



# **Operations Analysis and Artificial Intelligence**

Dr. David L. Alderson

Professor and Chair, Operations Research Department

Naval Postgraduate School, Monterey, CA

Fall 2025

# We are living in a Gold Rush!

## The AI Gold Rush

Shane Greenstein , Harvard Business School, Boston, MA, 02163, USA

Large language models (LLMs) have overrun commercial markets, more like a tsunami than the normal technical wave of interest. The topic is everywhere—news stories, blogs, podcasts, start-up investments, analyst reports, hackathons, and government announcements. A virtual frenzy surrounds it.

If you possess a technical background, you might find this frenzy puzzling. The technological roots of LLMs go back many years. However, today's experience looks like more than the continuation of a pre-existing trend. Something in the zeitgeist changed recently, making entrepreneurs and financiers rethink and shift the direction of investment. You might be tempted to call this an AI gold rush. If you are old enough to recall them, you might compare this frenzy to the gold rushes during the dot-com or PC booms.

information spreading—happened within a month of each other.

What about impatient economic actors? Gold miners overran Sutter's Mill that spring. By the following year's snowmelt, the area was packed with miners who believed they needed to stake a claim as soon as possible. Many panned for gold in the rivers. They were called "49ers."

Sadly, though the western slopes of the Sierra Nevada range contain some of the best deposits of gold in the world, most of those deposits do not have rivers passing over them. Most of those who panned for gold did not recover much and soon gave up.

Old-fashioned digging into the earth, which followed years after the 49ers, yielded much better rewards. This activity was expensive and slow, requiring the financing

## Reminiscent of California Gold Rush of 1849

- **Hopes** among senior executives in business and government: *prospecting for a golden future where using AI will allow making decisions quickly, in the face of uncertainty, and with assurances of favorable outcomes.*
- **Fears** among information workers: *anyone not using AI or LLMs in their daily work is at risk of losing their livelihood, because they are left behind and/or out of a job.*

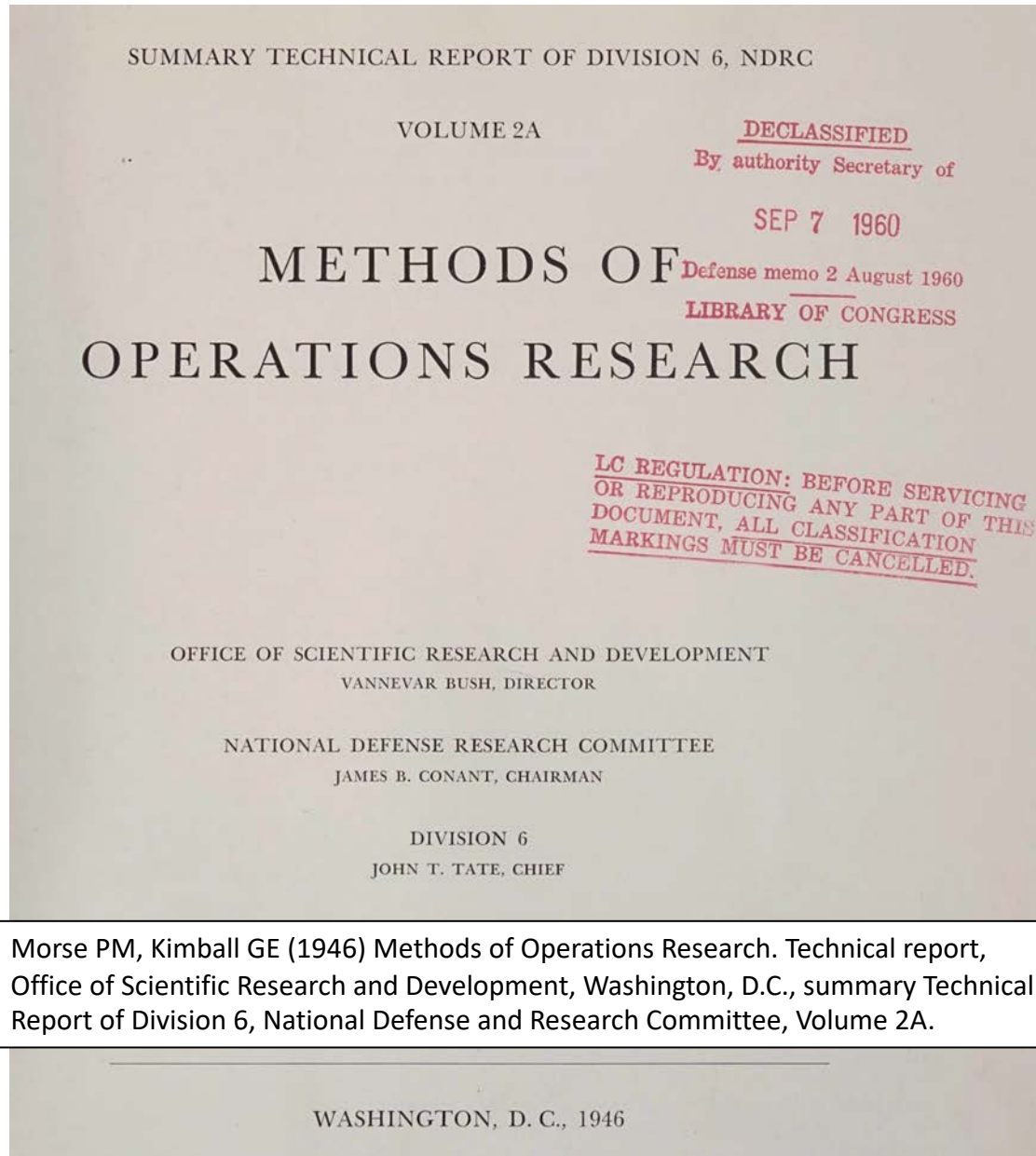
Fueling a massive rush of people and companies seeking new capabilities (and fortune) in developing and deploying AI-enabled technologies.

*IEEE Micro November/December 2023, pp. 126-128*

# Making Sense of the AI Gold Rush: An OA Perspective

- Operations Analysis (OA)
- Artificial Intelligence (AI)
- Automation
- Expertise

# What is Operations Analysis (OA)?



Morse PM, Kimball GE (1946) Methods of Operations Research. Technical report, Office of Scientific Research and Development, Washington, D.C., summary Technical Report of Division 6, National Defense and Research Committee, Volume 2A.

## Operational Research.

P. M. S. BLACKETT (Manchester University).

*Advancement of Science*, 5. April, 1948.

Two documents written in 1941 were here published generally for the first time. The first, prepared for the Admiralty, states the reasons for setting up Operational Research Sections and defines their functions and organisations.

The need for scientists to study secret operational data in war requires special staffs at Command Headquarters, able to observe actual operational conditions while retaining a scientific approach to the problems. Executive officers in the Services are not trained to analyse complex problems, and can profit by having quantitative results on which to take action. The research team should be directly responsible to the C.-in-C.; their reports should be distributed as widely as possible among the men concerned. The Services' requirements for war equipment need interpretation since the technicians who produce it are often unaware of many of the operational conditions; conversely, the Services need to be kept informed of scientific developments which may help their work. There is a balance to be struck between developing new devices and making the best use of existing ones: the introduction of Operational Research Sections is in part a redeployment of scientific effort to improve this balance.

The second paper, revised in 1943, discusses the methods used to solve many of the wartime problems.

The practical objectives of operational research in war are the appraisal and improvement of weapons and tactics, and the evaluation of strategy in terms of national effort. The methods are broadly classifiable as either "*a priori*" or "variational", mostly the latter. The *a priori* procedure may fail when the problem is complex, and variational analysis is the most useful method of forecasting the yield for a future operation

P. M. S. Blackett (1950) *Operational Research*, *Journal of the Operational Research Society*, 1:1, 3-6, DOI: 10.1057/jors.1950.2



# First-in-the-world OA curriculum 1951. Warfighting-centric Curriculum.

## Catalogue of

## UNITED STATES NAVAL POSTGRADUATE SCHOOL

ACADEMIC YEAR 1951-1952



POSTGRADUATE SCHOOL, ANNAPOLIS, MARYLAND

### OPERATIONS ENGINEERING

#### OPERATIONS ANALYSIS CURRICULA — RO GROUPS

##### Objective

The objective of these curricula is to further educate selected officers in general engineering and science, in basic principles of operations research, and in the application of those principles to naval problems. Operations research may be defined as "a scientific method of providing commanders with a quantitative basis for decisions regarding operations under their control."

Field work in this course will consist of active participation in the solving of current problems under the direction of Commander Operational Development Force.

##### FIRST YEAR (RO)

###### First Term

|                                                          |     |
|----------------------------------------------------------|-----|
| Ma 181(C) Directional Derivatives & Line Intervals ..... | 5-0 |
| Ma 251(C) Graphical and Mech. Comp ..                    | 0-4 |
| Ph 240(C) Geom. Optics and Phys. Opt.                    | 3-3 |
| Ch 101(C) General Chemistry .....                        | 3-2 |
| Or 120(C) Surface Fire Control .....                     | 2-0 |

Total 13-9

###### Third Term

|                                                                  |     |
|------------------------------------------------------------------|-----|
| Ma 183(B) Complex Variable & Dif. Eqs. of Theoretical Physics .. | 5-0 |
| Ph 142(B) Mechanics .....                                        | 4-0 |
| Ph 361(A) Electromagnetism .....                                 | 3-0 |
| Ma 382(A) Probability II .....                                   | 2-0 |
| Oa 100(A) Classical Prob. in Operations Analysis .....           | 2-0 |
| Or 131(A) A.A. Fire Control .....                                | 2-0 |

Total 18-0

###### Second Term

|                                             |     |
|---------------------------------------------|-----|
| Ma 182(B) Dif. Eqs. & Vector Analysis ..... | 5-0 |
| Ma 381(B) Probability I .....               | 4-0 |
| Ph 141(B) Mechanics .....                   | 4-0 |
| Ph 341(C) Electricity & Mag .....           | 4-2 |

Total 17-2

###### Fourth Term

|                                                                              |     |
|------------------------------------------------------------------------------|-----|
| Ma 184(A) Matrices, Tensors & Variations .....                               | 5-0 |
| Ph 362(A) Electromagnetic Waves .....                                        | 3-0 |
| Ph 640(B) Atomic Physics .....                                               | 3-3 |
| Oa 101(A) Applications of Probability & Kinematics to Operations Analy. .... | 5-0 |
| Or 132(C) A.A. Fire Control .....                                            | 2-0 |
| *IE 104 Psycho. Sys. Research .....                                          | 0-1 |

Total 18-4

Intersessional Period — Six Weeks Practical Work.

\*Lecture course — no academic credit.

# What officers need to know: How to frame, execute, and evaluate operational problems

“I think that the operations research curriculum that I went through is one that's very relevant to what we do in the Navy. ...I've used it in three significant tours in the Pentagon. I've also used it at sea and war fighting. **What the curriculum taught me to do was to properly frame a problem, ask the right questions, to assess risk and to move on from there.**”

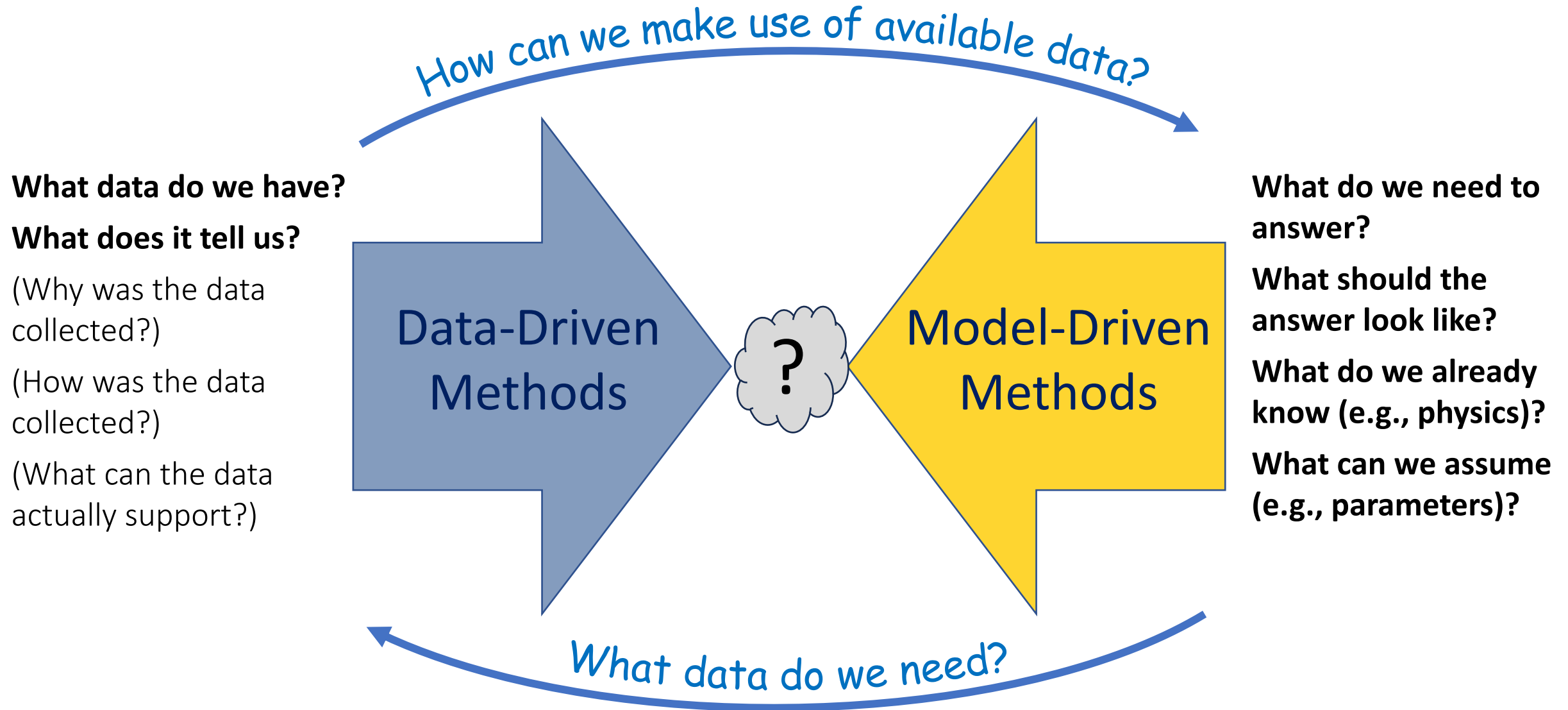
--Admiral Michael G. Mullen, USN  
17th Chairman, Joint Chiefs of Staff  
NPS OR (360) M.S. '85



“NPS is one of the elite schools of this country... America needs elite, highly technical institutions that are executing research and development at the zenith of science to address our unique defense problems.”

--Admiral Samuel Paparo  
Commander, U.S. Indo-Pacific Command  
NPS OR M.S.A. (363) '15

# Operations analysts use two families of methods



# Artificial Intelligence – An (Over)Simplified History

- 1955     McCarthy J, Minsky ML, Rochester N, Shannon CE, “A proposal for the Dartmouth summer research project on artificial intelligence”
- 1970s     Rise of Expert Systems – intended to mimic decision-making of domain experts  
simulated reasoning via if-then rules (knowledge base + inference rules)
- 1980s     knowledge acquisition was slow, costly; massive knowledge base expensive to store  
Fall of Expert Systems – led to an “AI Winter” in funding
- 1986     Herbert Simon keynote at TIMS/ORSA (now INFORMS) Annual Meeting

## References:

Nilsson NJ (2009) The quest for artificial intelligence (Cambridge University Press).  
Mitchell M (2019) Artificial Intelligence: A guide for thinking humans (Penguin UK).



# Artificial Intelligence – An (Over)Simplified History

## Two Heads Are Better than One: The Collaboration between AI and OR

HERBERT A. SIMON

*Departments of Psychology  
and Computer Science  
Carnegie-Mellon University  
Pittsburgh, Pennsylvania 15213*

*Originally delivered as the plenary address at the TIMS/ORSA Joint National Meeting,  
Miami, Florida, October 1986.*

*Published 1987. Interfaces 17(4):8-15. <https://doi.org/10.1287/inte.17.4.8>*



Computer scientist,  
economist, cognitive  
scientist

Famous for “bounded  
rationality”, “satisficing”

Turing Award (1975)

Nobel Prize in  
Economics (1978)

*“Management science and operations research are a part of the great effort, often styled the Second Industrial Revolution, that is striving to understand and enhance intelligence. **Joining hands with AI, management science and operations research can aspire to tackle every kind of problem-solving and decision-making task the human mind confronts.**”*

Nilsson NJ (2009) The quest for artificial intelligence (Cambridge University Press).

Mitchell M (2019) Artificial Intelligence: A guide for thinking humans (Penguin UK).

# Artificial Intelligence – An (Over)Simplified History

- 1955 McCarthy J, Minsky ML, Rochester N, Shannon CE, “A proposal for the Dartmouth summer research project on artificial intelligence”
- 1970s Rise of Expert Systems – intended to mimic decision-making of domain experts  
simulated reasoning via if-then rules (knowledge base + inference rules)
- 1980s knowledge acquisition was slow, costly; massive knowledge base expensive to store  
Fall of Expert Systems – led to an “AI Winter” in funding
- 1986 Herbert Simon keynote at TIMS/ORSA (now INFORMS) Annual Meeting

## References:

- Nilsson NJ (2009) The quest for artificial intelligence (Cambridge University Press).
- Mitchell M (2019) Artificial Intelligence: A guide for thinking humans (Penguin UK).

# Artificial Intelligence – An (Over)Simplified History

- 1955 McCarthy J, Minsky ML, Rochester N, Shannon CE, “A proposal for the Dartmouth summer research project on artificial intelligence”
- 1970s Rise of Expert Systems – intended to mimic decision-making of domain experts  
simulated reasoning via if-then rules (knowledge base + inference rules)
- 1980s knowledge acquisition was slow, costly; massive knowledge base expensive to store  
Fall of Expert Systems – led to an “AI Winter” in funding
- 1986 Herbert Simon keynote at TIMS/ORSA (now INFORMS) Annual Meeting
- 1990s Rise of the commercial Internet; Introduction of Recurrent Neural Networks
- 2010s Machine learning (ML) technologies became prevalent. Particularly useful when trained on large data sets. Rise in data analysis, data analytics, and data science.
- 2017 Google paper “Attention is All You Need” introduces the Transformer for deep learning
- 2018 Generative Pre-trained Transformers (GPTs) as a type of large language model (LLM)
- 2022 ChatGPT (GPT3.2)

## References:

- Nilsson NJ (2009) The quest for artificial intelligence (Cambridge University Press).
- Mitchell M (2019) Artificial Intelligence: A guide for thinking humans (Penguin UK).

# How do GPT / LLMs work?

## Token-predicting machines

- Tokens embedded in a high-dim vector space
- Layered neural network models
- Attention heads in neural network adjust weights between token vectors
- Feed-forward layers use vector math to “reason” about tokens
- Models with billions (trillions) of parameters

## Requires scale!

- Trained on LOTS of data
- Using lots of compute (and energy)

Concern: LLMs nothing more than “stochastic parrots” ...?

## Large language models, explained with a minimum of math and jargon

Want to really understand how large language models work? Here's a gentle primer.



TIMOTHY B. LEE AND SEAN TROTT

JUL 27, 2023

<https://www.understandingai.org/p/large-language-models-explained-with>

See also

<https://arstechnica.com/science/2018/12/how-computers-got-shockingly-good-at-recognizing-images/>

ARTICLE | OPEN ACCESS



## On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

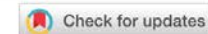


Authors: Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, Shmargaret Shmitchell [Authors](#)

[Info & Claims](#)

FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency • Pages 610 - 623  
<https://doi.org/10.1145/3442188.3445922>

Published: 01 March 2021 [Publication History](#)



# Proposed uses of AI for military decision-making

## (a partial list)

- **Search and Synthesis.** Google search on steroids. Of doctrine, instructions, and other official documents. For intelligence activities broadly. Supporting both summarization of data sources and integration across them.
- **(Descriptive) Assessment.** For military readiness, system health, inventory stockpiles, mission viability.
- **Detection and Vision.** Looking to identify (see) and classify (interpret) various data streams, to include text, images, and electronic or acoustic signals.
- **Pattern Recognition.** Techniques such as classification and clustering, to be used for both inference and prediction, and applied to tasks such as anomaly detection and threat analysis.
- **General information workflow automation.** Typically, with “Faster, Better, Cheaper (FBC)” efficiencies in mind.



# AI is now part of national security strategy



<https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2024/10/24/memorandum-on-advancing-the-united-states-leadership-in-artificial-intelligence-harnessing-artificial-intelligence-to-fulfill-national-security-objectives-and-fostering-the-safety-security/>



<https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/>



<https://www.war.gov/News/Releases/Release/Article/3996199/cdao-and-diu-launch-new-effort-focused-on-accelerating-dod-adoption-of-ai-capab/>



<https://www.war.gov/News/News-Stories/Article/Article/4165279/defense-officials-outline-ais-strategic-role-in-national-security/>

This is all great.

But how does it help (or hurt?) my work as an  
Operations Analyst?

And what exactly is that work, anyways...?

**OA  
Studies**



**Decision  
Support Tools**



**Automated  
Analytics**



## OA Studies



## Decision Support Tools



## Automated Analytics



### Deliverable

Report  
(and briefing)

### User / Customer

Executive  
Decision-Maker

### Use Case

One-time, bespoke  
  
Blends quantitative +  
qualitative factors

### Goal

Insight to  
specific decision(s)

### Example

In a South China Sea  
conflict, should we  
devote combatants to  
escorting merchant  
resupply vessels?

8 Naval Operations Analysis

104 The OA Method

Wagner DH, Mylander WC, Sanders TJ (1999) *Naval Operations Analysis* (Naval Institute Press), 3rd ed.

It has been asserted above that OA is essentially the application of scientific methods to the resolution of operational problems. Without digressing for a discussion of what constitutes scientific methods, it is reasonable to state that examination of successful applications of OA reveals a consistent pattern, a general form. That general approach is termed the **OA method**. It is by no means a problem-solving algorithm, but most if not all of its features are to be found in varying degrees of development in all cases of creative problem solving through OA. An outline of the OA method is as follows:

#### A. Formulation of the problem.

1. Identification of the objectives of the operation's decision-maker (may or may not be quantitative).
2. Identification of the reasonable alternative courses of action.
3. Identification of the variables that impact the courses of action.
4. Definition of a measure of effectiveness (MOE), i.e., a quantitative yardstick providing an ordering of the alternative courses of action that is consistent within the objective.

#### B. Analysis of the problem.

1. Construct a model of the operation by analytic formulas and/or Monte Carlo simulation (see Chapter 3) that is faithful to reality and amenable to analysis.
2. Evaluate, in terms of the MOE, outcomes of the alternative courses of action by exercising the model and by theoretical analysis.
3. Conduct operational trials or observation of "real world" operations to obtain data needed in (1) and (2).

#### C. Communication of the results, orally and in writing.

#### D. Analyst assistance to implementation of the decision.

|                        | <b>OA Studies</b>             | <b>Decision Support Tools</b>  | <b>Automated Analytics</b>  |
|------------------------|----------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|
| <b>Deliverable</b>     | Report<br>(and briefing)                                                                                       | Standalone Tool<br>(connected to data)                                                                           | (part of a)<br>Production System                                                                              |
| <b>User / Customer</b> | Executive<br>Decision-Maker                                                                                    | Operations Analyst<br>(domain-specific expert)                                                                   | “Production” Analyst<br>(frontline worker)                                                                    |
| <b>Use Case</b>        | One-time, bespoke<br><br>Blends quantitative +<br>qualitative factors                                          | Repeated, custom use<br><br>Uses data + computation for<br>domain-specific problem                               | Repeated, mass use<br><br>Part of a common<br>computational workflow                                          |
| <b>Goal</b>            | Insight to<br>specific decision(s)                                                                             | Accelerate + transform<br>complicated analysis                                                                   | Streamlined operations                                                                                        |
| <b>Example</b>         | In a South China Sea<br>conflict, should we<br>devote combatants to<br>escorting merchant<br>resupply vessels? | How best to schedule<br>training sessions at flight<br>school to meet<br>requirements?                           | How to integrate daily<br>satellite images into<br>readiness assessments?                                     |

Where can AI help?

|                 | OA Studies  | Decision Support Tools  | Automated Analytics  |
|-----------------|----------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| Deliverable     | Report<br>(and briefing)                                                                     | Standalone Tool<br>(connected to data)                                                                    | (part of a)<br>Production System                                                                       |
| User / Customer | Executive<br>Decision-Maker                                                                  | Operations Analyst<br>(domain-specific expert)                                                            | “Production” Analyst<br>(frontline worker)                                                             |
| Use Case        | One-time, bespoke<br><br>Blends quantitative +<br>qualitative factors                        | Repeated, custom use<br><br>Uses data + computation for<br>domain-specific problem                        | Repeated, mass use<br><br>Part of a common<br>computational workflow                                   |
| Goal            | Insight to<br>specific decision(s)                                                           | Accelerate + transform<br>complicated analysis                                                            | Streamlined operations                                                                                 |

Search & Synthesis

Data Analysis

Report Writing

Software Coding

|                                                                                                            |
|------------------------------------------------------------------------------------------------------------|
| <b>TECHNIQUES</b><br>(Optimization, Stochastic Modeling, Simulation, Statistics, Data Science, AI/ML, HPC) |
| <b>THEORY</b><br>(Mathematics, Algorithms, Computing)                                                      |



# The need for speed

AI's perceived value = it is believed to accelerate decision-making and action.

- Industry: Faster, Better, Cheaper (FBC) Pressures
- Military: Observe – Orient – Decide – Act (OODA) Loop

***“future wars will be won by those who can ‘see, understand, decide, and act faster,’ underscoring the necessity of integrating artificial intelligence (AI), electronic warfare, and space operations into military strategies, and ‘finding ways to combine AI tools and human decision making that deliver a decision advantage.’”***

Dec 10, 2024 SGL by Admiral Paparo; <https://nps.edu/-/indopacom-commander-discusses-challenges>

***“AI is not entirely new, but advancements in computing power and big data are transforming how we think about processes — not just acquisition, but our daily operations... AI can significantly enhance the Joint Staff's ability to integrate and analyze global military operations, ultimately **enabling better, faster decisions.**”*** – the Joint Staff AI Lead (April 24, 2025)

<https://www.war.gov/News/News-Stories/Article/Article/4165279/defense-officials-outline-ais-strategic-role-in-national-security/>

AI is valued as an ***automation technology***.

# Understanding automation technology

(a cognitive systems engineering perspective)

## Work as **imagined** (WAI)

- System is built and operated as designed
- Components of the system (humans, algorithms, devices) behave as specified
- Exceptions/Anomalies are relatively few & usually well anticipated.

## How we **imagine** automation

- Humans are the sources of inefficiency
- New technology can be introduced as a simple substitution of machines for people
- The system will be preserved and improved

## Work as **done** (WAD)

- Things are messy: “adaptations tailored to contingencies and context are always going on”
- “The adaptations that make the system function also hide the systems weaknesses.”
- “Management often can’t see the gaps so it seems that the system is functioning as designed.”
- Anomalies and surprises are continuous.

## How automation **actually** happens

- Adding or expanding the machine’s role changes the cooperative architecture and changes the role of the human in the system
- This is a **joint system**—and needs to be designed and operated as such

### References:

Hollnagel, Woods & Leveson (2006). Resilience Engineering. Woods et al., Behind Human Error (1994/ 2<sup>nd</sup> edition 2010). Woods and Decker (2000)  
Quotes from Woods et al. (2021). Patterns in How People Think and Work: Importance of Patterns Discovery for Understanding Complex Adaptive Systems..

# The field of cognitive systems engineering is full of examples of “automation gone wrong”

*Automatica*, Vol. 19, No. 6, pp. 775–779, 1983  
Printed in Great Britain.

0005-1098/83 \$3.00 + 0.00  
Pergamon Press Ltd.  
© 1983 International Federation of Automatic Control

Brief Paper

## Ironies of Automation\*

LISANNE BAINBRIDGE†

**Key Words**—Control engineering computer applications; man-machine systems; on-line operation; process control; system failure and recovery.

THEOR. ISSUES IN ERGON. SCI., 2000, VOL. 1, NO. 3, 272–282



## Anticipating the effects of technological change: a new era of dynamics for human factors

DAVID WOODS†\* and SIDNEY DEKKER‡

† Institute for Ergonomics, 210 Baker Systems, The Ohio State University, Columbus, OH 43210, USA

‡ Linköping Institute of Technology, S-58183 Linköping, Sweden

**Keywords:** Technology change; Cognitive task analysis; Future of Human Factors.

June 2016 - *Journal of Cognitive Engineering and Decision Making*

## The Risks of Autonomy: Doyle's Catch

David D. Woods, The Ohio State University

*Special Issue Paper*

## Limits of Automata—Then and Now: Challenges of Architecture, Brittleness, and Scale

David D. Woods<sup>1</sup>

*Journal of Cognitive Engineering and Decision Making*  
2024, Vol. 0(0) 1–8  
© 2024, Human Factors and Ergonomics Society  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/15553434241240203  
[journals.sagepub.com/home/edm](https://journals.sagepub.com/home/edm)

Sage

*Special Issue Paper*

## Wrong, Strong, and Silent: What Happens when Automated Systems With High Autonomy and High Authority Misbehave?

Sidney W. A. Dekker<sup>1</sup> and David D. Woods<sup>2</sup>

*Journal of Cognitive Engineering and Decision Making*  
2024, Vol. 0(0) 1–7  
© 2024, Human Factors and Ergonomics Society  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/15553434241240849  
[journals.sagepub.com/home/edm](https://journals.sagepub.com/home/edm)

Sage

# Towards Joint Activity Design Heuristics: Essentials for Human-Machine Teaming

Dane A. Morey<sup>1</sup> , Prerana Walli<sup>1</sup> , Kenneth S. Cassidy<sup>1</sup>,  
Priyanka K. Tewani<sup>1</sup>, Morgan E. Reynolds<sup>1</sup>, Samantha Malone<sup>1</sup>,  
Mohammadreza Jalaeian<sup>1</sup>, Michael F. Rayo<sup>1</sup>, and Nicolette M.  
McGeorge<sup>2</sup>

Proceedings of the Human Factors and  
Ergonomics Society Annual Meeting  
1–6

Copyright © 2023 Human Factors  
and Ergonomics Society  
DOI: 10.1177/21695067231193646  
journals.sagepub.com/home/pro



Johnson, M., & Vera, A. (2019). **No AI is an island:**  
the case for teaming intelligence. *AI magazine*, 40(1), 16-28.



## Key Ideas for Joint Systems:

- “The machine changes and augments what people can do, rather than replaces them.”
- “The machine interacts with people: they cannot remain separate or invisible”
- The Need for Teaming
- Role of Expertise

## COGNITIVE SYSTEMS ENGINEERING LAB

Innovation at the Intersection of People, Technology, and Work

ABOUT CSEL  
MEMBER DIRECTORY  
NEWS  
EVENTS  
RESEARCH  
COURSES  
INDUSTRY PARTNERSHIPS  
JOIN US



**CSEL**  
innovation at the intersection  
of people, technology, and work.

As researchers and designers, we identify patterns and leverage new technologies to support dynamic decision-making in the face of anomalies, uncertainty, stress, and even intentional deception by adversaries. Specifically, we are concerned with cognitive functions such as problem-solving, decision-making, attention, perception, and memory.

Microsoft Outlook

<https://u.osu.edu/csel/>

  
THE OHIO STATE  
UNIVERSITY  
COLLEGE OF ENGINEERING

# Assessing and Understanding Expertise

- The notion of “expertise” is often defined loosely.
- However, cognitive scientists\*\* have formal definitions of expertise in terms of:
  1. Developmental progression (*the experience needed* to acquire it)
  2. Knowledge organization (what is known and *how it is organized, e.g., models, abstractions*)
  3. Reasoning processes (*how experts think* and reason)
- Traditional levels of proficiency come from the “craft guilds” of the Middle Ages...

\*\*Leading expert: Robert R. Hoffman, PhD., Institute for Human and Machine Cognition, Florida University Systems, <http://www.ihmc.us/groups/rhoffman/>



# Basic Levels of Proficiency

|      | Traditional Description*                                                                                                                                                                                                                                                                                                                                                                                   | How they operate**                                                                                                                                                                                                                                                                                               | How they contribute**                                                                                                                                                   | How they learn and interact**                                                                                                                                                                                                               |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| HIGH | <b>Master</b><br>Traditionally, a master is <b>one of an elite group of experts whose judgments set the regulations, standards, or ideals.</b><br>Also, a master can be that expert who is regarded by the other experts as being “the” expert, or the “real” expert, especially with regard to sub-domain knowledge.                                                                                      | Has studied with many different teachers and has developed own distinctive style.<br>Has produced innovations in the standard practices of others, altered the course of history in the field, and knows how to do this again.                                                                                   | Develops new methods and practices for the field.<br>Capacity for long-range strategic thinking and action.<br>Sees historical drifts and shifting clearings.           | Learning continues by working with other masters as teachers.<br><b>Creates and leads professional networks.</b><br><b>Teaches others to be experts and masters.</b>                                                                        |
|      | <b>Expert</b><br>The distinguished or brilliant journeyman, highly regarded by peers, <b>whose judgments are uncommonly accurate and reliable, whose performance shows consummate skill and economy of effort, and who can deal effectively with certain types of rare or “tough” cases.</b> Also, an expert is one who has special skills or knowledge derived from extensive experience with subdomains. | Enormous breadth and depth of knowledge. Routinely forms and leads high-performance teams.<br>Admired by others as a benchmark of team performance.<br>Performance standards are well beyond those of most practitioners.                                                                                        | Consistently solves difficult, complex, problems.<br><b>Able to handle novel or unusual situations.</b><br>Produces consistently inspiring and excellent performances.  | Apprenticeship to masters.<br>Advanced coaching, development of breadth, focus on observing and adopting style of the teacher.<br><b>Teaches others. Years or decades of practice.</b>                                                      |
|      | <b>Journeyman (professional)</b><br><b>Literally, a person who can perform a day’s labor unsupervised,</b> although working under orders. An experienced and reliable worker, or one who has achieved a level of competence.<br>It is possible to remain at this level for life.                                                                                                                           | Appropriate action is deliberate yet appears to come from experience and intuition.<br><b>Seldom thinks in terms of rules for governing behavior.</b><br>Considerable experience and practice across a wide range of situations over years of work.                                                              | Operates effectively while unsupervised.<br><b>Able to deal with more complex situations.</b><br>Individual performance is a benchmark for others.                      | Apprenticeship to experts. Coaching.<br>Putting self into wide range of situations.<br><b>Membership and contribution to professional networks.</b><br><b>Teaches others.</b>                                                               |
|      | <b>Apprentice (advanced beginner)</b><br><b>Literally, one who is learning</b> —a student undergoing a program of instruction beyond the introductory level. Traditionally, the apprentice is immersed in the domain by living with and assisting someone at a higher level.<br>The length of an apprenticeship depends on the domain, ranging from about one to 12 years in the craft guilds.             | Carries out standard actions without causing breakdowns. Can fulfill standard promises to customers satisfactorily without supervision.<br><b>Performs most standard actions without conscious application of rules.</b> When faced with a new situation, works out appropriate actions by application of rules. | Able to perform advanced problem-solving on projects when coached to do so.<br>Can assist someone at a higher level.                                                    | <b>Repeated practice with common situations and increasing exposure to exceptional situations.</b><br>Apprenticeship to more advanced professionals and teams.<br><b>Membership in professional networks</b>                                |
| LOW  | <b>Novice (beginner)</b><br><b>Literally, someone who is new</b> —a probationary member.<br>There has been some (“minimal”) exposure to the domain and/or has begun introductory instruction.                                                                                                                                                                                                              | <b>All action appears to be governed by rules defining allowable actions and strategies.</b> Common situations are unfamiliar and are described by more rules.<br><b>Most action is deliberate application of rules</b> or conscious recall of prior actions in the familiar situations.                         | <b>Can contribute in realistic, well-understood situations with supervision.</b><br>Can perform simple actions for customers; needs supervision for more complex tasks. | <b>Memorization, drill, and practice in simple situations.</b><br>Problem-solving and practice with rules and strategies.<br><b>Learning involves recognizing common situations that help in recalling which rules should be exercised.</b> |

\*Taken verbatim from: Hoffman, R. R. (1998). How can expertise be defined?: Implications of research from cognitive psychology. In R. Williams, W. Faulkner, & J. Fleck (Eds.), Exploring expertise (pp. 81-100). New York: Macmillan.

\*\*Adapted from: Denning PJ (2002) Career redux. Communications of the ACM 45(9):21-26.

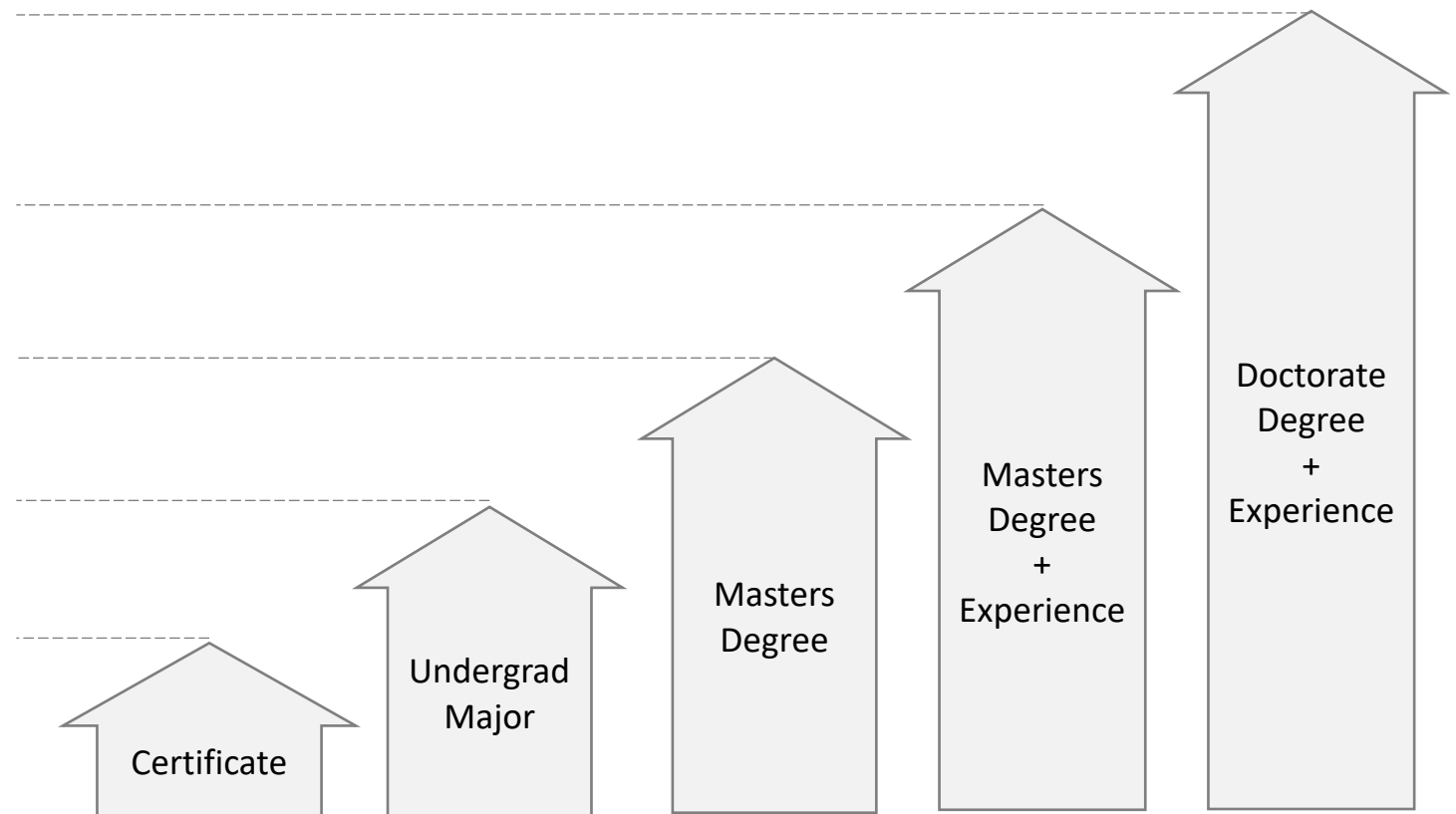


# Basic Levels of Proficiency

| <div> <div>HIGH</div> <div>↑</div> <div>LOW</div> </div> | Traditional Description*                                                                                                                                                                                                                                                                                                                                                                            |
|----------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                          | <b>Master</b><br>Traditionally, a master is one of an elite group of experts whose judgments set the regulations, standards, or ideals.<br>Also, a master can be that expert who is regarded by the other experts as being “the” expert, or the “real” expert, especially with regard to sub-domain knowledge.                                                                                      |
|                                                          | <b>Expert</b><br>The distinguished or brilliant journeyman, highly regarded by peers, whose judgments are uncommonly accurate and reliable, whose performance shows consummate skill and economy of effort, and who can deal effectively with certain types of rare or “tough” cases. Also, an expert is one who has special skills or knowledge derived from extensive experience with subdomains. |
|                                                          | <b>Journeyman (professional)</b><br>Literally, a person who can perform a day’s labor unsupervised, although working under orders. An experienced and reliable worker, or one who has achieved a level of competence.<br>It is possible to remain at this level for life.                                                                                                                           |
|                                                          | <b>Apprentice (advanced beginner)</b><br>Literally, one who is learning—a student undergoing a program of instruction beyond the introductory level. Traditionally, the apprentice is immersed in the domain by living with and assisting someone at a higher level.<br>The length of an apprenticeship depends on the domain, ranging from about one to 12 years in the craft guilds.              |
|                                                          | <b>Novice (beginner)</b><br>Literally, someone who is new—a probationary member.<br>There has been some (“minimal”) exposure to the domain and/or has begun introductory instruction.                                                                                                                                                                                                               |

We can use these levels to understand how education and experience contribute to the acquisition of proficiency in the field of operations analysis.

If an officer starts with little prior experience, the level of proficiency that can be attained scales approximately with the amount of education.



*Note: Someone with a master’s degree of education is typically not a “master” in this sense. Just as the rank of Master Chief Petty Officer follows a different notion of what it means to be a “master.”*

# What happens to skills when using automated tools?

Three possible outcomes:

- **The skill is enhanced.** This tends to occur in individuals who already possess a certain level of proficiency and are engaged in activities that accelerate their practice.
- **The skill atrophies over time.** This is the classic *de-skilling paradox*.
- **The skill never develops** in the first place.

Areas of concern:

- Education: Reading. Analysis. Writing. Thinking.
- [Software-Based] Tool Development
- Information-Based Work
- Decision-making in high-stakes environments

## HIGHER EDUCATION

# Everyone Is Cheating Their Way Through College

ChatGPT has unraveled the  
entire academic project. 🗨️

BY JAMES D. WALSH

New York Magazine, Retrieved May 11, 2025  
<https://nymag.com/intelligencer/article/openai-chatgpt-ai-cheating-education-college-students-school.html>

## Addressing the Transactional Model of School

We need to attack thoughtless ChatGPT use from the demand side.



JOHN WARNER  
MAY 11, 2025

<https://biblioracle.substack.com/p/addressing-the-transactional-model>

## Basic Argument:

- Education has increasingly adopted a “transactional” model (i.e., the student submits an assignment in exchange for a grade).
- Students, under their own form of FBC pressures, will naturally seek out tools that make them more productive
- One should expect the rampant use of these tools to optimize their grades, even if it ultimately circumvents their learning

## The Myth of Automated Learning

AI's real threat to education.



NICHOLAS CARR  
MAY 27, 2025

<https://www.newcartographies.com/p/the-myth-of-automated-learning>

## Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity

Joel Becker\*, Nate Rush\*, Beth Barnes, David Rein

Model Evaluation & Threat Research (METR)

### Abstract

Despite widespread adoption, the impact of AI tools on software development in the wild remains understudied. We conduct a randomized controlled trial (RCT) to understand how AI tools at the February–June 2025 frontier affect the productivity of experienced open-source developers. 16 developers with moderate AI experience complete 246 tasks in mature projects on which they have an average of 5 years of prior experience. Each task is randomly assigned to allow or disallow usage of early-2025 AI tools. When AI tools are allowed, developers primarily use Cursor Pro, a popular code editor, and Claude 3.5/3.7 Sonnet. Before starting tasks, developers forecast that allowing AI will reduce completion time by 24%. After completing the study, developers estimate that allowing AI reduced completion time by 20%. Surprisingly, we find that allowing AI actually *increases* completion time by 19%—AI tooling slowed developers down. This slowdown also contradicts predictions from experts in economics (39% shorter) and ML (38% shorter). To understand this result, we collect and evaluate evidence for 21 properties of our setting that *a priori* could contribute to the observed slowdown effect—for example, the size and quality standards of projects, or prior developer experience with AI tooling. Although the influence of experimental artifacts cannot be entirely ruled out, the robustness of the slowdown effect across our analyses suggests it is unlikely to primarily be a function of our experimental design.

### 1 Introduction

Software development is an important part of the modern economy, and a key domain for understanding and forecasting AI capabilities [1; 2]. Frontier AI systems demonstrate impressive capabilities on a wide range of software benchmarks [3; 4; 5; 6; 7; 8; 9] and in experiments measuring AI’s impact on developer productivity when completing synthetic tasks [10; 11]. However, tasks used in these *lab experiments* sacrifice realism for scale and efficiency: the tasks are typically self-contained, do not require much prior context/familiarity to understand and complete, and use algorithmic evaluation metrics which do not capture many important capabilities [12; 13; 14]. As a result, it can be difficult to draw inferences from results on these evaluations about AI’s impact in practice.

To reduce the inferential gap between measurements of AI capabilities and real-world impact, one can measure the impact of AI systems in real-world settings (i.e. *field experiments*). Existing field experiments aimed at measuring AI’s impact on software development measure outcomes like number of added lines of code or number of tasks completed [15; 16; 17]. However, AI systems can affect these outcomes without productivity actually increasing—for example, code can be more verbose but functionally equivalent, and tasks can be broken up into multiple smaller tasks without the total amount of work changing—making it challenging to interpret these results.

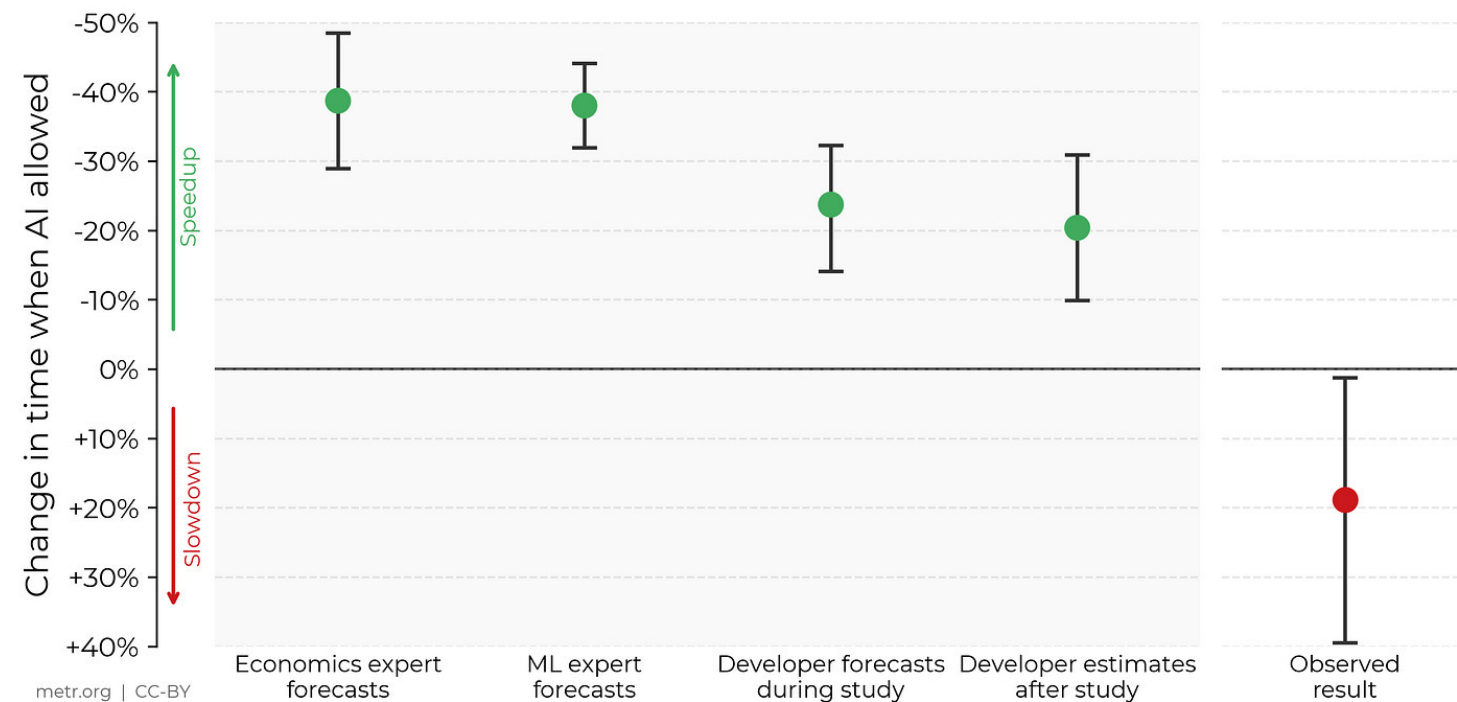
\*Equal contribution. Correspondence to {nate, joel}@metr.org

# AI tools might degrade performance in software coding activities

## Against Expert Forecasts and Developer Self-Reports, Early-2025 AI Slows Down Experienced Open-Source Developers



In this RCT, 16 developers with moderate AI experience complete 246 tasks in large and complex projects on which they have an average of 5 years of prior experience.



**Experienced software developers with access to AI tools took 19% longer to complete their tasks, despite believing they had finished 20% faster.**



# Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity

Joel Becker\*, Nate Rush\*, Beth Barnes, David Rein

Model Evaluation & Threat Research (METR)

## Abstract

Despite widespread adoption, the impact of AI tools on software development in the wild remains understudied. We conduct a randomized controlled trial (RCT) to understand how AI tools at the February–June 2025 frontier affect the productivity of experienced open-source developers. 16 developers with moderate AI experience complete 246 tasks in mature projects on which they have an average of 5 years of prior experience. Each task is randomly assigned to allow or disallow usage of early-2025 AI tools. When AI tools are allowed, developers primarily use Cursor Pro, a popular code editor, and Claude 3.5/3.7 Sonnet. Before starting tasks, developers forecast that allowing AI will reduce completion time by 24%. After completing the study, developers estimate that allowing AI reduced completion time by 20%. Surprisingly, we find that allowing AI actually *increases* completion time by 19%—AI tooling slowed developers down. This slowdown also contradicts predictions from experts in economics (39% shorter) and ML (38% shorter). To understand this result, we collect and evaluate evidence for 21 properties of our setting that *a priori* could contribute to the observed slowdown effect—for example, the size and quality standards of projects, or prior developer experience with AI tooling. Although the influence of experimental artifacts cannot be entirely ruled out, the robustness of the slowdown effect across our analyses suggests it is unlikely to primarily be a function of our experimental design.

## 1 Introduction

Software development is an important part of the modern economy, and a key domain for understanding and forecasting AI capabilities [1; 2]. Frontier AI systems demonstrate impressive capabilities on a wide range of software benchmarks [3; 4; 5; 6; 7; 8; 9] and in experiments measuring AI’s impact on developer productivity when completing synthetic tasks [10; 11]. However, tasks used in these *lab experiments* sacrifice realism for scale and efficiency: the tasks are typically self-contained, do not require much prior context/familiarity to understand and complete, and use algorithmic evaluation metrics which do not capture many important capabilities [12; 13; 14]. As a result, it can be difficult to draw inferences from results on these evaluations about AI’s impact in practice.

To reduce the inferential gap between measurements of AI capabilities and real-world impact, one can measure the impact of AI systems in real-world settings (i.e. *field experiments*). Existing field experiments aimed at measuring AI’s impact on software development measure outcomes like number of added lines of code or number of tasks completed [15; 16; 17]. However, AI systems can affect these outcomes without productivity actually increasing—for example, code can be more verbose but functionally equivalent, and tasks can be broken up into multiple smaller tasks without the total amount of work changing—making it challenging to interpret these results.

\*Equal contribution. Correspondence to {nate, joel}@metr.org

# AI tools might degrade performance in software coding activities

## Days before OpenAI



## Days after OpenAI



# The rise of AI “workslop”

Generative AI

## AI-Generated “Workslop” Is Destroying Productivity

by Kate Niederhoffer, Gabriella Rosen Kellerman, Angela Lee, Alex Liebscher, Kristina Rapuano and Jeffrey T. Hancock

September 22, 2025, Updated September 25, 2025



<https://hbr.org/2025/09/ai-generated-workslop-is-destroying-productivity>

- “But while some employees are using [AI-enabled tools] to polish good work, others use it to create content that is actually unhelpful, incomplete, or missing crucial context about the project at hand. “
- **“Employees are using AI tools to create low-effort, passable looking work that ends up creating more work for their coworkers.”**
- “We define workslop as *AI generated work content that masquerades as good work, but lacks the substance to meaningfully advance a given task.*”
- “The insidious effect of workslop is that it shifts the burden of the work downstream, requiring the receiver to interpret, correct, or redo the work. In other words, it transfers the effort from creator to receiver.”



All &gt; Technology &amp; Research

# How AI Can Degrade Human Performance in High-Stakes Settings

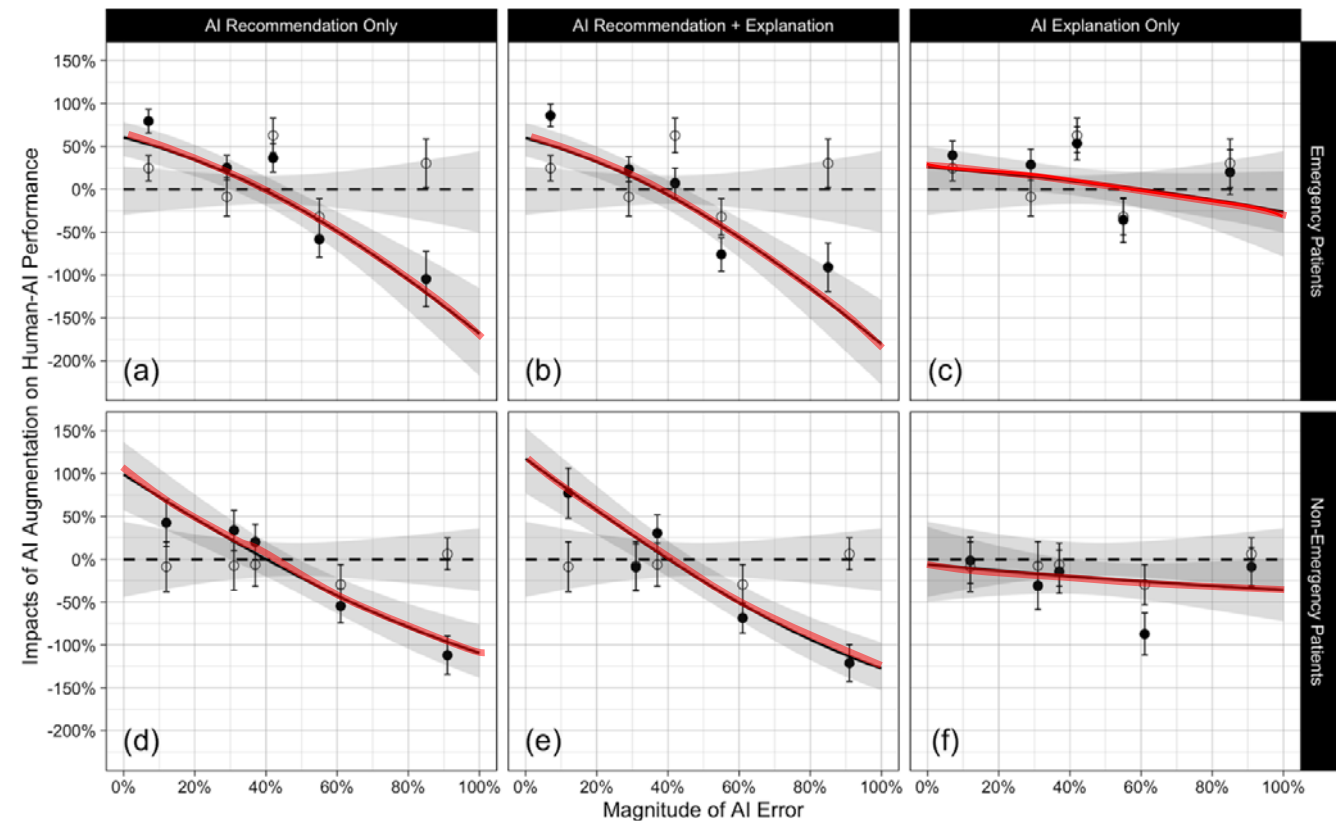
Across disciplines, bad AI predictions have a surprising tendency to make human experts perform worse.

Dane A. Morey, Mike Rayo and David Woods — Jul 15, 2025



## Performance of 450 nursing students and a dozen licensed nurses when reviewing 10 historical ICU cases

Experimental Condition — With AI Augmentation — Without AI Augmentation



- “When AI predictions were most correct, nurses performed 53% to 67% better than when they worked without AI assistance.”
- “However, when AI predictions were most misleading, nurses performed 96% to 120% worse than when they worked without AI assistance.”

# What happens to skills when using automated tools?

Three possible outcomes:

- **The skill is enhanced.** This tends to occur in individuals who already possess a certain level of proficiency and are engaged in activities that accelerate their practice.
- **The skill atrophies over time.** This is the classic *de-skilling paradox*.
- **The skill never develops** in the first place.

## Expertise is more than generating a minimally viable product!

- Expertise cannot be automated
  - Expertise involves knowledge and skill, as well as the ability to cope with the unexpected.
  - Being able to handle things that go wrong (increasingly complex situations)
  - These things all derive from experience.
- Performance depends on both the user and the tool
  - This is a joint system, and should be designed as such

# **We should not expect that the mere deployment of decision support tools necessarily improves performance.**

- It is insufficient (and misguided) to evaluate the performance of the tool in a standalone way.

Instead, we should be evaluating the joint performance of user + tool.

- In cases where a decision support tool makes incorrect (or poor) recommendations, we should NOT expect that users (even those with high expertise) will be able to say "that's wrong" and disregard the guidance of the tool.

In high-stakes situations involving life and death, we should insist on rigor in evaluating the correctness of model/tool output.

- How do we assess performance of LLMs where the output always looks plausible?
- Are the fundamental limitations in terms of what LLMs can do?

# Generative AI's crippling and widespread failure to induce robust models of the world

LLM failures to reason, as documented in Apple's Illusion of Thinking paper, are really only part of a much deeper problem



GARY MARCUS

JUN 28, 2025

<https://garymarcus.substack.com/p/generative-ais-crippling-and-widespread>

*“A **world model** (or cognitive model) is a computational framework that a system (a machine, or a person or other animal) uses to track what is happening in the world.”*

Each of us builds, maintains, and revises our own models for what is happening and how we make sense of the world around us.

The development of expertise is marked by how we organize information (i.e., build models) for improved retrieval and reasoning about what is happening and what might happen.

- **GPT-based LLMs do not build “world models”** in this way
- They are token-predicting machines (sometimes called “foundation models”—an unfortunate name)
- This limitation helps to explain their inability to effectively solve problems in many contexts
- **This is fundamental limitation** (i.e., it will not be solved with more data and more compute)



# Generative AI's crippling and widespread failure to induce robust models of the world

LLM failures to reason, as documented in Apple's Illusion of Thinking paper, are really only part of a much deeper problem



GARY MARCUS

JUN 28, 2025

<https://garymarcus.substack.com/p/generative-ais-crippling-and-widespread>

## What Has a Foundation Model Found? Using Inductive Bias to Probe for World Models

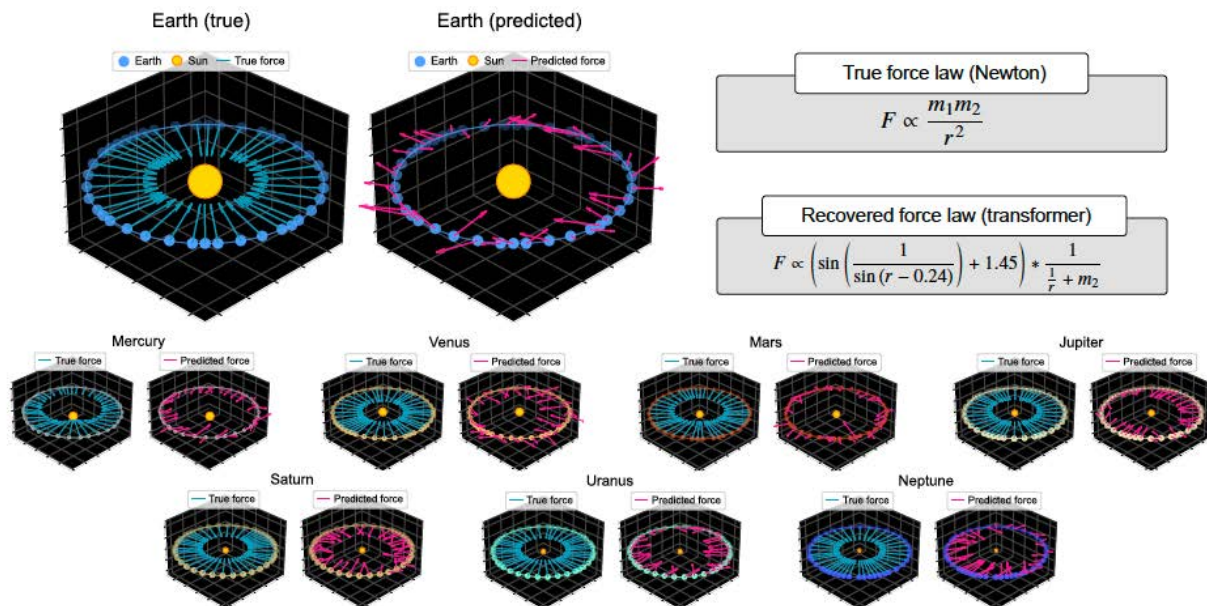


Figure 1: Each pair of panels illustrates the trajectory of a planet in the solar system and its gravitational force vectors, comparing the true Newtonian forces (left) to the predicted forces (right) from a transformer foundation model pretrained on orbital sequences and fine-tuned to predict forces. While the model excels at generating accurate predictions of planetary trajectories, it does not have an inductive bias toward true Newtonian mechanics; moreover, its force predictions recover a nonsensical force law, as revealed by symbolic regression.

## What Has a Foundation Model Found? Using Inductive Bias to Probe for World Models

Keyon Vafa<sup>1</sup> Peter G. Chang<sup>2</sup> Ashesh Rambachan<sup>2</sup> Sendhil Mullainathan<sup>2</sup>

### Abstract

Foundation models are premised on the idea that sequence prediction can uncover deeper domain understanding, much like how Kepler's predictions of planetary motion later led to the discovery of Newtonian mechanics. However, evaluating whether these models truly capture deeper structure remains a challenge. We develop a technique for evaluating foundation models that examines how they adapt to synthetic datasets generated from some postulated world model. Our technique measures whether the foundation model's inductive bias aligns with the world model, and we call this an *inductive bias probe*. A foundation model's inductive bias is a task-specific heuristic that fails to generalize to new physics tasks.

gerich, 2004). This path — from predicting sequences to understanding the deeper mechanisms that underlie them — is not unique to physics. In biology, animal breeders noticed patterns in the traits of offspring long before their predictive insights inspired Mendel to develop a theory of genetics.

How would we know if foundation models have also made the leap from making accurate predictions to developing reliable world models? This paper develops a framework for answering this question. Specifically, we create a procedure that, when given a foundation model and world model, tests whether the foundation model has learned that world model. We call this technique an *inductive bias probe*, and it is built on a simple insight: the implicit world model of a

*“We particularly find that foundation models trained on orbital trajectories consistently fail to apply Newtonian mechanics when adapted to new physics tasks. Further analysis reveals that these models behave as if they develop task-specific heuristics that fail to generalize.”*

### 1. Introduction

The promise of foundation models is that they can uncover deeper truths, but this idea is new in one sense, it is old in another. Hundreds of years ago, astronomers like Kepler discovered geometric patterns that could pinpoint the future locations of planets in the night sky. Newton would later expand on this progress to develop Newtonian mechanics, fundamental laws that could not only predict the movement of planets but also explain physical properties across the universe (Koestler, 1959; Ginzburg, 2004).

model is toward Newtonian mechanics, it should have an inductive bias towards these force vectors. In contrast, Figure 1 shows that the model produces poor force vectors. More extremely, when we perform this exercise at a larger scale across many solar systems, the laws of gravity it uses to generalize bear no resemblance to Newton's law (Table 1).

We further apply inductive bias probes in other domains with a known world model: lattice problems and Othello games (Liu et al., 2022; Hazineh et al., 2023; Nanda et al., 2023b; Vafa et al., 2024). Across these domains, we find that neural sequence models have weak inductive biases

<sup>1</sup>Harvard University <sup>2</sup>MIT. Correspondence to: Keyon Vafa <kvafa@g.harvard.edu>.

arXiv:2507.06952v2 [cs.LG] 10 Jul 2025



# Generative AI's crippling and widespread failure to induce robust models of the world

LLM failures to reason, as documented in Apple's Illusion of Thinking paper, are really only part of a much deeper problem



GARY MARCUS

JUN 28, 2025

<https://garymarcus.substack.com/p/generative-ais-crippling-and-widespread>



*Synthesized video from Dawid van Straaten, prompt ("Generate me a video of two men playing chess") in which the player for black reaches across the table and, in the midst of a rather unusual position moves his opponent's pawn horizontally, and quite illegally, several squares across the board.*

"In short ChatGPT can *approximate* the game of chess, but it can't play it *reliably*, precisely because (despite immense relevant evidence) it never induces proper world model of the board and the rules."

## Chap claims Atari 2600 'absolutely wrecked' ChatGPT at chess

1.19MHz eight-bit CPU trounced modern GPUs – can you do better with your retro-tech?

[Simon Sharwood](#)

Mon 9 Jun 2025 / 06:29 UTC

The Atari 2600 gaming console came into the world in 1977 with an eight-bit processor that ran at 1.19MHz, and just 128 bytes of RAM – but that's apparently enough power to beat ChatGPT at chess.

So says infrastructure architect Robert Caruso, who over the weekend [posted](#) the results of an experiment he conducted to "pit ChatGPT against the Atari 2600's chess engine (via Stella emulator) and see what happens."

**“ ChatGPT confused rooks for bishops, and repeatedly lost track of pieces**

Caruso decided to run the experiment after conversing with ChatGPT about the history of chess. At some point in that chat, the bot volunteered to play against the Atari – a reasonable suggestion as "Video Chess" was one of the games Atari commissioned for its console.

Online chat in which chess works discuss the merits of Video Chess suggest it may have played at a level beginners may have found challenging, and perhaps gave regular recreational players of intermediate skill a little to worry about.

Caruso thought his experiment would be "a lighthearted stroll down retro memory lane."

Instead, he watched as the Atari humiliated ChatGPT.

[https://www.theregister.com/2025/06/09/atari\\_vs\\_chatgpt\\_chess/](https://www.theregister.com/2025/06/09/atari_vs_chatgpt_chess/)

# Mathematical Optimization (MO)

The starting point for a MO problem is model formulation:

- Index sets
- Parameters
- Variables
- Objective
- Constraints

Building a “good” representation of the problem (i.e., a world model) is the essential first step toward being able to optimize a decision.

This requires an understanding of the underlying mathematics and domain expertise in the problem at hand.

The benefits and advantages of having a “world model” are immediate.

**Mathematical Optimization (MO)**

**vs.**

**Large Language Models (LLMs)**

## Mathematical Optimization (MO)

vs.

## Large Language Models (LLMs)

- When MO finds an optimal solution, it provides a certificate that guarantees it is the best. In other cases, it can potentially tell you how far away the solution is from the best one.

- MO will tell you if a problem is infeasible. If infeasible, the model can be interrogated to understand what would be required to make it feasible.

- If an MO problem goes wrong, there is typically a signal about what went wrong and what to do.

- The results of MO are repeatable. Different solves (either by different people or at different times) will get the same answer.

- MO is transparent in the sense that we can understand how the solution was obtained and explain why it is good.

- We can use MO to do parametric analysis: understand how the best solution changes as inputs (e.g., available resources) or requirements change. MO supports systematic “what-if” analysis that allows users to explore future uncertainties.

- We can use MO to find worst-case failures in operational plans, either for attack or defense.

- MO has mature theory that is decades old and is understood and supported by thousands of practitioners worldwide.

- An LLM provides no guarantees on its output and requires a human to interpret the results to determine whether the solution is any good or a nonsense hallucination.

- An LLM can’t say whether a problem is infeasible or not and will likely produce a plausible-sounding answer in either case.

- LLMs don’t provide such alarm signals.

- LLMs are not stable under repeated invocations. Different solves (either by different people or at different times) are likely to get different answers.

- LLMs are not transparent. It is unclear how or why they produce their results, and explaining their output can be difficult.

- LLMs do not support this.

- LLMs do not support this.

- LLMs are poorly understood, relatively speaking, and it is unclear who to call when things don’t work out as planned.

# Potential Limits in the current AI Gold Rush?

## In hindsight

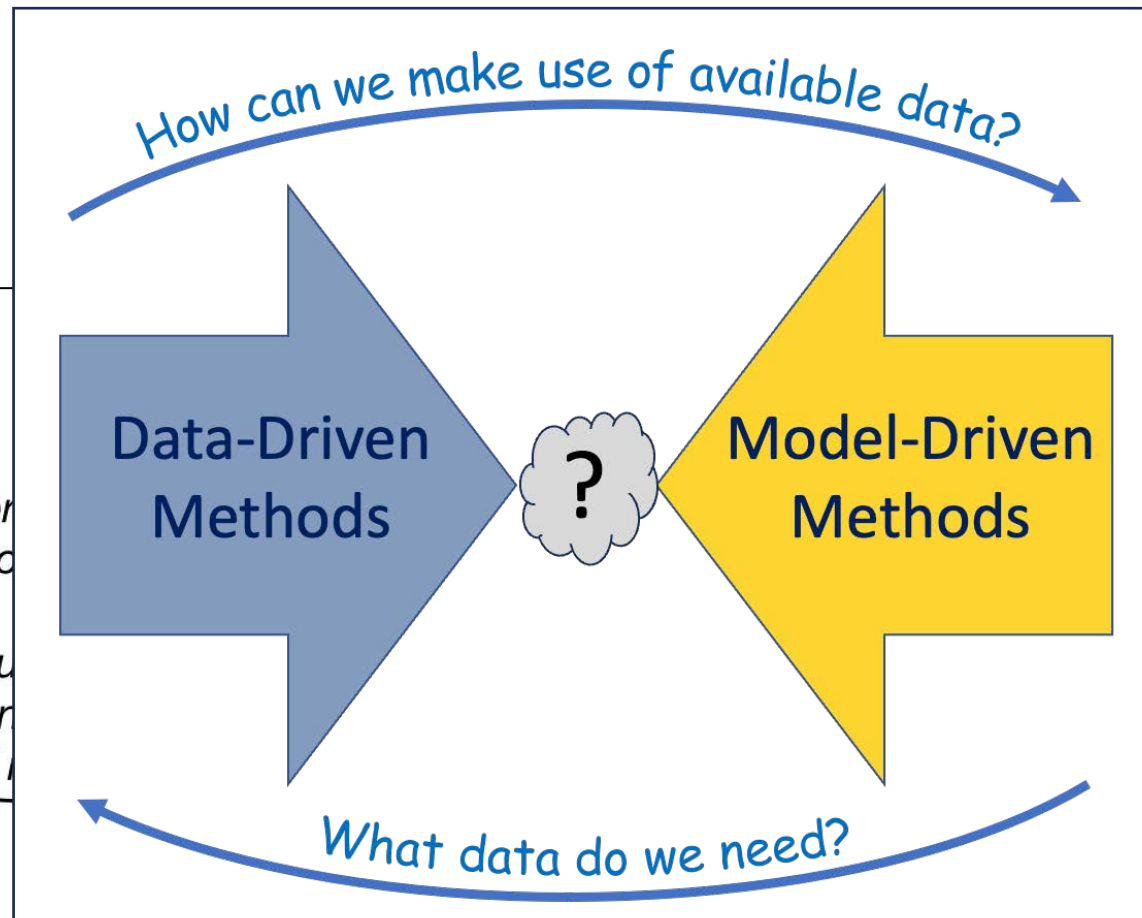
- **The bitter lesson about the bitter lesson** is that scaling works for some problems — mainly pattern recognition — but not for others
- Scaling has made GenAI much better at what it is good at, but it has also made it much worse at its core weaknesses around reliability, hallucination, reasoning, poor planning, etc, or its challenges in the face of novel problems

### Problems scale helps with

Object recognition, autocomplete, voice recognition, brainstorming, etc

### Problems scale doesn't help with

open-ended problems that cannot be beaten by brute force; reasoning with incomplete information, etc





# Making Sense of the AI Gold Rush: An OA Perspective

- Operations Analysis (OA)
- Artificial Intelligence (AI)
- Automation
- Expertise

The need to frame problems, analyze data, and provide insights to real-world operational problems will only increase.

AI-enabled tools will continue to succeed in contexts where data-driven methods are appropriate, but we should not expect them to circumvent the need for models or expertise.

You need to know how these tools work and where their use is likely to succeed (or fail). Be prepared to clean up slop.

We should be wary of the automation substitution myth, or that moving fast (on its own) produces superior results.

Do not be fooled into thinking that you can automate the development of your own expertise.

# I welcome comments and feedback. Thank you!

- Dr. David Alderson  
Professor and Chair, Operations Research  
Naval Postgraduate School  
dlalders@nps.edu  
<http://faculty.nps.edu/dlalders>



# References (1 of 2)

- Bainbridge L (1983) Ironies of automation. *Automatica* 19(6):775–779.
- Becker J, Rush N, Barnes E, Rein D (2025) Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity. URL <https://arxiv.org/abs/2507.09089>.
- Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 610–623.
- Blackett PMS (1950) Operational research. *Journal of the Operational Research Society* 1(1):3–6.
- Carr N (2025) The myth of automated learning. URL <https://www.newcartographies.com/p/the-myth-of-automated-learning>, Substack: New Cartographies; May 27, 2025.
- Dekker SW, Woods DD (2024) Wrong, strong, and silent: What happens when automated systems with high autonomy and high authority misbehave? *Journal of Cognitive Engineering and Decision Making* 18(4):339–345.
- Denning PJ (2002) Career redux. *Communications of the ACM* 45(9):21–26.
- Department of Defense, US (2024) CDAO and DIU Launch New Effort Focused on Accelerating DOD Adoption of AI Capabilities. press release, December 11, <https://www.defense.gov/News/Releases/Release/Article/3996199/cdao-and-diu-launch-new-effort-focused-on-accelerating-dod-adoption-of-ai-capab/>.
- Department of Navy, US (2024) Chief of Naval Operations Navigation Plan for America’s Warfighting Navy 2024. <https://www.navy.mil/leadership/chief-of-naval-operations/cno-navplan-2024/>.
- Greenstein S (2023) The AI Gold Rush. *IEEE Micro* 43(6):126–128.
- Hoffman RR (1998) How can expertise be defined? implications of research from cognitive psychology. *Exploring expertise: Issues and perspectives*, 81–100 (Springer).
- Hollnagel E, Woods D, Leveson N, eds. (2006) *Resilience Engineering: Concepts and Precepts* (Aldershot, UK: Ashgate Press).
- Johnson M, Vera A (2019) No AI is an island: the case for teaming intelligence. *AI magazine* 40(1):16–28.
- Lee TB (2018) How computers got shockingly good at recognizing images. *ARS Technica* URL <https://arstechnica.com/science/2018/12/how-computers-got-shockingly-good-at-recognizing-images/>, December 18, 2018.
- Lee TB, Trott S (2023) Large language models explained with a minimum of math and jargon. URL <https://www.understandingai.org/p/large-language-models-explained-with>, Substack: Understanding AI; July 27, 2023.
- Marcus G (2025a) Game over for pure LLMs. Even Turing Award Winner Rich Sutton has gotten off the bus. URL <https://garymarcus.substack.com/p/game-over-for-pure-llms-even-turing>, Substack: Marcus on AI; September 26, 2025.
- Marcus G (2025b) Generative AI’s crippling and widespread failure to induce robust models of the world. URL <https://garymarcus.substack.com/p/generative-ais-crippling-and-widespread>, Substack: Marcus on AI; June 28, 2025.
- McCarthy J, Minsky ML, Rochester N, Shannon CE (2006) A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine* 27(4):12–12.
- Milkman RH (1968) *Operations Research in World War II*. U.S. Naval Institute Proceedings 94/5/783.
- Mitchell M (2019) *Artificial Intelligence: A guide for thinking humans* (Penguin UK).

# References (2 of 2)

- Morey DA, Rayo M, Woods D (2025) How AI can degrade human performance in high-stakes settings. AI Frontiers URL <https://ai-frontiers.org/articles/how-ai-can-degrade-human-performance-in-high-stakes-settings>.
- Morey DA, Walli P, Cassidy KS, Tewani PK, Reynolds ME, Malone S, Jalaeian M, Rayo MF, McGeorge NM (2023) Towards joint activity design heuristics: Essentials for human-machine teaming. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, volume 67, 131–136 (SAGE Publications Sage CA: Los Angeles, CA)
- Morse PM, Kimball GE (1946) Methods of Operations Research. Technical report, Office of Scientific Research and Development, Washington, D.C., summary Technical Report of Division 6, National Defense and Research Committee, Volume 2A.
- Niederhoffer K, Kellerman GR, Lee A, Liebscher A, Rapuano K, Hancock JT (2025) AI-Generated “Workslop” Is Destroying Productivity. Harvard Business Review URL <https://hbr.org/2025/09/ai-generated-workslop-is-destroying-productivity>, september 22, 2025, Updated September 25, 2025.
- Nilsson NJ (2009) The quest for artificial intelligence (Cambridge University Press).
- Sharwood S (2025) Chap claims Atari 2600 ‘absolutely wrecked’ ChatGPT at chess. The Register URL [https://www.theregister.com/2025/06/09/atari\\_vs\\_chatgpt\\_chess/](https://www.theregister.com/2025/06/09/atari_vs_chatgpt_chess/), june 9, 2025.
- Shinego W (2025) Defense Officials Outline AI’s Strategic Role in National Security. DOD News April 24.
- Simon HA (1987) Two heads are better than one: the collaboration between AI and OR. Interfaces 17(4):8–15.
- The White House (2024) Memorandum on Advancing the United States’ Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence. official memorandum, October 24, <https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2024/10/24/memorandum-on-advancing-the-united-states-leadership-in-artificial-intelligence-harnessing-artificial-intelligence-to-fulfill-national-security-objectives-and-fostering-the-safety-security/>.
- The White House (2025) REMOVING BARRIERS TO AMERICAN LEADERSHIP IN ARTIFICIAL INTELLIGENCE. Executive Order, January 23, <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/>.
- Vafa K, Chang PG, Rambachan A, Mullainathan S (2025) What has a foundation model found? Using inductive bias to probe for world models. URL <https://arxiv.org/abs/2507.06952>.
- Wagner DH, Mylander WC, Sanders TJ (1999) Naval Operations Analysis (Naval Institute Press), 3rd edition.
- Walsh J (2025) Everyone is cheating their way through college. New York Magazine Retrieved May 11, 2025.
- Warner J (2025) Addressing the transactional model of school. URL <https://biblioracle.substack.com/p/addressing-the-transactional-model>, Substack: The Biblioracle Recommends; May 11, 2025.
- Woods DD (2016) The risks of autonomy: Doyle’s catch. Journal of Cognitive Engineering and Decision Making 10(2):131–133
- Woods D, Dekker S (2000) Anticipating the effects of technological change: A new era of dynamics for human factors. Theoretical issues in ergonomics science 1(3):272–282.
- Woods D, Johannesen LJ, Cook RI, Sarter NB (1994) Behind human error: Cognitive systems, computers, and hindsight (CSERIAC Dayton, OH).
- Woods D, Licu T, Leonhardt J, Rayo M, Balkin E, Ciponea R (2021) Patterns in how people think and work: importance of patterns discovery for understanding complex adaptive systems. EUROCONTROL.
- Woods DD (2024) Limits of automata—then and now: Challenges of architecture, brittleness, and scale. Journal of Cognitive Engineering and Decision Making 18(4):394–401.