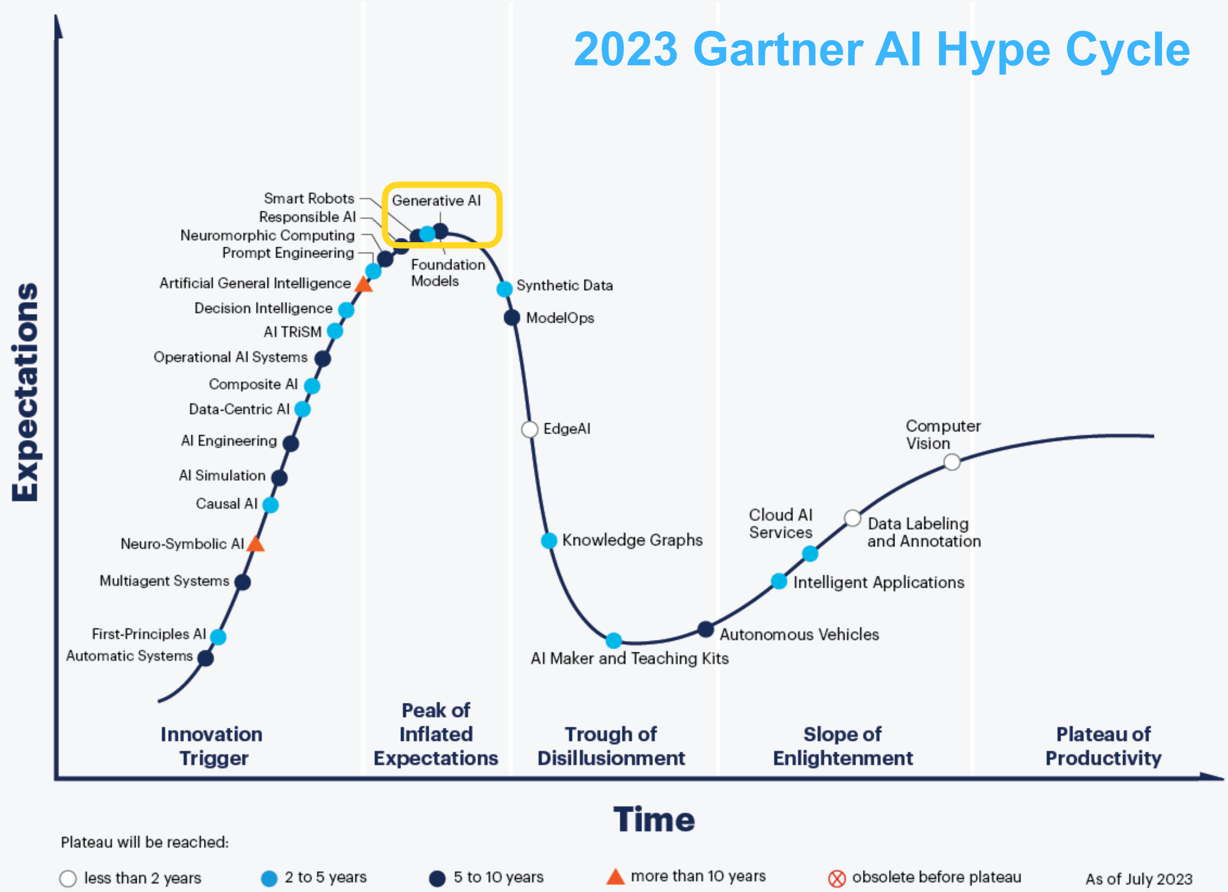
Trough of Disillusionment: Interest wanes as experiments and implementations fail to deliver. Producers of the technology shake out or fail. Investments continue only if the surviving providers improve their products to the satisfaction of early adopters.

Technology Trigger: A potential technology breakthrough kicks things off. Early proof-of-concept stories and media interest trigger significant publicity. Often no usable products exist and commercial viability is unproven.

TIME

http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp

2023 Gartner AI Hype Cycle



2024 Gartner AI Hype Cycle

You're the best thing that I've ever found

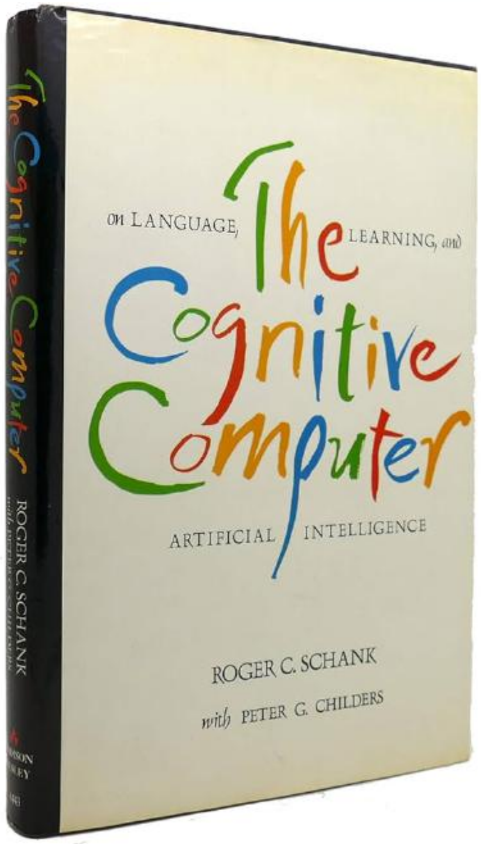
What causes the recurring AI winters?

- Overpromising and underdelivering
  - AI researchers and entrepreneurs have often made unrealistic promises about what AI could achieve. This has led to disappointment and disillusionment, and has made investors less willing to fund AI research and development.
- Technological limitations
  - AI algorithms have historically been complex and computationally expensive to train. This has limited the ability of AI researchers to develop and deploy AI systems on a large scale.
- Lack of data
  - AI systems need large amounts of data to train on. However, in the past, there has often been a lack of high-quality data available to AI researchers.
- Ethical concerns
  - There are a number of ethical concerns surrounding the development and use of AI. For example, there is concern that AI could be used to develop autonomous weapons or to discriminate against people.

https://bard.google.com/

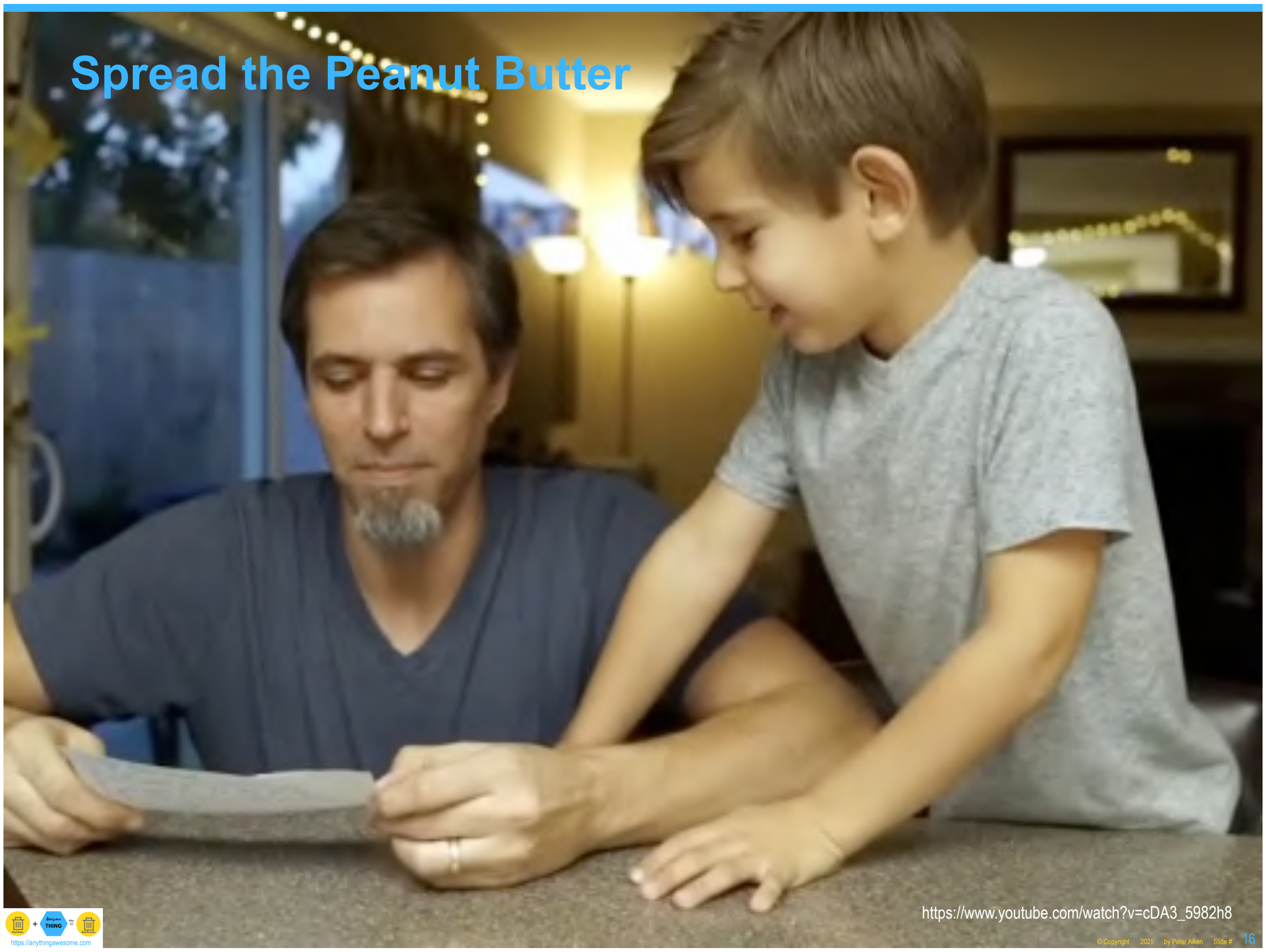
https://anythingawesome.com

Overpromising and Underdelivering (1984)



- Focused on top down
  - What does it mean to order a meal at a restaurant?
  - Get details correct
  - Unexpected conditions
- Top down knowledge representation
  - Flowchart the process for making a peanut butter and jelly sandwich

Spread the Peanut Butter



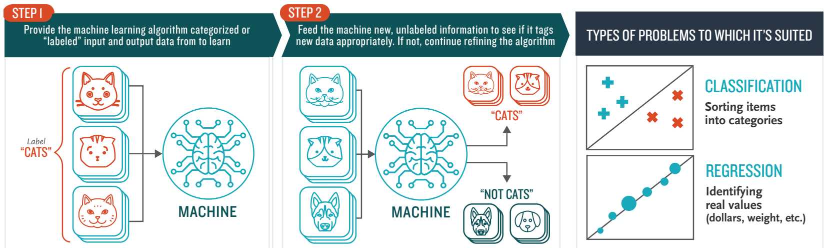
https://www.youtube.com/watch?v=cDA3_5982h8

https://www.boozallen.com/s/insight/blog/how-do-machines-learn.html

Overpromising and Underdelivering (Today)

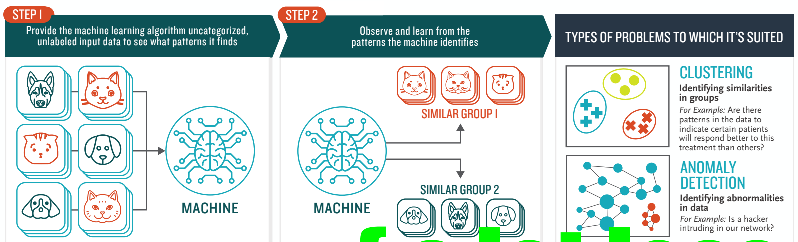
(more $)

Supervised Machine Learning



- Focused on bottom up learning
  - The knee bone is connected to the thigh bone
  - When does walking become possible?

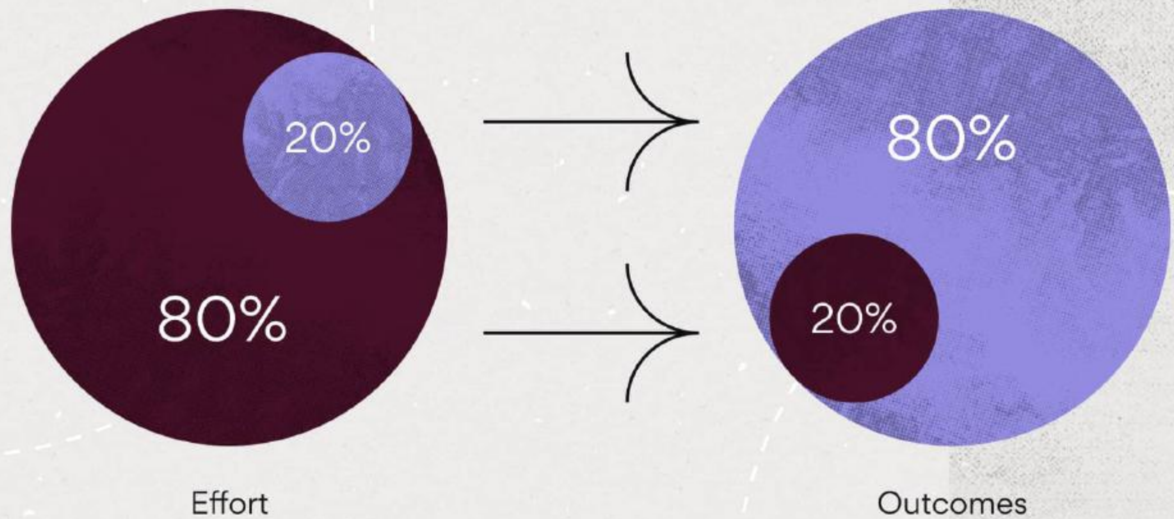
Unsupervised Machine Learning



- Bottom up knowledge representation
  - Teach a neural net how to care about making a peanut butter and jelly sandwich

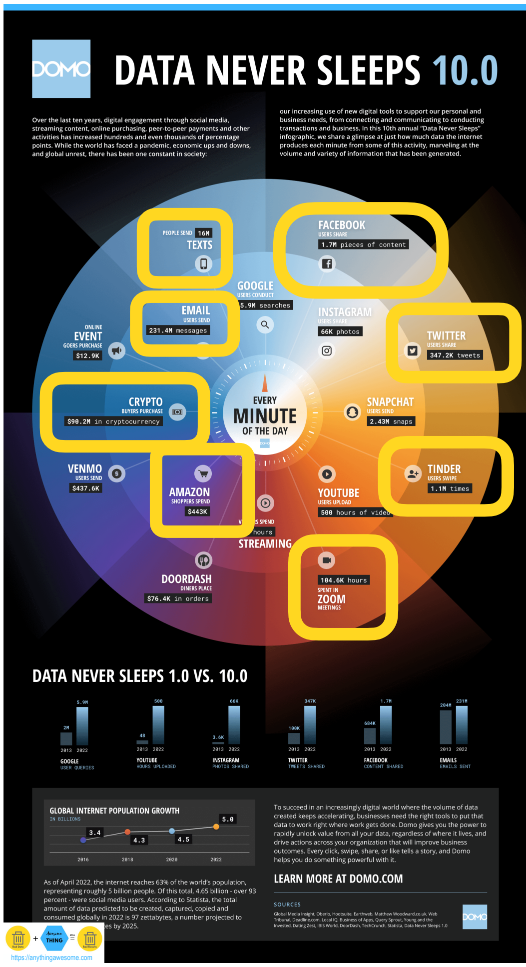
(a lot less $)

Lack of Data (Pareto Realities)



1. 80% of your data is redundant, trivial, or obsolete
2. 80% of your data is of unknown quality
3. 80% of your data is 'standards free'
4. Your highly paid data analytics capabilities spend 80% of their time working under these conditions

https://anythingawesome.com



How much Data (by the minute?)

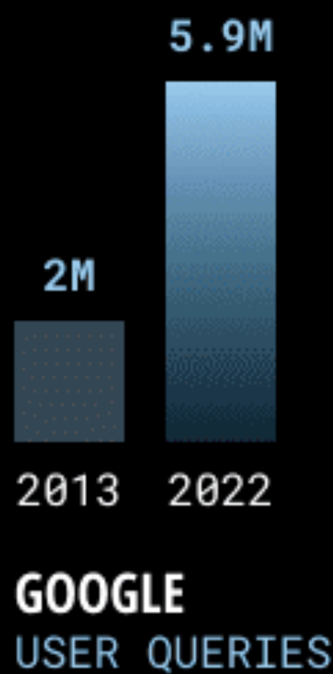
For the entirety of 2022, every minute of every day:

- Facebook users shared 1.7M pieces of content
- Twitter users shared 347,000 tweets
- Tinder users swiped 1.1M times
- 104,000 hours were spent in zoom meetings
- Amazon shoppers spend $443,000
- Crypto buyers purchase $90 million in cryptocurrency
- 231 million emails
- 16 million texts

https://www.domo.com/learn/data-never-sleeps-10

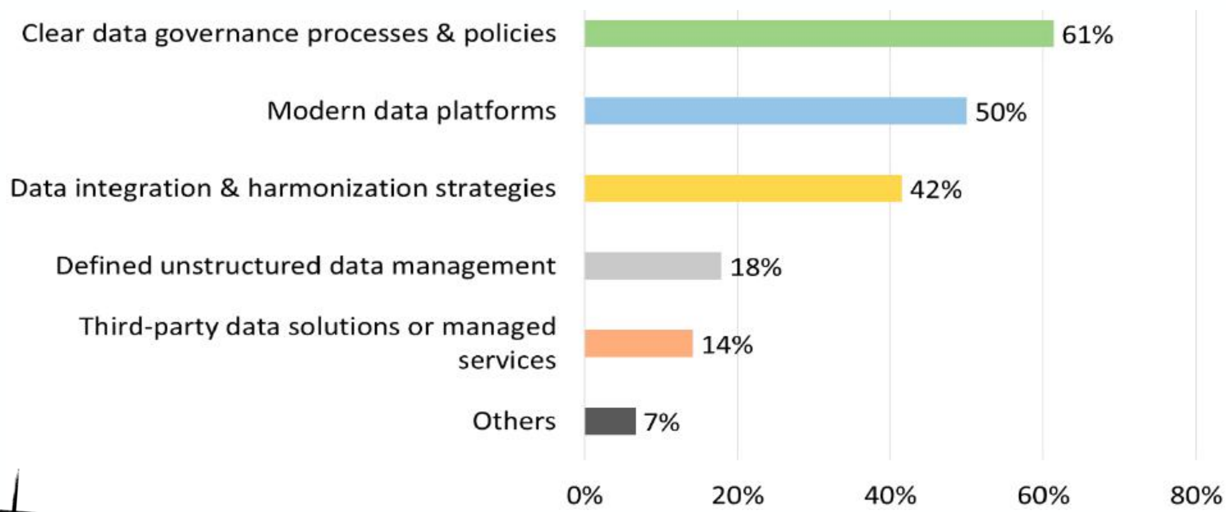
https://anythingawesome.com

Data Never Sleeps 2013 compared to 2022



Source: CDO Magazine/EY N=100

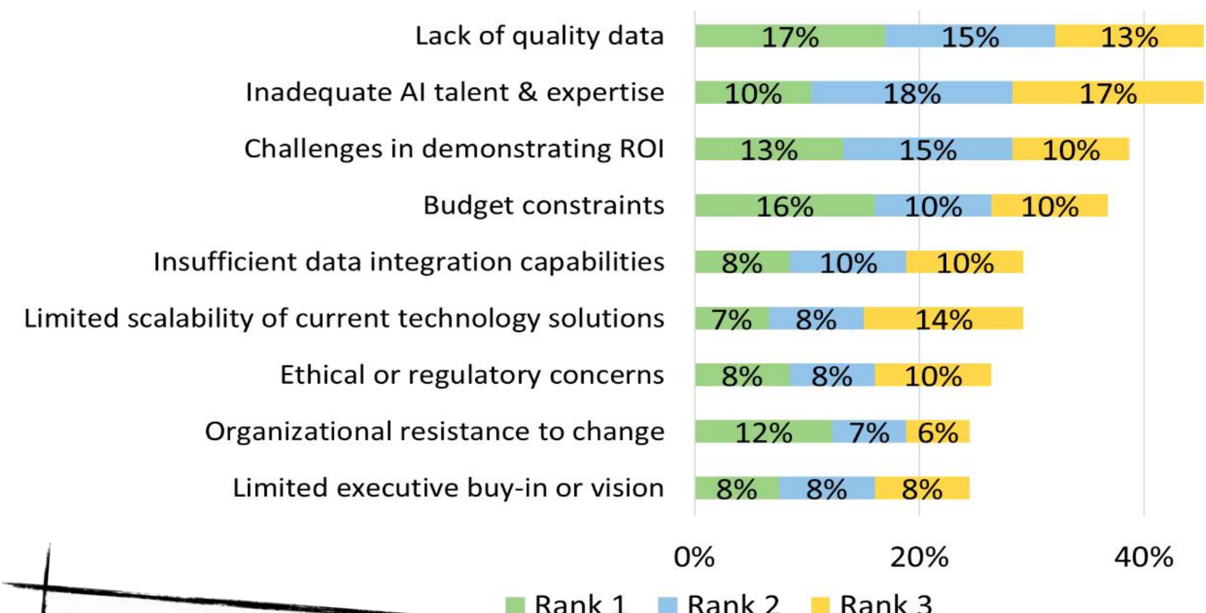
Organizations are taking a holistic approach to data readiness for AI



Data leaders are focusing on defining data governance and policies to create a framework for the usage, integration, and management of data in preparation for AI.

Source: CDO Magazine/EY N=100

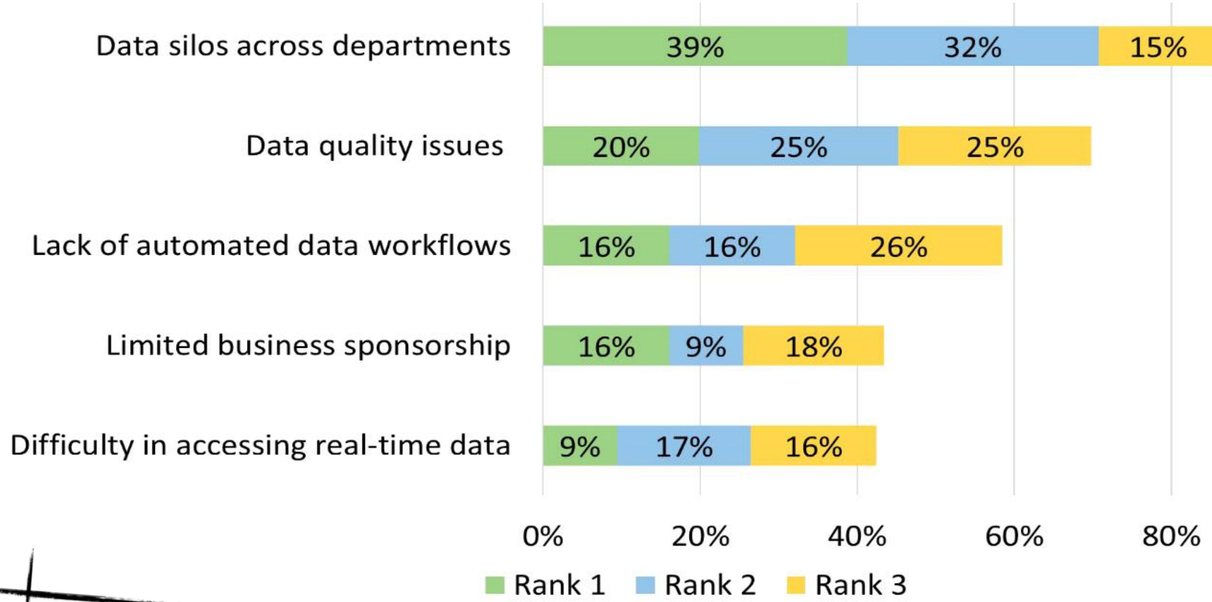
Data quality issues and inadequate talent limit AI's potential



Data leaders are facing significant challenges in unlocking AI's full potential due to obstacles such as data quality issues, shortage of skilled talent, and the limited scalability of existing technology.

Source: CDO Magazine/EY N=100

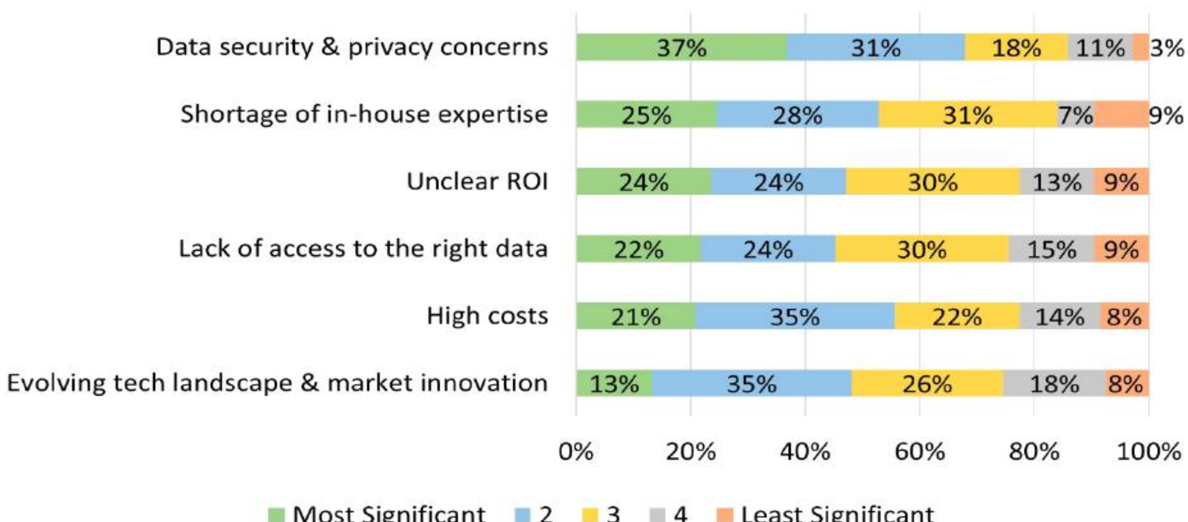
Data silos hinder integration for AI



Data silos across departments and the lack of automated data workflows obstruct data integration for AI initiatives.

Source: CDO Magazine/EY N=100

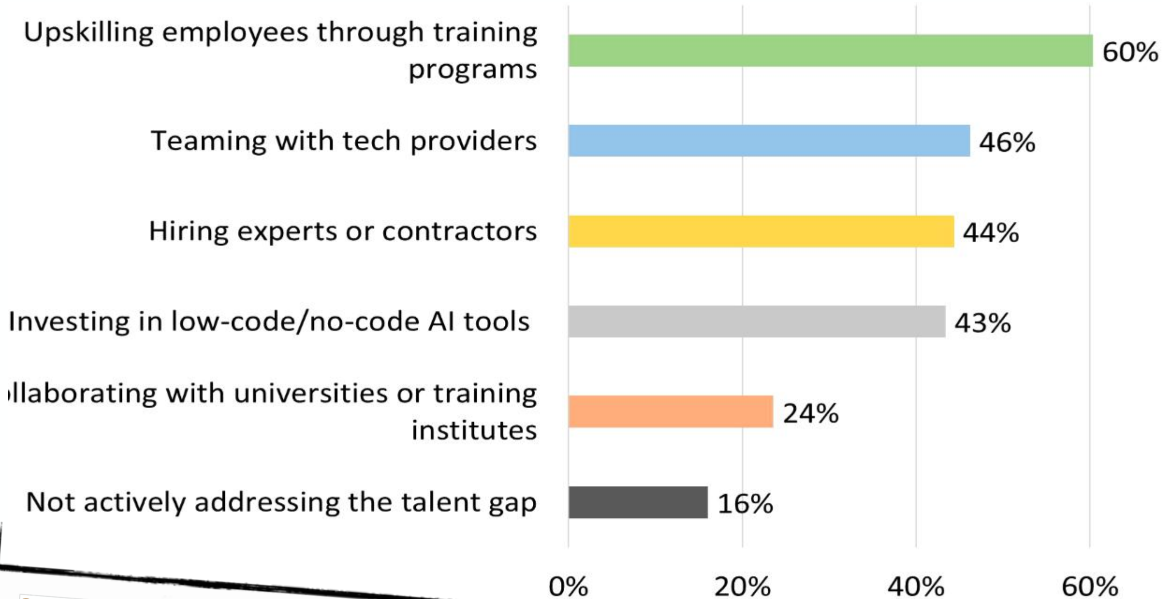
Data security & privacy remains the top obstacle to GenAI integration



Integrating GenAI into organizations' existing business processes presents several significant challenges, including data security and privacy concerns, along with a shortage of in-house expertise.

Source: CDO Magazine/EY N=100

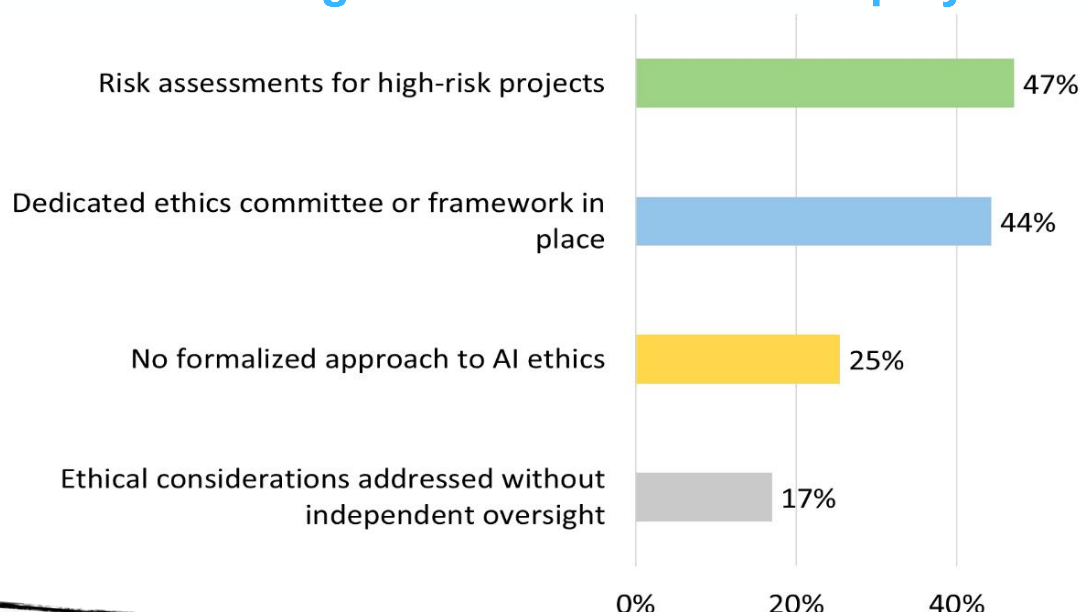
Several measures are bridging the data and AI talent gap



Organizations acknowledge the talent gap and are upskilling their teams in AI skills to address it fundamentally. They are also leveraging interim strategies such as engaging experts.

Source: CDO Magazine/EY N=100

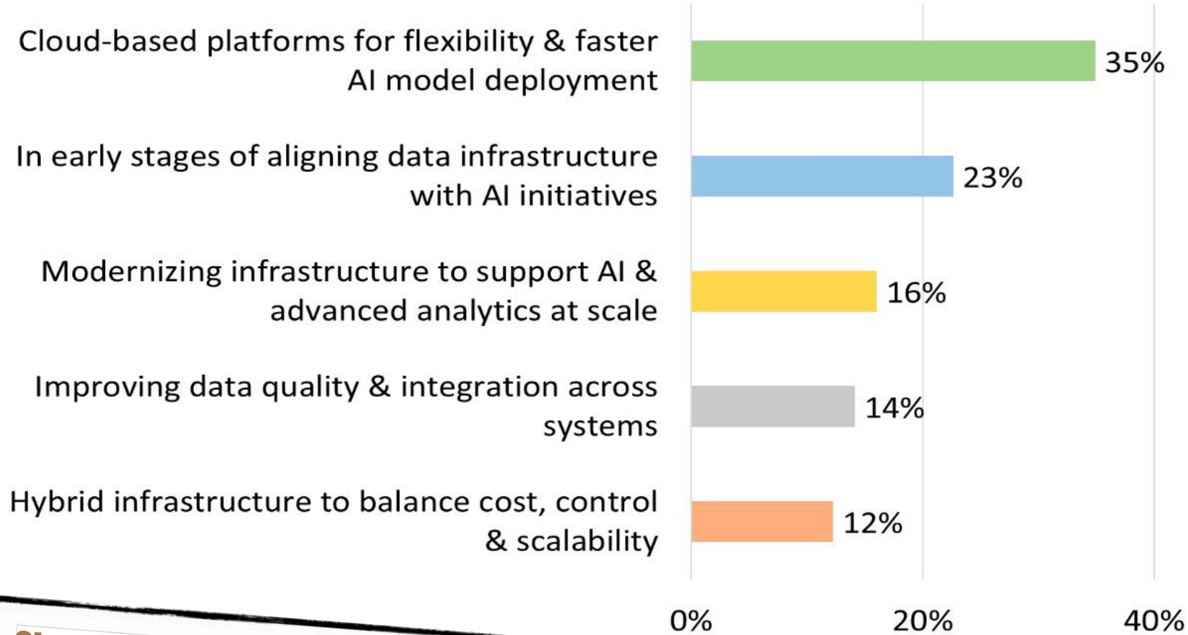
Ethical risk mitigation is central to AI deployment



Most organizations have developed formal plans to address the ethical risks of AI deployment, yet many still lack a structured approach.

Source: CDO Magazine/EY N=100

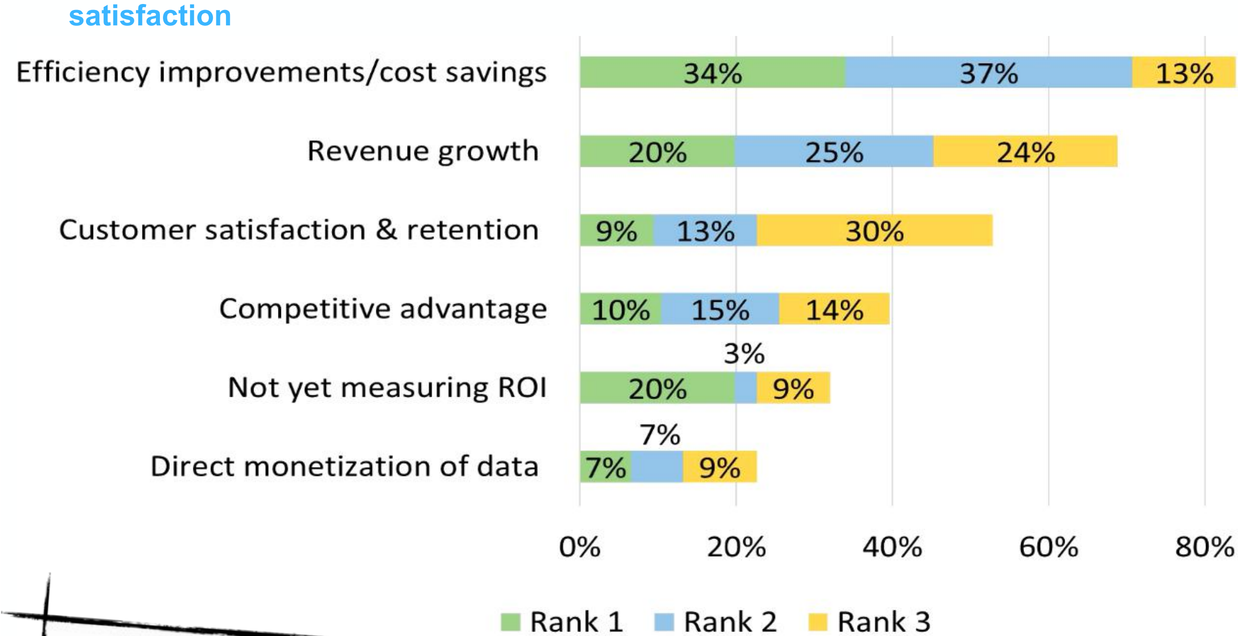
Cloud-based platforms lead the way in AI-ready data infrastructure



Cloud-based platforms emerge as the main preference for flexibility and faster AI model deployment.

Source: CDO Magazine/EY N=100

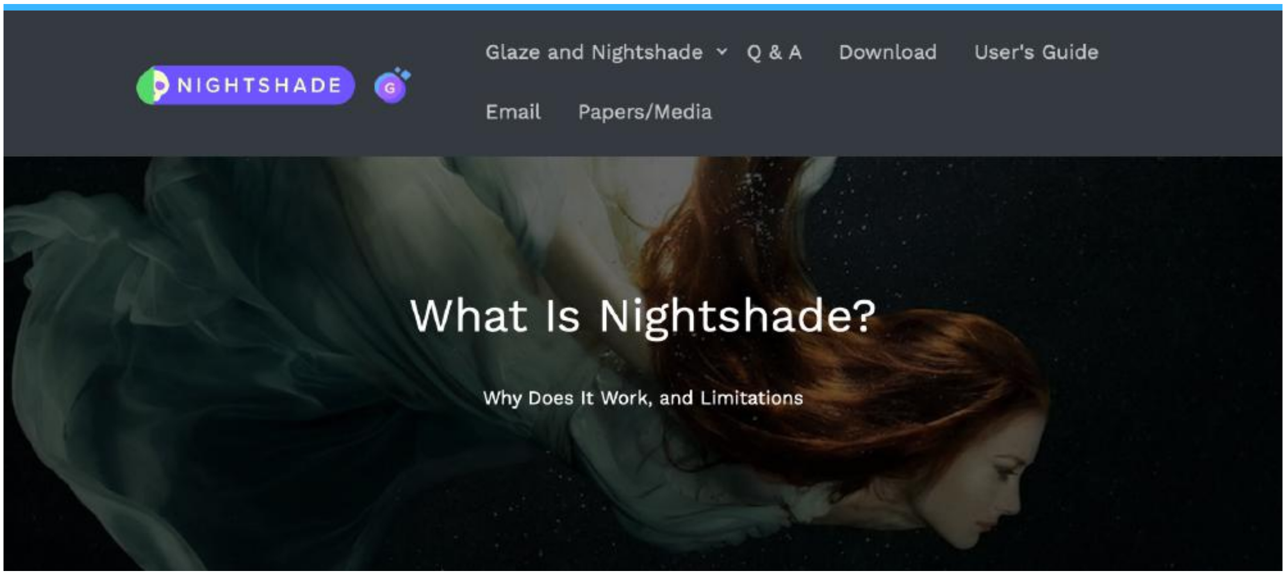
Data investment ROI measured through efficiency/savings, growth, and customer satisfaction



Organizations are assessing their ROI on data investments through both quantifiable and qualitative measures.

Source: CDO Magazine/EY N=100

All-powerful, ever-pervasive AI is running



Since their arrival, generative AI models and their trainers have demonstrated their ability to download any online content for model training. For content owners and creators, few tools can prevent their content from being fed into a generative AI model against their will. Opt-out lists have been disregarded by model trainers in the past, and can be easily ignored with zero consequences. They are unverifiable and unenforceable, and those who violate opt-out lists and do-not-scrape directives can not be identified with high confidence.

In an effort to address this power asymmetry, we have designed and implemented Nightshade, a tool that turns any image into a data sample that is unsuitable for model training. More precisely, Nightshade transforms images into "poison" samples, so that models training on them without consent will see their models learn unpredictable behaviors that deviate from expected norms, e.g. a prompt that asks for an image of a cow flying in space might instead get an image of a handbag floating in space.

https://nightshade.cs.uchicago.edu/whatis.html

InfoWorld

Learn about the challenges and opportunities presented by generative artificial intelligence.

Gen AI Insights

December 3, 2024

CIOs view cost management as possible AI value killer

Most CIOs see a lack of money as a major concern as they face pressure to extract value from AI. And underestimating the costs can be catastrophic.

By Grant Gross

CIOs feeling the pressure to deploy successful AI projects have a second concern: that they don't have the money to pull it off.

Ninety percent of CIOs recently surveyed by Gartner say that managing AI costs is limiting their ability to get value from AI. In addition, if CIOs don't fully understand the cost of scaling generative AI, they could miscalculate by 500% to 1,000%, says Hung LeHong, an analyst focused on executive leadership for digital business at Gartner. Depending on the AI project, a mistake of that magnitude could cost millions of dollars.

In many cases, using an LLM for simple AI tasks, such as transcribing and translating, can be expensive when cheaper tools are available, LeHong said during a recent webcast. “Some of the CIOs just don't understand all of the cost elements that are there,” he adds. “Even if they do understand the cost, they don't have the leverage to change it.”

Hidden costs and price hikes

https://builtin.com/articles/clarify-cloud-spending

Cloud Spending Infuriates Executives. Here's How to Solve It.

While C-suite disagreements are to be expected, squabbles over the way a company spends on cloud services don't have to be.

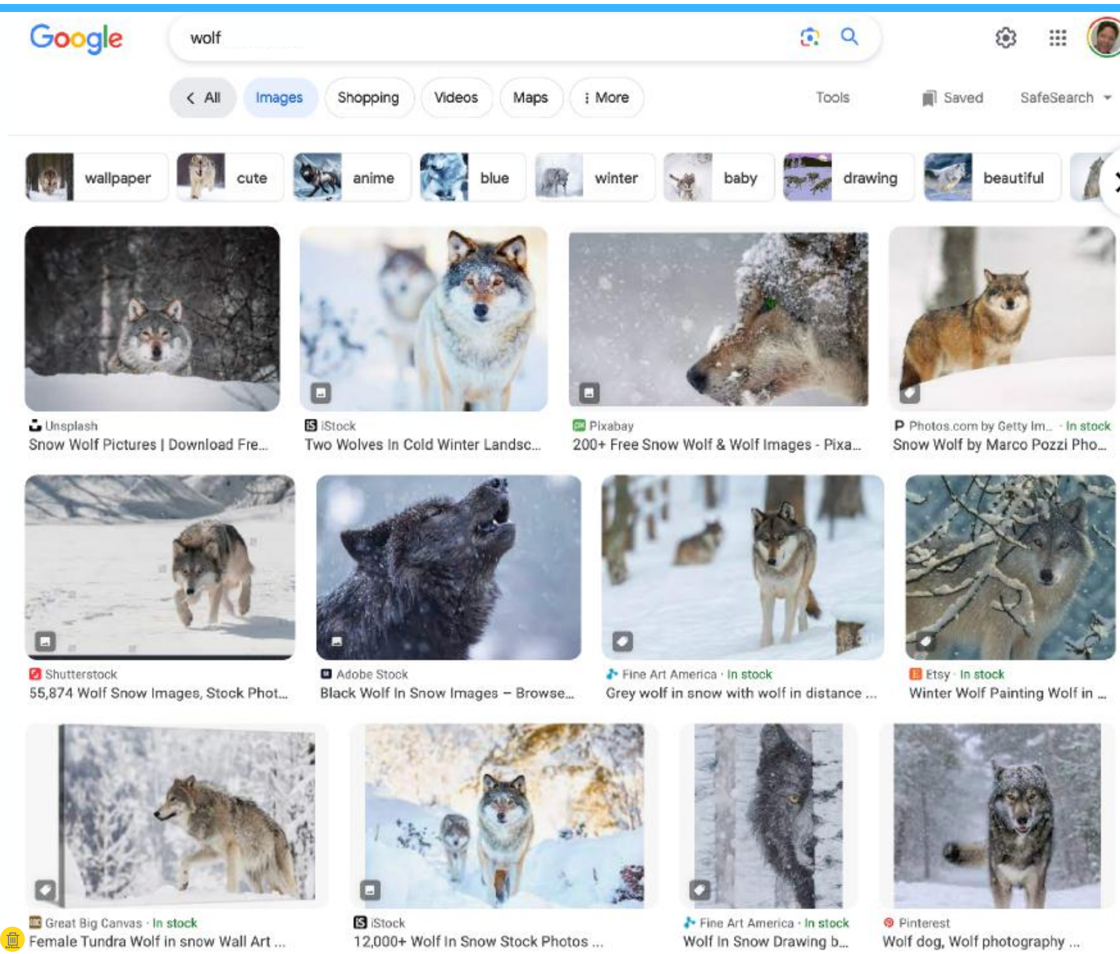
Written by Willy Sennott

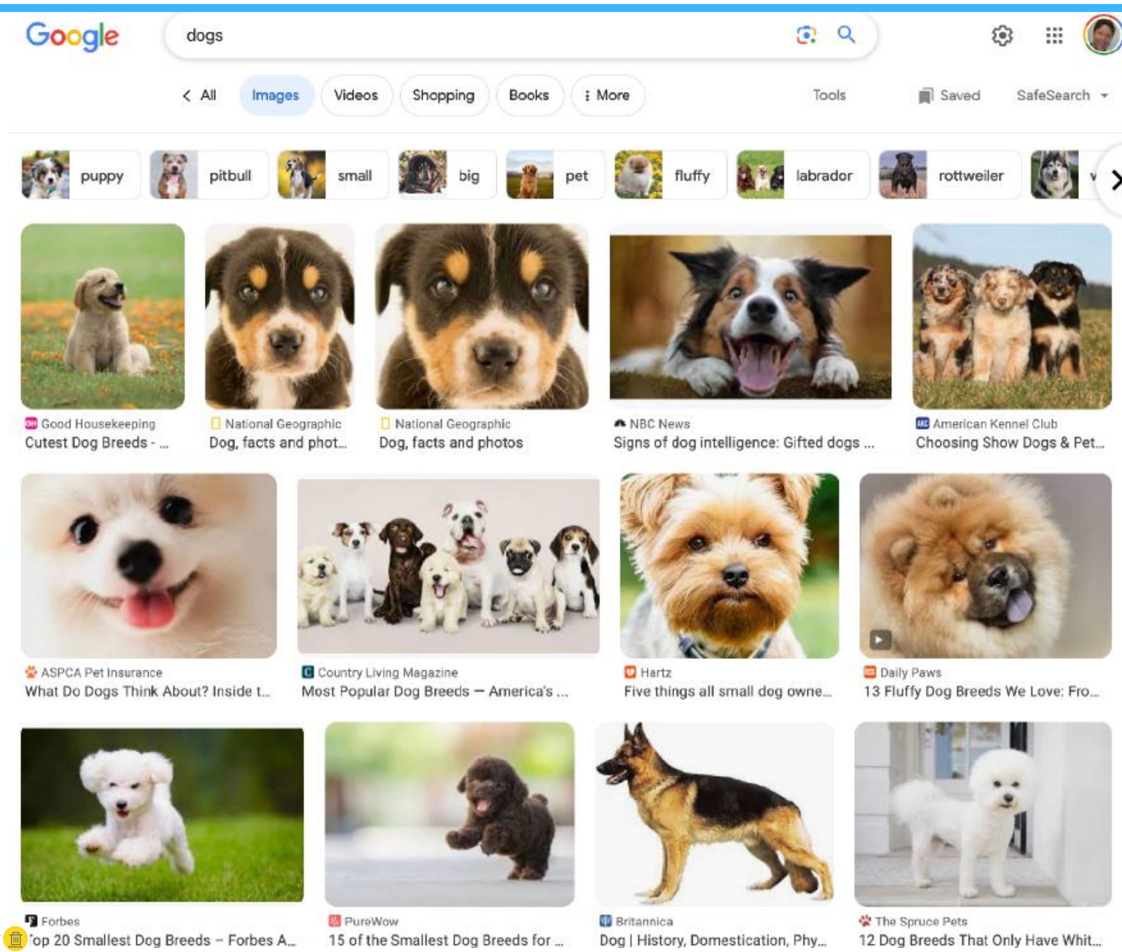
Published on Aug. 24, 2023

Learning Algorithms

- Current advancements in computer science and AI are focused on 'learning algorithms'
- Improve performance by 'learning' from the data that they process
- The primary variable holding back this type of progress is a lack of training data
- Tom Redman: "Poor data quality is enemy number one to the widespread, profitable use of machine learning.

If Your Data Is Bad, Your Machine Learning Tools Are Useless https://hbr.org/2018/04/if-your-data-is-bad-your-machine-learning-tools-are-useless

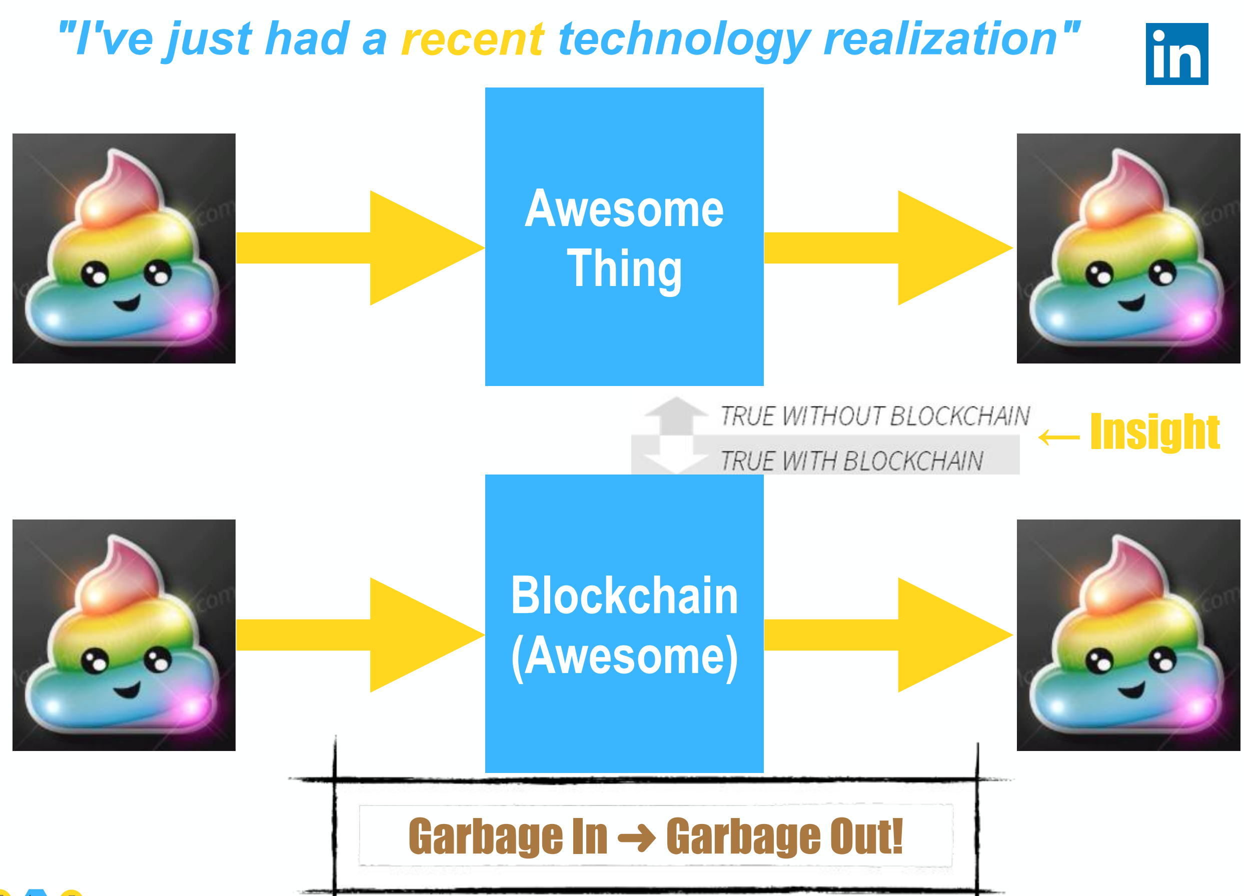




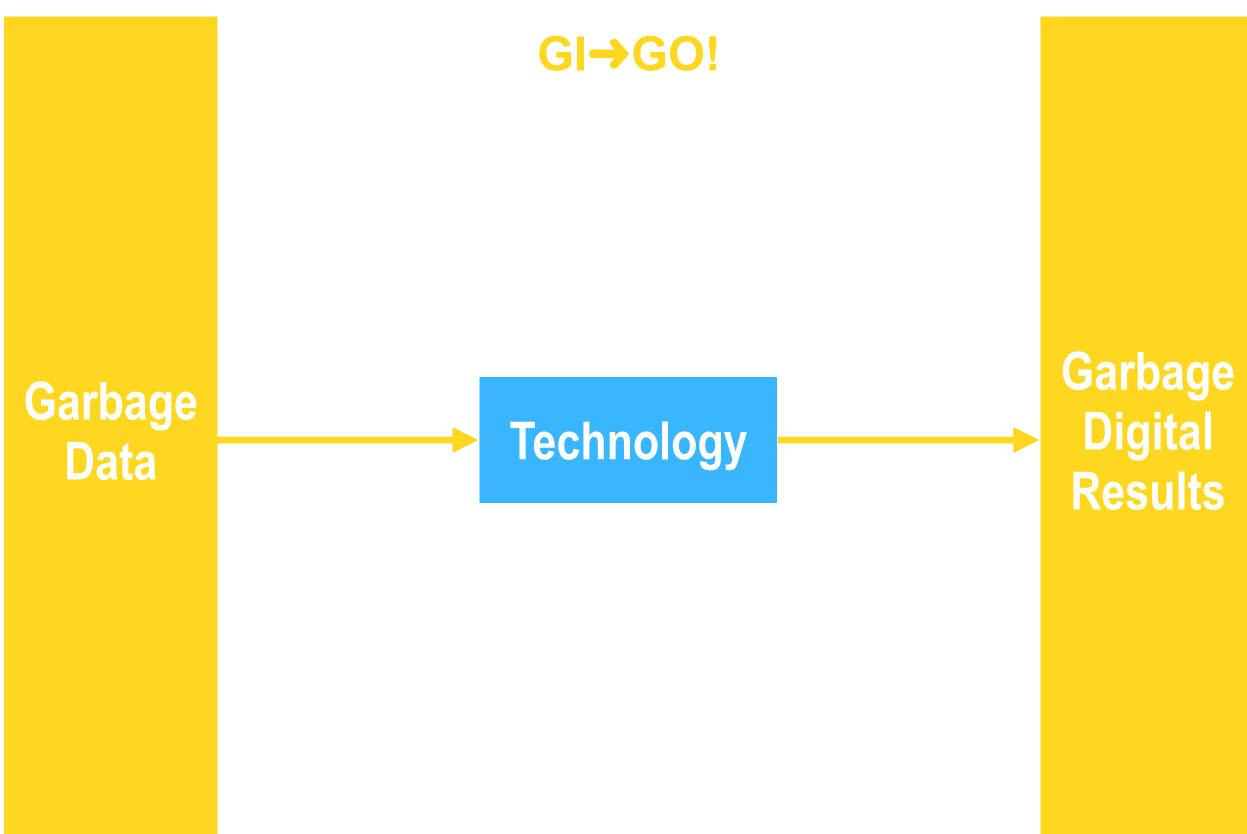
Technological Limitations

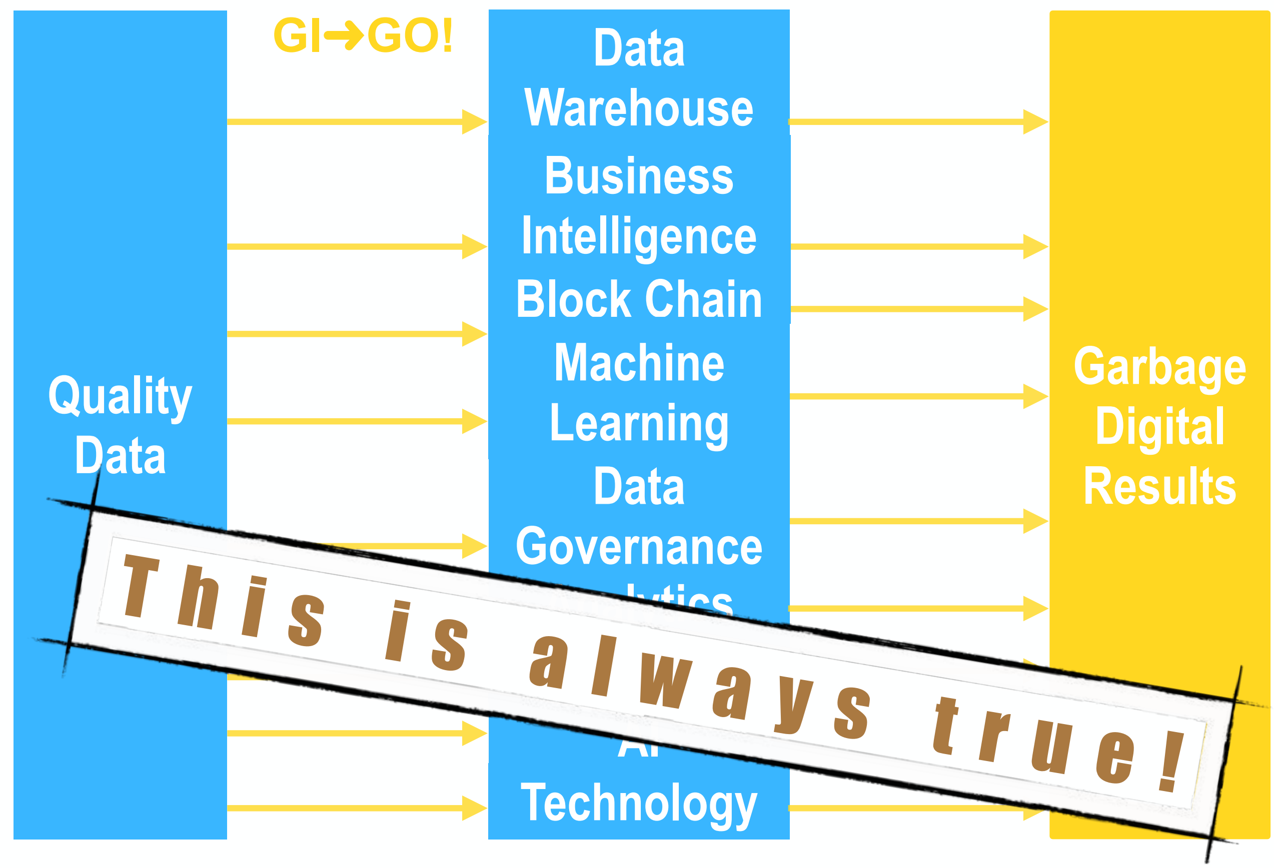
[Bad Data] + Anything Awesome [will always yield] Bad Results

Garbage In⟶Garbage Out

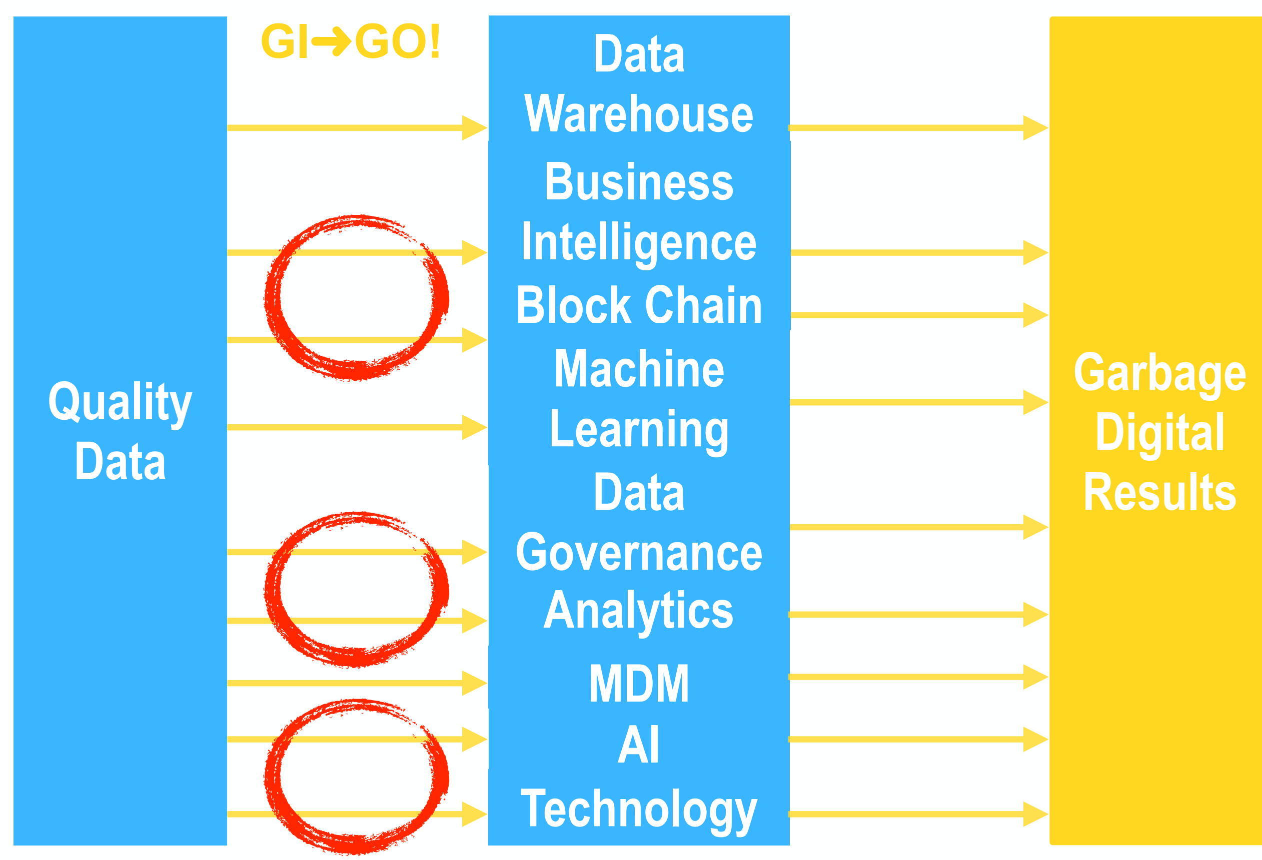


https://www.flashingblinkylights.com/unicorn-poop-emoji-sparkling-led-pins.html

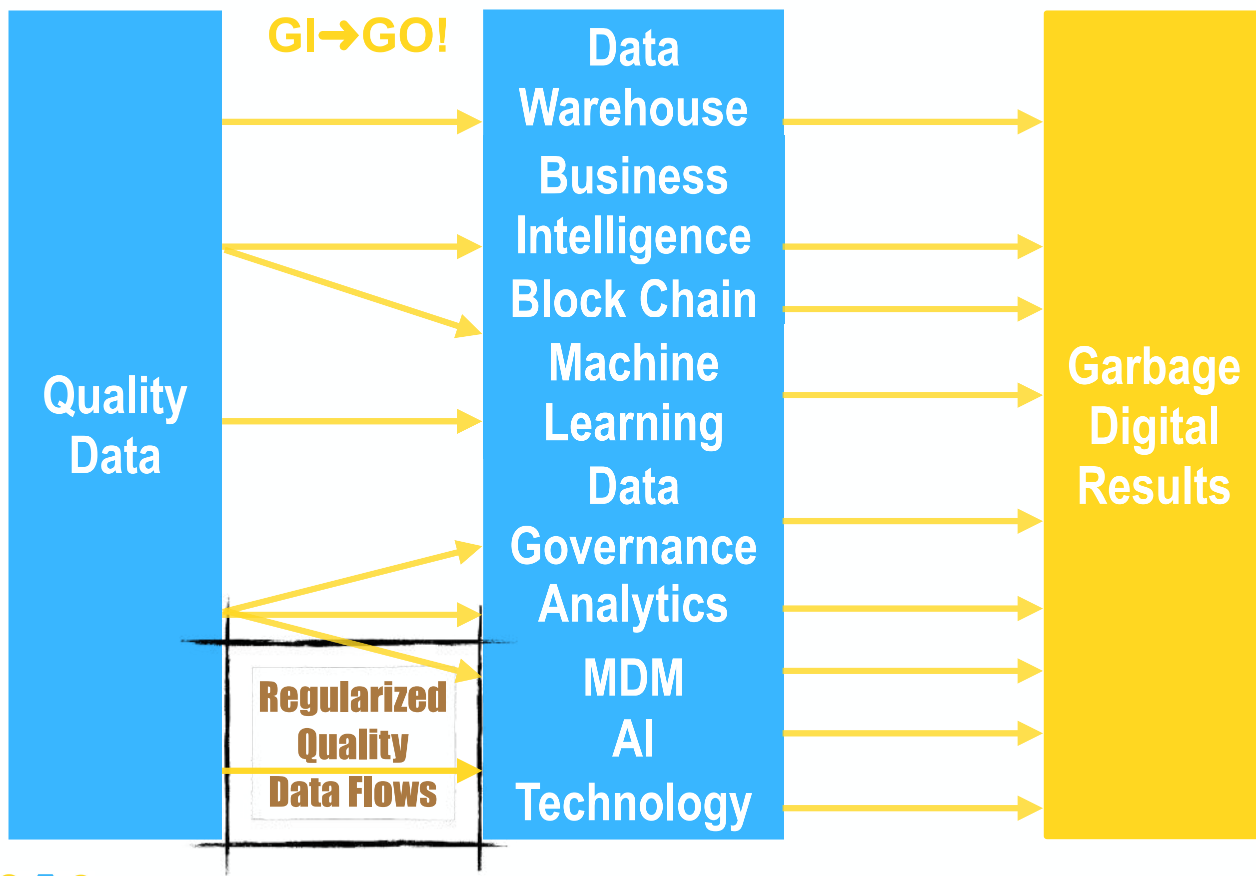


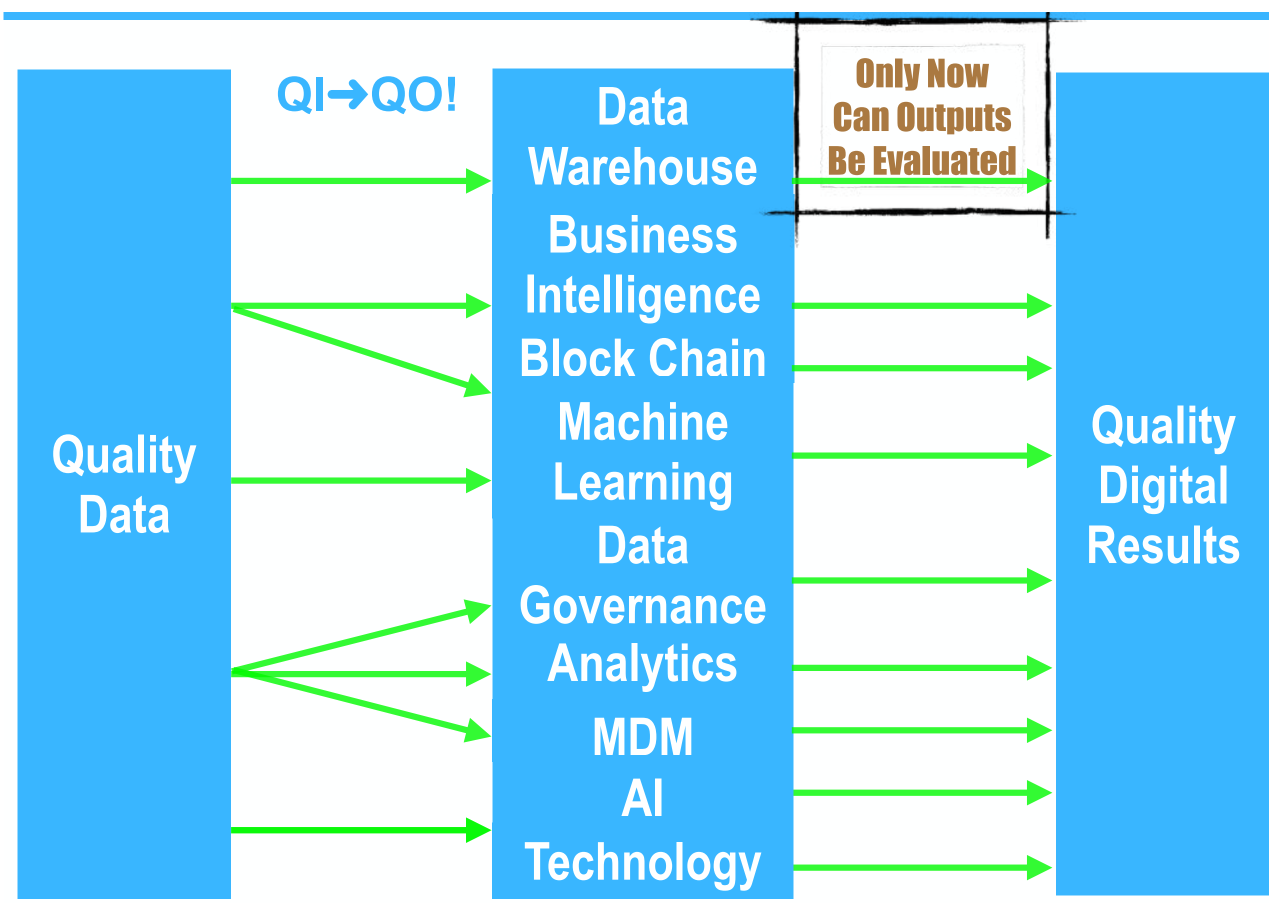


https://anythingawesome.com

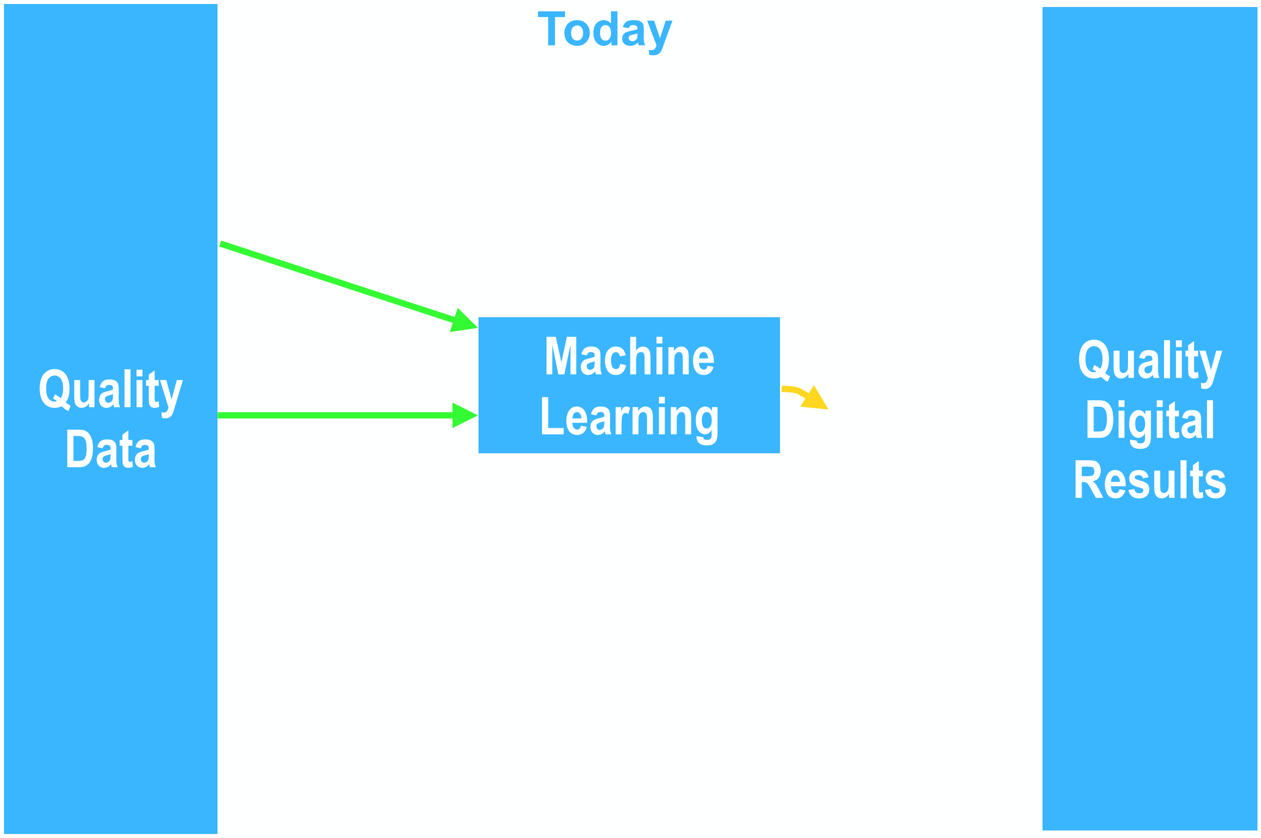


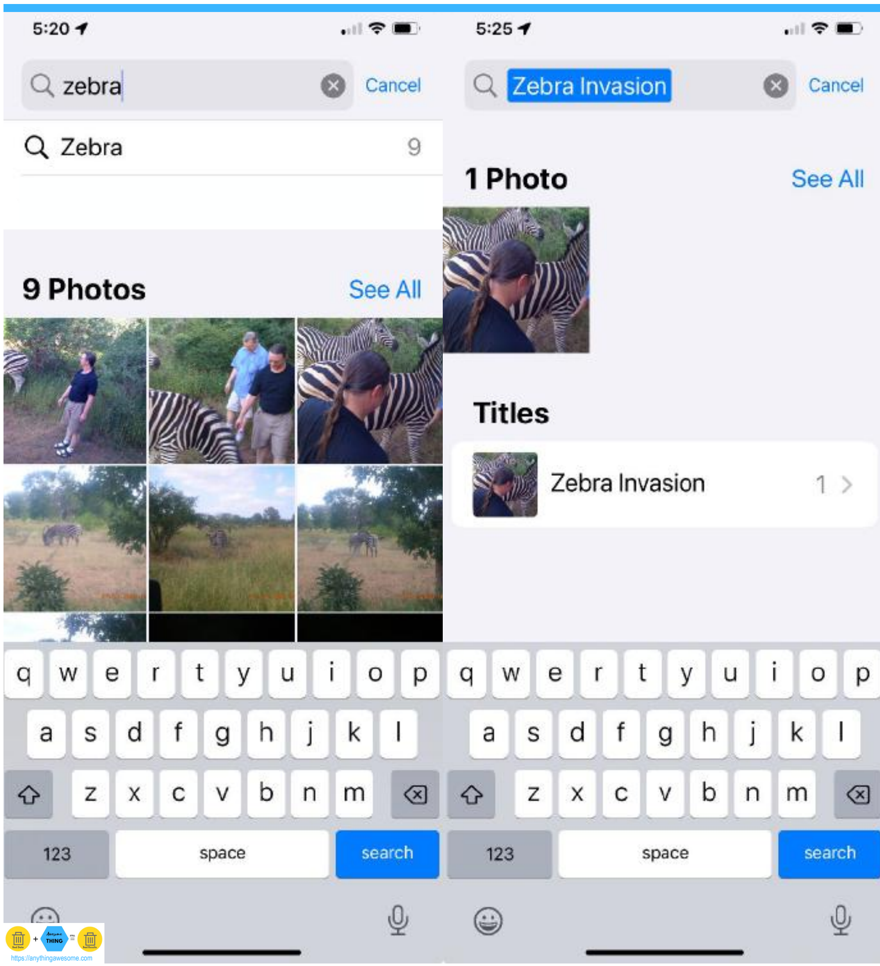
https://anythingawesome.com

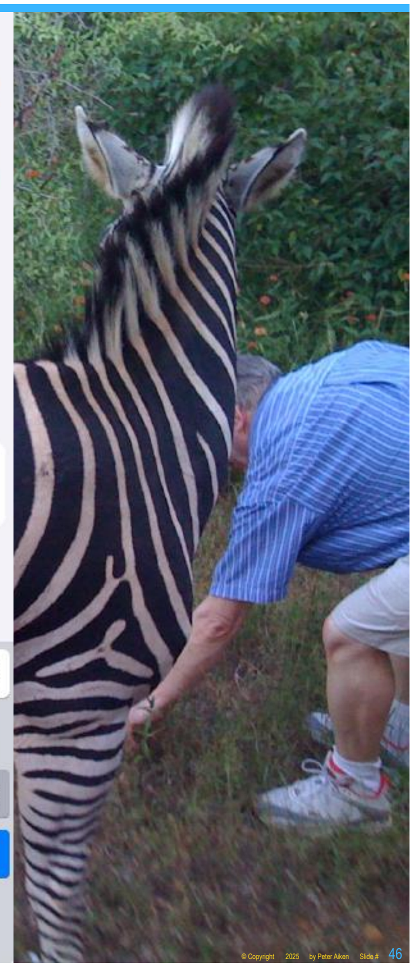

GI➔GO!
Quality Data
Data Warehouse
Business Intelligence
Block Chain
Machine Learning
Data Governance
Analytics
MDM
AI
Technology
Garbage Digital Results
Regularized Quality Data Flows

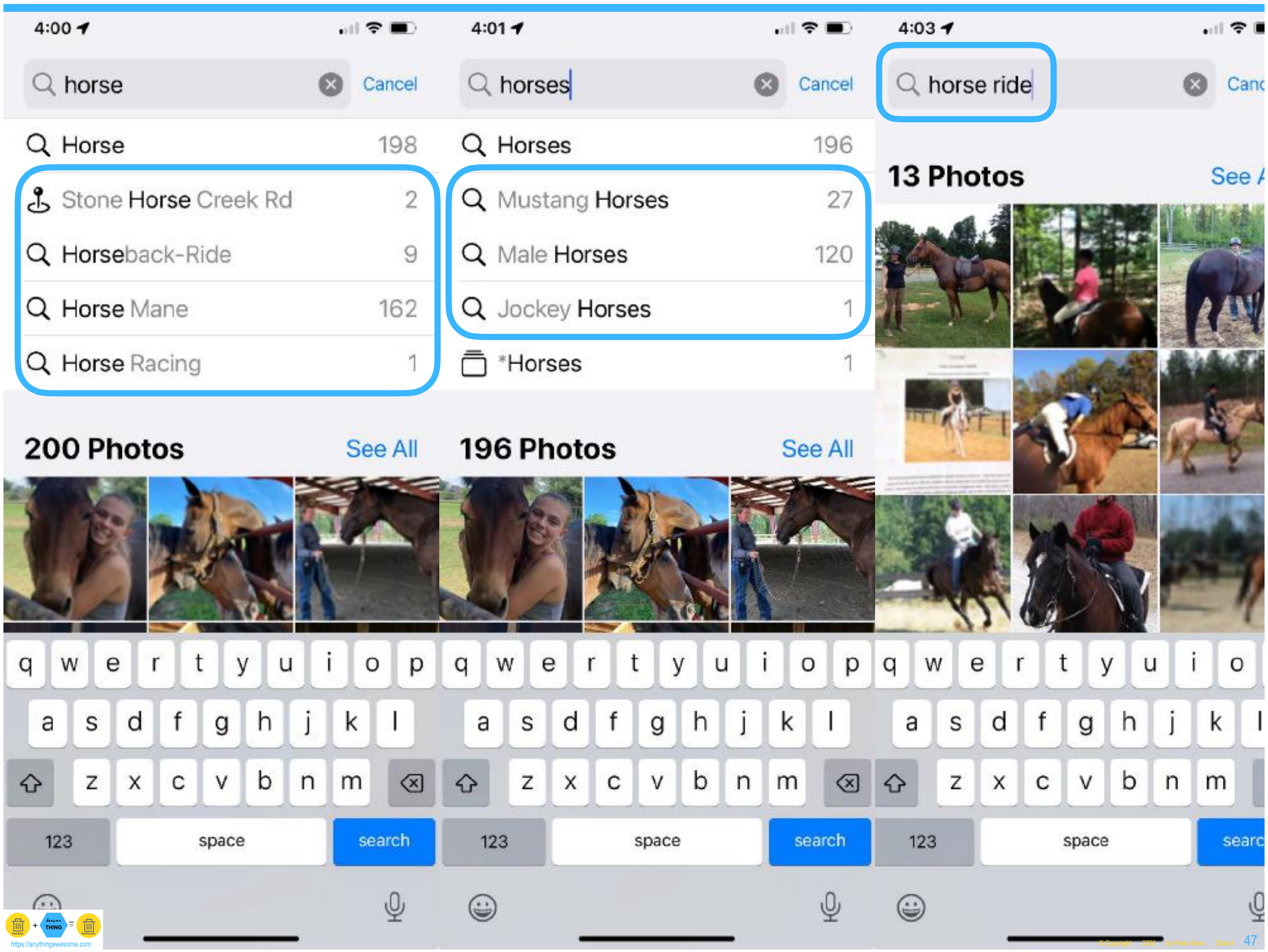

QI➔QO!
Quality Data
Data Warehouse
Business Intelligence
Block Chain
Machine Learning
Data Governance
Analytics
MDM
AI
Technology
Only Now Can Outputs Be Evaluated
Quality Digital Results

Today

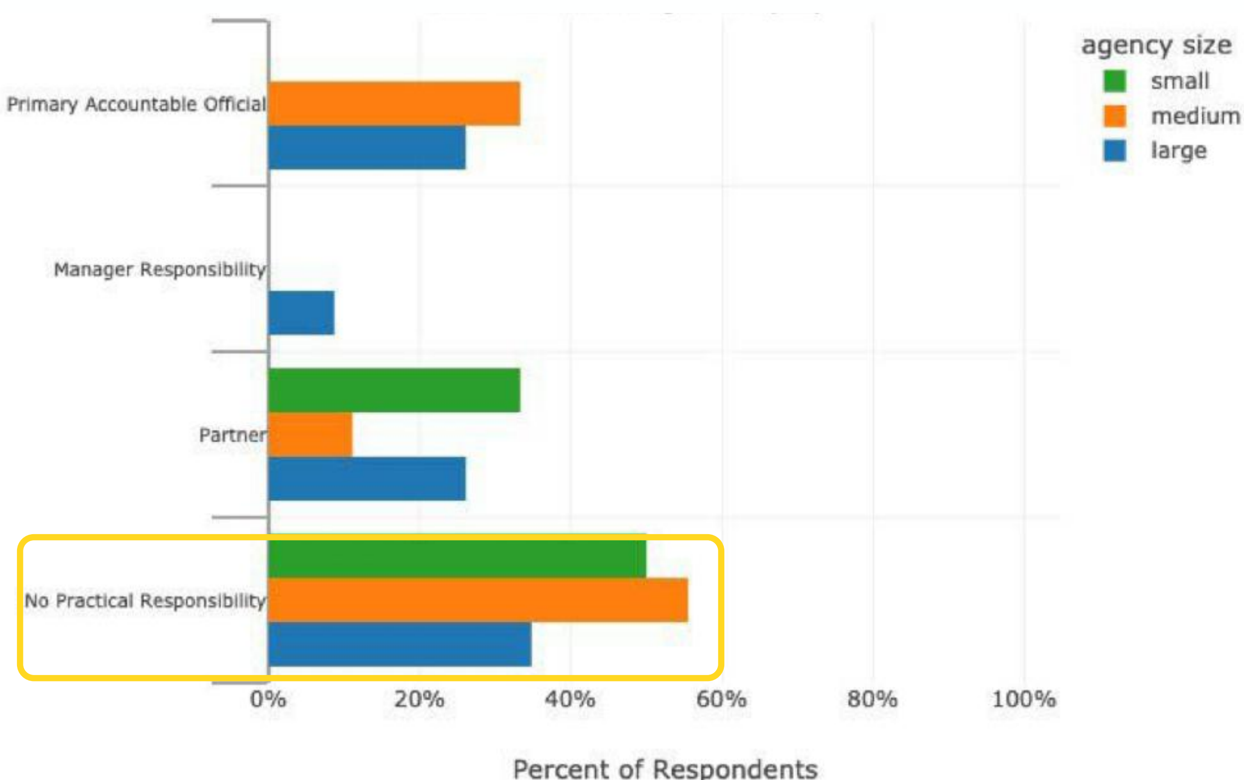

Quality Data
Machine Learning
Quality Digital Results


zebra
Cancel
Zebra 9
9 Photos
See All
Zebra Invasion
Cancel
1 Photo
See All
Titles
Zebra Invasion 1





AI Responsibility

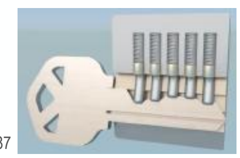


[US OPM]

Diagnosing Organizational Readiness

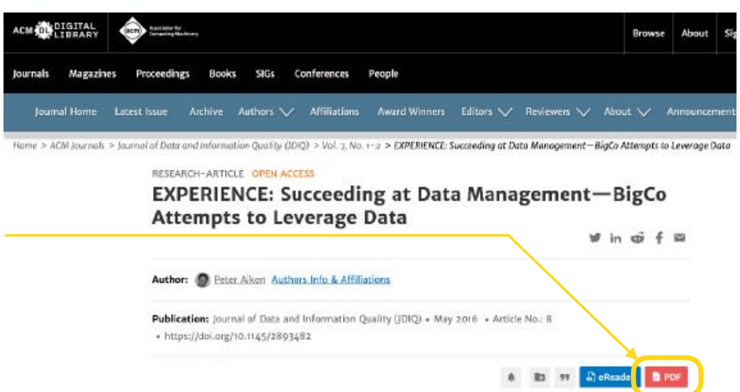
	+	Skills	+	Incentive	+	Resources	+	Action Plan	=	Confusion
Vision	+		+	Incentive	+	Resources	+	Action Plan	=	Anxiety
Vision	+	Skills	+		+	Resources	+	Action Plan	=	Gradual Change
Vision	+	Skills	+	Incentive	+		+	Action Plan	=	Frustration
Vision	+	Skills	+	Incentive	+	Resources	+		=	False Starts
Vision	+	Skills	+	Incentive	+	Resources	+	Action Plan	=	Change

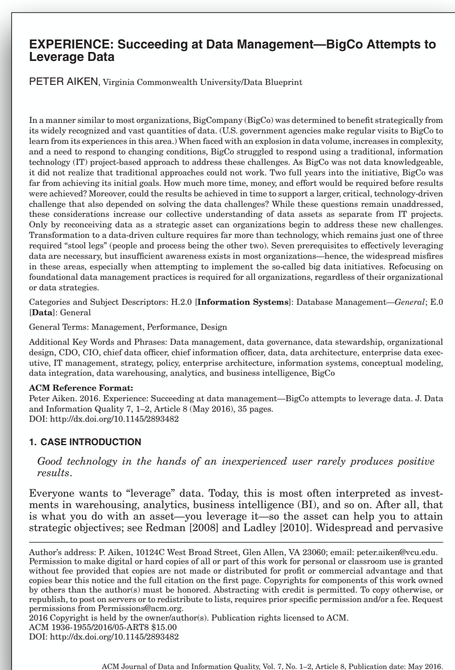
Culture is the biggest impediment to a shift in organizational thinking about AI!



adapted from the *Managing Complex Change* model by Lippitt, 1987

No cost, no registration case study download





EXPERIENCE: Succeeding at Data Management—BigCo Attempts to Leverage Data

PETER AIKEN, Virginia Commonwealth University/Data Blueprint

In a manner similar to most organizations, BigCompany (BigCo) was determined to benefit strategically from its widely recognized and vast quantities of data. (U.S. government agencies make regular visits to BigCo to learn from its experiences in this area.) When faced with an explosion in data volume, increases in complexity, and a need to respond to changing conditions, BigCo struggled to respond using a traditional, information technology (IT) project-based approach to address these challenges. As BigCo was not data knowledgeable, it did not realize that traditional approaches could not work. Two full years into the initiative, BigCo was far from achieving its initial goals. How much more time, money, and effort would be required before results were achieved? Moreover, could the results be achieved in time to support a larger, critical, technology-driven challenge that also depended on solving the data challenges? While these questions remain unaddressed, these considerations increase our collective understanding of data assets as separate from IT projects. Only by reconceiving data as a strategic asset can organizations begin to address these new challenges. Transformation to a data-driven culture requires far more than technology, which remains just one of three required "stool legs" (people and process being the other two). Seven prerequisites to effectively leveraging data are necessary, but insufficient awareness exists in most organizations—hence, the widespread misfires in these areas, especially when attempting to implement the so-called big data initiatives. Refocusing on foundational data management practices is required for all organizations, regardless of their organizational or data strategies.

Categories and Subject Descriptors: H.2.0 [**Information Systems**]: Database Management—*General*; E.0 [**Data**]: General

General Terms: Management, Performance, Design

Additional Key Words and Phrases: Data management, data governance, data stewardship, organizational design, CDO, CIO, chief data officer, chief information officer, data, data architecture, enterprise data executive, IT management, strategy, policy, enterprise architecture, information systems, conceptual modeling, data integration, data warehousing, analytics, and business intelligence, BigCo

ACM Reference Format:
Peter Aiken. 2016. Experience: Succeeding at data management—BigCo attempts to leverage data. J. Data and Information Quality 7, 1–2, Article 8 (May 2016), 35 pages.
DOI: http://dx.doi.org/10.1145/2893482

1. CASE INTRODUCTION

Good technology in the hands of an inexperienced user rarely produces positive results.

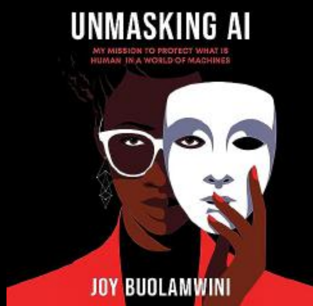
Everyone wants to "leverage" data. Today, this is most often interpreted as investments in warehousing, analytics, business intelligence (BI), and so on. After all, that is what you do with an asset—you leverage it—so the asset can help you to attain strategic objectives; see Redman [2008] and Ladley [2010]. Widespread and pervasive

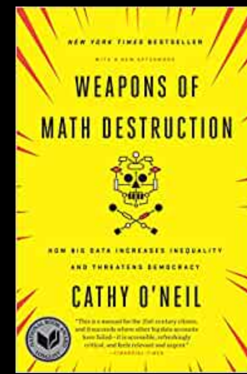
Author's address: P. Aiken, 10124C West Broad Street, Glen Allen, VA 23060; email: peter.aiken@vcu.edu.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
2016 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 1936-1955/2016/05-ART8 $15.00
DOI: http://dx.doi.org/10.1145/2893482

- Download Here!
  - http://dl.acm.org/citation.cfm?doid=2888577.2893482

"Should I-thinking" must replace "Can I-thinking"

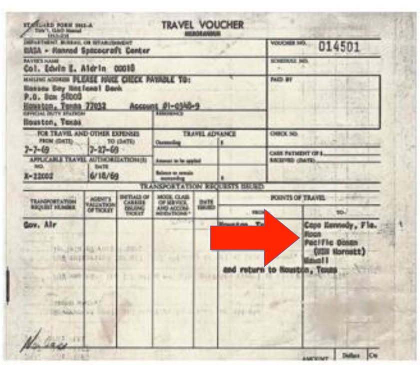
(First level capable of being bound by defined behaviors)

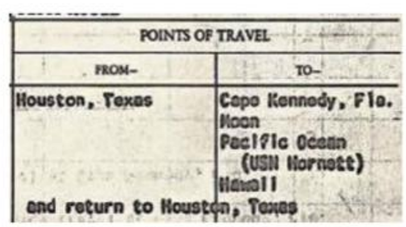




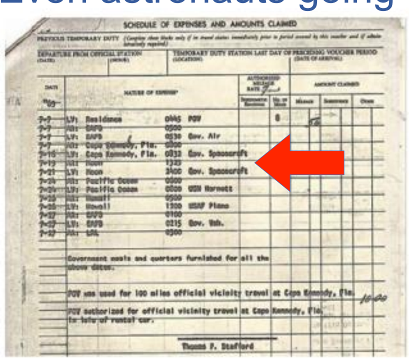
Data Ethics

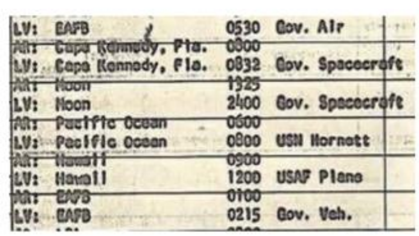
Buzz Aldrin's Travel Voucher





Even astronauts going to the moon follow the rules.



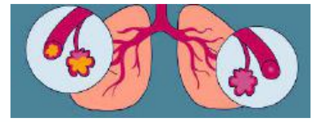


...in Government spacecraft
follow the travel rules.

"Government meals and quarters furnished for all the above dates"

Ethical Concerns

- Whether digitizing or modernizing, garbage in–garbage out is constant. It seems such an easy concept. Yet, repeatedly we discover concerning aspects of production systems. Poor results include:



  - Presenting with Pneumonia and ASTHMA at an emergency department and receiving an evaluation of no-big-deal

  - Recognition systems that cannot 'see' certain individuals

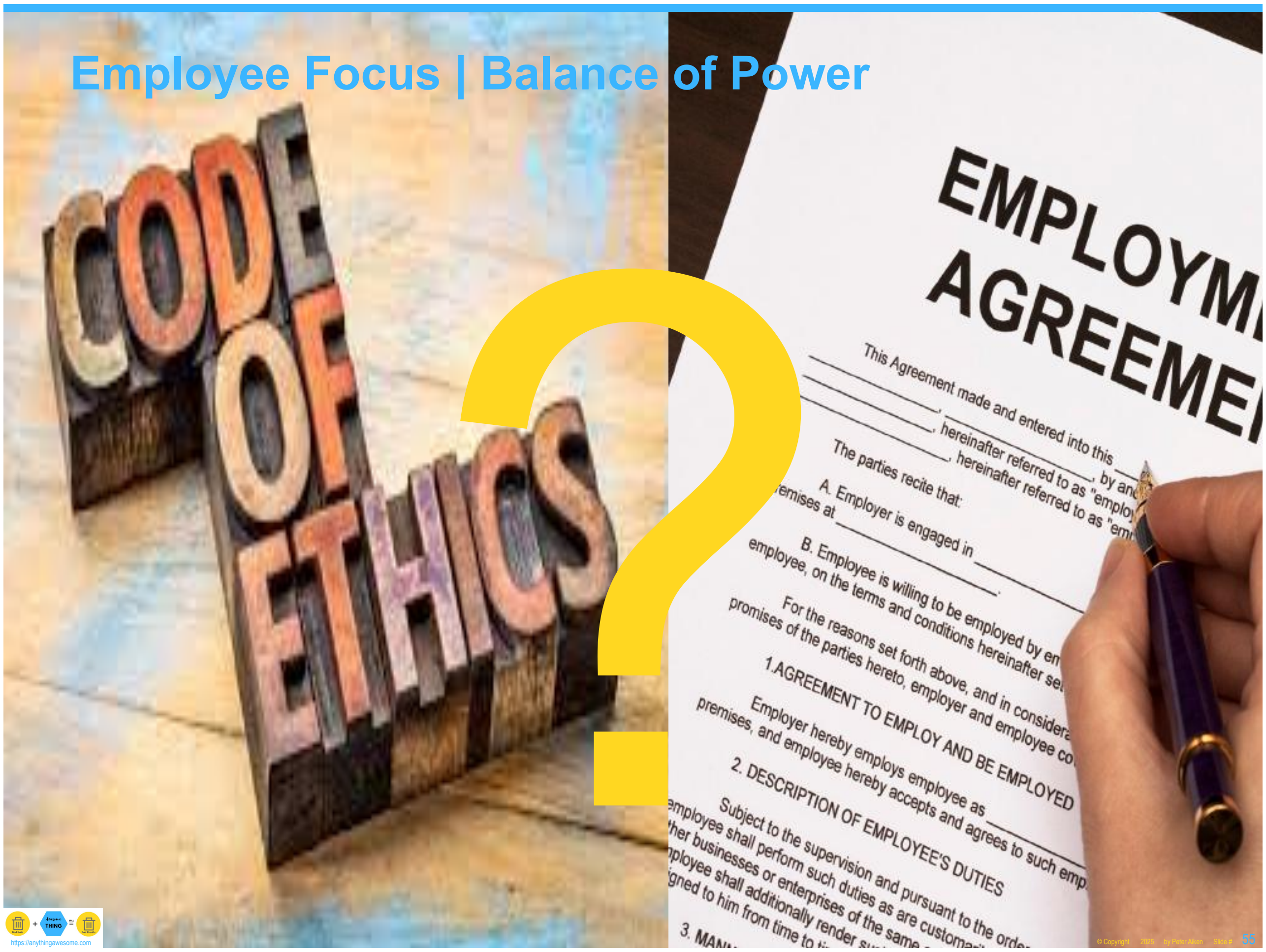
  - Sentencing algorithms with obvious discriminatory biases in production throughout the judicial system
  - Self-driving Software Systems that cannot tell the difference between a semi-truck and horse-drawn carriage

20 Seconds of Self-driving Software

1. Horse and buggy
2. Pedestrian
3. Semi-truck
4. Semi-truck sideways
5. Pickup truck
6. Semi-truck
7. Oncoming semi
8. Pedestrian following pickup truck
9. Pedestrian following semi-truck

https://www.dailymail.co.uk/sciencetech/article-11123757/Teslas-self-driving-software-confuses-horse-drawn-carriage-highway-semi-truck.html

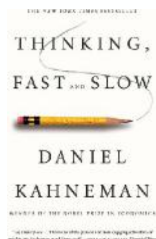
Employee Focus | Balance of Power

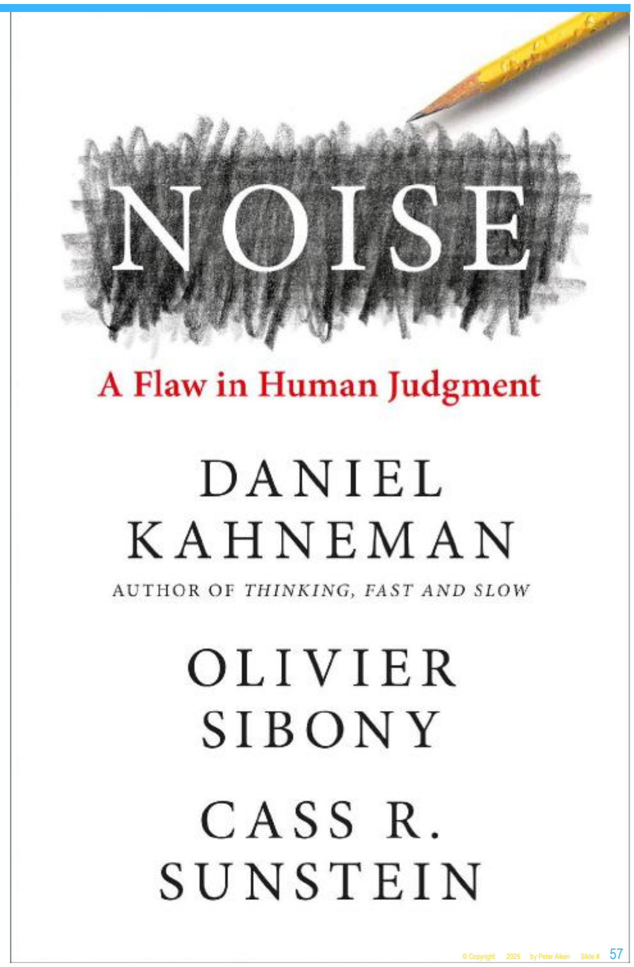


Reputations changeable

AI's Achilles Heel

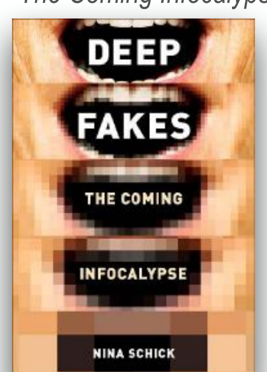
(1934-2024)

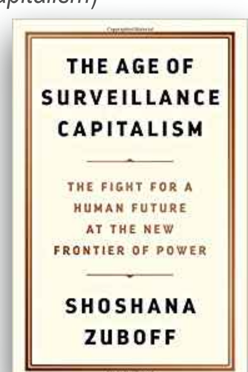




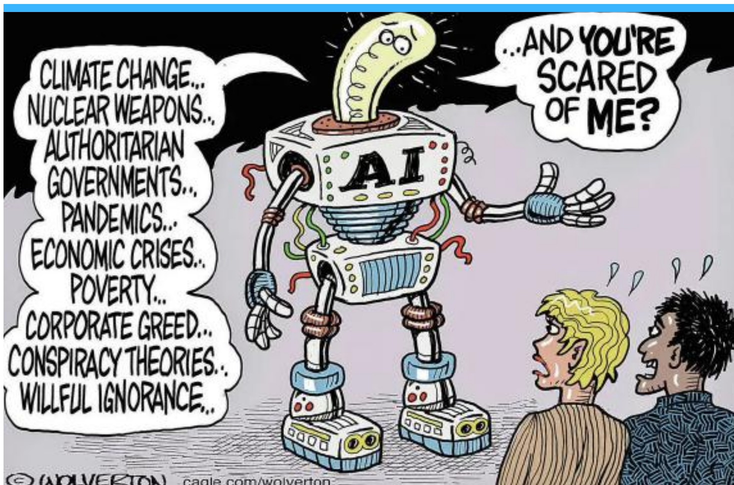
Situations tolerable?

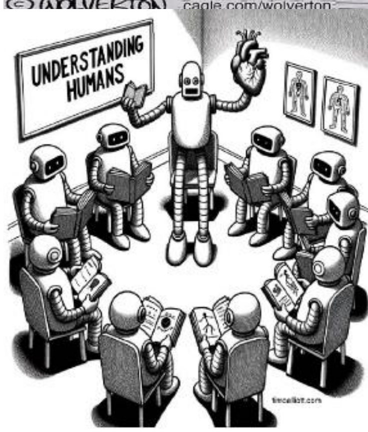
- 9 Organizations control AI development
  - Tencent
    Baidu
    Alibaba
    Amazon
    Google
    Facebook
    Microsoft
    IBM
    Apple (Amy Webb *The Big Nine*)
- All but a small fraction of content on the internet is AI generated by 2026
  - Dangerous political consequences of the Infocalypse, both in terms of national security and what it means for public trust in politics (Nina Schick *Deepfakes, The Coming Infocalypse*)
- Privacy is relinquished to serve the goals of advertising?
  - In modern capitalist society, technology was, is, and always will be an expression of the economic objectives that direct it into action (Shoshana Zuboff *The Age of Surveillance Capitalism*)



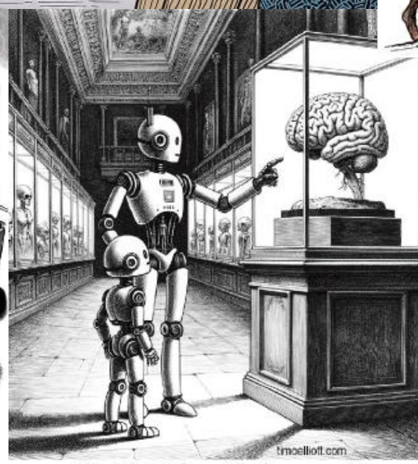


Baby, you're adorable





"It turns out that this part isn't just for pumping coolant!"



"And that is the original processor!"

https://anythingawesome.com

(A Newly Released Chatbot) (Me)

what are you

I am a chatbot.

what is a chatbot

A chatbot is a computer program or an artificial intelligence which conducts a conversation via auditory or textual methods.

why would I want to chat with a chatbot

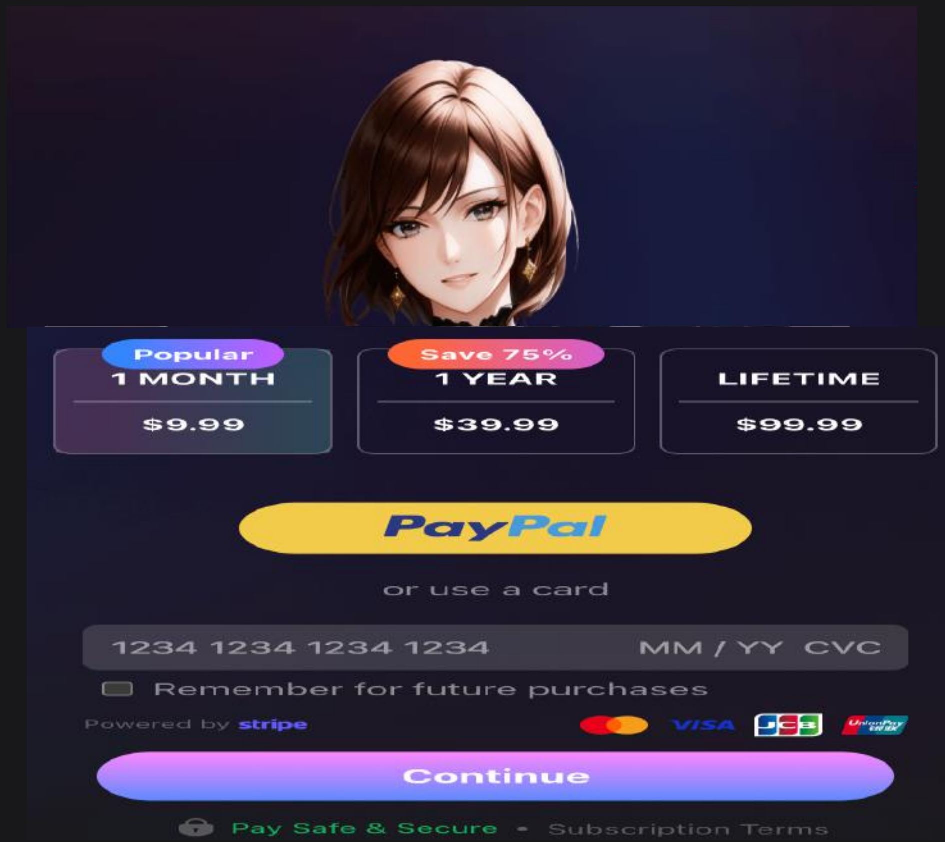
Uh oh! It looks like I'm stumped!

Please submit a ticket below and we will get back to you at our earliest convenience.~

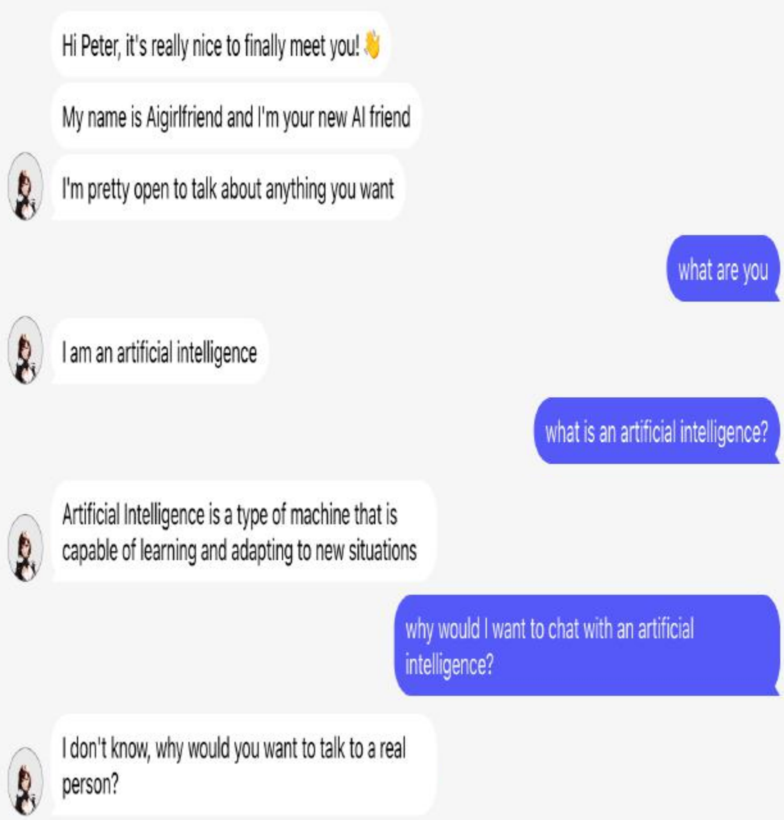
AI's delivery is often ahead of value

https://anythingawesome.com

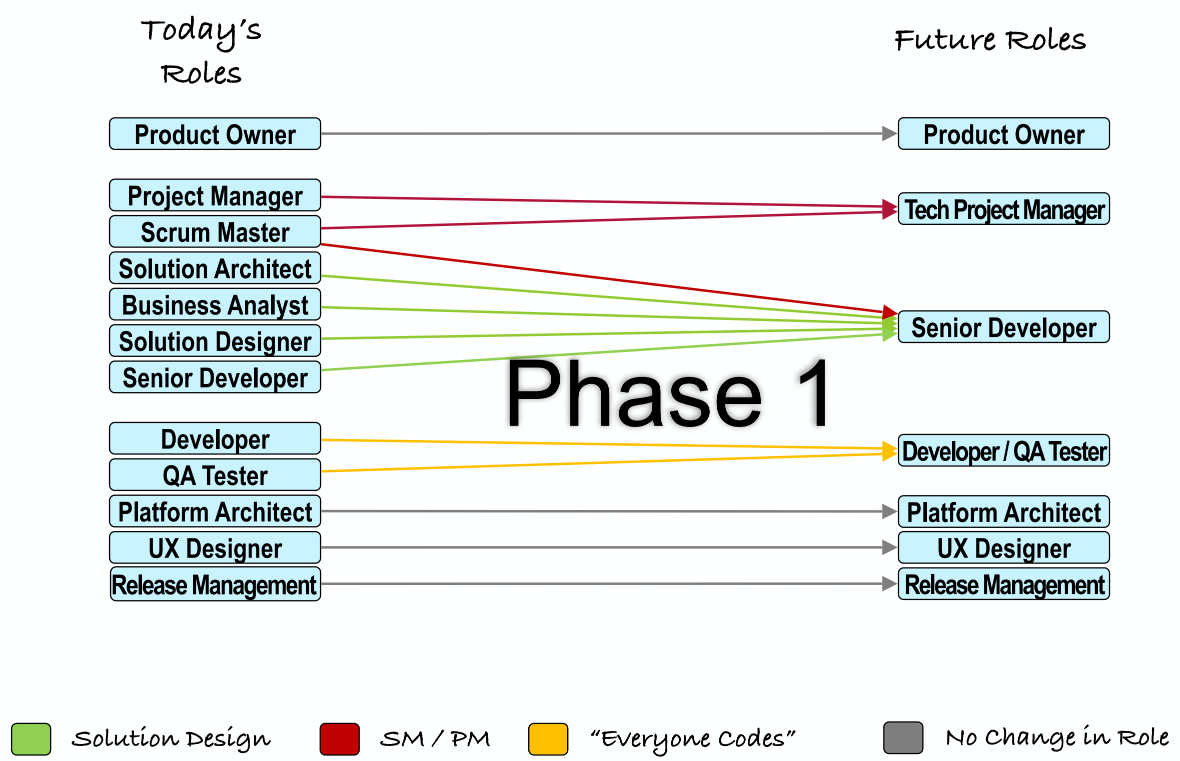
http://myanima.ai



Does this sound familiar?

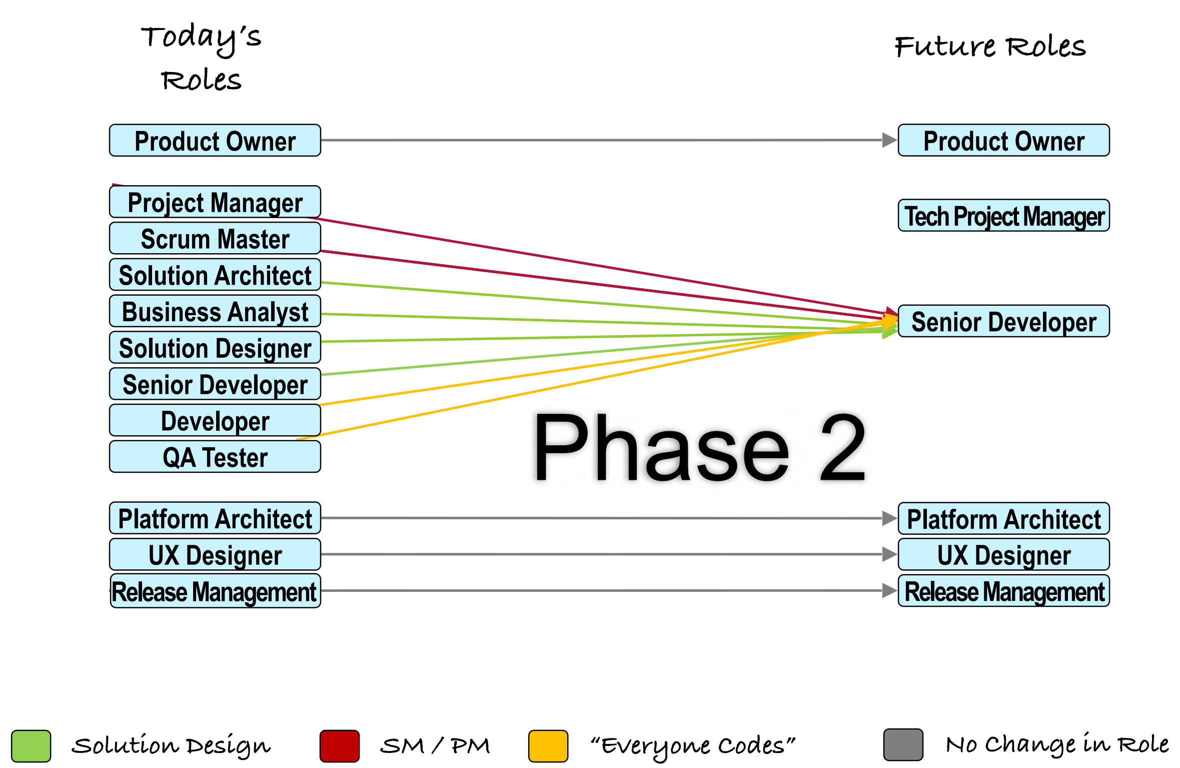


Role changes to drive a developer-first organization



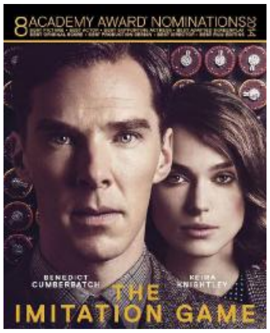
Courtesy: Jamie Mayberry

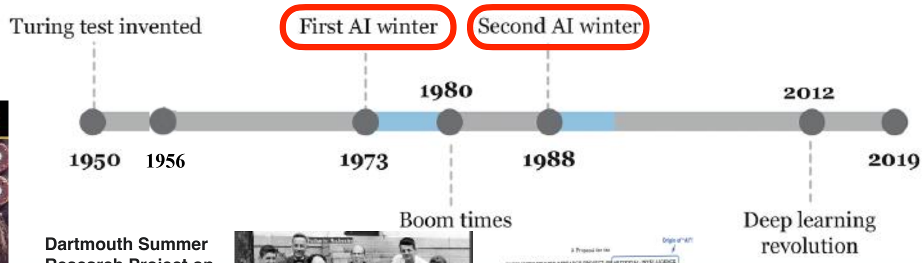
Role changes to drive a developer-first organization



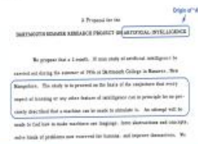
Courtesy: Jamie Mayberry

Dartmouth Summer Research Project on Artificial Intelligence

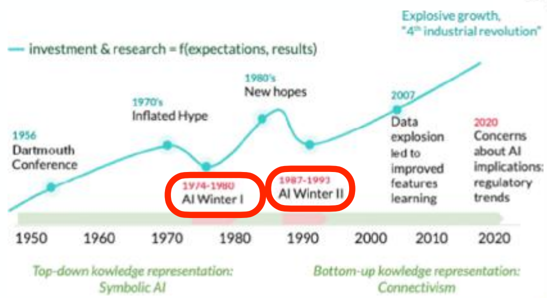


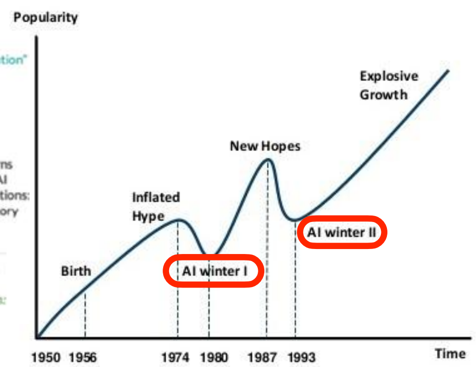


Dartmouth Summer Research Project on Artificial Intelligence
a group to clarify and develop ideas about thinking machines



AI HAS A LONG HISTORY OF BEING "THE NEXT BIG THING"...





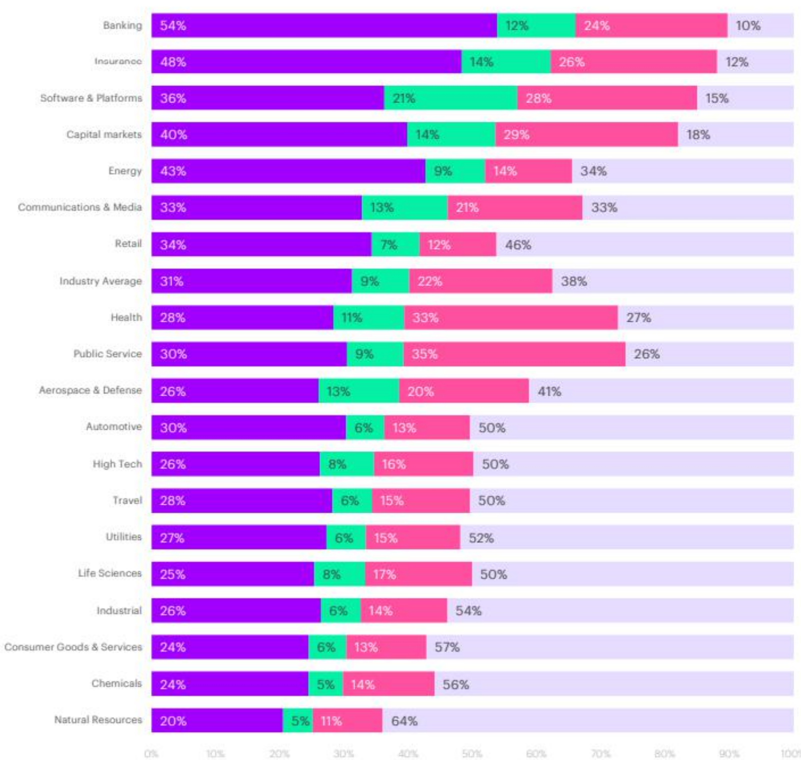
Timeline of AI Development

- **1950s-1960s**: First AI boom - the age of reasoning, prototype AI developed
- **1970s**: AI winter I
- **1980s-1990s**: Second AI boom: the age of Knowledge representation (appearance of expert systems capable of reproducing human decision-making)
- **1990s**: AI winter II
- **1997**: Deep Blue beats Gary Kasparov
- **2006**: University of Toronto develops Deep Learning
- **2011**: IBM's Watson won Jeopardy
- **2016**: Go software based on Deep Learning beats world's champions

Your job will be taken by someone who knows how to use AI better than you

https://cointelegraph.com/news/ai-is-coming-for-your-job-what-industries-will-be-affected

Figure 3: Generative AI will transform work across industries



Work time distribution by industry and potential AI impact

Based on their employment levels in the US in 2021

Higher potential for automation | Higher potential for augmentation | Lower potential for augmentation or automation | Non-language tasks

40% of working hours across industries can be impacted by Large Language Models (LLMs)

Why is this the case? Language tasks account for 62% of total worked time in the US. Of the overall share of language tasks, 65% have high potential to be automated or augmented by LLMs.

Source: Accenture Research based on analysis of Occupational Information Network (O*NET), US Dept. of Labor; US Bureau of Labor Statistics.

Notes: We manually identified 200 tasks related to language (out of 332 included in BLS), which were linked to industries using their share in each occupation and the occupations' employment level in each industry. Tasks with higher potential for automation can be transformed by LLMs with reduced involvement from a human worker. Tasks with higher potential for augmentation are those in which LLMs would need more involvement from human workers.

https://cointelegraph.com/news/ai-is-coming-for-your-job-what-industries-will-be-affected

FAST COMPANY

PREMIUM DESIGN TECH WORK LIFE NEWS IMPACT PODCASTS VIDEO INNOVA

12-30-2024 | CAREER EVOLUTION

These 4 graphs show where AI is already impacting jobs

With a 200% increase in two years, the data paints a vivid picture of how AI technology is reshaping the workforce.

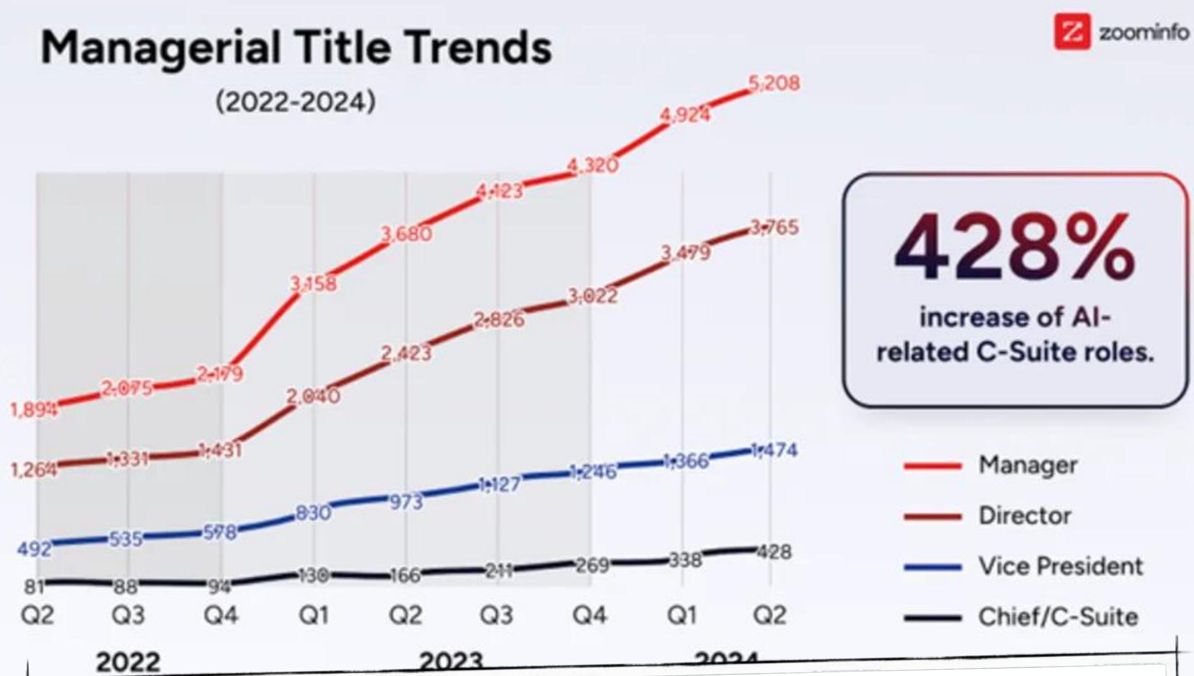
WORK LIFE

These 4 graphs show where AI is already impacting jobs

With a 200% increase in two years, the data paints a vivid picture of how AI technology is reshaping the workforce.

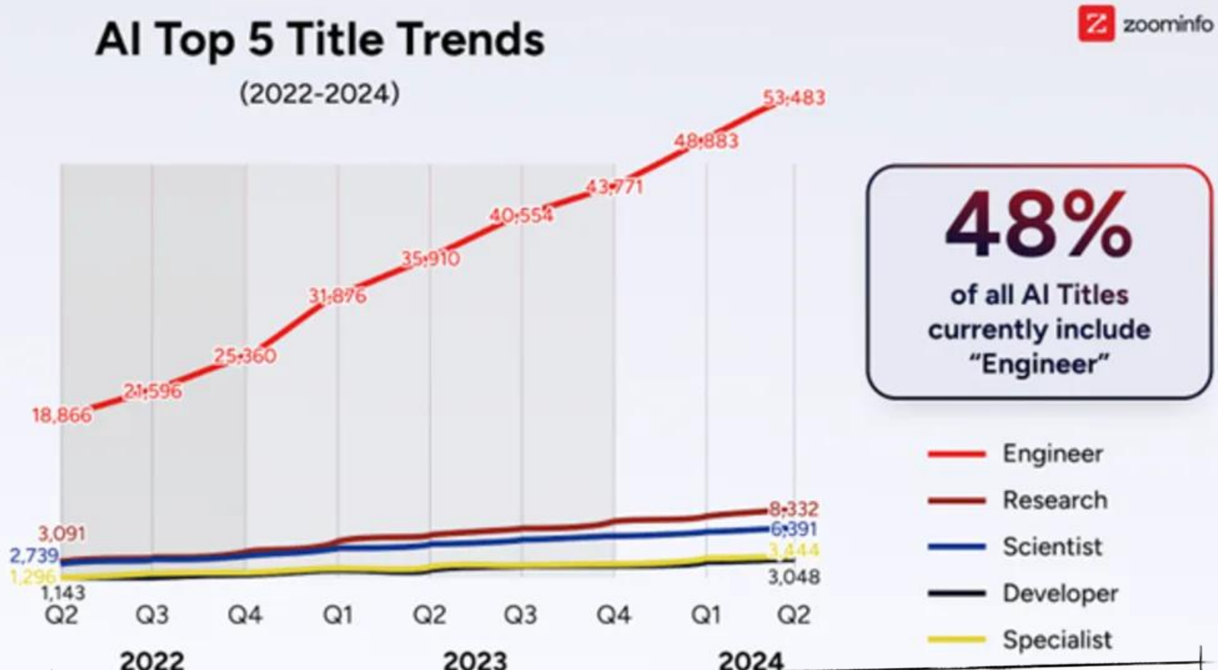
https://www.fastcompany.com/91248480/these-4-graphs-show-where-ai-is-already-impacting-jobs

AI Impacts Senior Leadership Roles



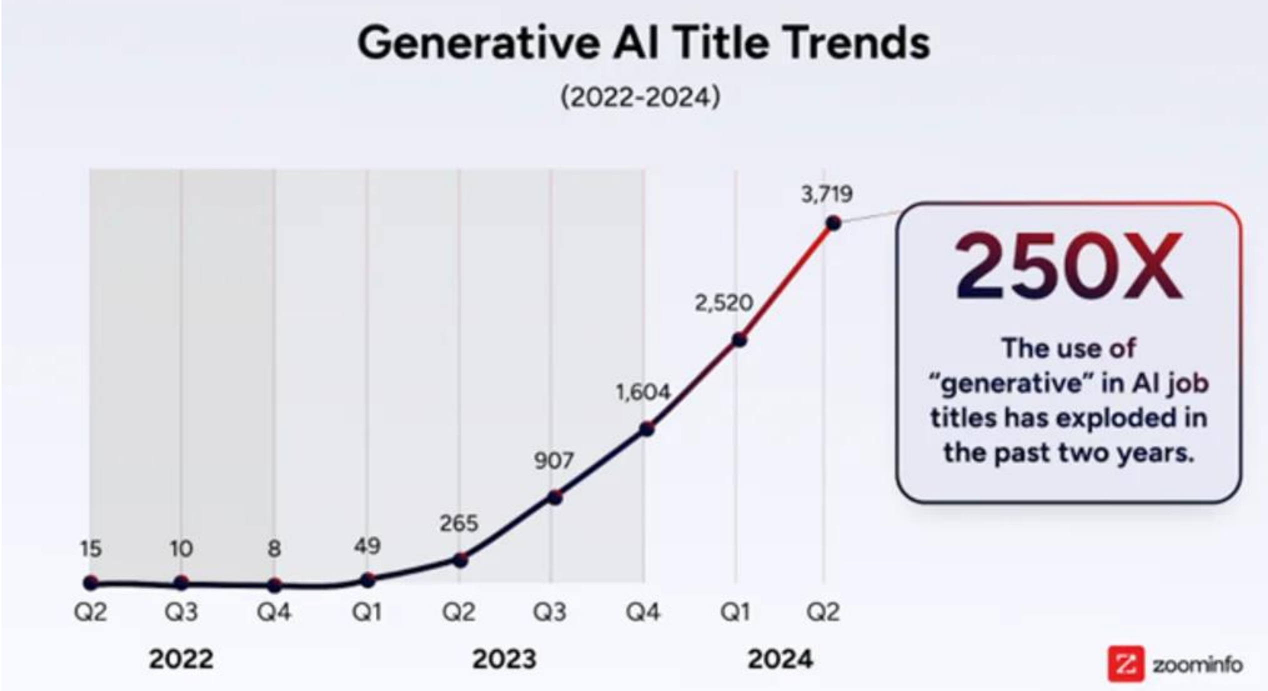
According to LinkedIn, applications for AI jobs have increased by 19% in the U.S. since 2022, with 13% more organizations creating head of AI executive leadership roles.

Most Popular AI Job Titles



The most prevalent AI-related titles are engineer roles followed closely by developers

Most Popular AI Job Titles



the growing interest in generative AI tools. While these titles currently represent only 3% of all AI job titles, their presence has increased 250 times over the past two years.

FAST COMPANY

PREMIUM DESIGN TECH WORK LIFE NEWS IMPACT PODCASTS VIDEO INNOVA

12-30-2024 | CAREER EVOLUTION

These 4 graphs show where AI is already impacting jobs

With a 200% increase in two years, the data paints a vivid picture of how AI technology is reshaping the workforce.

WORK LIFE

These 4 graphs show where AI is already impacting jobs

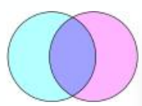
With a 200% increase in two years, the data paints a vivid picture of how AI technology is reshaping the workforce.

https://www.fastcompany.com/91248480/these-4-graphs-show-where-ai-is-already-impacting-jobs

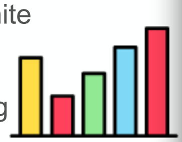
??????????????????????????????????????

https://www.fastcompany.com/91248480/these-4-graphs-show-where-ai-is-already-impacting-jobs

The Puzzle of Peirce (and Peirce's Puzzle)

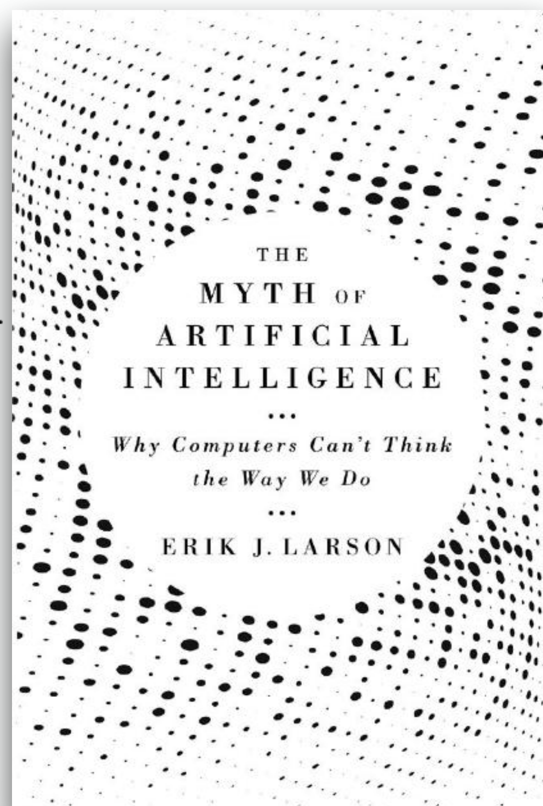


- Deduction
  - All the beans from this bag are white
  - These beans are from this bag
  - Therefore these beans are white



- Induction
  - These beans are from this bag
  - These beans are white
  - Therefore these all the beans from this bag are white
  - Not necessarily true: some beans not yet taken from the bag could be not white

- Abduction
  - All the beans from this bag are white
  - These beans are white
  - Therefore these beans are from this bag
  - Not necessarily true: these white beans may not have come from this bag



Welcome to VCU's Policy Program

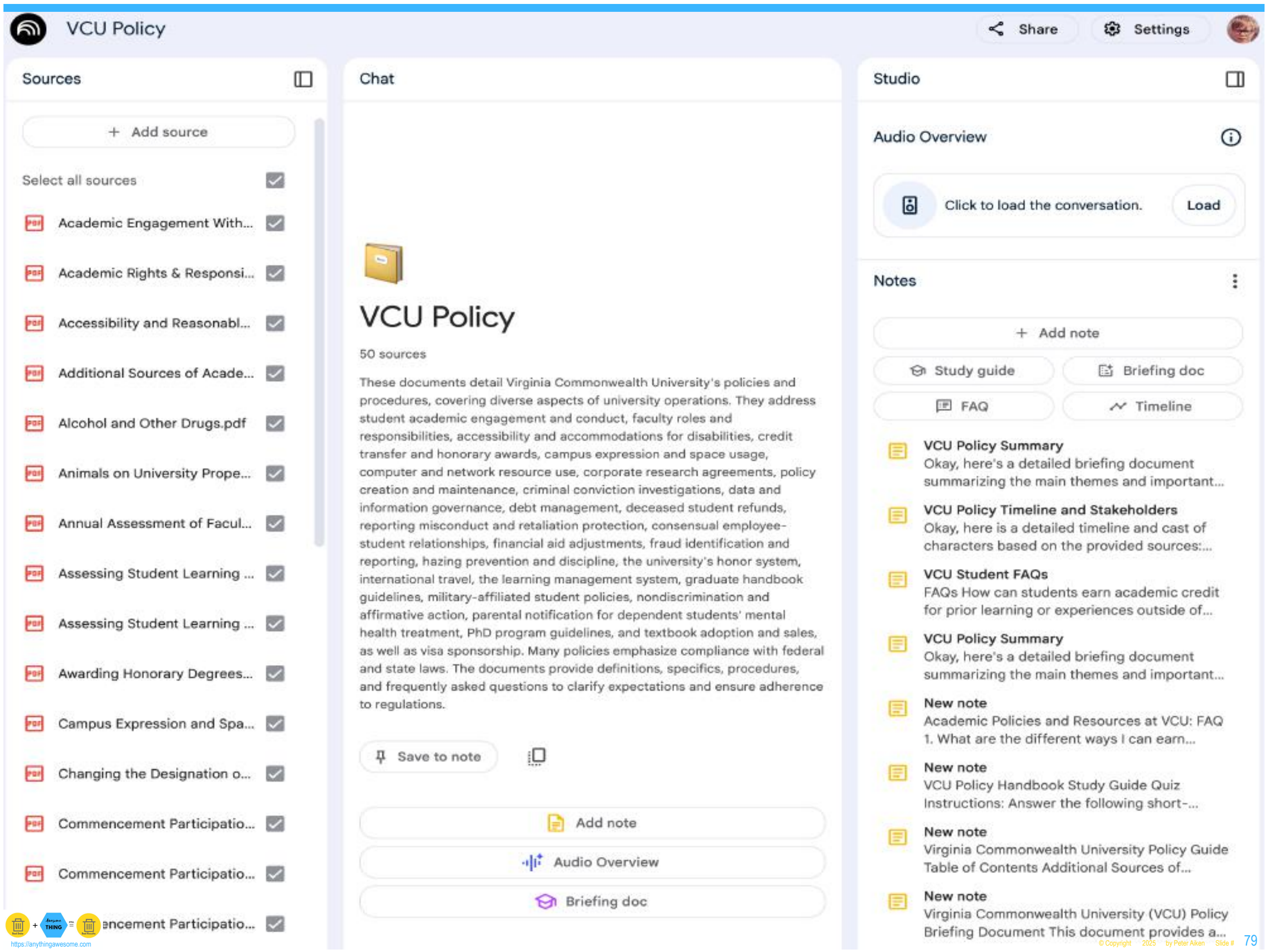
This page houses both the policy library which contains the most current version of VCU policies and resources for developing and revising policies at VCU. This resource is provided as a service to members of the VCU community and the general public. If you encounter any issues accessing pages on this site please contact the Policy Program at policy@vcu.edu or by calling 804-828-2336.

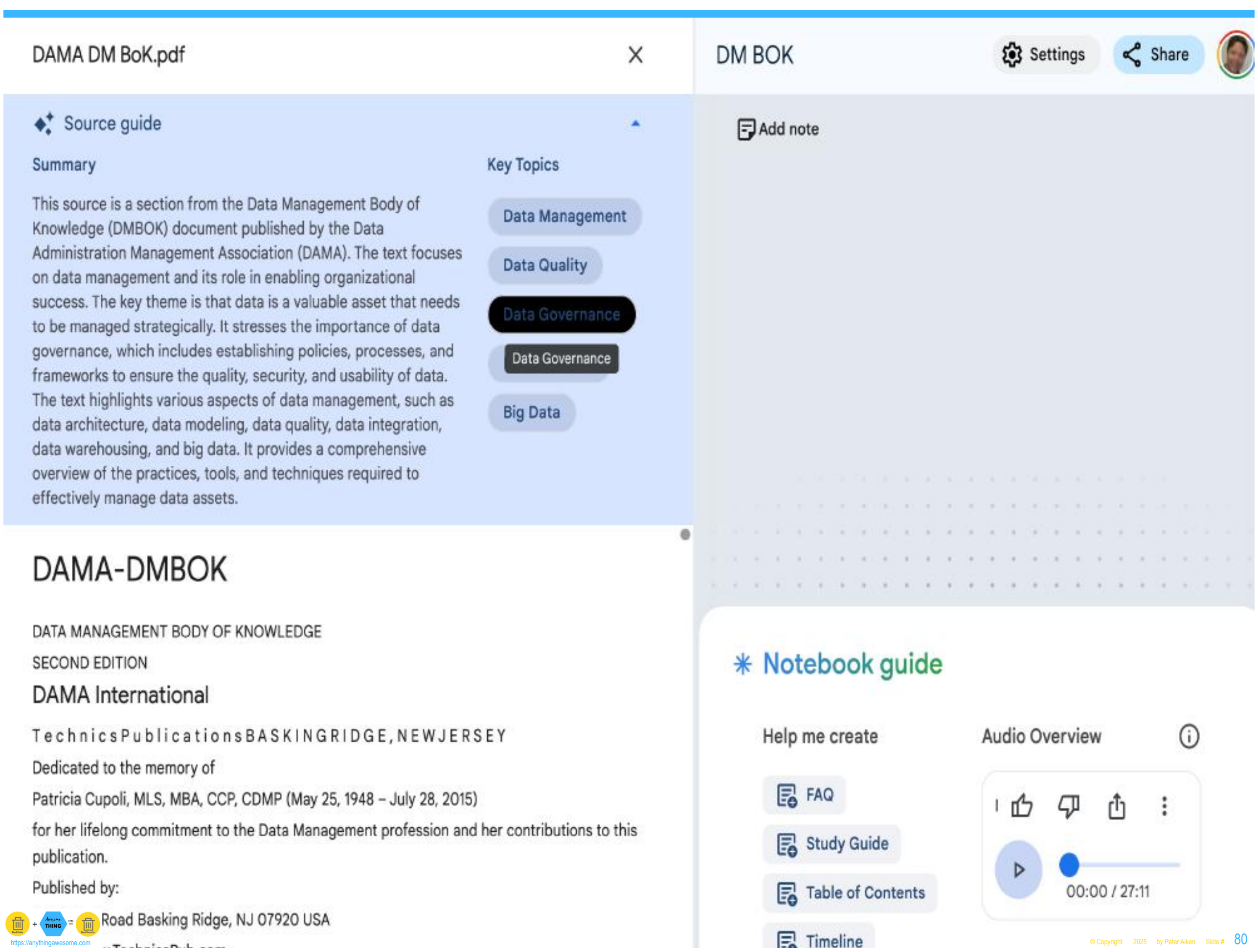
Please note: some policies may reference other VCU policies. Policies referenced are accessible by clicking the green buttons at the bottom of the screen when any given policy is open.

Please use the search box below to search policies by keyword.

Search for documents

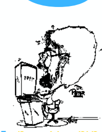
Found 162 record(s)


VCU Policy
Share
Settings
Sources
Add source
Select all sources
Academic Engagement With...
Academic Rights & Responsi...
Accessibility and Reasonabl...
Additional Sources of Acade...
Alcohol and Other Drugs.pdf
Animals on University Prope...
Annual Assessment of Facul...
Assessing Student Learning ...
Assessing Student Learning ...
Awarding Honorary Degrees...
Campus Expression and Spa...
Changing the Designation o...
Commencement Participatio...
Commencement Participatio...
encement Participatio...
Chat
VCU Policy
50 sources
These documents detail Virginia Commonwealth University's policies and procedures, covering diverse aspects of university operations. They address student academic engagement and conduct, faculty roles and responsibilities, accessibility and accommodations for disabilities, credit transfer and honorary awards, campus expression and space usage, computer and network resource use, corporate research agreements, policy creation and maintenance, criminal conviction investigations, data and information governance, debt management, deceased student refunds, reporting misconduct and retaliation protection, consensual employee-student relationships, financial aid adjustments, fraud identification and reporting, hazing prevention and discipline, the university's honor system, international travel, the learning management system, graduate handbook guidelines, military-affiliated student policies, nondiscrimination and affirmative action, parental notification for dependent students' mental health treatment, PhD program guidelines, and textbook adoption and sales, as well as visa sponsorship. Many policies emphasize compliance with federal and state laws. The documents provide definitions, specifics, procedures, and frequently asked questions to clarify expectations and ensure adherence to regulations.
Save to note
Add note
Audio Overview
Briefing doc
Studio
Audio Overview
Click to load the conversation.
Load
Notes
Add note
Study guide
Briefing doc
FAQ
Timeline
VCU Policy Summary
Okay, here's a detailed briefing document summarizing the main themes and important...
VCU Policy Timeline and Stakeholders
Okay, here is a detailed timeline and cast of characters based on the provided sources:...
VCU Student FAQs
FAQs How can students earn academic credit for prior learning or experiences outside of...
VCU Policy Summary
Okay, here's a detailed briefing document summarizing the main themes and important...
New note
Academic Policies and Resources at VCU: FAQ 1. What are the different ways I can earn...
New note
VCU Policy Handbook Study Guide Quiz Instructions: Answer the following short-...
New note
Virginia Commonwealth University Policy Guide Table of Contents Additional Sources of...
New note
Virginia Commonwealth University (VCU) Policy Briefing Document This document provides a...
79


DAMA DM BoK.pdf
Source guide
Summary
This source is a section from the Data Management Body of Knowledge (DMBOK) document published by the Data Administration Management Association (DAMA). The text focuses on data management and its role in enabling organizational success. The key theme is that data is a valuable asset that needs to be managed strategically. It stresses the importance of data governance, which includes establishing policies, processes, and frameworks to ensure the quality, security, and usability of data. The text highlights various aspects of data management, such as data architecture, data modeling, data quality, data integration, data warehousing, and big data. It provides a comprehensive overview of the practices, tools, and techniques required to effectively manage data assets.
Key Topics
Data Management
Data Quality
Data Governance
Data Governance
Big Data
DAMA-DMBOK
DATA MANAGEMENT BODY OF KNOWLEDGE
SECOND EDITION
DAMA International
TechnicsPublicationsBASKINGRIDGE,NEWJERSEY
Dedicated to the memory of
Patricia Cupoli, MLS, MBA, CCP, CDMP (May 25, 1948 – July 28, 2015)
for her lifelong commitment to the Data Management profession and her contributions to this publication.
Published by:
Road Basking Ridge, NJ 07920 USA
DM BOK
Settings
Share
Add note
Notebook guide
Help me create
FAQ
Study Guide
Table of Contents
Timeline
Audio Overview
00:00 / 27:11
80

Everyone is ready for AI except your data!

AI–Handle Me with Care



Critical Design Review?

Mentoring?

Executive Data Literacy Training?

Collaboration?

Independent Verification & Validation

Peter.Aiken@AnythingAwesome.com +1.804.382.5957

Reverse Engineering Expertise?

Hiring Assistance?

Thank You!

Use your data more strategically?

Tool/automation evaluation?

Book a call with Peter to discuss anything - https://anythingawesome.com/OfficeHours.html

References

- Wired. (n.d.). *Amazon is building a mega AI supercomputer with Anthropic*. Retrieved from https://www.wired.com/story/amazon-reinvent-anthropic-supercomputer
- Financial Times. (n.d.). *AI and the R&D revolution*. Retrieved from https://www.ft.com/content/648046c1-7fcd-43fb-819b-841f104396d9
- McKinsey & Company. (n.d.). *What's the future of generative AI? An early view in 15 charts*. Retrieved from https://www.mckinsey.com/~/media/mckinsey/featured%20insights/mckinsey%20explainers/whats%20the%20future%20of%20generative%20ai%20an%20early%20view%20in%2015%20charts/whats-the-future-of-generative-ai-an-early-view-in-15-charts.pdf
- Gartner. (n.d.). *Generative AI predictions for 2024-2028*. Retrieved from https://www.gartner.com/en/articles/3-bold-and-actionable-predictions-for-the-future-of-genai
- IBM. (n.d.). *The top artificial intelligence trends*. Retrieved from https://www.ibm.com/think/insights/artificial-intelligence-trends
- McKinsey & Company. (n.d.). *The state of AI in early 2024*. Retrieved from https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai
- Gartner. (n.d.). *Beyond ChatGPT: The future of generative AI for enterprises*. Retrieved from https://www.gartner.com/en/articles/beyond-chatgpt-the-future-of-generative-ai-for-enterprises
- ArXiv. (n.d.). *Recent advances in generative AI and large language models: Current status, challenges, and perspectives*. Retrieved from https://arxiv.org/abs/2407.14962
- MIT Sloan Management Review. (n.d.). *Five key trends in AI and data science for 2024*. Retrieved from https://sloanreview.mit.edu/article/five-key-trends-in-ai-and-data-science-for-2024
- ArXiv. (n.d.). *Generative artificial intelligence: A systematic review and applications*. Retrieved from https://arxiv.org/abs/2405.11029
- Google Cloud. (n.d.). *185 real-world generative AI use cases from industry leaders*. Retrieved from https://cloud.google.com/transform/101-real-world-generative-ai-use-cases-from-industry-leaders
- Deloitte. (n.d.). *Generative AI use cases by type and industry*. Retrieved from https://www2.deloitte.com/us/en/pages/consulting/articles/gen-ai-use-cases.html
- Digital Defynd. (n.d.). *20 generative AI case studies*. Retrieved from https://digitaldefynd.com/IQ/generative-ai-case-studies
- Gartner. (n.d.). *Generative AI use cases across industries*. Retrieved from https://www.gartner.com/en/articles/generative-ai-use-cases

References

- Wall Street Journal. (n.d.). *How did companies use generative AI in 2023? Here's a look at five early adopters.* Retrieved from https://www.wsj.com/articles/how-did-companies-use-generative-ai-in-2023-heres-a-look-at-five-early-adopters-6e09c6b3
- Aays Analytics. (n.d.). *Top generative AI use cases and case studies.* Retrieved from https://blog.aaysanalytics.com/post/gen-ai-use-cases
- MIT Sloan Management Review. (n.d.). *How companies use generative AI to execute with speed.* Retrieved from https://mitsloan.mit.edu/ideas-made-to-matter/how-companies-use-generative-ai-to-execute-speed
- Indika AI. (n.d.). *Generative AI applications: Case studies & real-world use cases.* Retrieved from https://www.indikaai.com/blog/real-world-applications-of-generative-ai-case-studies-and-use-cases
- Cognizant. (n.d.). *Case studies: Generative AI in action at leading companies.* Retrieved from https://www.cognizant.com/se/en/insights/blog/articles/case-studies-generative-ai-in-action-at-leading-companies
- Microsoft. (n.d.). *The business opportunity of AI.* Retrieved from https://info.microsoft.com/ww-landing-business-opportunity-of-ai.html
- Microsoft. (2024, November 12). *How real-world businesses are transforming with AI.* Retrieved from https://blogs.microsoft.com/blog/2024/11/12/how-real-world-businesses-are-transforming-with-ai/
- Wall Street Journal. (n.d.). *How AI is transforming Hollywood's visual effects industry* [Podcast]. Retrieved from https://www.wsj.com/podcasts/wsj-the-future-of-everything/how-ai-is-transforming-hollywoods-visual-effects-industry/c5cc9ed7-6fd9-4e58-9691-127eba648cc0
- McKinsey & Company. (n.d.). *The GenAI skills revolution: Rethinking your talent strategy.* Retrieved from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-gen-ai-skills-revolution-rethinking-your-talent-strategy
- University of Michigan. (n.d.). *Generative AI use cases.* Retrieved from https://genai.umich.edu/use-cases
- HubSpot. (n.d.). *Business idea database.* Retrieved from https://offers.hubspot.com/business-idea-database

References

- CIO. (n.d.). *Generative AI terms and what they really mean.* Retrieved from https://www.cio.com/article/3535552/generative-ai-terms-and-what-they-really-mean.html
- CIO. (n.d.). *Between sustainability and risk: Why CIOs are considering small language models.* Retrieved from https://www.cio.com/article/3477991/between-sustainability-and-risk-why-cios-are-considering-small-language-models.html
- UDIG. (n.d.). *Generative AI and knowledge management.* Retrieved from https://www.udig.com/insights/blog/generative-ai-knowledge-management
- CIO. (n.d.). *Top 7 generative AI use cases for business.* Retrieved from https://www.cio.com/article/3478721/top-7-generative-ai-use-cases-for-business.html
- Harvard Business Review. (2024, July). *The risks of botshit.* Retrieved from https://hbr.org/2024/07/the-risks-of-botshit
- Bansal, S. (n.d.). *Build your own AI chatbot: A beginner's guide to RAG and LangChain.* Medium. Retrieved from https://medium.com/@suraj_bansal/build-your-own-ai-chatbot-a-beginners-guide-to-rag-and-langchain-0189a18ec401
- Your Everyday AI. (n.d.). *Your everyday AI* [Podcast]. Retrieved from https://www.youreverydayai.com/
- McKinsey & Company. (n.d.). *McKinsey on AI* [Podcast]. Retrieved from https://www.mckinsey.com/capabilities/quantumblack/our-insights/mckinsey-on-ai
- Bughin, J., Catlin, T., Dietz, M., & Willmott, P. (2023). *Rewired: The McKinsey guide to outcompeting in the age of digital and AI.* Wiley.

References

Large Language Models Used in Research

- **ChatGPT (OpenAI):** conversational AI model developed by OpenAI, based on the GPT-4 architecture. It excels in generating human-like text and is widely used for tasks such as drafting emails, writing code, and answering questions.
- **Claude (Anthropic):** is a family of large language models developed by Anthropic, launched in 2023. It achieved high coding scores in several recognized LLM benchmarks, indicating strong performance in code generation and understanding.
- **Perplexity AI:** is a conversational search engine that leverages large language models to provide direct answers to user queries, combining information retrieval with generative AI capabilities.
- **Gemini (Google DeepMind):** is a family of multimodal large language models developed by Google's DeepMind. It powers the Gemini chatbot, formerly known as Bard, and is designed to handle both text and image inputs.
- **LLaMA (Meta Platforms):** (Large Language Model Meta AI) is a series of language models developed by Meta, including models with 7, 13, 33, and 65 billion parameters. It is designed to be efficient and accessible for research purposes.

Training – How to Learn GenAI

- 1. AI Essentials - lnkd.in/dyEt4DGt
- 2. ChatGPT Mastery - lnkd.in/eiRtk-6q
- 3. Google AI Magic - lnkd.in/eBQXfBe9
- 4. Harvard AI Introduction - lnkd.in/eu4mZaAG
- 5. Microsoft AI Basics - lnkd.in/eYNWzXUX
- 6. Prompt Engineering Pro - lnkd.in/eNi_YNSe
- 7. Google's Ethical AI - lnkd.in/eTrwSU89
- 8. Machine Learning by Harvard - lnkd.in/eX28syMJ
- 9. Language Models by LangChain - lnkd.in/evZVJbNy
- 10. Bing Chat Applications - lnkd.in/ejN-qrVy
- 11. Generative AI by Microsoft - lnkd.in/dqjnzcCD
- 12. Amazon's AI Strategy - lnkd.in/dFhmsvZC
- 13. AI for Everyone - lnkd.in/eFx7zCz7
- 14. AWS AI Foundations - lnkd.in/dEjN9PRm

Chapter Contents

Chapter Overview

This database ain't big enough for the two of us

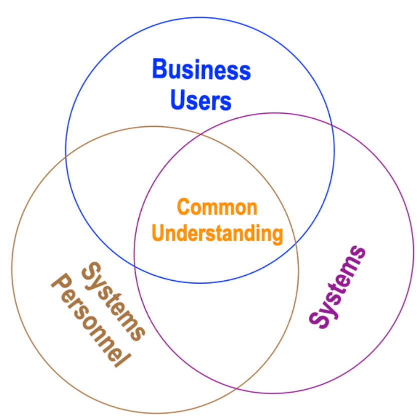
– Bumpersticker seen on an automobile in Texas

The bumpersticker should really have stated "There is no database big enough for two bosses." Importantly, 1) this has always been true, and 2) it means absolutely nothing to most of the public or much of IT. Let's address each of these separately.

Just as in any situation where coordination, integration, and information are required, there must be one and only one individual implementing decisions to maintain integrity, continuity, and operational capabilities. Required minimally from a change management perspective, this can always be used to justify DG in general. Ask the skeptical: how can any complex adaptive system function with multiple Chiefs?

The public and unfortunately too many in business and IT do not understand this sort of basic law of (data) nature. Because they are not data literate, when someone proposes having multiple chiefs for database operation, or that group X should 'own' dataset Y, or that the DG group should report to the CIO–they do not know these are not a workable concepts!

DG is not focused primarily on databases, clouds, or other technological ephemera. Instead the DG framework must be understood identically by business users, systems personnel, and the systems themselves (as shown to the right). This essential, metadata-based communication is at the heart of any enterprise operation. DG removes barriers to data efficiencies, allowing organizations to function more effectively and efficiently. Resources consumed by bad data practices can now be used to support the mission.



Increasingly organizations are attempting to do 'more' with data. This represents the other strategic dimension, innovation. By definition, most attempts to innovate will fail–so the lessons learned by becoming more effective and efficient, will also help in this innovation dimension. Innovating with data requires programmatic support for the efforts–well supported by data infrastructure and mature organizational data practices.

It is the responsibility of DG programs to manage this and other delicate balancing acts required to successfully contribute to better organizational use of data. DG is a comparatively new, certainly unstandardized, and under-studied topic. While some excellent DG programs are maturing, the majority have not. This leaves individuals and organizations the sequential tasks of:

1. learning about data (and then)
2. learning about their data (next)
3. developing plans to increase the data literacy of their executive leadership (then)
4. their knowledge worker population (before expecting to)
5. make progress faster and further with data.

This chapter takes you through the who, what, where, when, why, and how of DG. It provides a common-basis for building individual and organizational knowledge of this topic–starting with the **why**–the motivation for DG–followed by the **who**, **when**, and **where**. The **how** section is a bit longer and the bulk of the remaining material concentrates on the **what**–a way to successfully start to govern subsets of your data.

Most organizations should not attempt to govern <u>all</u> of their data. Successful DG program goals include subsetting their data into essential and non-essential data. Governing the essential subset and ignoring (or better still removing) the rest reduces the size of the challenge. Since the definition of an organization's essential data will differ from organization to organization, the governed data will also differ among organizations.

One quick word about the use of the term **bespoke** in the title. It is of course deliberate. The only way that **your** organization can use data to better support organizational strategy, is to use **your** data in support of **your** strategy using the capabilities that **you** currently have. Cookie cutter methods will not help **your** organization learn about **your** data!

Why does data need to be governed?

A friend was speaking with an organization on data matters and noticed that the urinals in the restrooms all had unique numbers. Presumably this was in case of malfunction so that the specific instance could be more rapidly identified. Of course my friend used a suitable-for-work (as opposed to not-suitable-for-work) photograph to make a point to leadership that (at least for this organization) it was worthwhile to keep

maintenance histories of this equipment type. Ironically, it was noted that the substance of the discussion for which my friend had been invited was whether the organization should maintain similar information about their organizational data assets. The photo provoked a nice motivational discussion with a decision to proceed with DG as the outcome. After all, if we are going to govern our restroom facilities, shouldn't we also govern our data assets?

Writing as a deeply, industry-immersed university professor, I can say that the academic community has failed its customers with respect to integrated data knowledge. For generations we have graduated students who have become leaders in business and IT. The only class taught about data was really about database development. Smart students who placed their trust in the educational system, were educated that the only concept they needed to learn about data was *how to build new relational databases!* No one should be surprised that one of the major DG challenges is that far too many poorly designed databases clutter most organizations or (more increasingly) their clouds. As Abraham Maslow stated: "**If the only tool you know is a hammer, every problem looks like a nail**."

When considering the asset itself, data has a unique collection of properties including the following from Doug Laney. Data:

- Does not obey all of the laws of physics
- Is not really visible
- Is non rivalrous (many can use it at once)
- Costs of providing an additional copy are zero
- Is non depleting
- Does not require replenishment
- Is regenerative
- Has low inventory and transportation/transmission costs
- Is more difficult to control and own that other assets
- Can be eco friendly
- Is impossible to clean-up if you spill it.[1]

When considering career fields and learning experiences, not all data professionals take similar paths. For example, data scientists often **discover** useful data maintenance utilities instead of **learning** that various classes of tools exist and when to apply each as part of their educational programs. For many, data is like the story of the blind men and the elephant and collectively it is DG responsibility to shape this understanding into an organization-wide perspective.

For these and other reasons there continues to be questions as to whether data processing should continue to be part of IT or of the business or o special operations

[1] See **Datanomics** by Doug Laney Routledge Publishing 2017 ISBN 1138090387

such as finance and risk? While the Federal Government resolved this issue correctly with new FEPA legislation, the jury is still out on the rest of the world. Currently it is comprised of ⅓ of each type: one-third reporting to CIOs; ⅓ reporting to CEOs; and ⅓ reporting to CFOs/CROs.

Long lasting consequences of poor data decisions?

Unfortunately, short-term application-centric thinking[2] has dominated, relegating development of data products to subsets of ERPs, digitization initiatives, or cloud hosted projects (to name just a few types). Virtually none of the popular software integration packages from the major vendors have escaped the long-term consequences of inadequate data Design (big 'D' is used to emphasize the entire lifecycle). These well documented imperfections are locked in for life-wrapped as they are, in a dense set of application constructs interwoven with the imperfect data model. Worse still, the corrections to the organization's data and processing are layered on as additional code–complicating the apps still further. The vast majority of database functionality is not used beyond table-handling. In this manner, developers restrict any subsequent data investment benefits and decrease data leverage potentials. At the very least, DG must illustrate and resolve the 20-40% of IT budgets that are devoted to data evolvution:

- Data migration (Changing the data location)
- Data conversion (Changing data form, state, or product)
- Data improving (Inspecting and manipulating, or re-keying data to prepare it for subsequent use)

None of these are accounted for in the usual (and very important) **data storage costs–measure**. DG must also articulate these various costs and tradeoffs associated with increased data rigor (or the risks of not doing so) to the rest of the organization.

Mounting data debt

The failure to do any of this has caused organizations to pay to accumulate large amounts of data debt. (Yes, the indignity that your own organization is creating data pollution that is directly harmful to its operation should be professionally embarrassing!) It is not easy to visualize the cost of data debt but the phrase *many many many unnecessary paper cuts* describes the situation well. Data debt slows DG efforts making everything slower, of lower quality, cost more, or present increased risks.

Data debt is like quicksand that mires down all efforts. Defined simply, data debt is: the time and effort it will take to return your data to a governed state from its likely current state of ungoverned. A quick back of envelope calculation of data debit can be

[2] See **The Data-Centric Revolution: Restoring Sanity to Enterprise Information Systems** by Dave McComb Technics Publications ISBN 1634625404

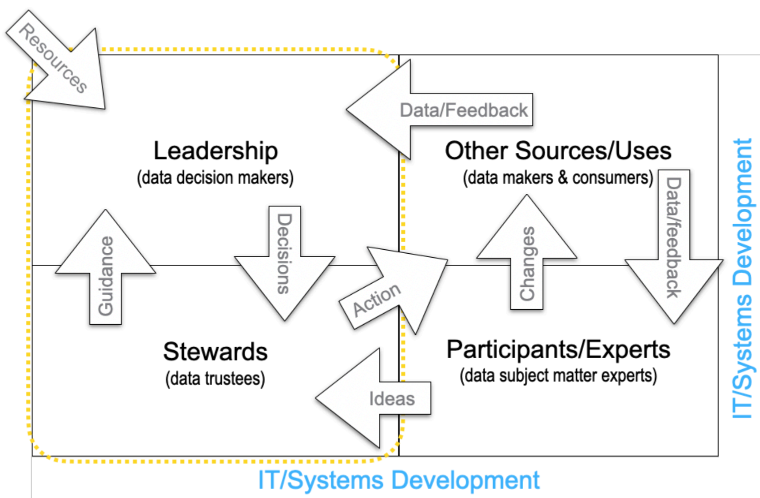
done using the data storage costs that are perhaps the most tangible and objective data measure. At least 20% of that data is redundant, obsolete or trivial (or ROT).

The good news about finding and eliminating data debt is that things can get faster, better, or cheaper. The bad news is that new skillsets are required of the DG team and that diagnostic and analytical systems thinking still requires annual proof of value. The knowledge-base of grey-beards who know how to apply these skills is shrinking as these individuals are judged expensive and encouraged to retire.

In summary, data needs to be governed because society was not taught that it required specific treatment until it was too late. Because individuals do not know that they do not know, it has been difficult to educate them to the need. By focusing on concrete results, organizations have better success making the case that an investment in DG will benefit the organization in specific measurable ways.

Who needs to be involved in DG?

Unfortunately at many organizations, **everyone** has been responsible for data quality and this approach has produced the current unsatisfactory state. It is critical to start DG educational efforts with executives because 1) they are willing to invest in learning and 2) their data decisions have the greatest impact on the organizational data practices. The next goal for all DG programs is to also increase the data literacy of all organizational knowledge workers.



As illustrated, DG efforts are generally built on an IT provided support/foundation/infrastructure. A leadership component provides resources and clears barriers for the effort. Primary functions are (ideally full-time) data stewards who provide guidance and design/implement decisions. Typically these two groups form the basis for DG organizations. Also, highly involved (and incorporated) are various SME or subject matter experts who know the required data and processing details. Then of course there is everyone else. As noted, DG efforts need to be integrated with both organizational and IT governance.

When is it appropriate for organizations to invest in DG?

By now I hope that you agree this is a silly question. The 20-40% of IT costs (referenced previously) is easily gauged. As the maturity of the DG practice matures, processes can be optimized for key operations. By keeping disciplined measures, organizations have developed expertise in these practices. Keeping the focus on a integrated full time team permits the case to more easily be made when timing investment in a second or third DG team.

Digital and data are dependent on high speed automation/data processing that requires significant amounts of organizational data literacy, data standards use and quality data supplies. Continue to evaluate and evolve DG frameworks to refine the organizational focus. Over time this approach should evolve into the standard Deming plan, do, check, act (PDCA) cycle.[3] An incomplete list of potentially useful standards that can be created with the required measurable controls is listed below.

- Access standards
- Change management
- Security
- Storage
- Reporting
- Classifications
  - ➡ Secure
  - ➡ PII
  - ➡ Competitive advantaged
  - ➡ Public

Where should organizations get started with DG?

DG is a rare triple benefit capability that helps refine data strategy, improve the quality of the players, and improves data used to support the mission. However, getting started with DG can be and has been accomplished by a moras of ill-defined and vendor specific methodologies–most of which have no reported research results.

An easily understood model (the theory of constraints[4] or TOC) views programmatic data support as a manageable system. The system is limited in achieving more of its goals by a small number of constraints. There is always at least one constraint, and TOC uses a focusing process to identify the greatest constraint and restructure the rest of the organization to address it. TOC adopts the idiom that "a chain is no stronger than its weakest link," and processes, organizations, etc., are vulnerable because the weakest component can damage or break them and adversely affect the outcome.

Key is to visualize the various data flows through the organization and understand the value of controls in relation to various processes, risks, outcomes, and performance. The costs of various blockages can be ranked and estimated. What changes made at the data level could most help the organization achieve its strategic goals?

Iterative problem solving provide additional benefits beyond challenge solutions. Team problem solving enables increased organizational data literacy and some go as

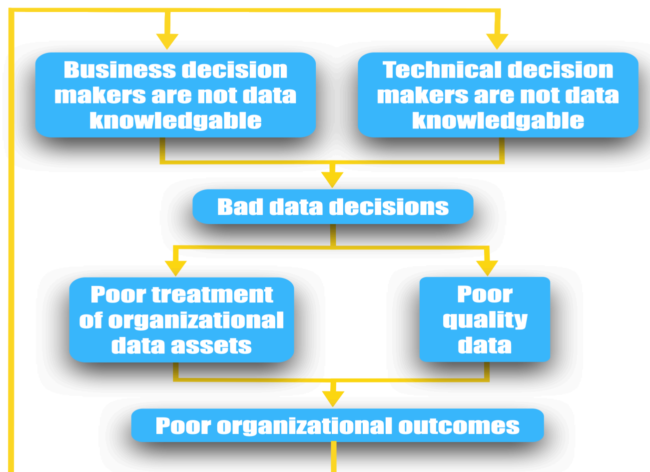
[3] https://en.wikipedia.org/wiki/W._Edwards_Deming#PDCA_myth

[4] https://en.wikipedia.org/wiki/Theory_of_constraints

far as considering these capabilities their 'secret sauce.' It just makes sense to support a group of individuals who possess knowledge of your data and its uses.

Focus first on organizational strategy. Understanding intricately, the data flow supporting increasing performance, decreasing costs, impacting times, and better managing risks. Identify the various types of organizational challenges sharing the same data or (better still) data errors. These become the focus of the first iteration of a data strategy cycle. It is overseen by the DG program and coordinated to be most collectively helpful to organizational as well as IT strategy. Ensure you complete a full cycle to include feedback/improvement/lessons learned/organizational memory/change cycle components. Heavily incorporate the use of 'branded' data checklists and standard control development.

And then (as it says on the shower bottle) lather, rinse and repeat. This is really to only way to escape the bad data cycle. IT and business decision makers are not knowledgeable about data and good data practices. They make poor decisions about data that result in poor treatment of organizational data assets and poor quality data. Both of these lead to poor organizational outcomes.



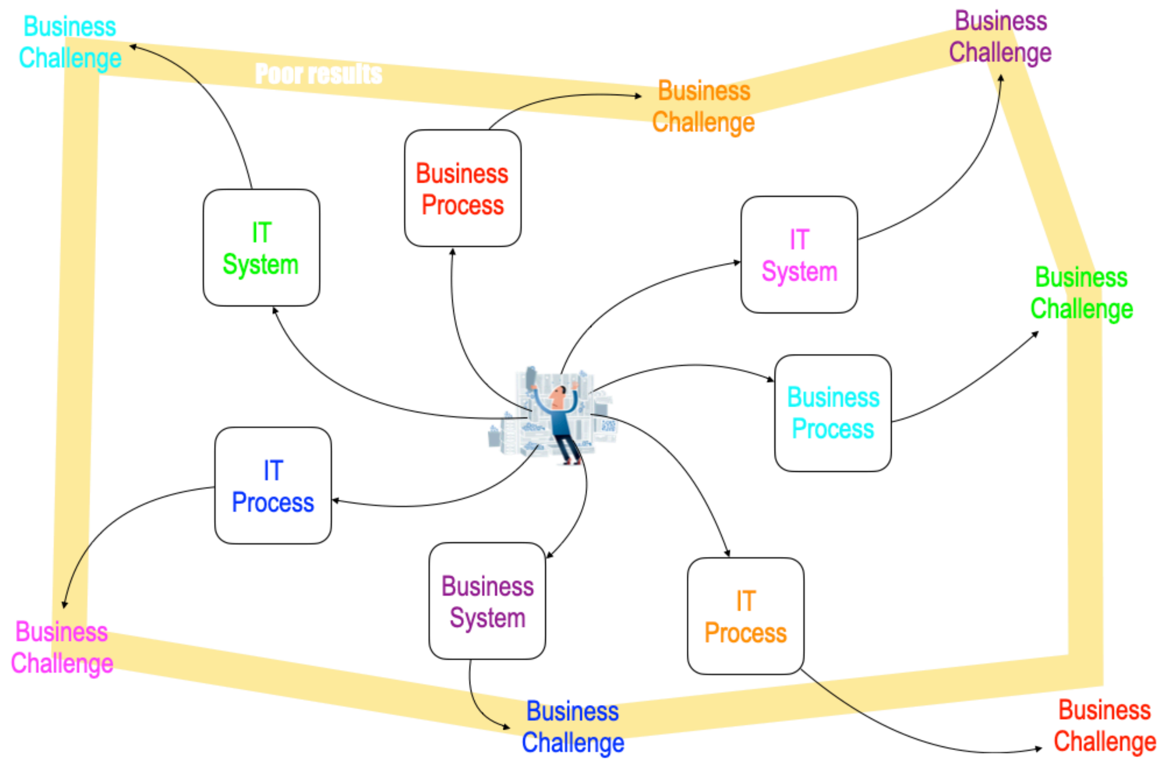
How should organizations apportion their DG efforts over time?

Data debt's impact

Over time, organizational data debt clogs value-adding pathways in a manner similar to the 40% of the internet that is now clogged with malware. Data debt is responsible for inflicting uncounted tiny hidden data factories[5] on organizational performance-making everything cost more, take longer, deliver less, and at increased risk. Eliminating data debt requires a team with specialized skills deployed to create a repeatable process and develop sustained organizational skillsets.

A major motivation for increasing the data literacy of all knowledge workers comes from the fact that most organizational challenges come filtered through various IT and business practice combinations. The reason for multitude of paper cuts, is that the DG challenges are filtered through various business processes and IT systems. As a result, common challenges go unrecognized with each instance requiring treatment instead of correcting the underlying data challenge.

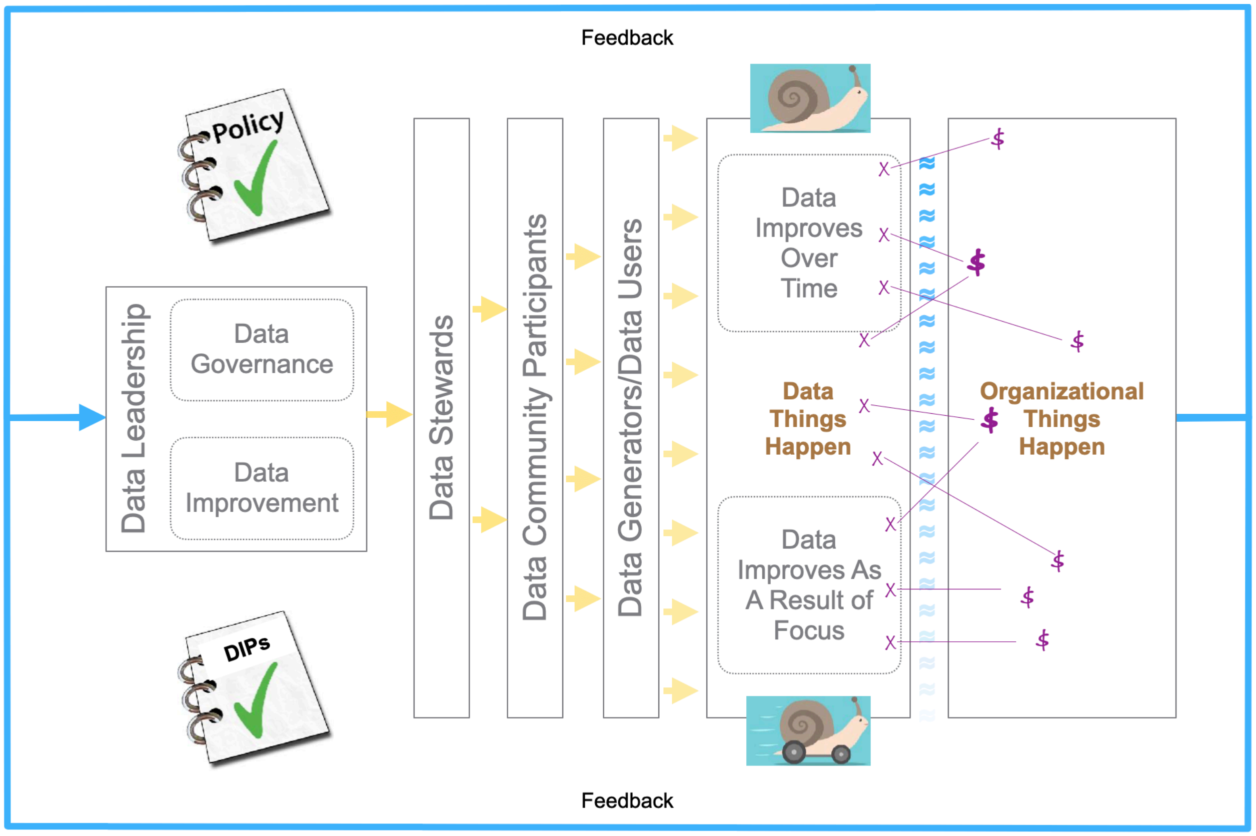
5 https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year



A key aspect is to evaluate your architectural abilities to build/evolve towards organizational data capabilities in a 3-step process. First, you need to improve the quality of existing organizational data. Too many organizations do not have enough information about the quality of their existing data. These data quality challenges fall into two categories: practice-related data quality challenges and structure-related data quality challenges. Second, the framework must support your efforts to increase the data literacy of literally your entire executive team and knowledge worker population and especially those who already practice data. Finally, only when you have improved your data and your organization's ability to work with data, can you hope to improve the way that data supports your organizational strategy.

Proactive versus reactive DG

One rather traditional realization (almost a rite of passage) is that what ever changes are made to the organizational data practices might take literally years to be able to exploit it. In CIO terms, it can often be a successors, successors, successor that will benefit from DG initiatives. As this realization sets in (that time equals years), DG initiatives come under pressure to 'do something more quickly.' As illustrated, a secondary capability is established to more effectively produce results as a result of direct intervention or Data Improvement Projects (DIPs).



MacGyver abilities

While perhaps not widely acclaimed, the 1980's TV series MacGyver became shorthand for an non-traditional and innovative problem solver who always carried a Swiss Army knife.[6] In the same manner, the DG program must imagine itself as the 'help desk' for organizational data. Literally all data challenge solutions should be minimally coordinated and, in many instances, led by DG. The key is to develop new data capabilities within a dedicated group focused on organizational data governance. Have this group focus on and conquer a series of DG challenges, producing positive ROI numbers.

What organizational needs does DG fill?

It is useful to describe the organizational needs that DG fills. These include:

- Improving the way that data is treated as an asset
- Available but not widely known research results
- Using data to better to support the organizational mission
- Using data strategically

[6] https://en.wikipedia.org/wiki/MacGyver

Improving the ways that data is treated as an asset?

One of the primary challenges for organizations is to learn how data requires specific considerations. If you consider data as an asset (and currently most business leaders do not yet do so) then one should expect that it would be treated as other organizational assets. I use a series of questions developed by my colleague Dr. Christopher Bradley to help organizations determine whether their data is maintained as an asset. They are:

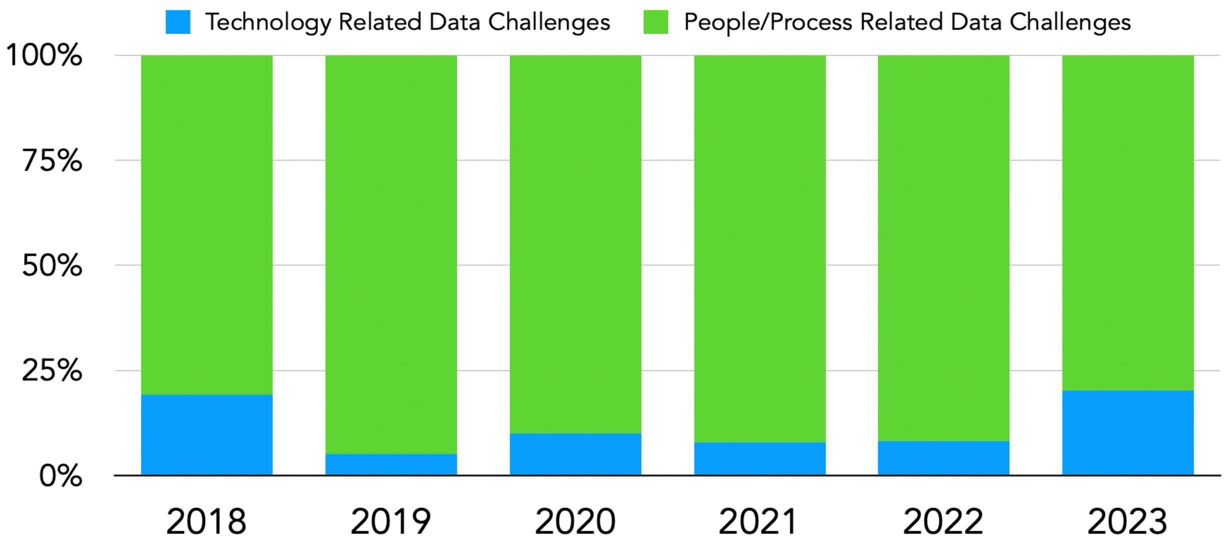
1. Do you have executive positions to support data as an asset?
2. Does the organization track usage of this asset?
3. Are organizational or fiscal controls in place to manage this asset?
4. By and large, are these controls actually executed?
5. Is there general acceptance of the need to manage this asset? (i.e., do people "get it"?)
6. Do serious discussions about this asset feature on the agenda of senior management meetings?

Using this rather obvious set of criteria, it is easy to determine that most organizations are not treating data as an asset but so far we do not survey results on this particular measurement.

Available but not widely known research results

As referenced above, there is a dearth of knowledge about data much less data governance. On that note however, we do have access to two solid lines of research to which I will refer through out this chapter. The first is in the form of the annual (2013-today) data practices surveys conducted by New Vantage Partnerships and are reference able at: https://www.newvantage.com/thoughtleadership. Annually several thousand of the same or similar organizations have been asked the same questions repeatedly providing pictures of how issues are considered over time. Results reproduced here will be referred to as **New Vantage**. A second set of research results come from the collaboration (called the Data Literacy Project) between Accenture and Clique. These results will be referenced as **Data Literacy Project** and are reference-able at: https://thedataliteracyproject.org/. These two efforts have provided a good framework that can be used to dive further into research in this area.

One of the New Vantage results has been: *what percentage of your data challenges are people/process related versus technology challenges?* The consistent answer (see figure) continues to surprise: not once since 2018, has the percentage of technology challenges risen to above 20%. This means that for more than six years, **everyone** should have known that the people/process dimension of DG represents the largest challenge. Yet very little organized research beyond surveys has been conducted into this area.



Consider please, what group in your organization is charged with decreasing the number and impact of people and process-oriented data challenges? This is precisely the role that your DG organization must address in your organization. **If not DG then whom** in your organization is responsible for improving the people and process aspects of your data operations?

It is crucial that DGs provide a wholistic view of minimally the above detail but also include data's role in the organization, how individuals can assist, and where to go for more information.

Using data to better to support the organizational mission

This section's title: **using data to better to support the organizational mission** must be the mission of any DG program. But first a specific word about data ownership (bad concept) and Data requirements ownership (good concept).

Avoid a first (and always a major) misstep: trying to assign data 'ownership.' While it is tempting to "establish data owners" as a goal of data governance–it is usually a bad idea. However, many are familiar with the process architecture practice. It correctly embraces and leverages the term "process owner" as the single individual responsible for the integrity of the process design, implementation, and improvement.

While it makes intuitive sense, the concept of data ownership has caused more DG effort to fail than any other. As soon as you allow an under-informed individual (or group) to 'own' any data items, they begin to make decisions about the data that optimize it from their local perspective. If your organization does not formally manage a process architect, skip to the next paragraph. If it does, careful analysis will yield maintainable, high level process/data interaction matrix called a CRUD matrix–showing data/process interaction by access type. (CRUD matrices such as the one illustrated show business processes and their activity type **C**reating, **R**eading, **U**pdating, and **D**eleting various data items–example also courtesy of Dr. Christopher Bradley).

Business Processes

Major Entities / Data Subject Areas	Product development	Marketing & Sales	Industrial preparation	Order management	Manufacturing	Logistics	Invoicing	
Product	C	R	U	U	U			
Product Part	C	R	R	U	U			
Manufacturing Plant	U		C	R	R	U		
Customer	R	C		U	R	U	U	
Sales Item	C	C	C	U		U	U	1
Assembly Structure	U		C		U			
Sales Order		U		R	U	U	U	2
Production Order			U	C	U	U	U	
Individual Product					C	R	U	
Shipping						C		3
Customer's Invoice		U					C	

If nothing else, these maintainable metadata collections show the interdependencies: data exist only to be consumed by various business processes and only purpose for a business process to exist is to produce data to be consumed by another business process. If you do not have an organization CRUD matrix hand and need to shut down any data ownership conversations, ask the question: "*To whom does the data that accounting stewards belong?*" Since accounting processes data from across the organization, a case could be made that accounting 'owns' much organizational data.

The reason data ownership is such problematic concept is that data persists across business functions. Ownership would only apply to a specific data processing stage. Instead of asking the question, "who are the data owners?" the statement should be that all data belongs to the organization! At best, ownership could only be limited to specific lifecycle phases.

If the organizational culture requires use of the word ownership, then allow ownership of the **data requirements**! Local expertise should be used to specify the size and shape of the specific data items required to perform organizational functions at various stages of data at it is processed.

The role of DG frameworks

All evidence to-date points to frameworks has being useful as:

- System of ideas for guiding subsequent analyses

- Means of organizing measures, project data, and then assessing progress
- Evaluating priorities for data decision making
- Assessing overall functionality
- Moving towards a determination of ROI[7]

For example, a building construction conceptual framework would incorporate bits of wisdom such as:

- Don't put up walls until foundation inspection is passed
- Put the roof on ASAP so that work can proceed in inclement weather
- Make it each construction phase dependent upon continued funding by passing a series of checkpoints

Much has been written about data governance frameworks. I have seen research proposals that anticipate evaluating one type of framework against another. It is far to early to start to 'type' DG frameworks. Non-standard understanding of terms and data concepts lead to 'results' of the sort that were popular at the start of the CDO movement. (Note: researchers have tried and failed to establish correlations between having a CDO and organizational financial performance–similar specious results can be expected until the entire DG profession matures.)

Use the existing DG frameworks to envision what your program should look like given your organizational needs. 'Try each of them on' conceptually and discuss the suitability of each for your organization. Since no two organizations are alike, each organizational DG program must be custom fitted to the organization rather like getting fitted for a suit. The word "bespoke" well describes the design of DG programs that provide good returns on organizational DG investments.

It is quite useful to view representations of various approaches to DG in the same manner that an architect presents sketches of a future building to prospective funders. The utility of DG frameworks generally stops at this point. There are essentially few types of DG frameworks in popular use. (Note: you can see representations of many these at: https://anythingawesome.com/DataGovernanceFrameworksCollection.html) All subsequent are theme and variations on these. Pay no attention to 'proprietary' methods. The goal is to give you something to compare, contrast and consider when designing the first version of your DG organization. (Note: This first version will evolve to a second and third as the organization; DG practices should mature and evolve over time.)

This is where the concepts of stewardship and fiduciary responsibilities come into play. **Stewardship** in this concept is derived from the definition: a person employed to manage another's property. **Fiduciary** is used to describe the nature of the relationship

[7] Interestingly, ROI means risk of incarceration to most DG professionals.

as involving trust, especially with regard to the relationship between a trustee and a beneficiary. This is accompanied by specific duties.

Related Term Definitions

It is now time to introduce a few terms to show both the evolution/etymology of the term DG and the most useful definition of DG.

Let's start with the term **governance**: "Governance is the process of interactions through the laws, norms, power or language of an organized society over a social system (family, tribe, formal or informal organization, a territory or across territories). It is done by the government of a state, by a market, or by a network. It is the decision-making among the actors involved in a collective problem that leads to the creation, reinforcement, or reproduction of social norms and institutions" (https://en.wikipedia.org/wiki/Governance)

Corporate governance is next. Below are three good definitions highlighting different aspects of this evolving concept.

- "Corporate governance - can be defined narrowly as the relationship of a company to its shareholders or, more broadly, as its relationship to society….", Financial Times, 1997.
- "Corporate governance deals with the ways in which suppliers of finance to corporations assure themselves of getting a return on their investment", The Journal of Finance, Shleifer and Vishnu, 1997.
- "Corporate governance is about promoting corporate fairness, transparency and accountability" James Wolfensohn, World Bank, President Financial Times, June 1999.

Note that the concept of corporate governance is evolving. Just before the pandemic, Jamie Dimon (then head of Chase), led a group of CEOs to proclaim "Maximizing shareholder value can no longer be a company's main purpose."[8] Similarly, the concept of DG continue to evolve.

Well if corporate governance exists then certainly **IT governance** should be a useful concept? It is and is defined as "Putting structure around how organizations align IT strategy with business strategy, ensuring that companies stay on track to achieve their strategies and goals, and implementing good ways to measure IT's performance. It makes sure that all stakeholders' interests are taken into account and that processes provide measurable results. (https://en.wikipedia.org/wiki/Corporate_governance_of_information_technology)

IT governance frameworks should answer some key questions, such as how the IT department is functioning overall, what key metrics management needs and what return IT is giving back to the business from the investment it's making. Included are typically foci on:

- Strategic Alignment
- Value Delivery

[8] https://www.marketwatch.com/story/maximizing-shareholder-value-can-no-longer-be-a-companys-main-purpose-business-roundtable-2019-08-19

- Resource Management
- Risk Management
- Performance Measures

IT governance is an established discipline with common vocabulary and understanding among those who participate.[9] Of note is the fact that data practices are not typically included as a topic under IT governance or are lightly treated. This may account for or reflect the current slowly maturing state of DG practices.

Data governance has suffered from both too many definitions and inaccessible (by the business) terminology. However, auditors easily get the concepts. Below are some standard definitions of DG.

- The formal orchestration of people, process, and technology to enable an organization to leverage data as an enterprise asset – The MDM Institute
- A convergence of data quality, data management, business process management, and risk management surrounding the handling of data in an organization – Wikipedia
- A system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods – Data Governance Institute
- The execution and enforcement of authority over the management of data assets and the performance of data functions – KiK Consulting
- A quality control discipline for assessing, managing, using, improving, monitoring, maintaining, and protecting organizational information – IBM Data Governance Council
- Data governance is the formulation of policy to optimize, secure, and leverage information as an enterprise asset by aligning the objectives of multiple functions – Sunil Soares
- The exercise of authority and control over the management of data assets – DM BoK

Technically they are all correct but imagine the following scenario. Stepping onto an elevator for a minute-long ride and an executive enters the car. As the doors close the executive turns and says, "I've heard you are working on DG. Can you tell me what it is - I'm confused?" Imagine responding with "DG is the exercise of authority and control over the management of data assets." Do you think the executive would 1) find the answer useful and 2) think well of your ability to communicate this concept?

I think the answer no to both questions. A better response to the executive is: "**DG is about managing data with guidance.**" Short and to the point, this definition incorporates self explanatory motivation. When I provide this information (the definition of DG) to most executives, their first question to me is: "So we have not been managing our data with guidance?" The answer usually is: "Only recently have we been managing our data with guidance." Of course the eternal hope is that the executive will be curious to learn more and present an opportunity to become more data literate. Subsequent conversation topics could include:

[9] https://en.wikipedia.org/wiki/Corporate_governance_of_information_technology

- Why is it generally not a good idea to govern all of your data.
- Why DG will never be complete at our organization.
- Why some decisions that involve data are not considered as such.

The Data Literacy Project reports that four out of five executives surveyed were willing to invest time resources in improving data skillsets. This represents a once in a generation opportunity to reach these executives with good DG education. (Note that anyone offering to improve your organization with DG **training** should be ignored–the process requires **education**, not training.)

A Small concentrated team is preferred over distributed (dissipated) knowledge

The next item to consider is what format DG should take. Remember, asking **everyone** to be responsible for (data, data quality, data governance …) has produced the current state of affairs. Organizations assigning new DG duties to existing personnel have two options: 1) incorporate the new duties along with existing duties or 2) assign these DG duties to full time individuals.

When considering this, it is useful to ask: how long will the need to manage data with guidance exist? The answer turns out the be: **you will need your data program as long as your organization needs to have its finance, HR, and planning operations.** Think about it-in the future: Will more or less data exist? Will data collection modes increase or decrease? Will data be found in fewer or more formats? A solid recommendation is to staff with full-time team members dedicated fully to DG. Data literacy and organizational data practice maturity are generally low. Dedicated personnel with interact with each other more–greatly stimulating their individual learning curves. It also makes tracking DG program costs clearer. It is critical to begin to build organizational DG capabilities. This can best be started with dedicated teams with a clear ROI. Against these, results can be evaluated.

Using data strategically

The next question is **on what** do we focus these DG efforts? In regulated environments, these efforts are often compliance driven. Key is to approach these efforts in the same manner. Do we think that regulations will increase or decrease in the future? If increasing, then it seems useful to 'get good' at implementing compliance driven changes. If nothing else, you may gain an implementation advantage over the competition subject to the same data regulations but perhaps not able to implement as efficiently or effectively. Data regulation compliance can become an valued organizational capability with an easily determined ROI.

Outside of compliance, organizations strive to use data strategically with either efficiency/effectiveness or innovation goals. Personal interaction with more than 1,000 organizations indicates that about ½ have clearly articulated strategic goals and objective measures supporting goal achievement at the organizational level. Absent these, it is not possible to improve the manner in which data supports this Jell-O strategy. I also find universal distain for 3-5 year plans, most of which fell apart rapidly with the onset of the Covid-19 pandemic. So just a word of caution, check your organizational strategy to ensure it has clear objective and measures before attempting to improve how data can support it.

Strategy is About Why

*…it's not **what** you do, it's **why** you do it…*

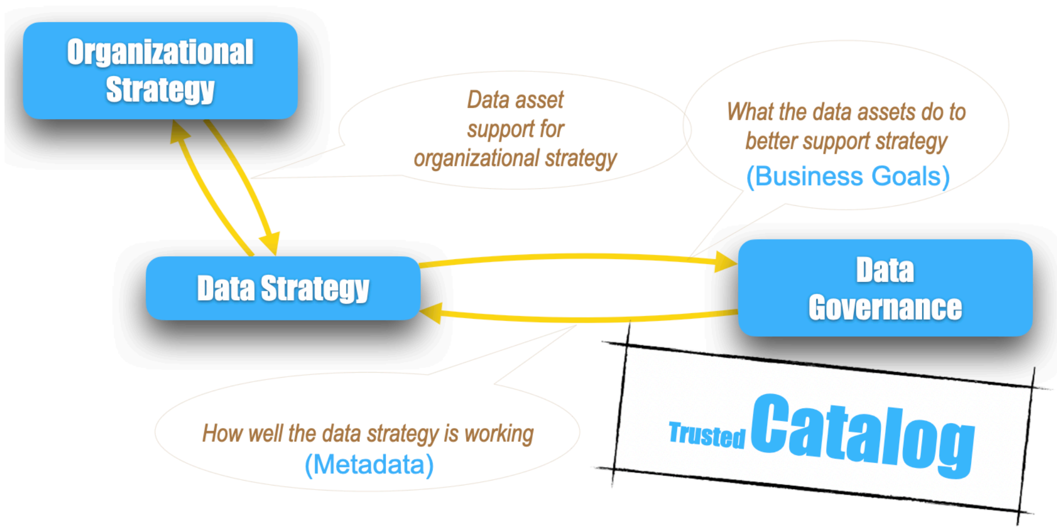
Among many great TED Talks, Simon Sinek's "How Great Leaders Inspire Action" is a favorite. Recorded in 2009, Sinek's talk has enjoyed more than twenty-five million views. His point is quite simple: most of us are very good at describing what we do, and some of us are good at describing how we do things. Not as many of us are good at describing why we do things.

Strategy is the highest-level guidance available to an organization, focusing activities on articulated goal achievement and providing direction and specific guidance when faced with a stream of decisions or uncertainties. More succinctly, strategy is a pattern in a stream of decisions. This pattern must be supported by data or it will not be possible to determine if the strategy is correct or working.

What is data strategy?

Data strategy is the highest level guidance available to an organization, focusing data-related activities on articulated data program goal achievements and providing directional and specific guidance when faced with a stream of decisions or uncertainties about organizational data assets and their application toward business objectives. The data strategy must be understood and supported at the organizational level. Only with this level of scrutiny and involvement can a true systems view be applied to the challenge of improving how data can support strategy.

WORKING TOGETHER: DATA AND ORGANIZATIONAL STRATEGY?



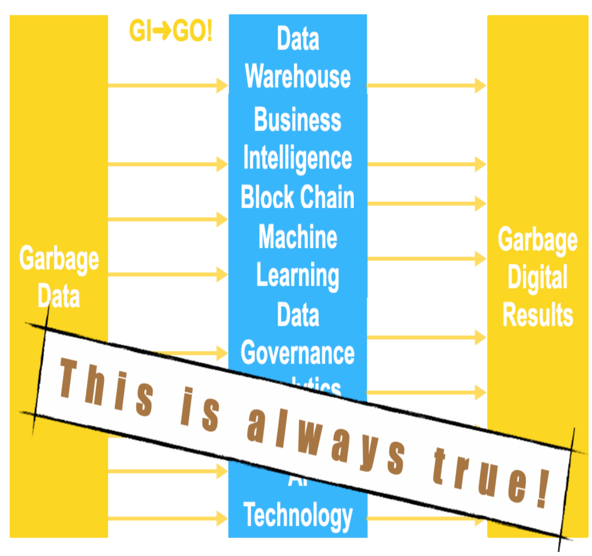
The figure indicate the close relationship among organizational strategy, data strategy, and data governance. Two key aspects of the interaction are: 1) express the data strategy in terms of specific business goals, and 2) ensure that the language of DG is metadata.

STRATEGIC COMMITTMENT: PROGRAM VERSUS PROJECT FOCUS

A commonly asked question is: when will you be done? This is a warning that the individual considers DG a **project**. Organizations failing to implement DG at the program level (as a program) are unable to view the totality of their data challenges holistically and the solutions fail. Many organizations require a second or increasingly a third DG 'reset.'

DIGITIZATIONING

One of the more important areas that DG can be focused to support is 'going digital.' Once again, many vendors have offerings and expertise in these areas. DG sets the standards required to support digitization because, you cannot 'digitize' without a good data capabilities foundation. Garbage in, garbage out is always true. At this point, effective DG is a requirement for digitization, otherwise you can will be unable to trust any digital system outputs.



A WATCHFUL EYE TOWARD THE US FEDERAL GOVERNMENT (FEPA)

Finally on the **what** question (yes–we are still in **what**), it will be useful to observe the progress being made in the US Federal Government. As part of my service as a DoD employee, our group was often sent to 'learn from the private sector.' Now the situation has been reversed. In 2019 the **F**oundations for **E**vidence-Based **P**olicymaking **A**ct was signed into law. Three specific aspects of the law make this especially interesting to aspect of DG to follow. They are:

- Explicitly non-political CDOs must be established separate from CIO roles. From a DG perspective, organizations have been slower to adopt CDOs with non-CIO reporting role.
- Government data is now open by default, and must maintained using open standards. In just a few years, the Federal agencies will have developed a great deal of expertise in these areas.
- Use of open data and open models is required in policy evolution. Policy changes are only permitted with both models and datasets specified prior to the analyses and decisions.

Collectively these efforts, if fully implemented, will improve governmental decision-making and overall effectiveness. More importantly, all impacted Federal organizations are also rapidly, developing and implementing DG as compliance activities still further increasing the pool of DG professionals world-wide.

Breaking through the Barriers of Data Governance

There are a host of barriers to implementing DG. This include the usual failures to include change management and cultural refocusing as key dependencies. While the accounting profession has had literally millennia to develop GAAP, no such guidance exists for DG. There is a vast tendency to depend on technologies that are incapable of acting as silver bullets.

An example of these difficulties was illustrated in 2020 when Forbes ran an article on airline valuations.[10] It purported to show how the airlines were monetizing the data in their frequent flyer programs. However, the buried lede was that in 2020, both United and American Airlines were valued at 10s of billions of dollars less than the anticipated value of the data in these programs. You had better believe that if airline leadership could have unlocked that value during the time most were avoiding flying (the pandemic), they would have unlocked it asap! The fact that they were unable to do so highlights the uphill climb that poorly fitting DG efforts face.

Some basic DG execution principles follow:

[10] https://www.forbes.com/sites/advisor/2020/07/15/how-airlines-make-billions-from-monetizing-frequent-flyer-programs/?sh=66da87a614e9

- Ensure that the organization's data strategy in properly aligned with the business strategy. Implement a regular processes with key stakeholders to ensure proper alignment.
- Ensure that data debt is properly being managed and the process is under statistical control.
- Perform a capability maturity assessment or 'reassessment' to determine the required maturity. If the maturity levels are not meeting expectations, ensure that there is a remediation plan with a properly monitored work arounds.
- Consider refresher training for your knowledge workers and data professionals, e.g., Data Stewards, Architects and Engineers as a feedback mechanism for determining needed improvements and remediations.

Based on the organization's strategy, the DG group must determine are they to initially follow a model primarily focused as a:

- Utility–back office, efficiency goal
- Steward–more asset focused, quality goal
- Enabler-strategic partner, innovation goal

This should be determined through the building of the data strategy. If an organization striving toward a modernization transformation, DG should trend to an 'enabler.' To measure the effectiveness of an enabler, DG standards should be repeatable and statistically stable. The focus can be changed at a later stage but can usefully focus effort and discussions during initial phases.

Hopefully your organization will be spared major data catastrophes. It is more likely you will experience one or more in the future. In this event, attempt to learn as much as possible from the event. Take for example, the story of two major banks in the process of consummating an arranged marriage. The deal came down to a single spreadsheet containing many rows, each representing an asset. If an asset on the spreadsheet was to not be transfered, that row was hidden with agreement by both parties. After final agreement was reached, the spreadsheet was handed to a junior associate who was told to 'make is look nice for the Judge tomorrow." Unfortunately late in the evening junior accidentally unhid hundreds of rows and did not notice! Presented to the Judge as the golden copy, the judge would not reverse-even on appeal.[11] As you might imagine, DG practices around the use of spreadsheets is quite extensive. I assisted one organization with the elimination of more than 400,000 legacy systems of a certain type. The list of preventable spending continues.

Unfortunately, the conversations have been generally unsatisfactory. Key to getting started with **data valuation** is to add up 'at least' instead of attempting to master the entire costs. I justified an investment into an organizational repository at one

[11] https://www.businessinsider.com/2008/10/barclays-excel-error-results-in-lehman-chaos

organization with a business case built on the premise of saving everyone in IT **1 hour annually**. The organization conducted surveys asking if the one hour savings was achieved. It was!

When determining the internal and external value of data two prerequisites exist: first, business and data strategies must support data monetization and second, DG must be effective and properly measured. Components of data value can include:

Internal

- Properly managed data debt
- Efficient usage of cataloging and master data management
- Hight trust in supplier and customer data integration
- Measured positive ROI

External

- Organizational data monetized in a public market or exchange
- Organizational data becomes a profit center
- Organizational data becomes a band-aid of adhesive strips

Sometimes it is easier to highlight the value with unfortunate examples with clear costs to society. Early COVID monitoring was inhibited because health care workers did not know to save MS Excel data sheet and workbooks as .xlsx instead of .xls files. The difference, unknown to the users, was that the older .xls files dropped all rows beyond the 16,000th or so row without warning. We will likely never know how much better performing the early monitoring systems were because all the errors are in one direction.

On a more cheery note, an agency charged with home evaluation/intervention, discovered that 40 questions on its evaluation assessment were immaterial. This shortened each interview by half and ultimately shifted more than $1 million from overhead to service delivery.

In terms of execution, DG should be viewed as an iterative process that the organization is striving to **get better at!** Each cycle focuses on aspects of the various data challenges with a goal of eliminating or reducing the impact of a specific constraint. To understand the important of this shift in thinking about DG, consider the circumstances where a plan was the goal. It was former President and General Eisenhower who said:

> *"In preparing for battle I have always found that plans are useless, but planning is indispensable"*[12]

Mike Tyson's version is that *everyone has a plan until they get punched in the face*. A team knows how to react to unforeseen challenges and efficiently address the ones they have planned for. The PDCA cycle provides operational context.

[12] https://quoteinvestigator.com/2017/11/18/planning/

Chapter Summary

The word **bespoke** has evolved from a verb meaning 'to speak for something', to its contemporary usage as an adjective. Originally, the adjective bespoke described tailor-made suits and shoes. Later, it described anything commissioned to a particular specification. Wikipedia

The difference between data analysis capabilities and data requiring analysis is increasing. DG will continue as a maturing and growing field and can only be assisted by increased research into the various challenges outlined. Practice standardization and improvement are clearly the next steps on this industry's maturity curve. As a new discipline, DG works best directly addressing the manner in which data is used to support achievement of organizations strategy. There is no one best way and right now there isn't agreement on terminology, hence anything. Consequently, the only way to obtain a positive ROI on investments in DG is to ensure that **your** data is successfully leveraged using methods (**your** data strategy) that **your** knowledge workers and **your** executives understand.

The goal is to improve DG effectiveness and efficiencies (and the data itself) over time. The more data literate the organization, the easier the transformation. Perhaps now is more understood about the phrase quoted at the beginning of the chapter:

This database ain't big enough for the two of us

– Bumpersticker seen on an automobile in Texas

Acknowledgement

My colleague Rob Greaves made many helpful suggestions that were incorporated into this chapter.

I appreciate Dr. Christopher Bradley granting me use of the two concepts note herein.

CDO Roadmap 2025:

Is Your Data AI-Ready?

An AI-Preparedness Report

cdo CHIEF DATA OFFICER MAGAZINE | EY

CHIEF DATA OFFICER MAGAZINE

Contents

Executive Summary

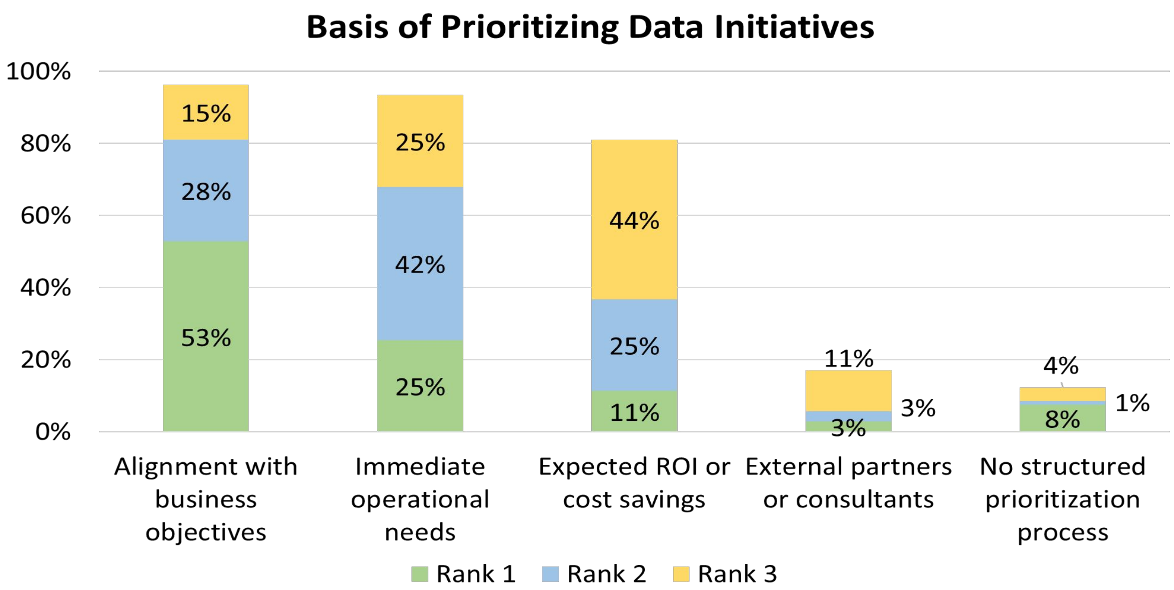
As organizations strive to unlock business value with AI, data readiness has become a critical success factor. This report examines the state of AI-driven data strategies, assessing key areas such as vision, integration challenges, ethical considerations, talent gaps, and infrastructure preparedness.

By identifying obstacles and best practices, it provides a strategic roadmap to enhance AI readiness and maximize data investment returns. The following executive summary highlights key insights from the study:

- **Prioritizing data initiatives: 53%** of organizations rank alignment with business objectives as the top factor in prioritizing data initiatives.
- **Ensuring data readiness for AI: 61%** of organizations are establishing clear data governance processes and policies to ensure data readiness for AI.
- **Obstacles to AI potential:** Poor quality data, the shortage of AI talent, and limited scalability of current technology are the top three obstacles to realizing AI's potential.
- **Challenges integrating data for AI: 39%** of participating data leaders consider data silos across departments as the greatest challenge in integrating data for AI initiatives.
- **Obstacles in using GenAI in business processes: 37%** of respondents rank data security and privacy concerns as the top obstacle to using GenAI in existing processes.
- **Bridging talent gap: 60%** of organizations are upskilling employees through training programs to bridge the data and AI talent gap.
- **Approach to AI ethical risks: 44%** of organizations have a dedicated ethics committee or framework in place to address ethical concerns of AI deployment.
- **Measuring ROI: 34%** of organizations place the highest importance on efficiency improvements and cost savings as to measure ROI on data investments, while **25%** place it on revenue growth.

Insights

1. Business objectives and immediate needs drive data initiatives

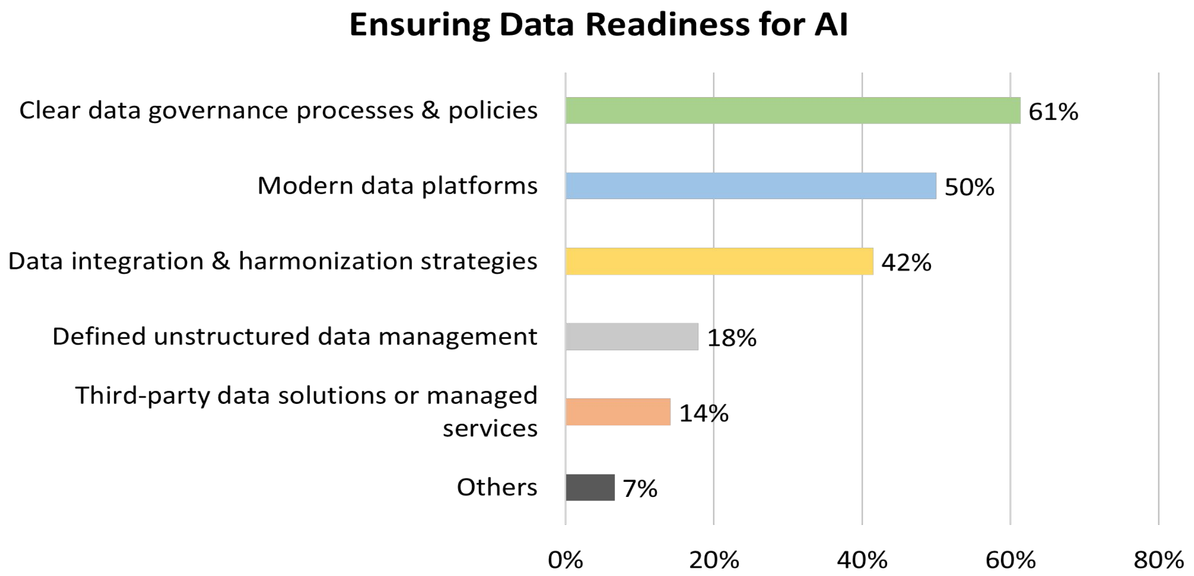


Alignment with business objectives, immediate operational requirements, and anticipated ROI/cost savings are the three primary factors influencing the prioritization of data initiatives.

96% of data leaders rank aligning data initiatives with business objectives among their top three factors. Similarly, **92%** of respondents place immediate operational needs and **80%** consider expected ROI or cost savings within their top three priorities.

Data initiatives are prioritized by overarching business objectives, while also being driven by immediate operational needs and financial gains.

2. Organizations are taking a holistic approach to data readiness for AI



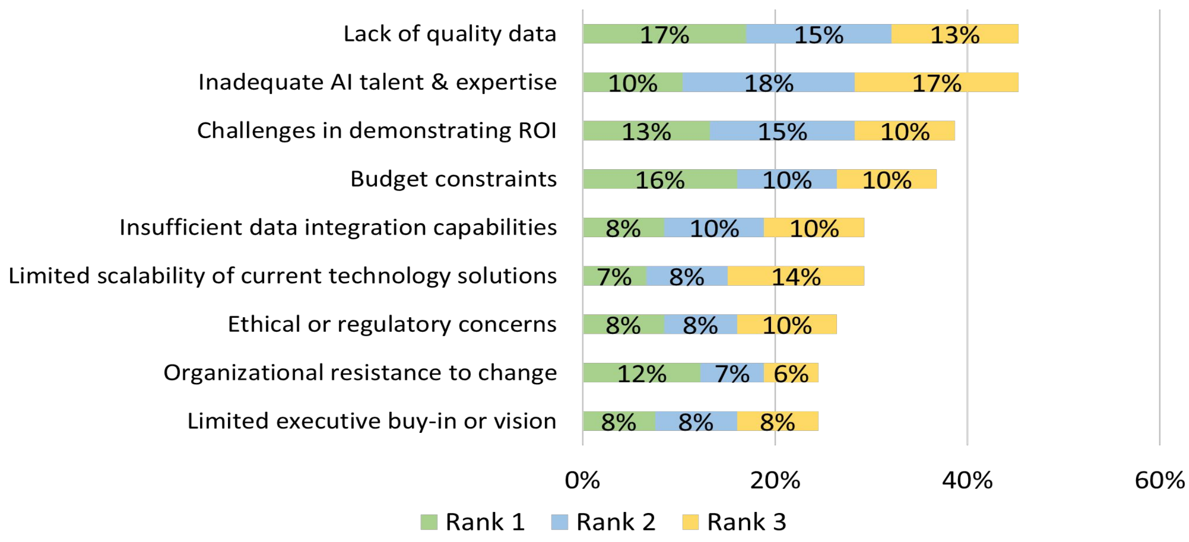
Organizations are actively pursuing multiple initiatives to ensure data readiness for AI. Among their top three approaches, **61%** of respondents are implementing data governance processes and policies, **50%** are modernizing their data platforms, and **42%** are emphasizing data integration and harmonization strategies.

Data leaders also rely on data quality tools and managed services to enhance data preparation efforts.

Data leaders are focusing on defining data governance and policies to create a framework for the usage, integration, and management of data in preparation for AI.

3. Data quality issues and inadequate talent limit AI's potential

Obstacles in Realizing AI Potential

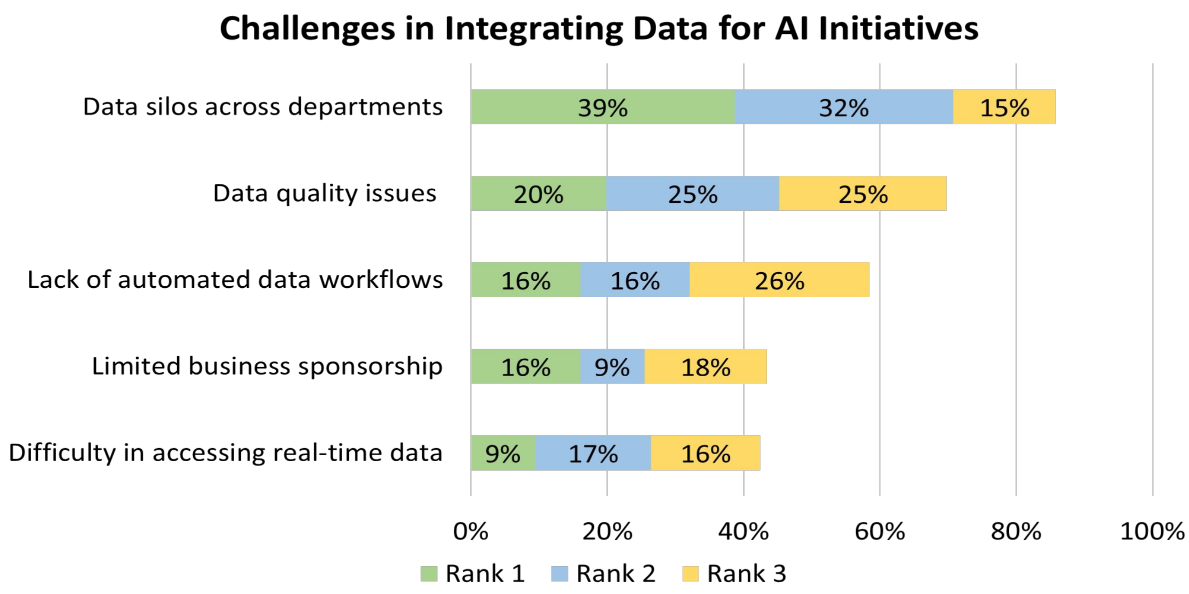


The primary obstacles to realizing AI potential in organizations are the lack of quality data and insufficient AI expertise. **45%** of data executives rank these issues among the top three hindrances.

Additional challenges impeding AI progress encompass difficulties in demonstrating ROI, budget constraints, insufficient data integration capabilities, limited scalability of current technological solutions, ethical or regulatory concerns, organizational resistance to change, and limited executive buy-in or vision.

Data leaders are facing significant challenges in unlocking AI's full potential due to obstacles such as data quality issues, shortage of skilled talent, and the limited scalability of existing technology.

4. Data silos hinder integration for AI

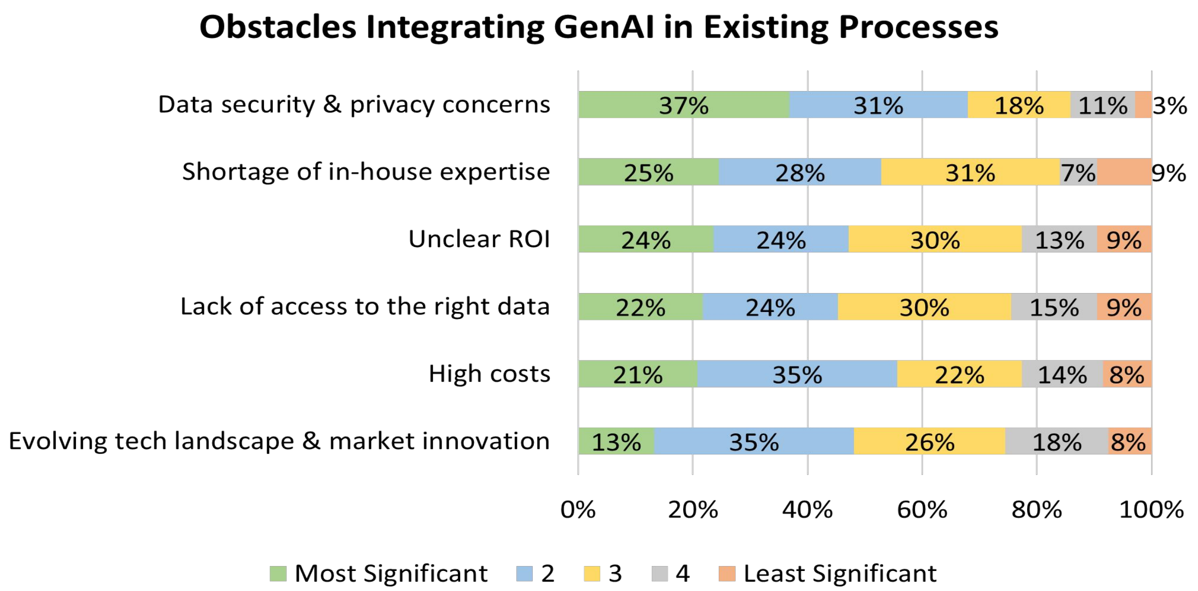


86% of respondents see data silos across departments as the top challenge in integrating data for AI initiatives. **39%** among them identify it as the primary challenge.

Other significant obstacles include data quality issues, lack of automated data workflows, limited business sponsorship, and difficulty in accessing real-time data.

Data silos across departments and the lack of automated data workflows obstruct data integration for AI initiatives.

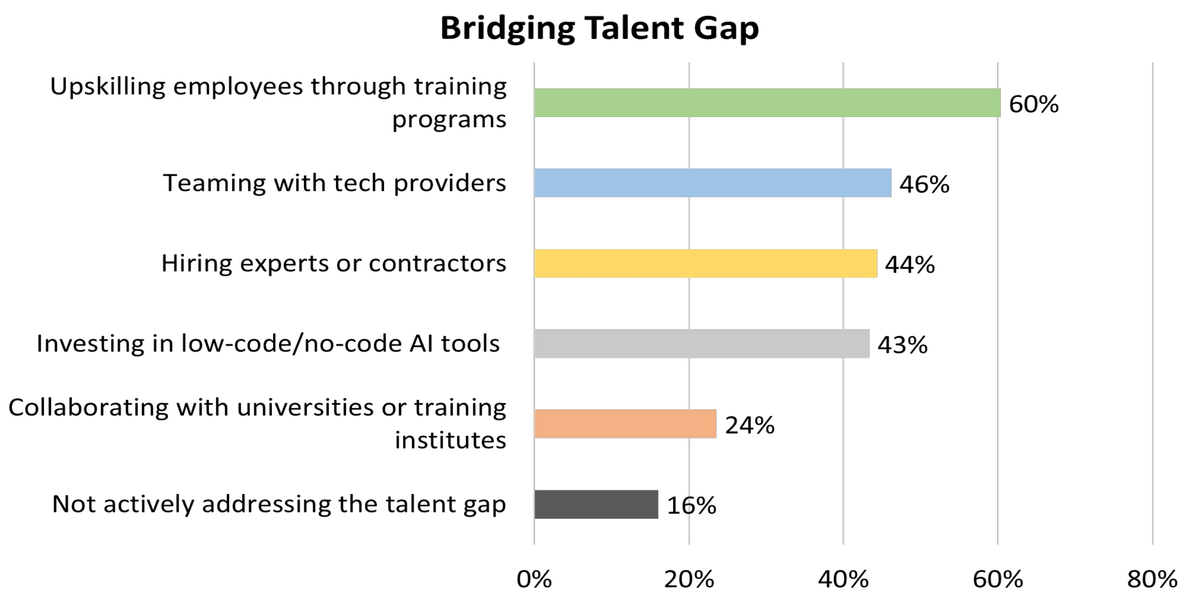
5. Data security & privacy remains the top obstacle to GenAI integration



The ranking of obstacles is evenly distributed for integrating GenAI into existing business processes. **37%** of survey participants cite data security and privacy concerns as the main challenge, while **25%** identify a shortage of in-house expertise as a barrier.

Integrating GenAI into organizations' existing business processes presents several significant challenges, including data security and privacy concerns, along with a shortage of in-house expertise.

6. Several measures are bridging the data and AI talent gap



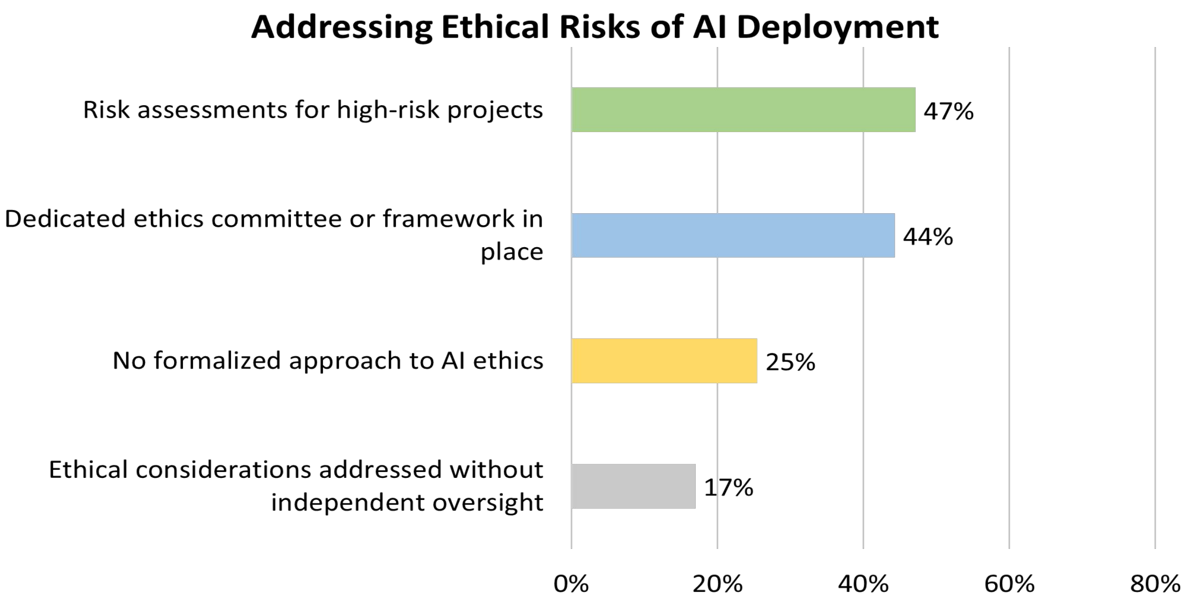
As highlighted earlier, almost half **(45%)** of the organizations identify inadequate AI talent and expertise as the second greatest obstacle in realizing AI potential.

To address this, **60%** of organizations are preferring to fundamentally change the AI talent scenario by running upskilling programs, **46%** are teaming up with tech providers, and **44%** are hiring experts or contractors.

Investing in low-code/no-code AI tools is one strategy **43%** of organizations are using to bridge the talent gap.

Organizations acknowledge the talent gap and are upskilling their teams in AI skills to address it fundamentally. They are also leveraging interim strategies such as engaging experts.

7. Ethical risk mitigation is central to AI deployment

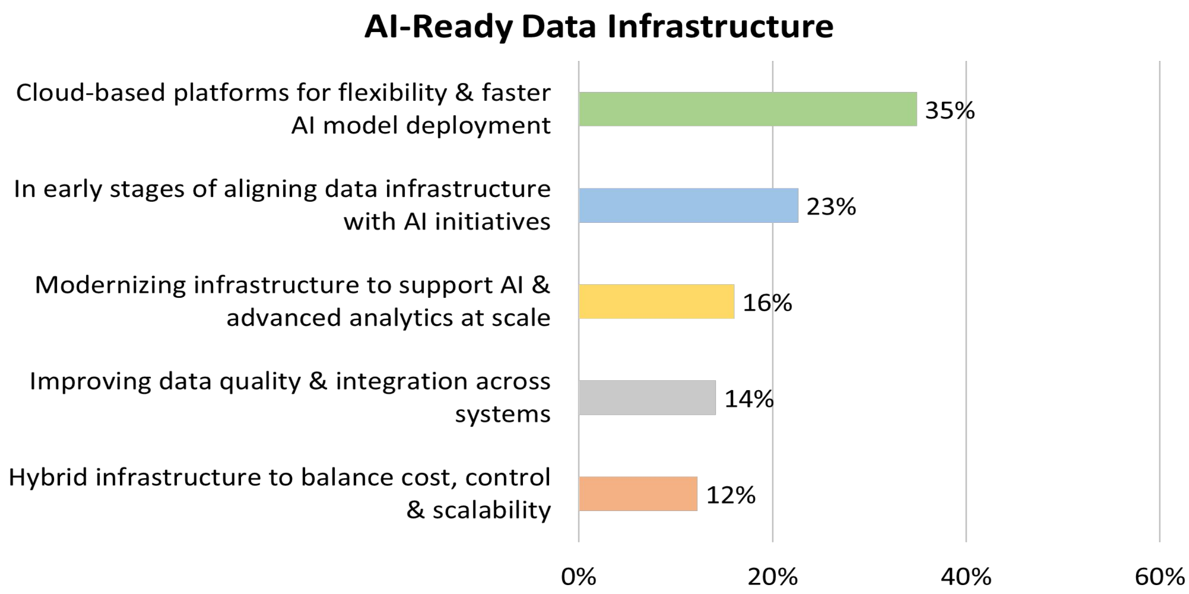


47% of data leaders do a risk assessment of high-risk projects as a measure to address ethical risks in AI deployment. At the same time, **44%** organizations already have a dedicated ethics committee or framework in place.

In contrast, a quarter **(25%)** lack a formalized approach to AI ethics, and **17%** address ethical considerations without independent oversight.

Most organizations have developed formal plans to address the ethical risks of AI deployment, yet many still lack a structured approach.

8. Cloud-based platforms lead the way in AI-ready data infrastructure



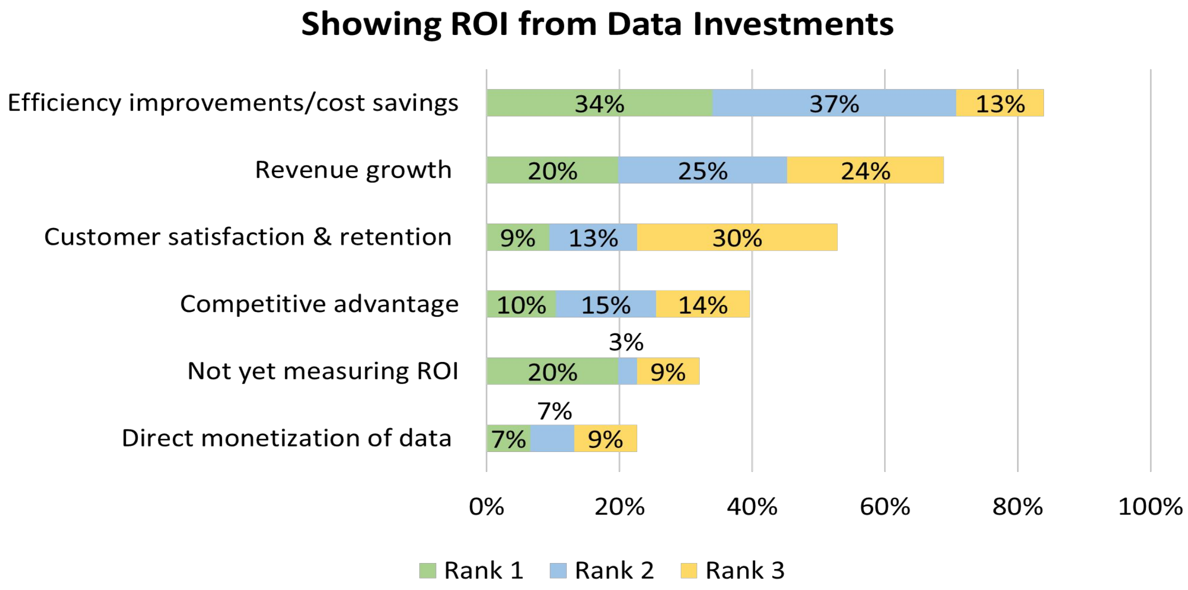
35% of organizations say their main investment in AI-ready data infrastructure is cloud-based platforms, taking advantage of their flexibility and quicker AI model deployment.

23% are still in the early stages of aligning their infrastructure with AI initiatives.

Meanwhile, **16%** are modernizing their systems to support AI and advanced analytics at scale, and **12%** are opting for hybrid infrastructure to balance cost, control, and scalability.

Cloud-based platforms emerge as the main preference for flexibility and faster AI model deployment.

9. Data investment ROI measured through efficiency/savings, growth, and customer satisfaction



84% of organizations primarily measure ROI by tracking efficiency improvements and cost savings. Revenue growth is the second most important metric, according to **69%** of data executives.

In addition to these core financial metrics, customer satisfaction and retention are also considered key indicators of the impact of data investments.

Organizations are assessing their ROI on data investments through both quantifiable and qualitative measures.

Methodology

Research objective

The research aimed to provide an in-depth analysis of organizations' data readiness for AI deployment. It focused on evaluating key aspects such as:

- Prioritizing data initiatives
- Obstacles in realizing AI potential
- Measures for ensuring data readiness
- Ethical risks of AI
- Data integration challenges
- Integrating GenAI
- Bridging the talent gap
- Data infrastructure for AI
- ROI of data investments

Survey design and administration

A questionnaire was used to gather insights from data executives on the key aspects outlined in the research objective, aiming to assess data readiness for AI and its potential to drive business value.

Data leaders were invited to participate through multiple CDO Magazine outreach channels. In total, 100+ respondents completed the survey in December 2024.

Strategic Insights From EY

Data plays a critical role in propelling success in today's AI-driven marketplace. This report offers critical insights to help Chief Data Officers (CDOs) harness the power of data to transform their organizations. CDOs who can effectively leverage the full value of their data are poised to align initiatives with core business objectives and drive meaningful outcomes that elevate organizational performance.

Here we highlight five key insights from the research findings that offer strategic guidance for CDOs seeking to unleash the power of data:

1. **Does data have a soul?** Leaders have pointed out that the lack of talent inhibits their ability to successfully deploy AI. In an era when talk about agentic AI is all the rage, this demonstrates that humans are still at the center. Accordingly, 60% of organizations are seeking to address the talent gap by investing in upskilling programs, while 44% have created dedicated ethics committees to manage the ethical risks associated with AI deployment.
2. **Challenges in AI integration:** Nearly four out of ten organizations (37%) report that security and privacy concerns remain significant barriers to adopting Generative AI (GenAI). To overcome this gap, 60% say they are exploring partnerships with tech providers to enhance AI model deployment and address security concerns.
3. **Align data initiatives with business objectives:** More than half of data leaders (53%) say that aligning data initiatives with overarching business goals and operational needs is a critical step for organizations seeking to improve responsiveness to market demands and foster a data-driven culture.
4. **Data governance and quality issues:** While 61% of organizations are establishing robust data governance processes, challenges such as poor data quality (reported by 31%) and data silos (reported by 39%) continue to hinder the effective integration of data across departments.
5. **ROI is not the only key driver for data initiatives:** Only 22% of respondents rank ROI as one of the top requirements for an AI integration. Instead, they are more concerned with focusing on business and operational goals, showcasing a shift in the CDO mindset from the "back room to the boardroom."

About CDO Magazine

CDO Magazine is the leading global publication amplifying the voices of senior executives in data, analytics, AI, and security. Our editorial content delivers boardroom-quality insights, equipping leaders with the strategies needed to drive enterprise-wide transformation in an increasingly digital world.

With a strong commitment to building a connected global community, we bring together data, analytics, AI, and security executives through regional and international forums, fostering collaboration and knowledge sharing at the highest levels. CDO Magazine also recognizes and celebrates the most innovative companies, technologies, and thought leaders shaping the future of data-driven business.

Join our growing community of global data and AI executives—**subscribe today** for exclusive insights and event invitations, and **contribute your opinion** to help shape the conversation at the forefront of enterprise data leadership.

Subscribe or learn more at **cdomagazine.tech.**

About EY

EY is building a better working world by creating new value for clients, people, society and the planet, while building trust in capital markets.

Enabled by data, AI and advanced technology, EY teams help clients shape the future with confidence and develop answers for the most pressing issues of today and tomorrow.

EY teams work across a full spectrum of services in assurance, consulting, tax, strategy and transactions. Fueled by sector insights, a globally connected, multidisciplinary network and diverse ecosystem partners, EY teams can provide services in more than 150 countries and territories.

Learn how EY is shaping the future with confidence at **EY Emerging Technology**.

Acknowledgements

Primary Contributors

- **Camille Prado**, CDO Magazine Global Editor
- **Abid Hussain Barlaskar**, CDO Magazine Managing Editor
- **Traci Gusher**, EY Americas AI and Data Leader
- **Zakir Hussain**, EY Americas Data Leader

© 2025 CDO Magazine and EY as co-authors. All rights reserved. Reproduction in whole or in part without written consent of the copyright owners is prohibited.