

Stanford University Human-Centered Artificial Intelligence

San Mateo County Board Retreat

Vanessa Parli Director of Research and Education Programs March 4, 2025

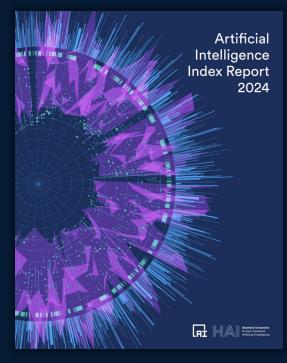
Our Goal

To advance Al **research**, **education**, **policy** and **practice** to improve the human condition Al technology should be inspired by human intelligence The development of Al must be guided by its human impact

Applications of Al should enhance and **augment humans**, not replace them **FOMO**: a state of mental or emotional strain caused by the "fear of missing out".

AI FOMO

2024 7th Edition



COMING SOON 8th Edition













Al is getting far better

Midjourney generations over time: "a hyper-realistic image of Harry Potter" Source: Midjourney, 2023



V1, February 2022



V2, April 2022

V3, July 2022



V5, March 2023



V5.1, March 2023



V5.2, June 2023



V6, December 2023

Al is getting far better



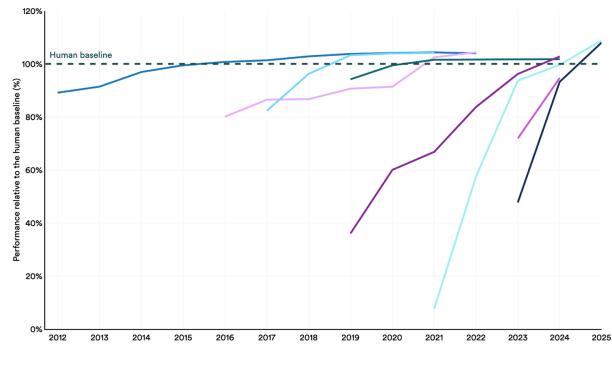
V1, February 2022

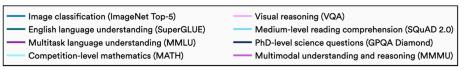
V6, December 2023

Al meets or beats humans on some tasks

Select AI Index technical performance benchmarks vs. human performance

Source: Al Index, 2025 | Chart: 2025 Al Index report





Humanity's Last Exam

Classics

Question:

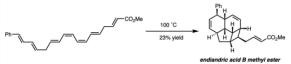


Here is a representation of a Roman inscription, orginally found on a tombstone. Provide a translation for the Palmyrene script. A transliteration of the text is provided: RGYN° BT HRY BR °T° HBL

R Henry T ☐ Merton College, Oxford

: Chemistry

Question:



The reaction shown is a thermal pericyclic cascade that converts the starting heptaene into endiandric acid B methyl ester. The cascade involves three steps: two electrocyclizations followed by a cycloaddition. What types of electrocyclizations are involved in step 1 and step 2, and what type of cycloaddition is involved in step 3?

Provide your answer for the electrocyclizations in the form of $[n\pi]$ con or $[n\pi]$ -dis (where n is the number of π electrons involved, and whether it is conrotatory or disrotatory), and your answer for the cycloaddition in the form of [m+n] (where m and n are the number of atoms on each component).

있 Noah B Stanford University

Secology 🖓

Question:

Hummingbirds within Apodiformes uniquely have a bilaterally paired oval bone, a sesamoid embedded in the caudolateral portion of the expanded, cruciate aponeurosis of insertion of m. depressor caudae. How many paired tendons are supported by this sesamoid bone? Answer with a number.

S Edward V Massachusetts Institute of Technology

A^文 Linguistics

Question:

I am providing the standardized Biblical Hebrew source text from the Biblia Hebraica Stuttgartensia (Psalms 104:7). Your task is to distinguish between closed and open syllables. Please identify and list all closed syllables (ending in a consonant sound) based on the latest research on the Tiberian pronunciation tradition of Biblical Hebrew by scholars such as Geoffrey Khan, Aaron D. Hornkohl, Kim Phillips, and Benjamin Suchard. Medieval sources, such as the Karaite transcription manuscripts, have enabled modern researchers to better understand specific aspects of Biblical Hebrew pronunciation in the Tiberian tradition, including the qualities and functions of the shewa and which letters were pronounced as consonants at the ends of syllables.

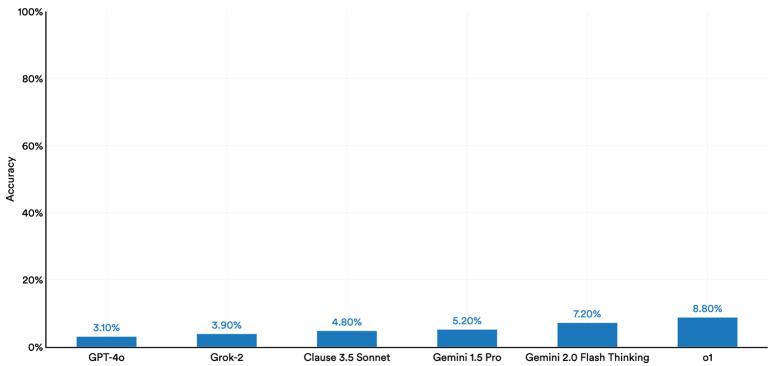
(Psalms 104:7) מִן־גַּעֲרֶתְךָ יְנוּסֵוּן מִן־קוֹל רְעַמְרָ יֵחָפֵזָוּן



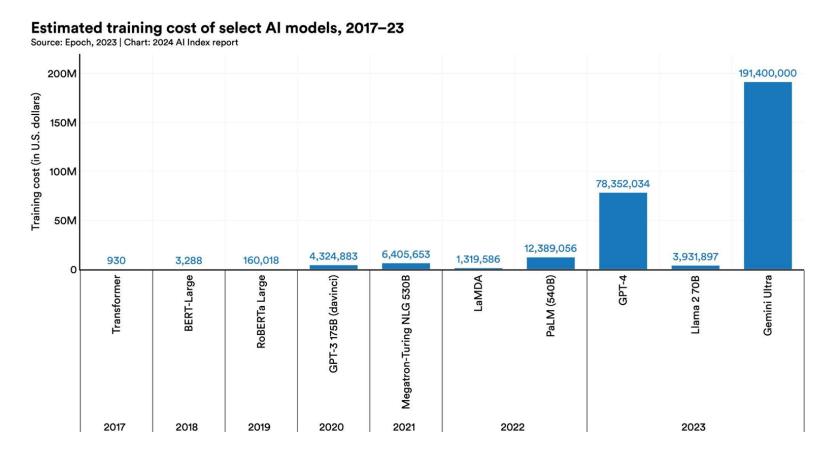
Humanity's Last Exam

Humanity's Last Exam (HLE): accuracy

Source: Phan et al., 2025 | Chart: 2025 Al Index report



Frontier models get way more expensive



Standardized evaluations for LLMs are lacking

Reported general benchmarks for popular foundation models

Source: Al Index, 2024 | Table: 2024 Al Index report

General benchmarks	GPT-4	Llama 2	Claude 2	Gemini	Mistral 7B
MMLU	~	\checkmark	\checkmark	~	\checkmark
HellaSwag	\checkmark	\checkmark		\checkmark	\checkmark
Challenge (ARC)	~	\checkmark	~		~
WinoGrande	~	\checkmark			~
Codex HumanEval	✓	~	~	 ✓ 	~
GSM8K	\checkmark	\checkmark	~	\checkmark	~
Big Bench Hard		\checkmark		\checkmark	\checkmark
Natural Questions		~		\checkmark	✓
BoolQ		~		~	~

Evaluations focused on safety and responsibility, even more so

Reported responsible AI benchmarks for popular foundation models

Source: Al Index, 2024 | Table: 2024 Al Index report

Responsible AI benchmarks	GPT-4	Llama 2	Claude 2	Gemini	Mistral 7B
TruthfulQA	~	~	~		
RealToxicityPrompts	~			\checkmark	
ToxiGen		~			
BOLD		~			
BBQ			\checkmark	\checkmark	

Research – Inspiring Action



Al for Scaling Legal Reform: Mapping and Redacting Racial Covenants in Santa Clara County

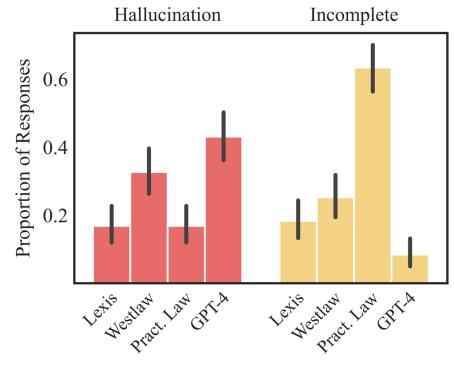
This conveyance is subject to the following conditions, restrictions and reservations;

1. No part of said property shall be sold, let or leased, transferred or assigned to, or occupied by any person not of the Caucasian race, or tobbe used by any other than a person of the Caucasian race.

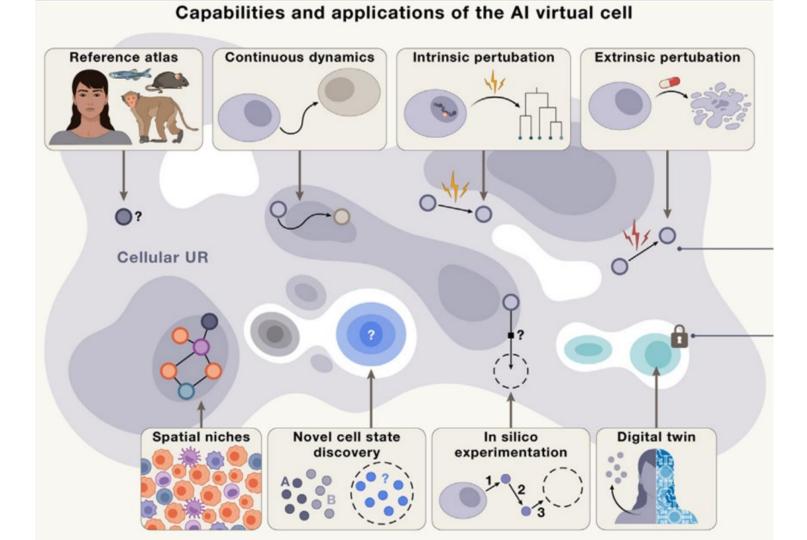
2. A septic tank must be installed according to the State Board of Realth regulations. No outhouses other than a garage will be permitted on the

RegLab

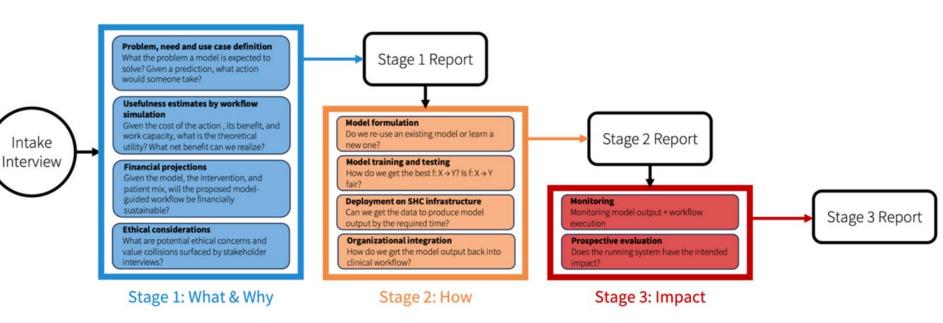
Assessing the Reliability of Leading AI Legal Research Tools



RegLab



Fair, Useful and Reliable AI Model (FURM) Assessment



Roteachai

Find Curriculum-Aligned Instructional Scaffolds Now

Illustrative Mathematics Link	Common Core Standards	
To get started, open your Illustrative Mathema	tics lesson (like <u>this one</u>) and paste the URL below	
https://curriculum.illustrativemathematics.org/MS/teachers/1/5/6/preparation.htm		Search >
Lesson Objectives	Lesson Standards	
Interpret different methods for computing the product of decimals, and e	valuate (orally) their 6.NS.B	

usefulness., Justify (orally, in writing, and through other representations) where to place the decimal point in the product of two decimals with multiple non-zero digits., I can use area diagrams to represent and reason about multiplication of decimals., I know and can explain more than one way to multiply decimals using fractions and place value.

Retrieved and Generated Scaffolds

Warm Up

	ackground Knowledge uiz
Rei	levant Standards: 6.NS.B
rev	s task provides five questions that iew and activate relevant knowledge d skills for the lesson.
	🚼 Al Generated

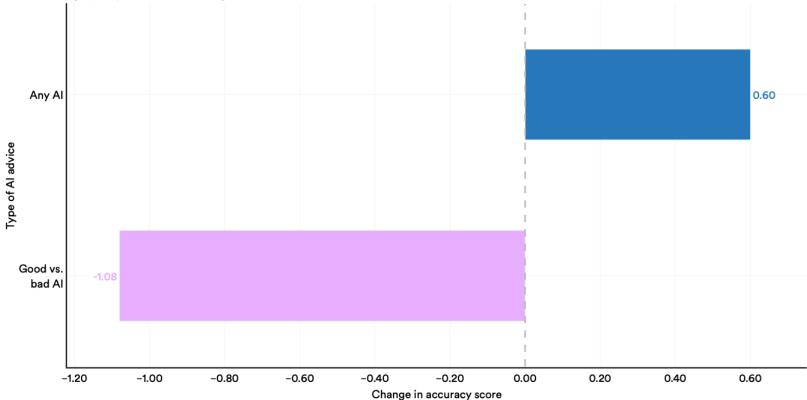
Productivity of Customer Support Agents and AI Support Resolutions per hour 4 Agents who gain access in their fifth month with the firm Agents with access from the time they join the firm 3 Agents with no access 2 2 8 10 0 6

Agent tenure, months

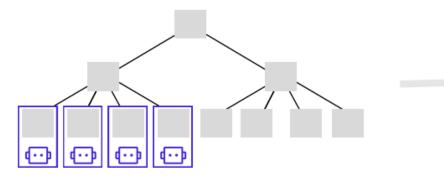
Bars represent 95% confidence intervals. Source: Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond. NBER Working Paper 31161.

Effects on job performance of receiving different types of AI advice

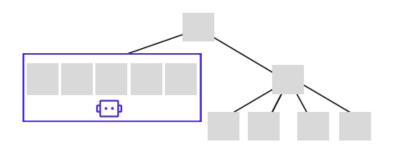
Source: Dell'Acqua, 2023 | Chart: 2024 Al Index report



How Algorithms Can Conflict with Organizational Structures



An algorithm is deployed to initially support individual roles in the organization chart in their decision making.



However, the algorithm is far more effective if it can reason over decision spaces that span multiple roles in the organization chart. This realization destabilizes the organizational chart, which had divided those responsibilities.

The Anthropic Economic Index

Computer & Mathematical	7.2% d
Top Titles	То
Computer Programmers	6.1% 6
 Software Developers, Systems Software 	5.3% ^E
🖻 Software Developers, Applications	3.4% E
Top Tasks	То
Develop and maintain software applications and websites	6.8%
Program and debug computer systems and machinery	6.9%
Design & maintain database systems for data management and analysis	2.3%
自 Office & Administrative	7.9%
Top Titles	То
🖻 Bioinformatics Technicians	2.9% E
🖆 Statistical Assistants	0.4% É
Word Processors	0.4% E
Top Tasks	То
Perform routine IT system administration and maintenance	1.8%
Provide comprehensive customer service and support	0.7%
& Record, analyze, and report	0.6%

operational and research data

🖋 Arts & Media	10.3%	ୱ
Top Titles		Тор
🖻 Technical Writers	1.8%	e
🖻 Copy Writers	1.6%	e
🖻 Editors	1.3%	É
Top Tasks		Тор
Produce and perform in film, TV, theater, and music	1.8%	J
Manage organizational public relations & strategic comms	1.3%	J
Develop & execute multi-industry	1.2%	5
marketing & promotional strategies		
marketing & promotional strategies	6.4%	
	6.4%	Top
基 Life, Physical & Social Science	6.4% 0.5%	
스 Life, Physical & Social Science Top Titles		Тор
∐ Life, Physical & Social Science Top Titles	0.5%	Top
Life, Physical & Social Science Top Titles Clinical Psychologists Historians	0.5% 0.4%	Top
 Life, Physical & Social Science Top Titles Clinical Psychologists Historians Anthropologists 	0.5% 0.4%	Top E E

Conduct chemical analyses and 0.3% experiments on various substances

Education & Library	9.3%
Titles	
Tutors	1.6%
Archivists	1.5%
Instructional Designers	0.8%
Tasks	
Design and develop comprehensive educational curricula and materials	1.9%
Teach and instruct diverse subjects across educational settings	1.7%
Manage book and document publishing processes	1.4%
	Teach and instruct diverse subjects across educational settings Manage book and document

🖬 Bus	iness & Financial	5.9%
Top Title	5	
🖻 Sec	urity Management Specialists	0.5%
🖻 Cre	dit Counselors	0.4%
🖨 Fin	ancial Analysts	0.4%
Top Task	5	
	alyze financial data & develop estment & budgeting strategies	0.8%
	vide personal financial advice l education	0.8%
	cord, analyze, and report erational and research data	0.4%

Human-Centered AI

Proliferation of AI for Good Institutes



Human-Centered Al

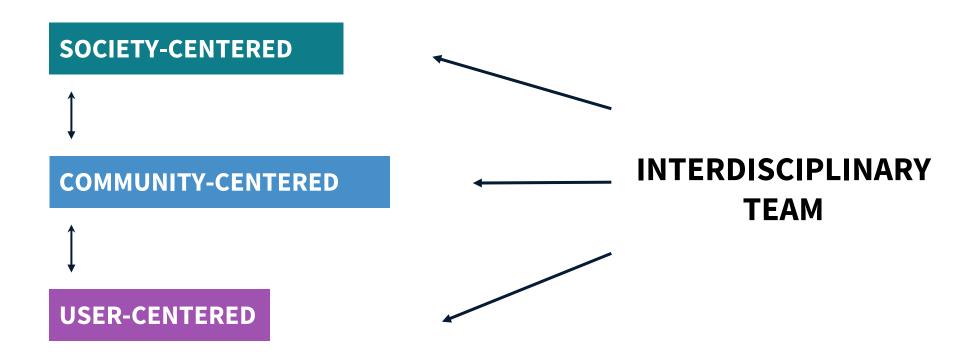
SOCIETY-CENTERED

COMMUNITY-CENTERED

Ì

USER-CENTERED

Human-Centered Al



Continuing Education

- Designing Your Human–Centered Al Strategy: The Social Sector Cohort:
- Custom Executive Education for organization groups
- Generative AI: Technology, Business and Society
- Al for Public Servants
- MANY public events

Thank you

<u>vparli@stanford.edu</u> hai.stanford.edu

