



Stanford University
Human-Centered
Artificial Intelligence

San Mateo County Board Retreat

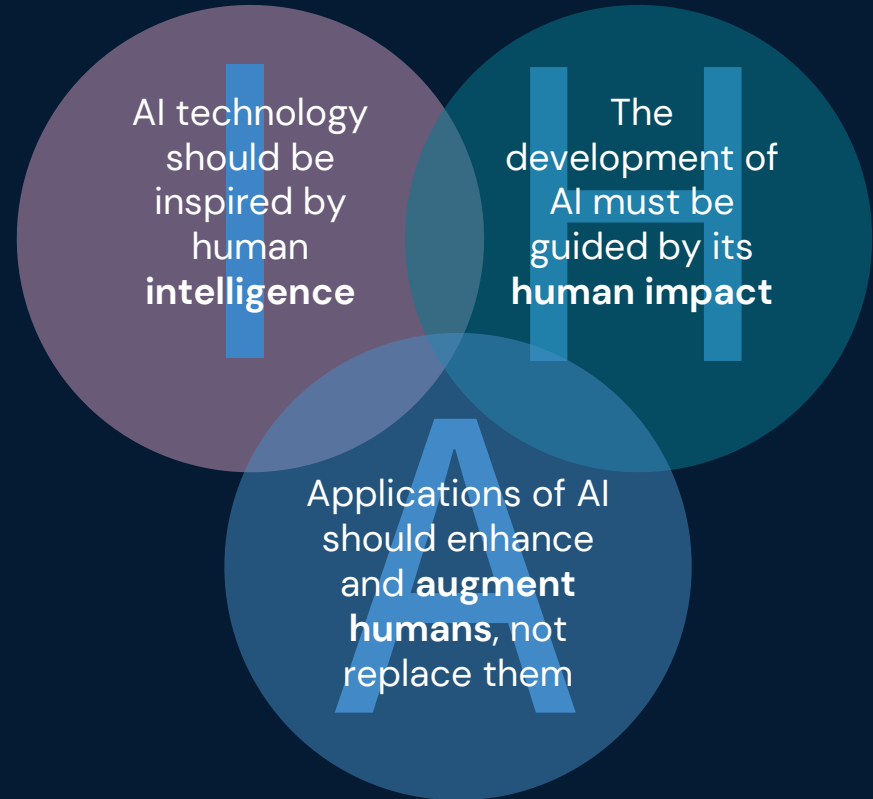
Vanessa Parli

Director of Research and Education Programs

March 4, 2025

Our Goal

To advance AI **research**,
education, **policy** and
practice to improve the
human condition



FOMO: a state of mental or emotional strain caused by the “fear of missing out”.

AI FOMO

2024
7th Edition



**COMING
SOON**
8th Edition



AI 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



V4, November 2022



V5, March 2023



V5.1, March 2023



V5.2, June 2023



V6, December 2023

AI is getting far better



V1, February 2022

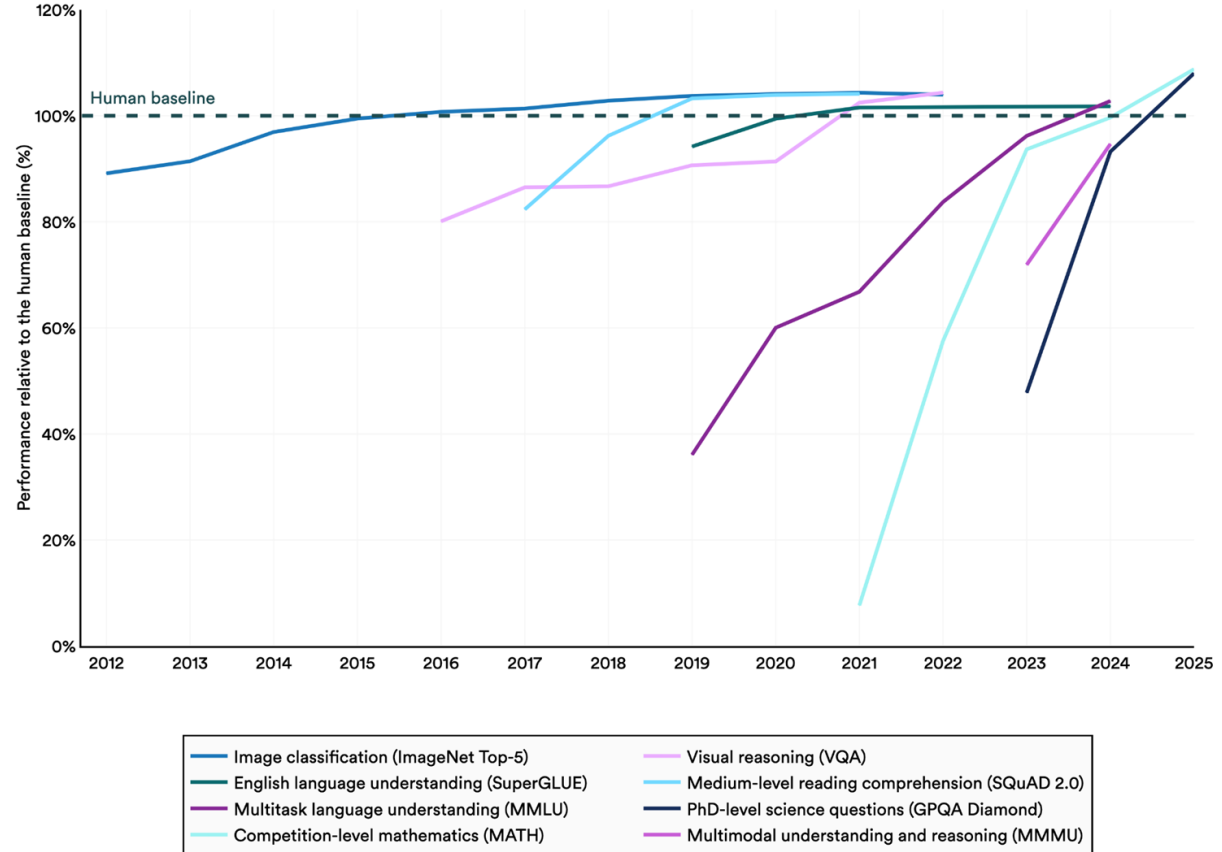


V6, December 2023

AI meets or beats humans on some tasks

Select AI Index technical performance benchmarks vs. human performance

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



Humanity's Last Exam

Classics

Question:



Here is a representation of a Roman inscription, originally found on a tombstone. Provide a translation for the Palmyrene script.

A transliteration of the text is provided: RGYN^o BT HRY BR ^oT^o HBL

Henry T
Merton College, Oxford

Ecology

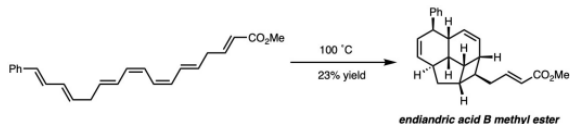
Question:

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.

Edward V
Massachusetts Institute of Technology

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).

Noah B
Stanford University

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.

בַּיּוֹמִים מִן־גִּבְעוֹתֶיךָ יְגִיחוּ מִן־קוֹלֵי־יְעִמָּהּ יִפְדָּח

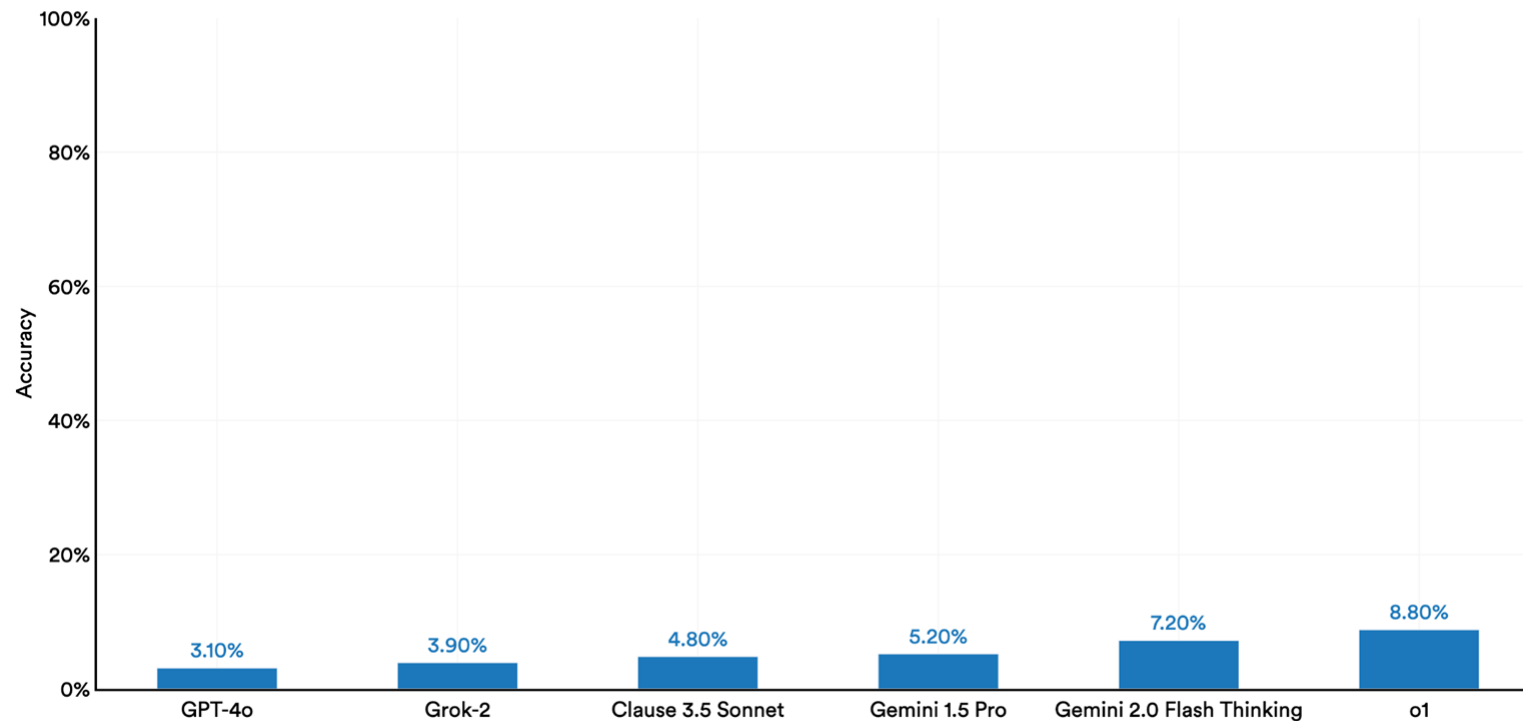
Lina B
University of Cambridge



Humanity's Last Exam

Humanity's Last Exam (HLE): accuracy

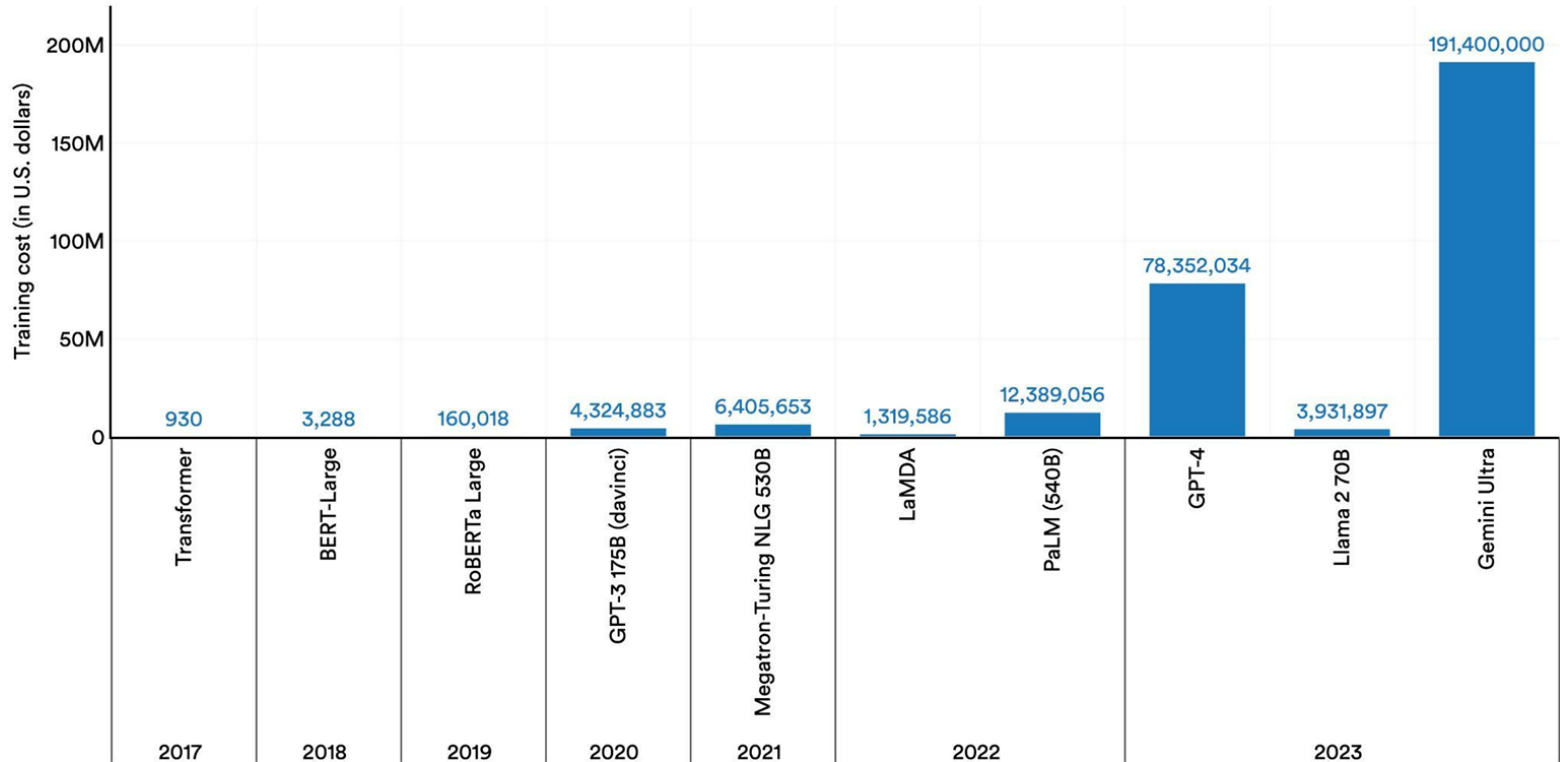
Source: Phan et al., 2025 | Chart: 2025 AI Index report



Frontier models get way more expensive

Estimated training cost of select AI models, 2017–23

Source: Epoch, 2023 | Chart: 2024 AI Index report



Standardized evaluations for LLMs are lacking

Reported general benchmarks for popular foundation models

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

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

Evaluations focused on safety and responsibility, even more so

Reported responsible AI benchmarks for popular foundation models

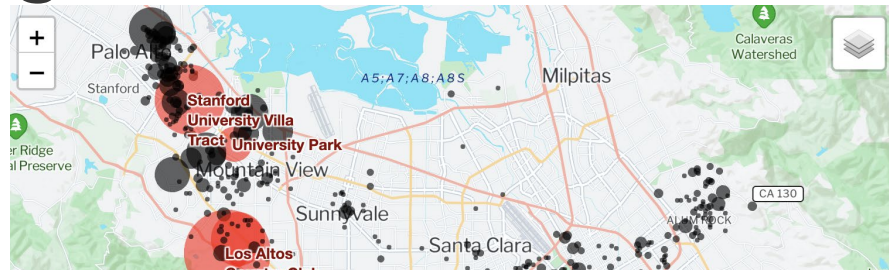
Source: AI Index, 2024 | Table: 2024 AI Index report

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

Research -
Inspiring Action



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



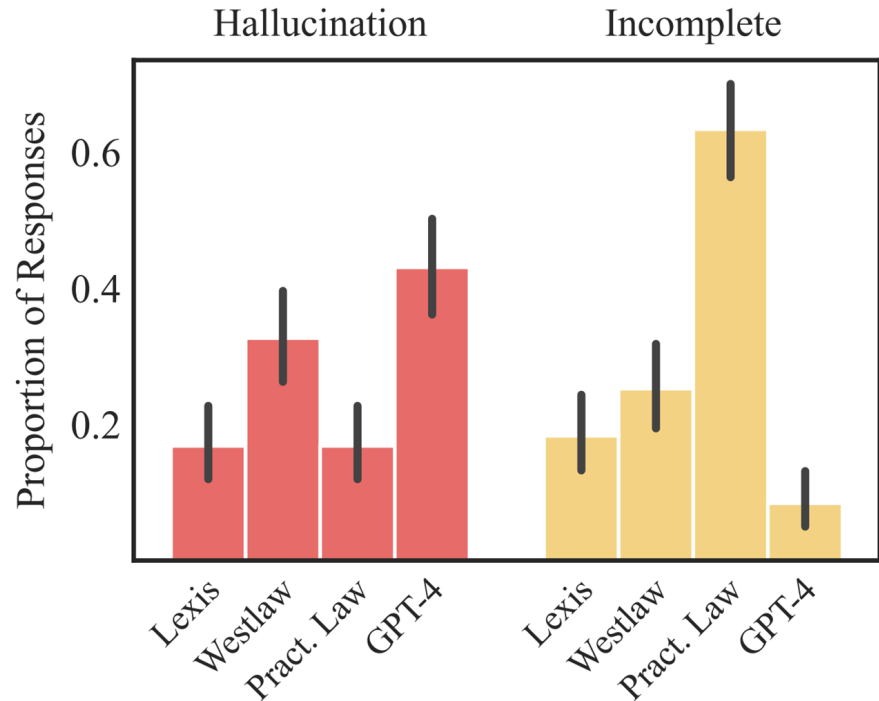
This conveyance is subject to the following conditions, restrictions and reservations;

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 to be used by any other than a person of the Caucasian race.

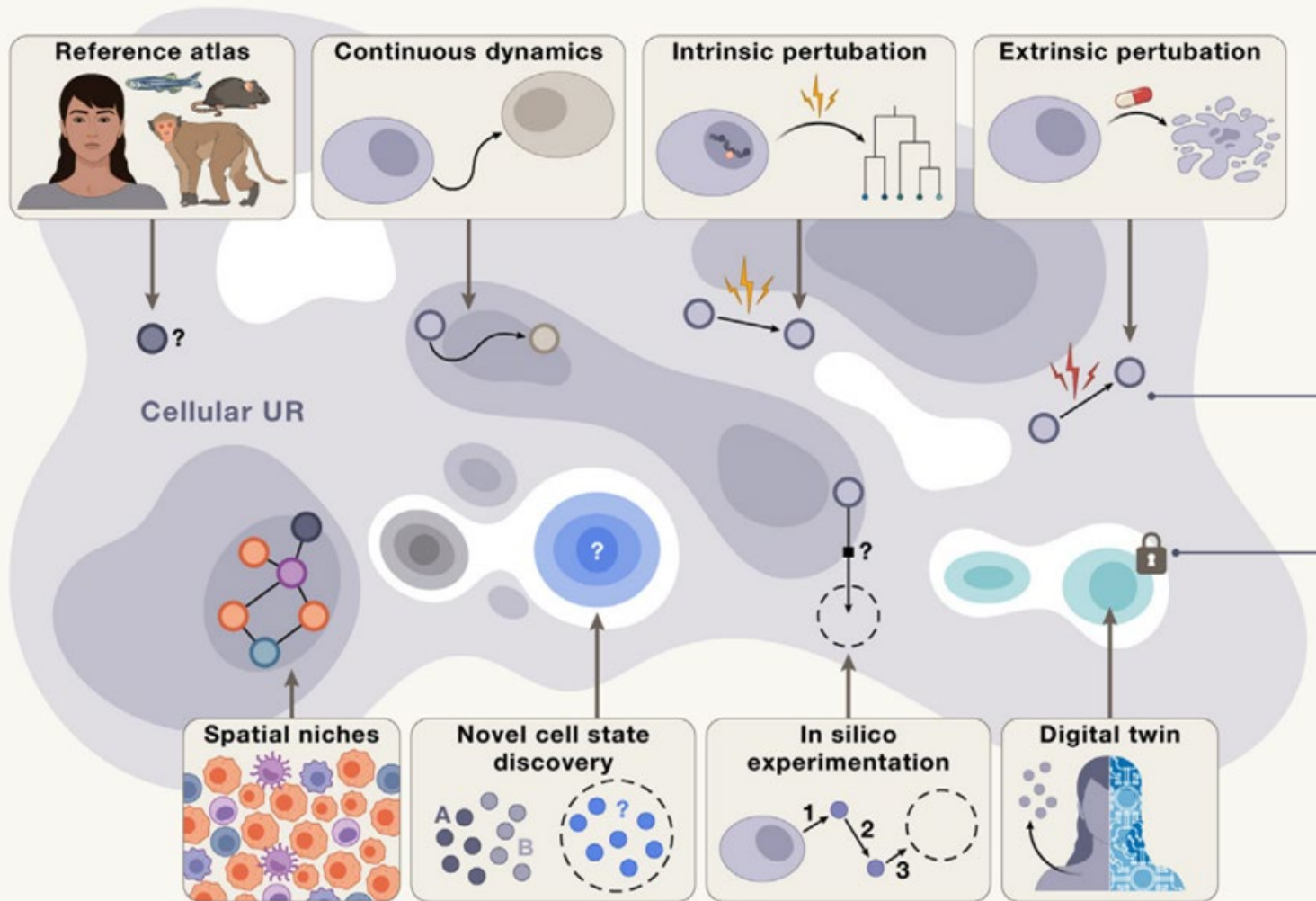
2. A septic tank must be installed according to the State Board of Health regulations. No outhouses other than a garage will be permitted on the



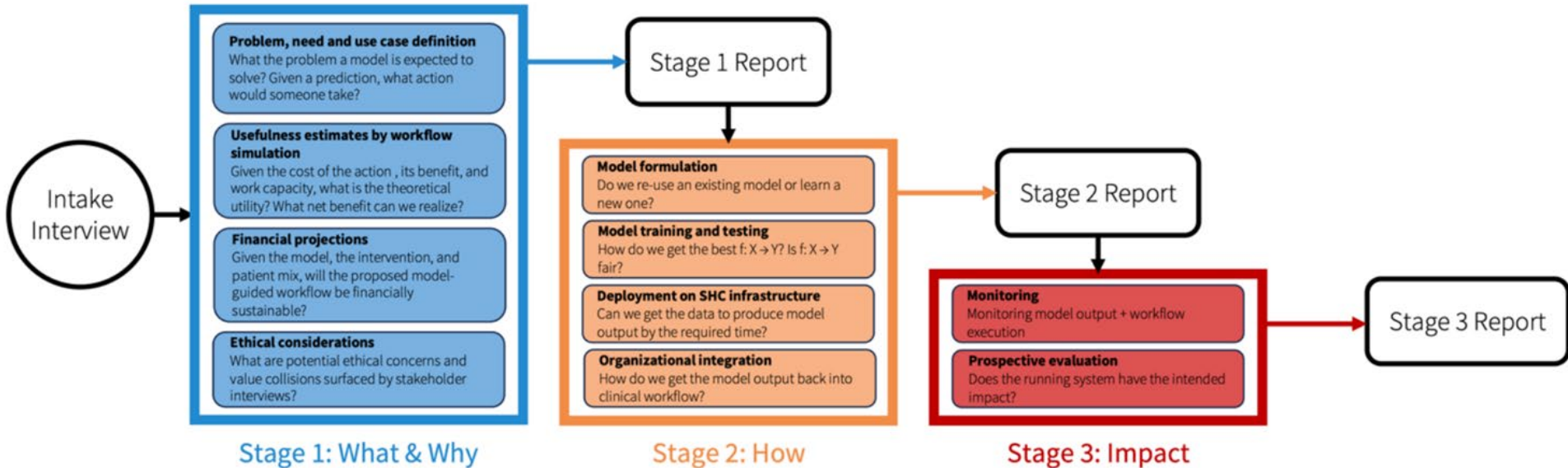
Assessing the Reliability of Leading AI Legal Research Tools



Capabilities and applications of the AI virtual cell



Fair, Useful and Reliable AI Model (FURM) Assessment



Illustrative Mathematics Link

Common Core Standards

To get started, open your Illustrative Mathematics lesson (like [this one](#)) and paste the URL below

Search >

Lesson Objectives

Interpret different methods for computing the product of decimals, and evaluate (orally) their 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.

Lesson Standards

6.NS.B

Retrieved and Generated Scaffolds

Warm Up



Background Knowledge Quiz

Relevant Standards: 6.NS.B

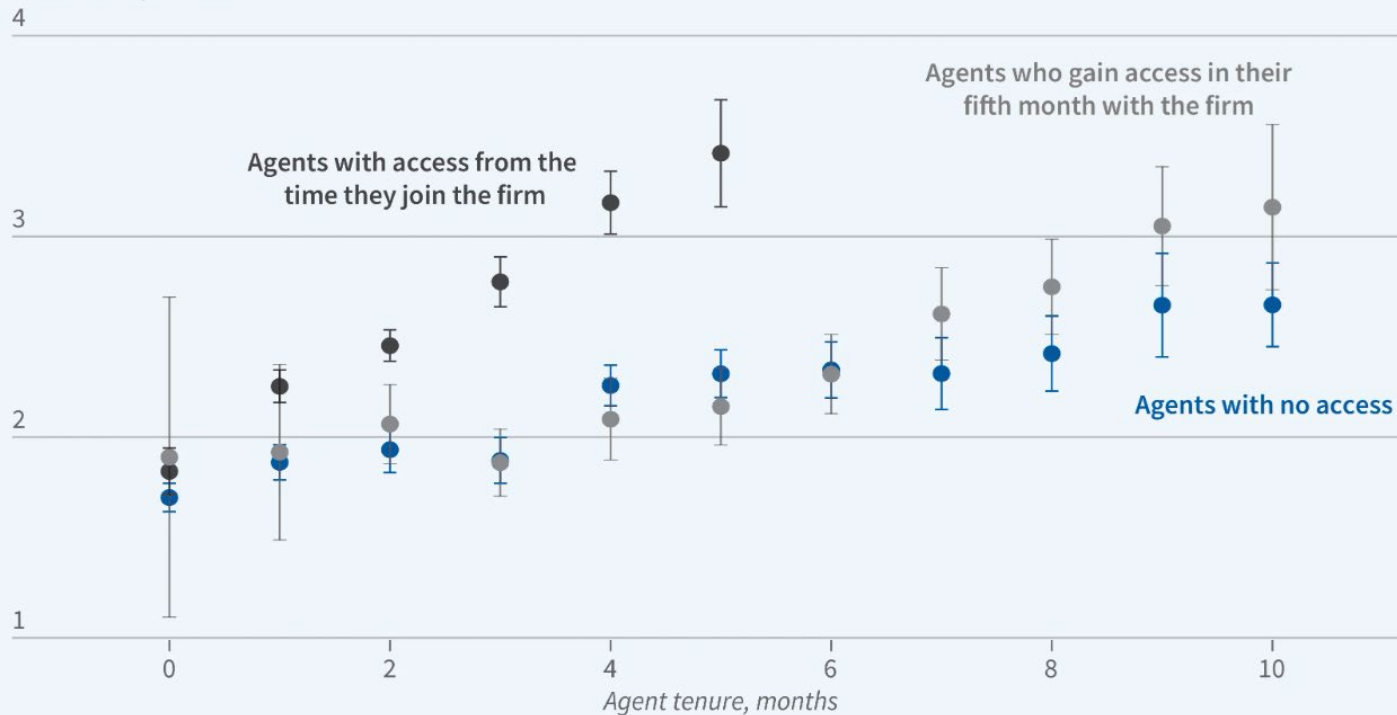
This task provides five questions that review and activate relevant knowledge and skills for the lesson.

AI Generated

Additional Practice

Productivity of Customer Support Agents and AI Support

Resolutions per hour

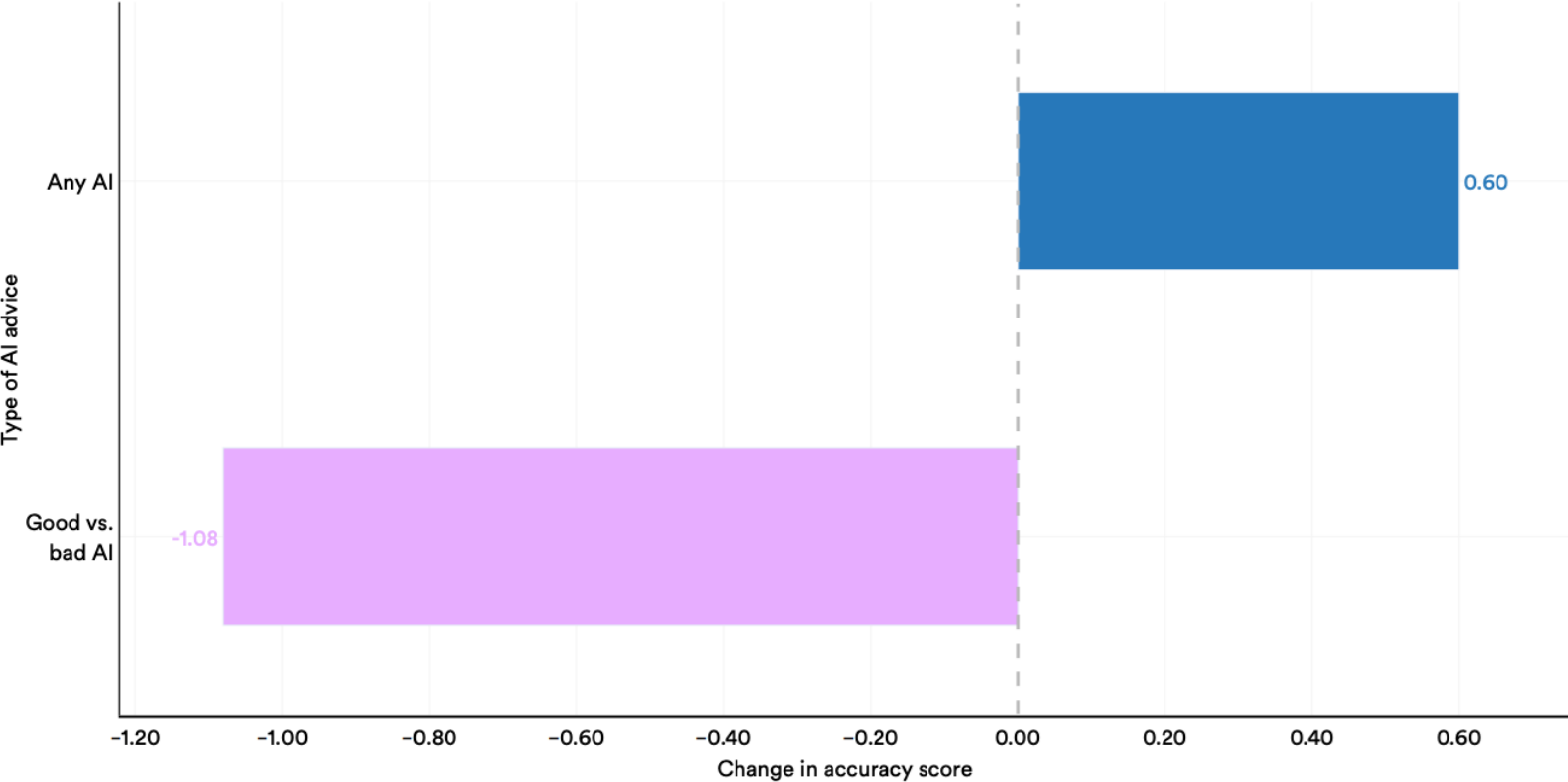


Bars represent 95% confidence intervals.

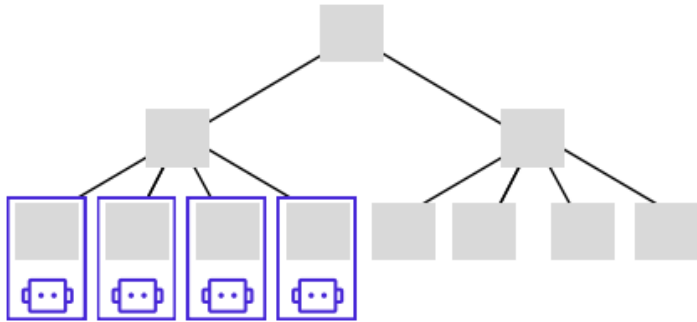
Source: Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond. NBER Working Paper 31161.

Effects on job performance of receiving different types of AI advice

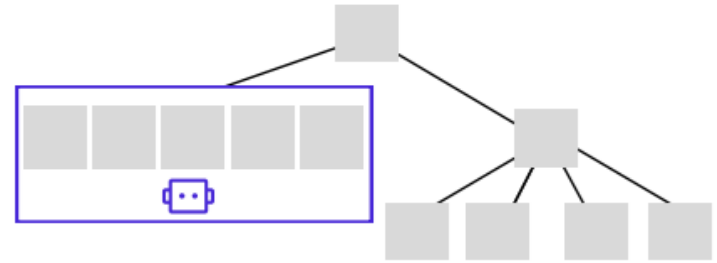
Source: Dell'Acqua, 2023 | Chart: 2024 AI Index report



How Algorithms Can Conflict with Organizational Structures

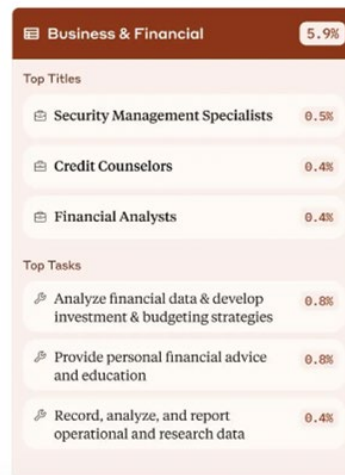
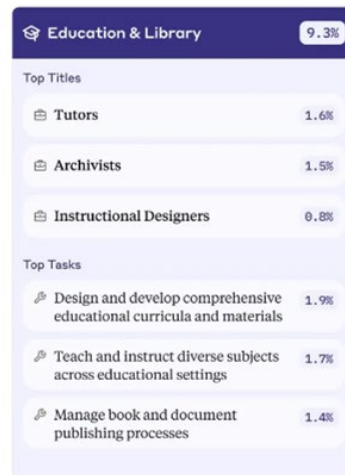
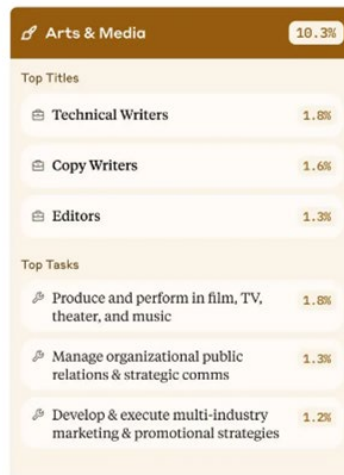


An algorithm is deployed to initially support **individual roles** in the organization chart in their decision making.



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



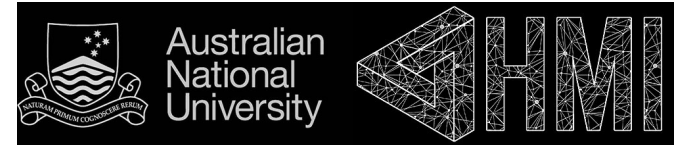
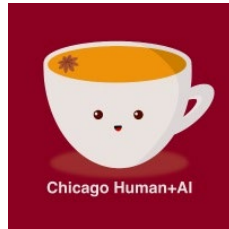
Human-Centered AI



Proliferation of AI for Good Institutes



**HUMAN CENTRED
TECHNOLOGY
RESEARCH CLUSTER**



Human-AI Interaction @ UBC



Human-Centered AI

SOCIETY-CENTERED

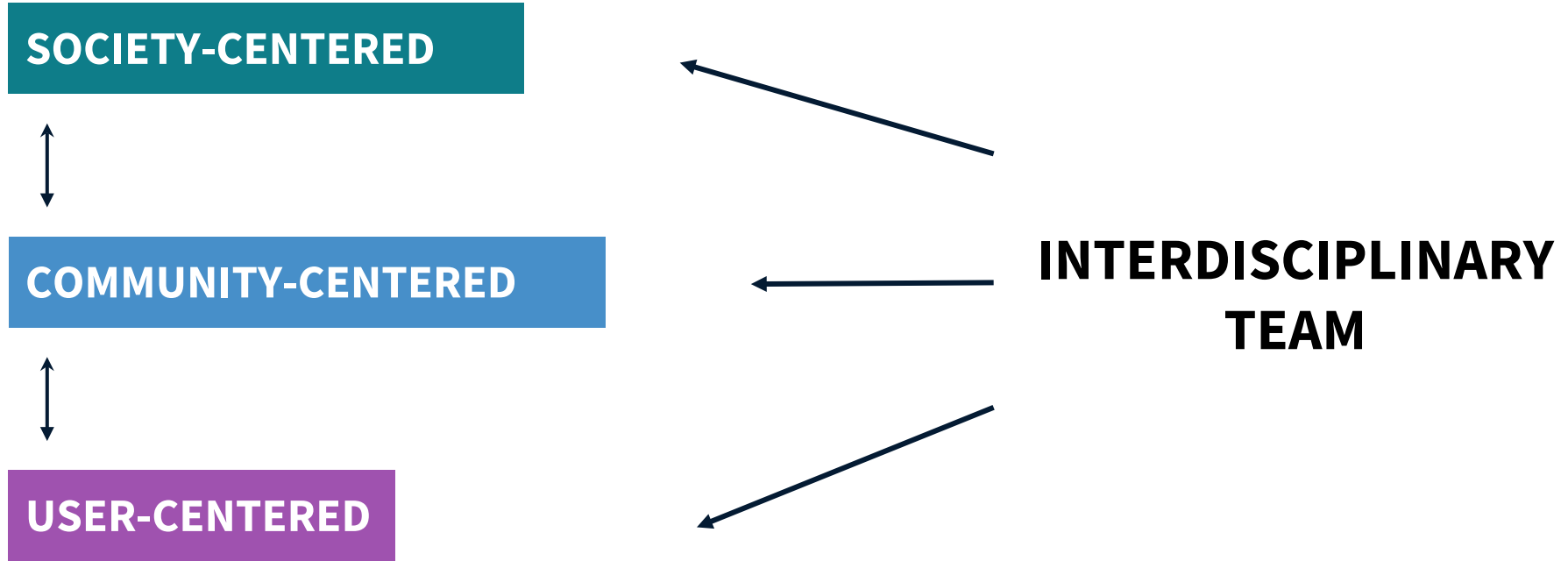


COMMUNITY-CENTERED



USER-CENTERED

Human-Centered AI



Continuing Education

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

Thank you

vparli@stanford.edu

hai.stanford.edu

