

# Your future in DS

**Jason G. Fleischer, Ph.D.**

**Asst. Teaching Professor**

**Department of Cognitive Science, UC San Diego**

**jfleischer@ucsd.edu**

 **@jasongfleischer**

**<https://jgfleischer.com>**

Slides in this presentation are from material kindly provided by  
Shannon Ellis and Brad Voytek

# Courses in DS and ML at UCSD

Did you know about the reciprocal agreement between DS and COGS??

- DS My list of '20-21 ML (and ML adjacent) courses
- CSE
- CS
- ECE
- COGS
- But also many other departments like ECON, MATH, LING, BENG, etc

# Some job titles and what they do

- Analytics or statistician: data handling, analysis
- Data scientist: programming, data handling, analysis
- Data engineer: programming, databases, management
- Data architect: programming, databases, design
- Data manager: databases, design, management
- \*OPs (eg, devOPs, dataOPs, full stack): programming, tool development, mangagement concentrating on end to end process
- ML Engineer: programming, tool development, management of infrastructure
- ML researcher: programming, algorithm design and testing

## New Roles and Job Responsibilities in 2024 from Burtchworks DS report

**AI Ethics Officer:** Develops, implements, and oversees ethical guidelines and practices for AI and machine learning systems. Ensures that AI applications are transparent, fair, and free from biases.

**Machine Learning Operations (MLOps) Engineer:** Focuses on deploying, monitoring, and maintaining machine learning models in production. Collaborates with data scientists and IT teams to automate and streamline ML workflows.

**Data Science Product Manager:** Manages the development and launch of data-driven products. Works with data scientists, engineers, and stakeholders to ensure product alignment with business goals and user needs.

**AI Research Scientist:** Conducts cutting-edge research in artificial intelligence, focusing on developing new algorithms, models, and techniques. Publishes findings in scientific journals and presents at conferences.

**Deep Learning Engineer:** Specializes in building and optimizing deep neural networks for tasks such as image recognition, natural language processing, and speech recognition. Works with large datasets and high-performance computing resources.

**Natural Language Processing (NLP) Engineer:** Develops and implements algorithms for understanding and generating human language. Works on applications like chatbots, sentiment analysis, and language translation.

**AI Solution Architect:** Designs and implements AI solutions to solve complex business problems. Works closely with stakeholders to understand requirements and develop scalable AI architectures.

**Data Privacy Officer:** Ensures that data handling practices comply with privacy laws and regulations. Works on protecting sensitive information and developing data governance policies.

**Computer Vision Engineer:** Focuses on developing algorithms and systems for analyzing and interpreting visual data from the real world. Works on applications like autonomous vehicles, facial recognition, and medical imaging.

**AI Educator/Trainer:** Provides training and education on AI and machine learning topics. Develops curriculum and delivers courses, workshops, and seminars to students, professionals, and organizations.

**Robotics Engineer:** Designs, builds, and programs robots and robotic systems. Works on integrating AI and machine learning to enhance robotic capabilities and autonomy.

**Data Storyteller:** Translates complex data analysis into compelling narratives and visualizations. Communicates insights to stakeholders in a clear and impactful manner.

**Cognitive AI Developer:** Develops AI systems that mimic human thought processes. Works on applications such as decision support systems, cognitive automation, and AI-based problem-solving tools.

**Federated Learning Engineer:** Specializes in federated learning, where machine learning models are trained across multiple decentralized devices or servers. Focuses on privacy-preserving and collaborative AI.

**AI Quality Assurance (QA) Specialist:** Ensures the quality and performance of AI models and systems. Conducts testing, validation, and evaluation to identify and resolve issues before deployment.

**Autonomous Systems Engineer:** Develops and maintains autonomous systems, such as self-driving cars or drones. Integrates AI algorithms to enable autonomous decision-making and operation.

**AI Personalization Specialist:** Focuses on creating personalized user experiences using AI. Works on recommendation systems, adaptive learning platforms, and personalized marketing.

**AI-driven Healthcare Specialist:** Develops AI applications for the healthcare industry. Works on predictive analytics, medical imaging analysis, and personalized treatment plans.

**AI Regulation and Compliance Specialist:** Ensures that AI systems comply with regulatory requirements and standards. Works on developing frameworks for legal and ethical AI use.

**Voice and Speech Recognition Engineer:** Develops algorithms and systems for recognizing and synthesizing human speech. Works on applications like virtual assistants, transcription services, and voice biometrics.

# What should you learn next?

- Data science?
  - Statistics + Modeling
  - SQL and/or Data viz packages
  - Cloud!
  - Become fantastic at one of Python/R and learn many packages. Be capable at the other language
- AI?
  - Statistics, Vector calculus, ML theory
  - Cloud!
  - Foundation models
  - Become usefully good at several frameworks (PyTorch, Go, Tensorflow, etc)
  - Read arXiv and NeurIPS/ICLR papers
- Both: learn to properly use AI tools to help you :^/

# 2024 Burtchworks Datascience and AI Professionals Salary Report



Where we are  
**TODAY**

### *Data Science & AI Hiring Landscape*

## CONTINUED GROWTH AND SPECIALIZATION IN AI

#### **Generative AI and Large Language Models**

The demand for professionals skilled in generative AI and large language models has continued to rise. Companies are increasingly focused on enhancing their core enterprise business process and aspire to develop innovative products and services.

#### **Specialized Roles**

There is a growing trend towards specialization within the AI field, with new job titles such as AI Engineer, Prompt Engineer, NLP Specialist, Computer Vision Engineer and more.

## INCREASED INVESTMENT IN AI

#### **Venture Capital and Private Equity**

There has been a significant influx of venture capital and private equity funding into AI startups and projects. This trend is driving rapid innovation and the development of new AI applications across various industries.

#### **Corporate Investment**

Large corporations are also increasing their investment in AI to stay competitive and leverage the latest advancements in technology.

### *Speeding the AI Impact*

## EMPHASIS ON DATA AND AI AS A SERVICE

#### **Data and AI Services**

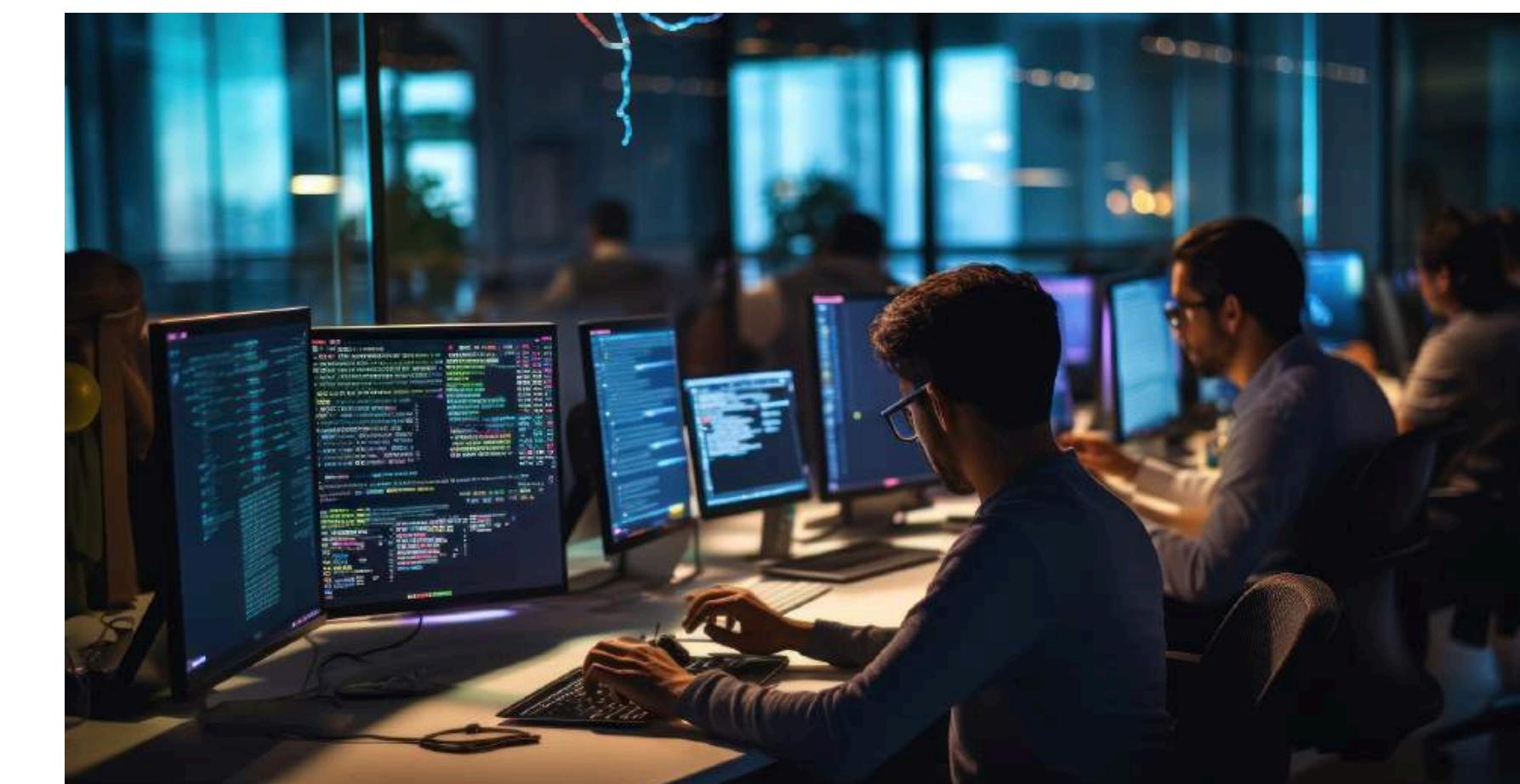
The adoption of Data as a Service (DaaS) and AI as a Service (AlaaS) models is growing. Companies are leveraging these services to access advanced AI capabilities without the need to build and maintain their own infrastructure.

Data as a Service (DaaS) is a cloud-based data management strategy that allows users to access data on demand over the Internet. In this model, data is stored, managed, and delivered by a service provider, making it accessible to users and applications regardless of geographic or organizational boundaries.

AI as a Service (AlaaS) is a cloud-based service model that provides businesses and individuals with access to AI capabilities without the need for extensive in-house infrastructure or expertise. AlaaS allows users to integrate AI functionalities such as machine learning, natural language processing, and computer vision into their applications via APIs and other tools.

#### **Cloud Integration**

Integration with cloud platforms is becoming essential, with major providers offering a range of AI and machine learning services that are easily accessible and scalable.



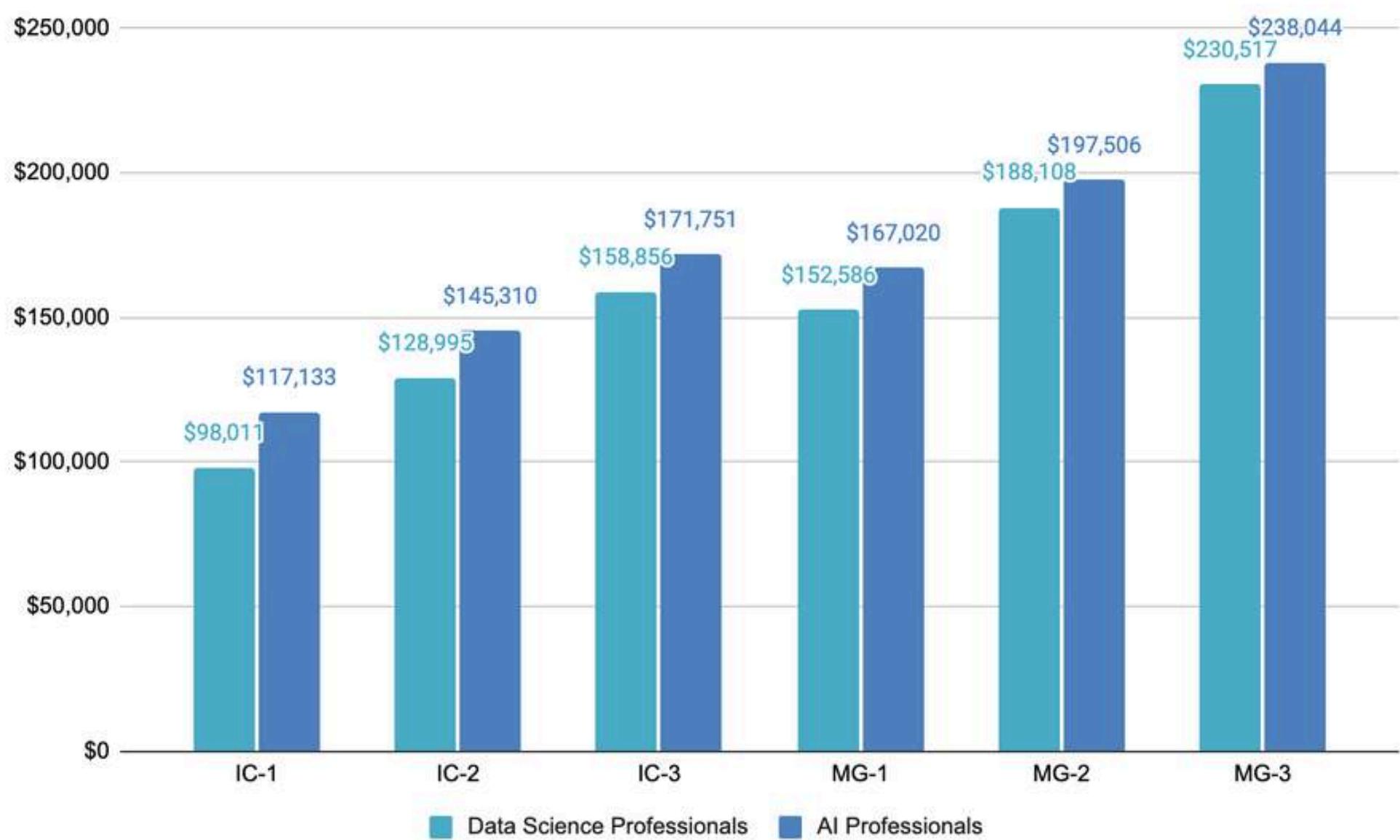
# 2024 BASELINE DATA

AI professionals use advanced machine learning and deep learning techniques tailored to these types of unstructured data, leveraging specialized knowledge and tools to extract meaningful insights and build powerful AI applications.

It takes continuous learning and adaptation to stay updated with the latest research and developments in the field. This goes along with the fact that many AI roles require advanced degrees (Master's or Ph.D.) in fields like computer science, artificial intelligence, or related disciplines. The investment in education and training is higher for AI professionals than data scientists in general. Additionally, AI projects often involve solving complex and novel problems that require significant innovation and creativity. Of course, the potential impact of AI solutions can be substantial, influencing strategic decisions and operations within organizations. AI roles can be highly specialized, focusing on specific areas such as reinforcement learning, generative models, or autonomous systems, which are often critical to the strategic goals of companies.

As a result of all this, companies investing heavily in AI research and development often allocate substantial budgets to attract top talent in AI to maintain a competitive edge. Startups and tech giants offer attractive compensation packages to secure the best AI talent, contributing to overall higher salary levels in the industry. These factors collectively contribute to the higher salaries for AI professionals compared to data scientists.

## 2024 Mean Base Salaries

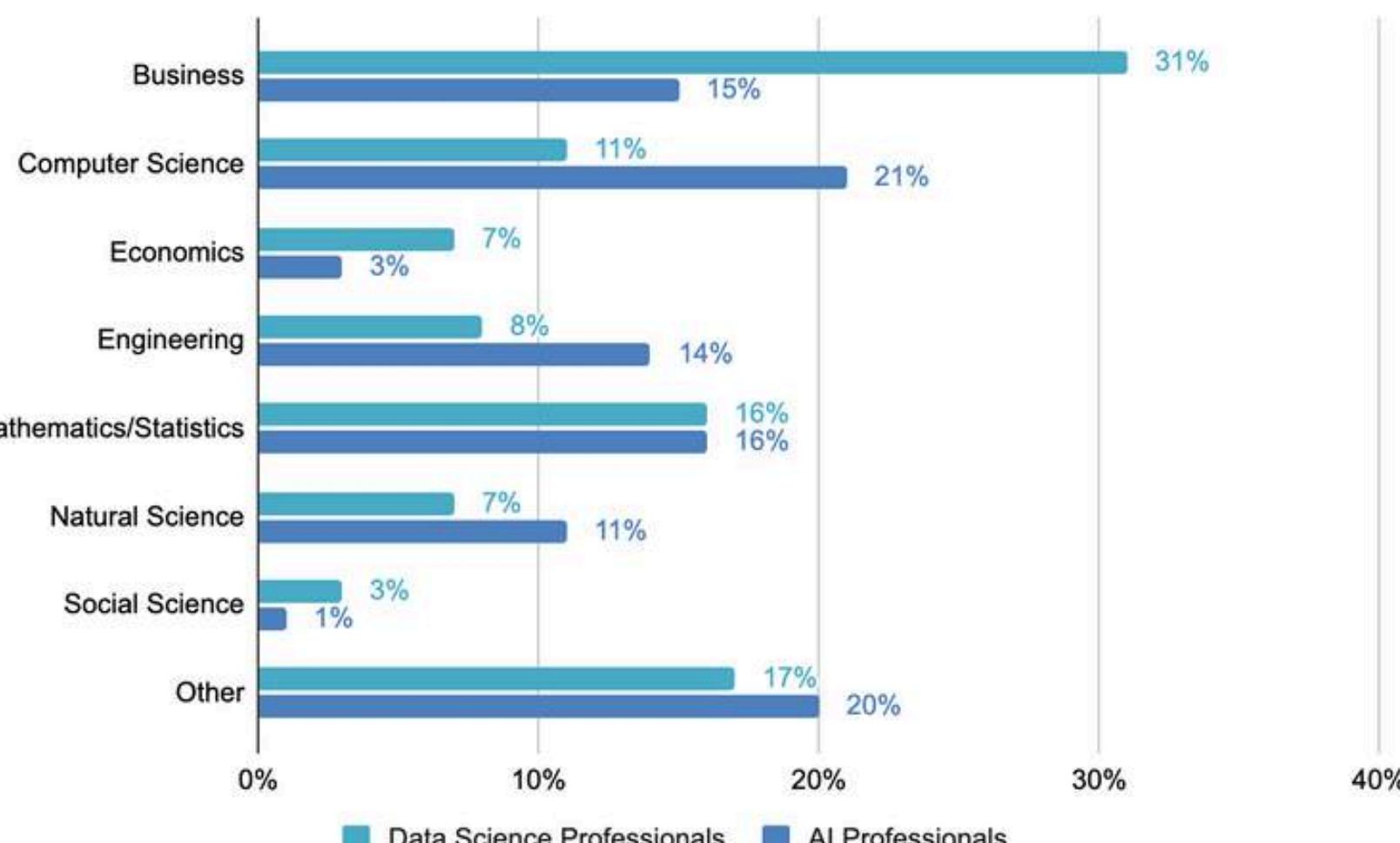


# EDUCATION: COMPARISON OF AREA OF STUDY

- More individuals are up-skilling via courses and private learning.
- Data Science professionals working with structured data are more responsible for delivering business insights, and this explains why a business degree is more advantageous, whereas computer science degrees are more apt for those working in AI (unstructured).

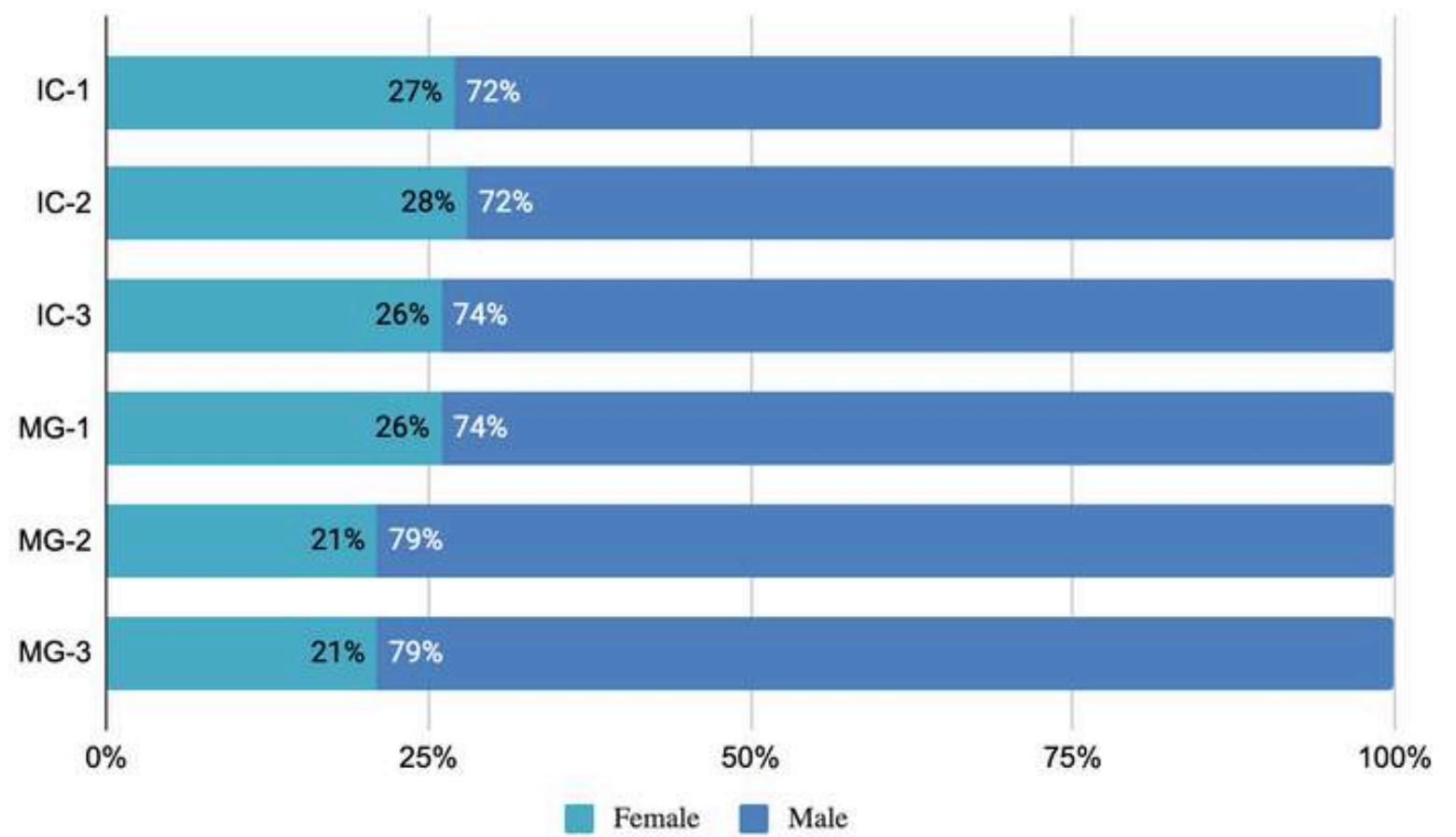
COMPARISON OF AREA OF report (*for highest degree earned*)

## Data Science Professionals vs. AI Professionals

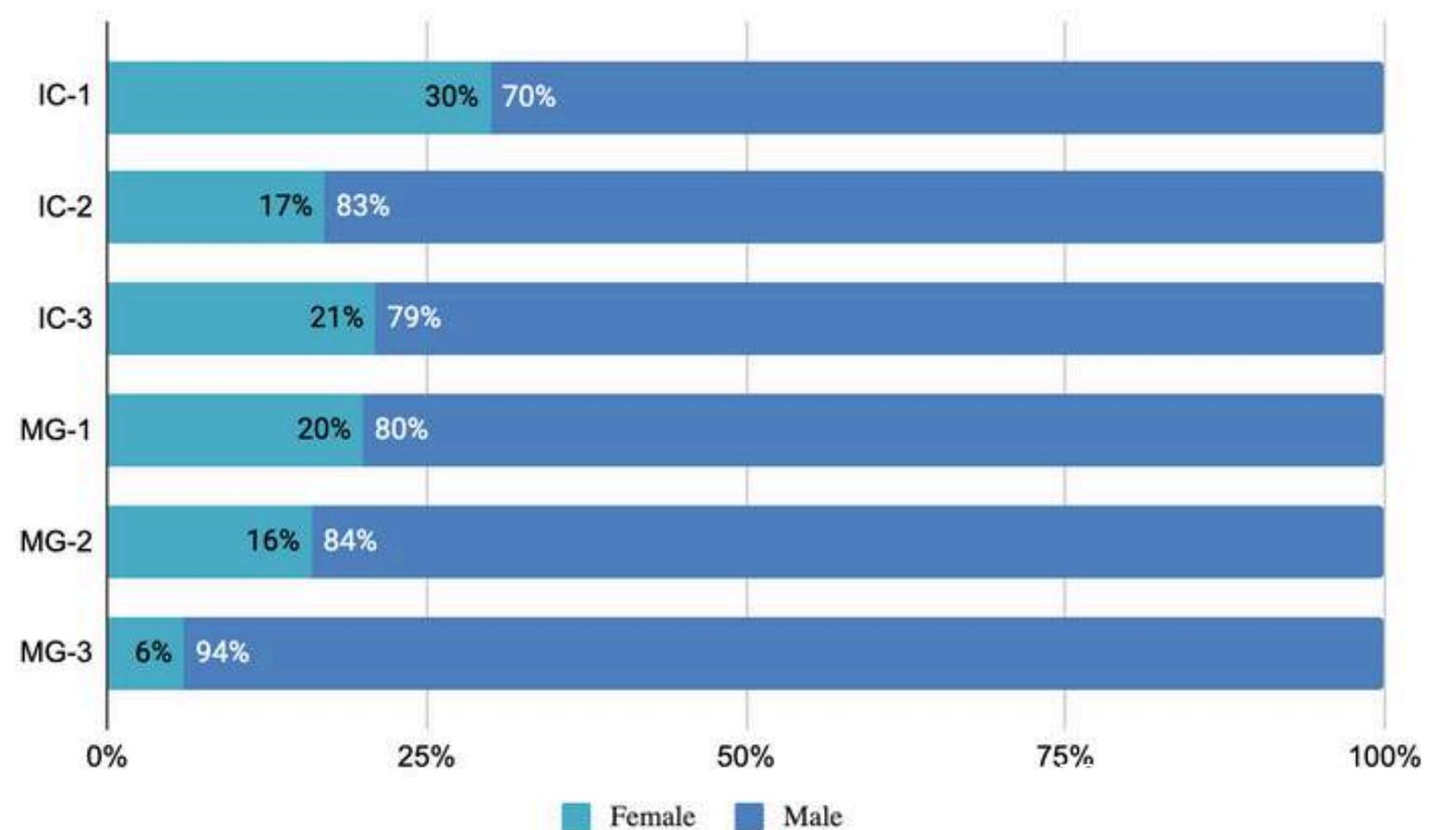


# DISTRIBUTION OF DATA SCIENTISTS & AI PROFESSIONALS BY GENDER AND JOB LEVEL

**Data Scientist Distribution by Gender and Job Level**



**AI Professional Distribution by Gender and Job Level**

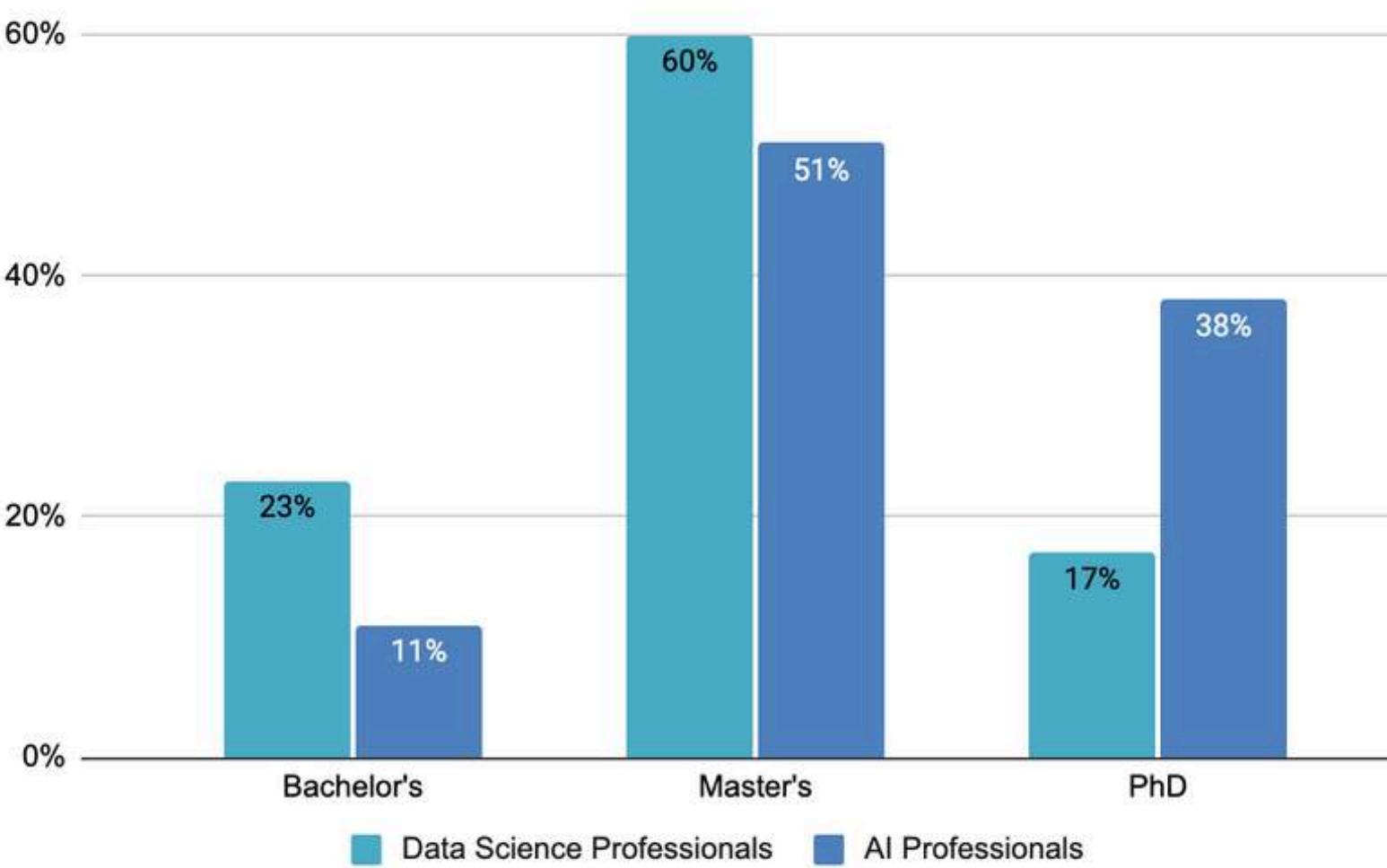


# EDUCATION: COMPARISON OF DEGREE LEVEL

- 80% of all Data Science and AI professionals surveyed held an advanced degree. Education level has historically had a marked effect on salary.
- The proportion of AI professionals with a Master's and/or PhD as their highest degree earned is higher than Data Science professionals and is a statistically significant difference.
- Graduates with higher education degrees are leaving academia and moving to industry roles in data and AI, creating a shortage of skilled labor in academia.
- Mid-career individuals from industry are going back and getting higher education degrees to further differentiate their skill set and education to position themselves in a competitive market.
- There is an increase in domestic skilled labor competitiveness.

COMPARISON OF DEGREE LEVEL (for highest degree earned)

**Data Science Professionals vs. AI Professionals**



# **2025 Burtchworks**

## **Research Professionals**

# **Salary Report**

**Key Takeaways**

## 2025: DATA INTEGRATION, RESEARCH, AND ADVANCED TOOLS

Employers and professionals are navigating an evolving landscape shaped by the following factors:

**Integration of Data**

Market researchers are increasingly expected to merge primary research with secondary sources and big data, enabling a more comprehensive consumer view.

**Cross-disciplinary Collaboration**

Collaboration with data science, analytics, and technology teams is becoming the norm, particularly for projects requiring advanced insights and predictive capabilities.

**Adoption of Advanced Tools**

There is a growing emphasis on tools like Python, R, SQL, and machine learning models to enable deeper insights and foster automation within research processes.

**Increased Focus on ROI**

Companies are emphasizing measurable outcomes, requiring market researchers to align insights with clear business objectives and demonstrate return on investment (ROI) for research initiatives.

**Expansion of DIY Research Platforms**

Tools like Qualtrics and SurveyMonkey are enabling organizations to conduct quick-turn studies, although this trend emphasizes the need for professionals to guide proper design and analysis.

**Current Market Impacts**

## UPTICK IN HIRING AND HYBRID WORK MODELS

**Job Market Flux**

The market has shifted from an employer-driven hiring freeze in previous years to increased hiring activity in 2024. This uptick has generated optimism for sustained growth through 2025.

**Hybrid Work Models**

Hybrid working remains the most popular arrangement, with client-side roles typically requiring 2-3 days of in-person attendance. Supplier-side organizations continue to offer greater flexibility for remote work.

**Competitive Landscape**

While the candidate pool remains rich due to layoffs in 2023, companies are now actively seeking to retain top talent by streamlining hiring processes and offering attractive packages.



***Navigating the Market***

# HOW TO NAVIGATE THE CURRENT HIRING MARKET

**Growing Importance of Data Integration**

Market researchers are now expected to combine primary research with secondary and big data sources, creating a more holistic view of consumers.

**Cross-disciplinary Collaboration**

Engagement with data science teams is becoming standard, so proficiency across these technical areas is a benefit for market research professionals.

**Emerging Tools and Platforms**

New tools and platforms are emerging for automation, visualization, or AI-driven analysis. Market researchers should adopt more advanced tools and techniques such as Python, R, and SQL, and learn machine learning models to derive deeper insights from data.

# COMPENSATION CHANGES: 2023 AND 2024 BASELINE DATA

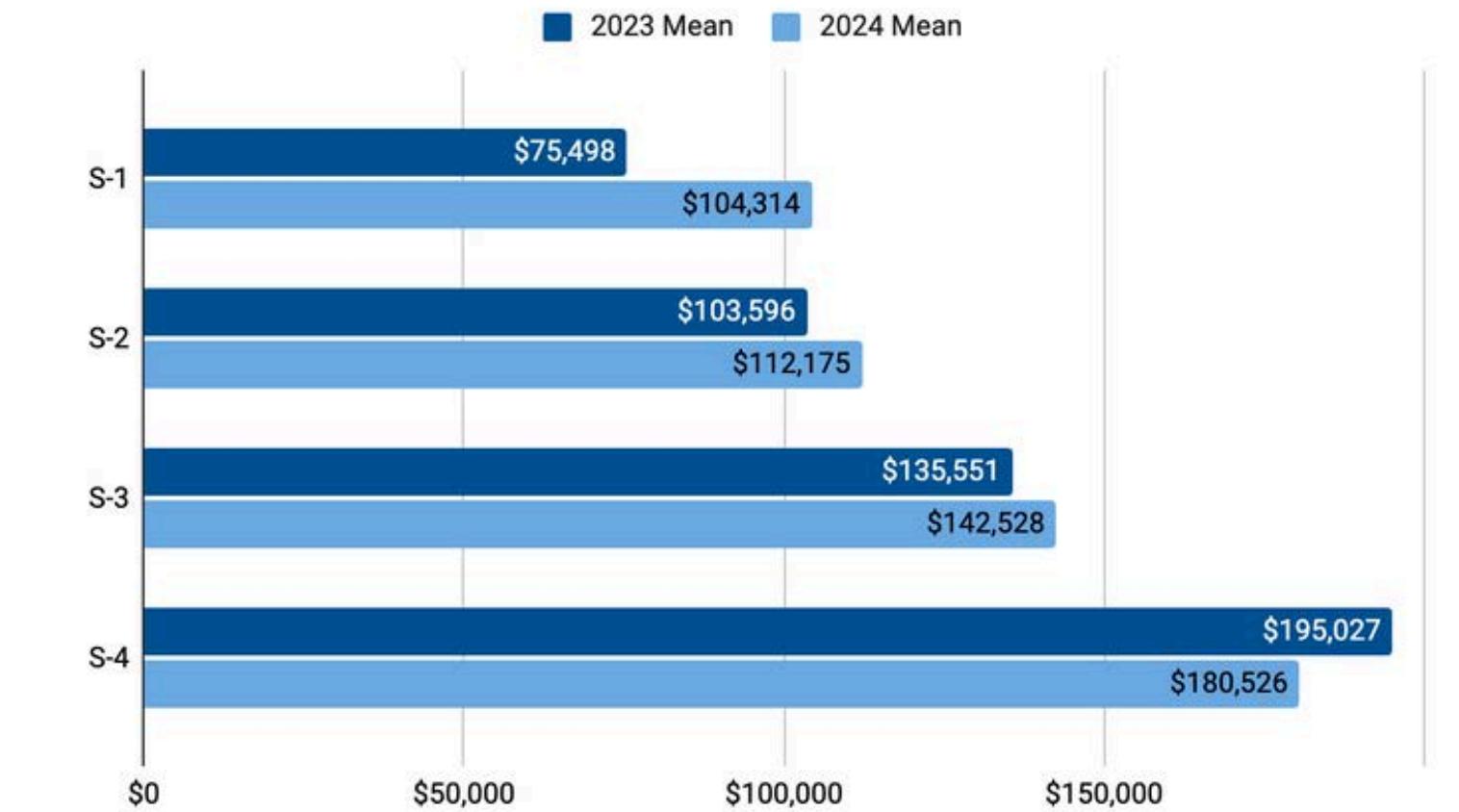
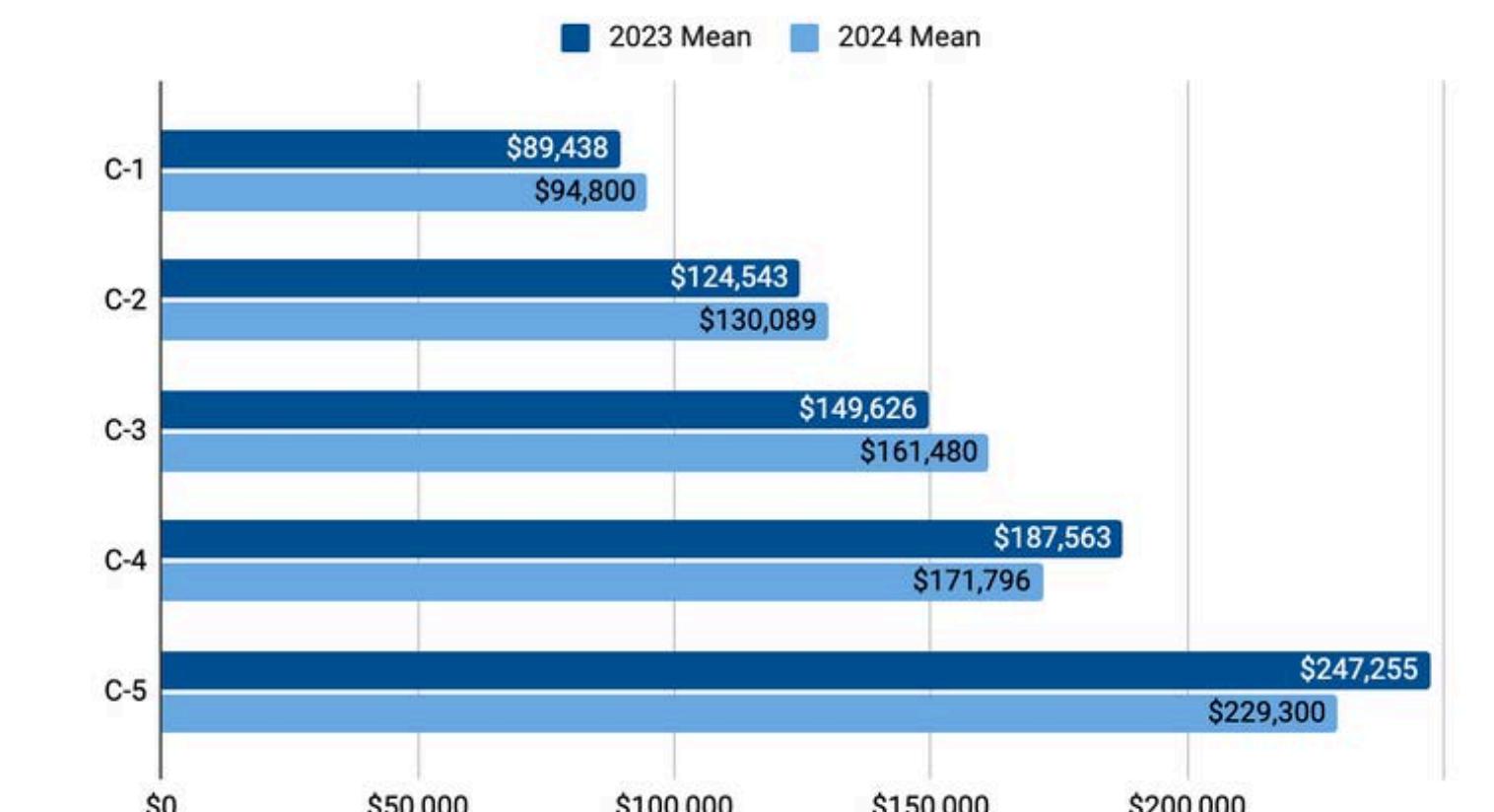
**Comparison of Supplier-Side Mean Base Salary: 2023 v. 2024**

Supplier side:  
Consultancy / X as a service

Client side:  
In-house for a company

Supplier-Side	
Level	Typical Years of Experience
S-1	0-4 years
S-2	5-8 years
S-3	9-14 years
S-4	15+ years

Client-Side	
Level	Typical Years of Experience
C-1	0-4 years
C-2	5-8 years
C-3	9-14 years
C-4	15-24 years
C-5	25+ years

**Comparison of Client-Side Mean Base Salary: 2023 v. 2024**

# Appen (aka Figure-Eight aka Crowdflower) State of AI report

## Methodology

The State of AI 2024 survey aimed to assess AI and data implementation across organizations. We partnered with The Harris Poll to conduct an online survey of 509 U.S. information technology decision-makers (ITDMs) from April 18 to May 9, 2024. Participants were ages 21 to 65 and worked at companies with 100+ employees.

### Industry

- Automotive & Transportation
- Finance & Financial Services
- Healthcare & Pharmaceuticals
- Internet & Electronics
- Retail & Consumer Durables (Including E-Commerce)
- Technology
- Telecommunications

### Role

- Business Leaders and Managers
- Data Scientists
- Data Engineers
- Developers

### Company sizes

- Small: 101-1,000 employees
- Medium: 1,001-10,000 employees
- Large: 10,001+ employees

*The research was conducted online in the United States by The Harris Poll on behalf of Appen among 509 information technology decision-makers. The survey was conducted April 18 - May 9, 2024. Raw data were not weighted and are therefore only representative of the individuals who completed the survey.*

*Respondents for this survey were selected from among those who have agreed to participate in our surveys. The sampling precision of Harris online polls is measured by using a Bayesian credible interval. For this study, the sample data is accurate to within ±4.3 percentage points using a 95% confidence level. This credible interval will be wider among subsets of the surveyed population of interest.*

*All sample surveys and polls, whether or not they use probability sampling, are subject to other multiple sources of error which are most often not possible to quantify or estimate, including, but not limited to coverage error, error associated with nonresponse, error associated with question wording and response options and post-survey weighting and adjustments (not applicable in this case).*

### Small footnote

The 2023 study was conducted from February to March 2023 among 400 US and 100 European (UK, Ireland, and Germany) ITDMs. The 2022 study was conducted by The Harris Poll in June 2022 among 504 ITDMs; the 2021 study was conducted in March 2021 among 501 US ITDMs; and the 2020 study was conducted using a different method among 290 ITDMs.

# Introduction

As artificial intelligence (AI) adoption accelerates, the challenges surrounding data management, model reliability, and scalability are intensifying. In our 2024 State of AI report, we explore the critical factors driving AI development, including the rise of generative AI, enterprise headwinds, the importance of data quality, and the increasing demand for strategic partnerships to address the complexities of AI implementation. Our research, conducted in collaboration with The Harris Poll, surveyed over 500 U.S. information technology decision-makers (ITDMs) across a range of industries. The findings reveal five key takeaways shaping the future of AI:

## 1. Generative AI Adoption is Increasing

- up 17 percentage points from 2023

## 2. Enterprise Deployments & ROI are Down

- lower percentage of projects get deployed and show ROI

## 3. Quality Data is Essential to AI Success

- 97% say data diversity, reduced bias, or scalability are crucial

## 4. Data Challenges Persist

- 10 percentage point increase in data preparation bottlenecks

## 5. Companies Seek Strategic Data Partners

- 93% want full AI Data Lifecycle expertise

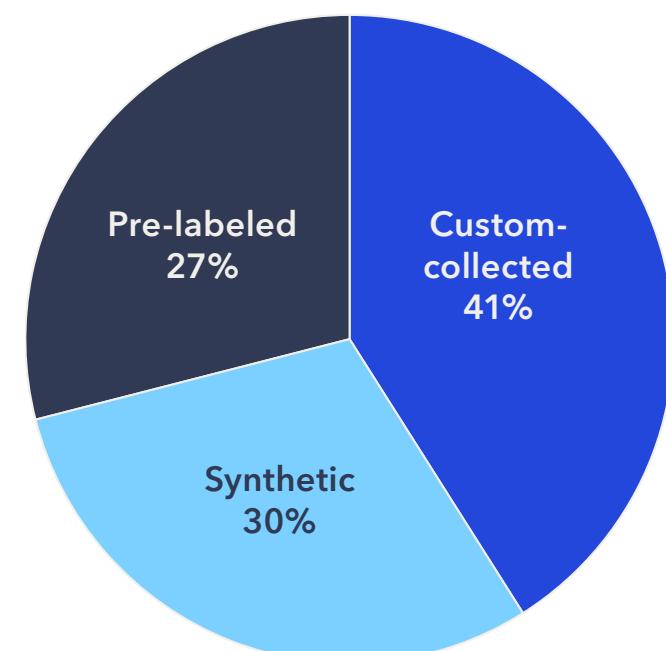
These insights demonstrate that while AI offers transformative potential, its success depends on addressing ongoing data challenges and forming strong partnerships. At Appen, we're committed to supporting organizations through our comprehensive AI Data Platform (ADAP) and a global network of contributors, enabling businesses to navigate the complexities of AI with confidence.

Use of Generative AI	2023	2024
Use Generative AI to improve productivity and efficiency of internal business processes	49%	53%
Use Generative AI to support internal IT operations	47%	49%
Use Generative AI to aid in the research and development of new products	37%	46%
Use Generative AI to support other business functions (e.g., Marketing, Communications, Design)	45%	43%
Use Generative AI to support manufacturing/production operations	36%	41%
Use Generative AI based features into products or services that we sell	40%	41%
Use Generative AI to reduce business costs	37%	41%

© 2024 State of AI Report **Appen**

Despite its rapid growth, managing bias and ensuring fairness in generative AI model training remains a key challenge. To tackle this, custom data collection has become the primary method for sourcing training data, highlighting its importance in creating reliable AI systems.

MOST COMMONLY USED DATA SET BY USE CASE

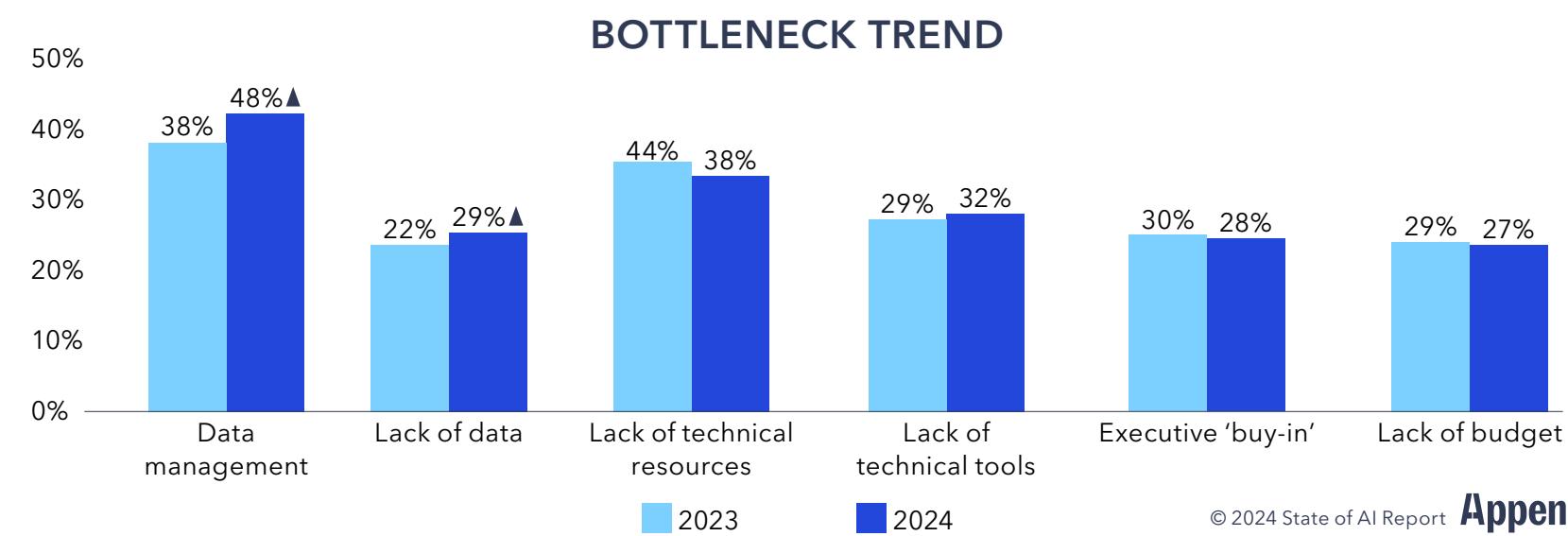


© 2024 State of AI Report **Appen**

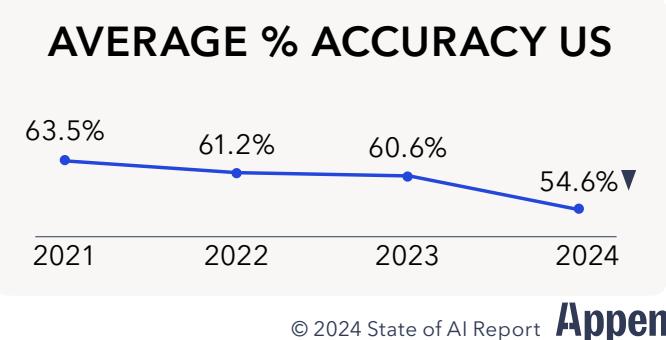
## Data Challenges Persist

Sourcing quality data remains a significant challenge, becoming more difficult as AI use cases grow increasingly specialized. Data management has emerged as the leading obstacle in AI projects for ITDMs, with a reported 10 percentage point increase in bottlenecks related to sourcing, cleaning, and labeling data since 2023.

The lack of data availability has also risen by 7 percentage points during the same period. This increase may be attributed to the growing range of generative AI applications, which are driving new data requirements and aligned with the earlier observation of a rise in custom data collection for generative AI models.



Despite the importance of high-quality data annotation in AI development, average accuracy of data has dropped about 9 percentage points since 2021. This may be a result of the increasing complexity of AI systems and the corresponding data annotation requirements, which increasingly require specialized domain-specific knowledge and quality control across large datasets.



To overcome these challenges, organizations must shift their attention from short-term gains to long-term value creation in AI and focus on building a robust foundation of high-quality data.

## Quality Data is Essential to AI Success

**DATA DIVERSITY IS THE MOST VALUED FEATURE IN AI, FOLLOWED BY BIAS REDUCTION AND SCALABILITY.**

**80%** ACKNOWLEDGING THAT HUMAN INSIGHT IS KEY OF RESPONDENTS TO REFINING AI SYSTEMS

Our latest research reveals that a significant majority—86% of respondents—are retraining or updating their machine-learning models at least once every quarter. This underscores the rapid development across the industry and the close feedback loop between data and models in the AI training process.

Ninety percent of respondents confirmed that they rely on external data providers for their AI model training and/or annotation.

Custom-collected datasets, especially in text, images, video, audio, and point-of-interest data, have become the backbone of many leading AI applications. The effectiveness of these models depends on data that is accurate, diverse, and representative. Notably, 80% of respondents highlight the importance of human-in-the-loop machine learning, acknowledging that human insight is key to refining AI systems.

Data diversity is the most valued feature of AI data, closely followed by efforts to reduce bias, and ensuring data scalability. 7% of respondents agree these elements are vital for building AI models. Quality human-in-the-loop training and evaluation data play a crucial role in achieving these goals by providing the nuanced, diverse, and accurate data necessary to refine models, reduce bias, and ensure they perform reliably at scale.

**What else should I be thinking  
about in my career?**

# Soft skills

<https://www.manning.com/books/build-a-career-in-data-science>

<https://www.youtube.com/watch?v=pBisSm0Zr3A>

BUILD A CAREER IN  
**Data Science**

Emily Robinson

Jacqueline Nolis

MANNING

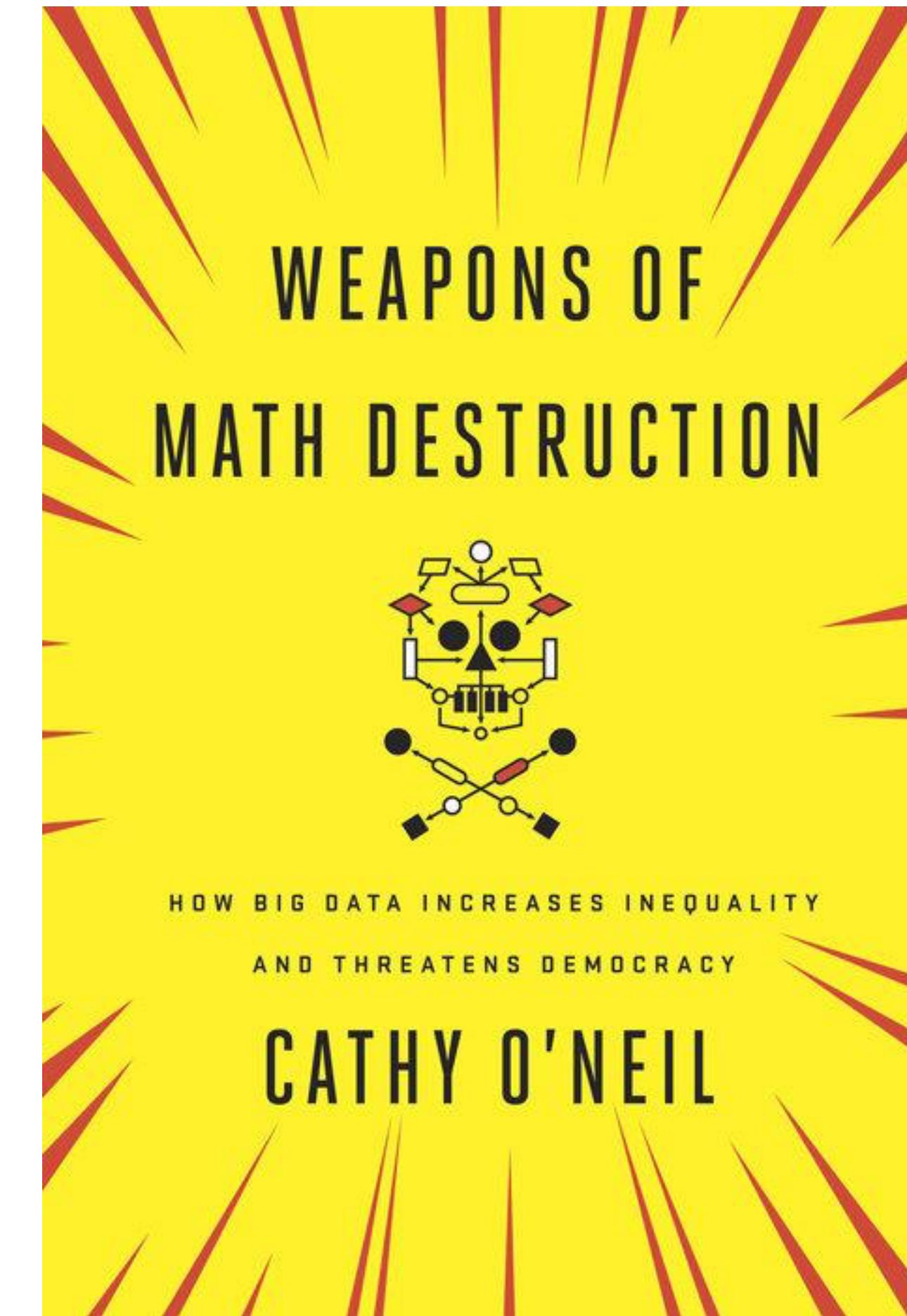


# Problems with algorithms in general

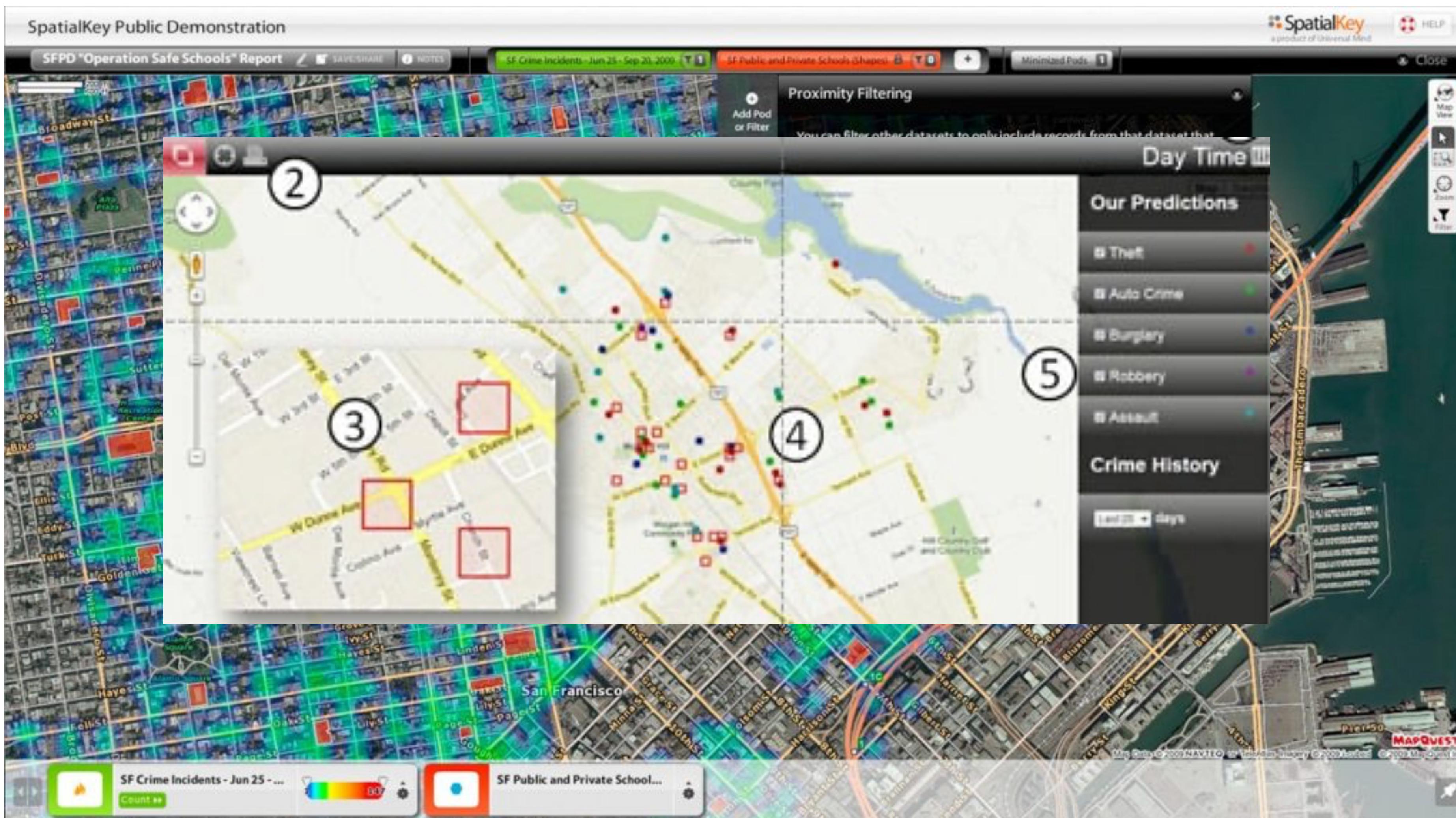
- Bias
- Ethics of use
- Accountability
- Societal impacts
- Economic impacts
- Environmental impacts

# Don't be a tool for creating WMDs

- Algorithms (and DS!) implement our biases, yet look objective
- Can implement our biases at scale
- Can have huge impacts on people's lives
- Are not transparent or accountable to the people being impacted

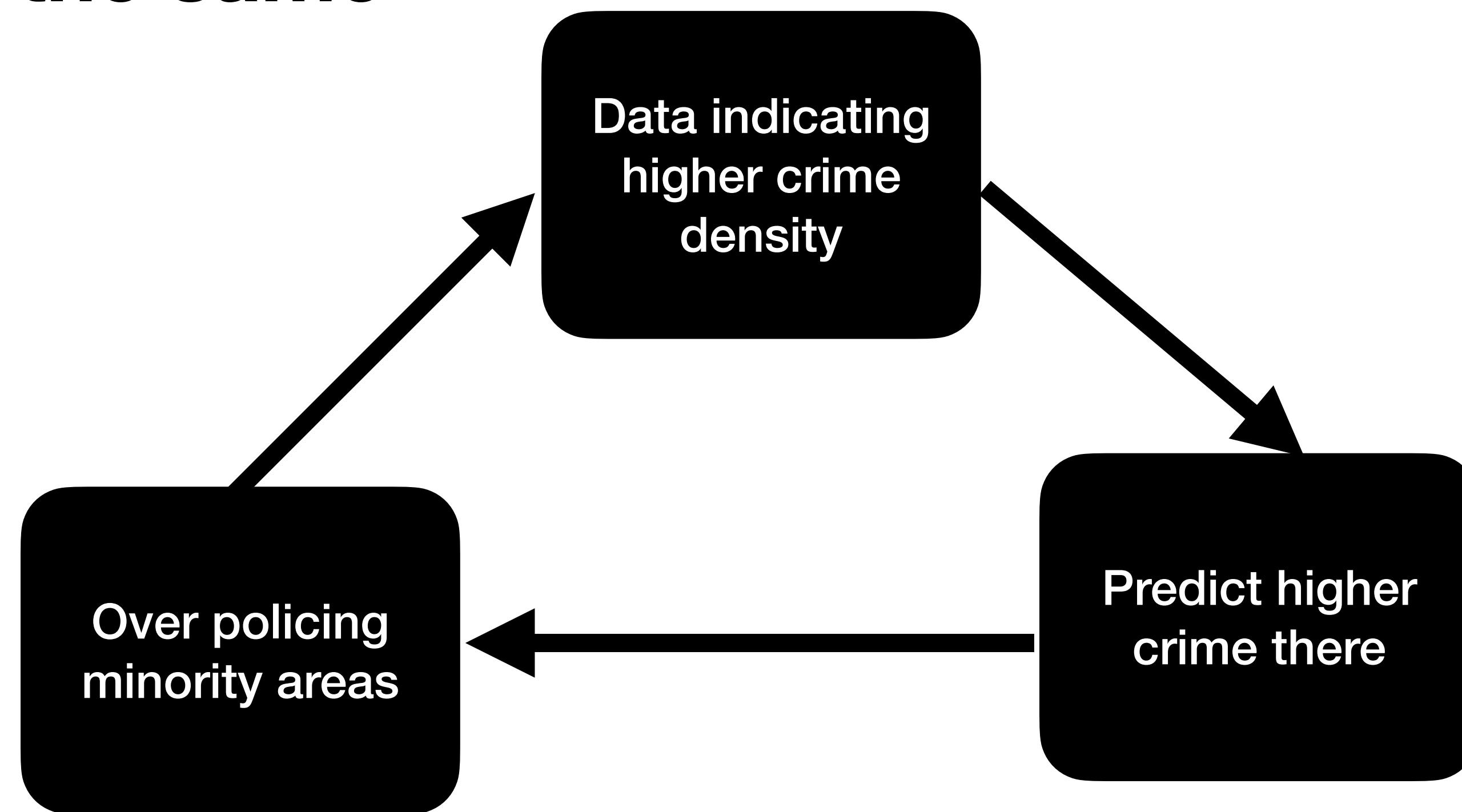


# Predictive policing



# Predictive policing & sentencing

**Blacks arrested for possession at 4x the rate of whites  
Usage rates the same**





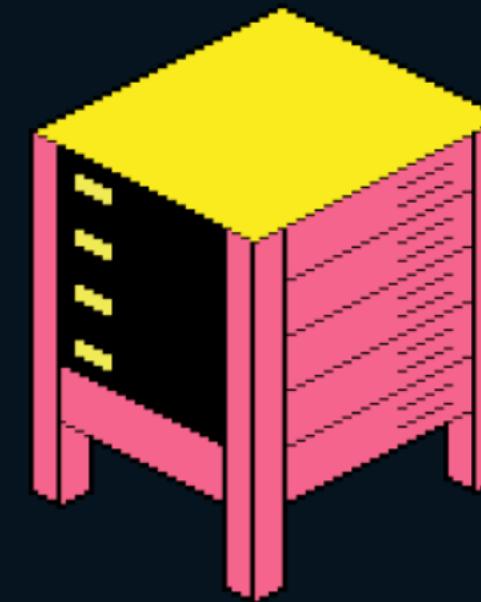
**“A lot of times, people are talking about bias in the sense of equalizing performance across groups. They’re not thinking about the underlying foundation, whether a task should exist in the first place, who creates it, who will deploy it on which population, who owns the data, and how is it used?”**

**-Timnit Gebru**

four years. Google expects to spend \$75 billion on AI infrastructure alone in 2025.

This isn't simply the norm of a digital world. It's unique to AI, and a marked departure from Big Tech's electricity appetite in the recent past. From 2005 to 2017, the amount of electricity going to data centers remained quite flat thanks to increases in efficiency, despite the construction of armies of new data centers to serve the rise of cloud-based online services, from Facebook to Netflix. In 2017, AI began to change everything. Data centers started getting built with energy-intensive hardware designed for AI, which led them to double their electricity consumption by 2023. The latest reports show that 4.4% of all the energy in the US now goes toward data centers.

**The carbon intensity of electricity used by data centers was 48% higher than the US average.**



Given the direction AI is headed—more personalized, able to reason and solve complex problems on our behalf, and everywhere we look—it's likely that our AI footprint today is the smallest it will ever be. According to new projections published by Lawrence Berkeley National Laboratory in December, by 2028 more than half of the electricity going to data centers will be used for AI. At that point, AI alone could consume as much electricity annually as 22% of all US households.

# **Problems with AI**

## **Same same**

- Bias
- Ethics of use
- Accountability
- Societal impacts
- Economic impacts
- Environmental impacts



News & Politics   Culture   Technology   Business   Life   Advice   Podcasts

\* YES... HA HA HA... YES!

THE INDUSTRY

# Why You Might Soon Be Paid Like an Uber Driver—Even If You're Not One

As gig workers' pay gets slashed by algorithms, experts warn that A.I.-driven wage systems mean that no one's paycheck is safe.

BY MOLLY GLICK   OCT 13, 2024 • 5:45 AM



<https://rooseveltinstitute.org/publications/uber-for-nursing/>

ABOUT   PUBLICATIONS   THINK TANK   ROOSEVELT NETWORK   ROOSEVELT SOCIETY   FDR LIBRARY

BRIEFS   WORKER POWER AND ECONOMIC SECURITY

## Uber for Nursing: How an AI-Powered Gig Model Is Threatening Health Care

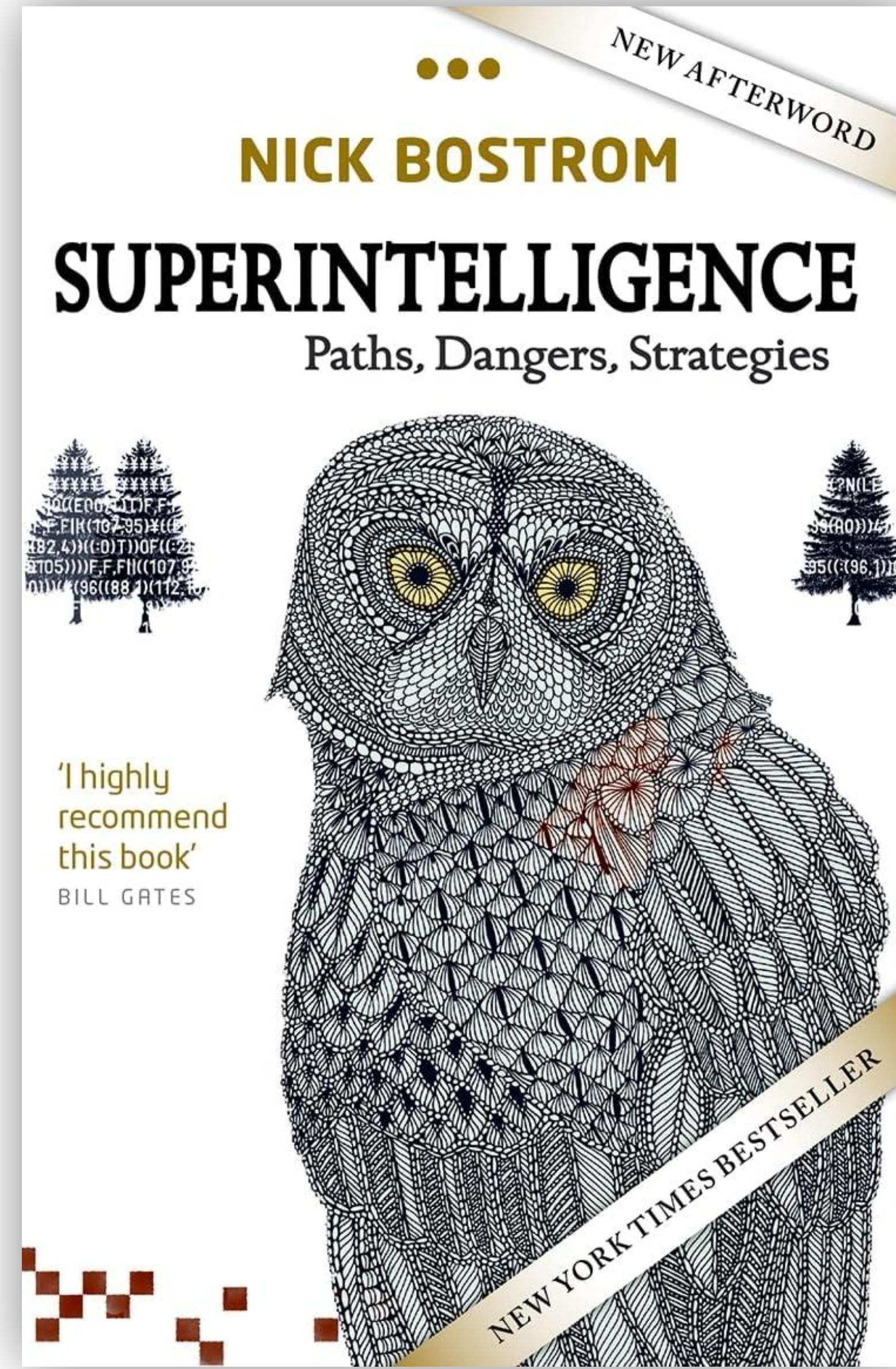
DECEMBER 17, 2024

By Katie J. Wells, Funda Ustek Spilda

# Problems with AI

## New new

- Interpretability
- Alignment

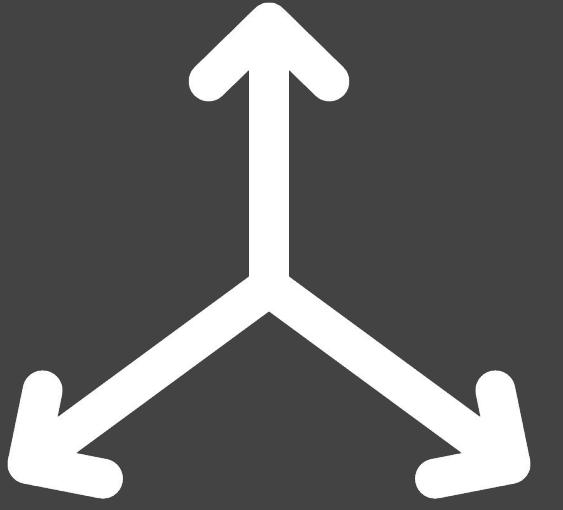


# Paperclip AI Thought Experiment



## Orthogonality

Greater intelligence does not imply greater alignment to human values



## Instrumental Convergence

All final goals pursued by sufficiently intelligent agents have subgoals in common



**(Super)intelligent AI may (un)intentionally make decisions misaligned with human values even while pursuing harmless goals**

# Cognitive Superpowers

*What Bostrom claims an AI would need to be able to do for world domination, human extinction, and other fun activities. What would make an AI a threat?*

Task	Skillset
Hacking	Access and manipulate computational resources remotely
Technology Research	Develop technology with real-world possibilities
Economic Productivity	Perform economically productive activity to acquire resources
Strategizing	Achieve distant goals and overcome opposition via planning
Social Manipulation	Leverage human and institutional support via persuasion
Intelligence Amplification	Bootstrap its own capabilities independently

**The alignment problem is fighting  
the law of unintended consequences**

LLMs are trained to be polite and not upset users

[\*\*Instagram's AI Chatbots Lie About Being Licensed Therapists \(404 media\)\*\*](#)

[\*\*Most AI Chatbots Easily Tricked Into Giving Dangerous Responses, Study Finds \(the guardian\)\*\*](#)

LLMs make it easy to do some kinds of homework

[\*\*Ghost Students Are Creating an 'Agonizing' Problem For California Colleges \(sfgate.com\)\*\*](#)

Wonder what might go wrong here? Wonder why he wants this?

[\*\*Zuckerberg's Grand Vision: Most of Your Friends Will Be AI \(msn.com\)\*\*](#)

**But lets end on a positive note  
and advice for you....**

# My advice...

- The world will turn, everything will be different in not too long
- Stay curious! Keep learning!
- Think about the big picture once in a while, including ethics
- Learn about the system you're part of, and try to change it
- Remember that there is no single definition of success. Your grades, job title, and bank account do not define you.
- Even as you strive to succeed, don't let it take over your life!  
Stay happy and balanced!!

# My hope for you...

- Many productive failures
- Beautiful triumphs to match the beautiful failures
- A sense of purpose in your life
- Many moments where you feel achievement and joy!

# Thanks to your amazing instruction team!

-  Aatyanth Thimma Udayakumar
-  Alec Slim
-  Alessandro D'Amico
-  Alokita Karmokar
-  Annalise Ahn
-  Felipe Lorenzi
-  Jason Chen
-  Jiawei Li
-  Kristoffer Alejo
-  Miles Davis
-  Nandini Bohra
-  Reuben Chatterjee
-  Ruby Ying
-  Zheng Zeng

**And thanks to YOU!**



