The Art of Artificial Intelligence:
How AI Can Enhance the Practice of Palliative Care

Deborah M. Unger, MD
FAAHPM

5.1.2026

Providence

29M
Patient Visits

122k
Employees

1,000
Clinics

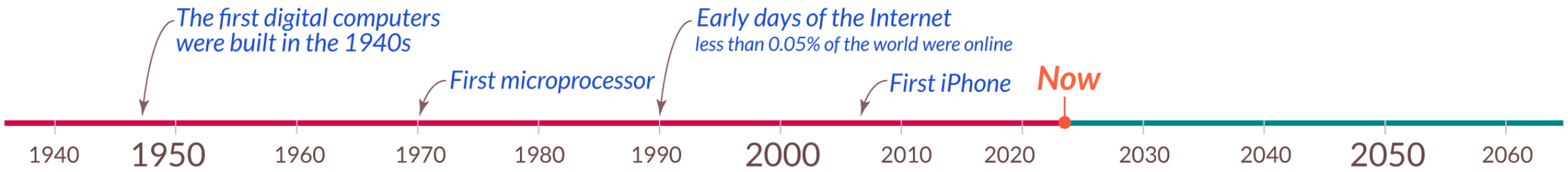
51
Hospitals

Providence
Institute for Human Caring

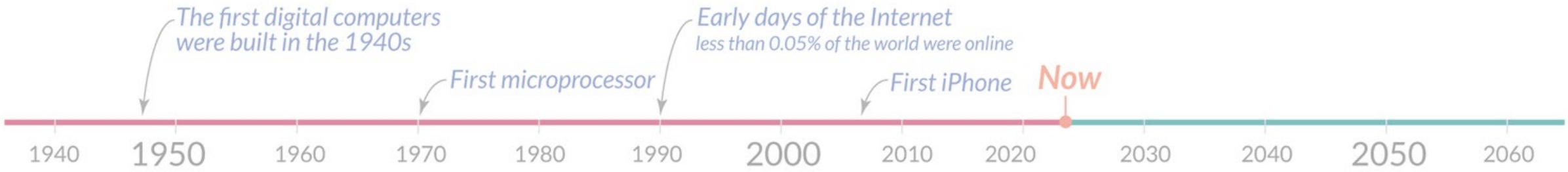
Overview

AI and Machine Learning

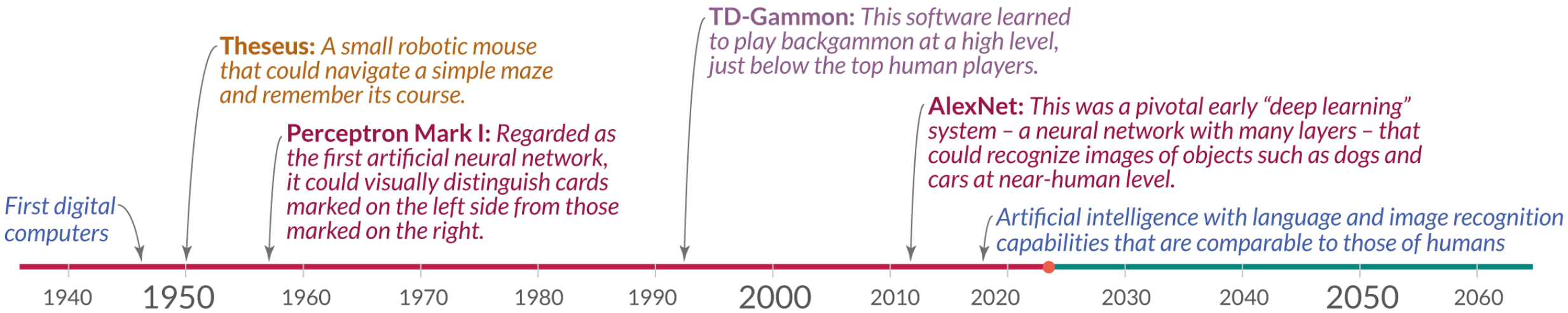
Our Digital World



The brief history of artificial intelligence: the world has changed fast — what might be next? - Our World in Data

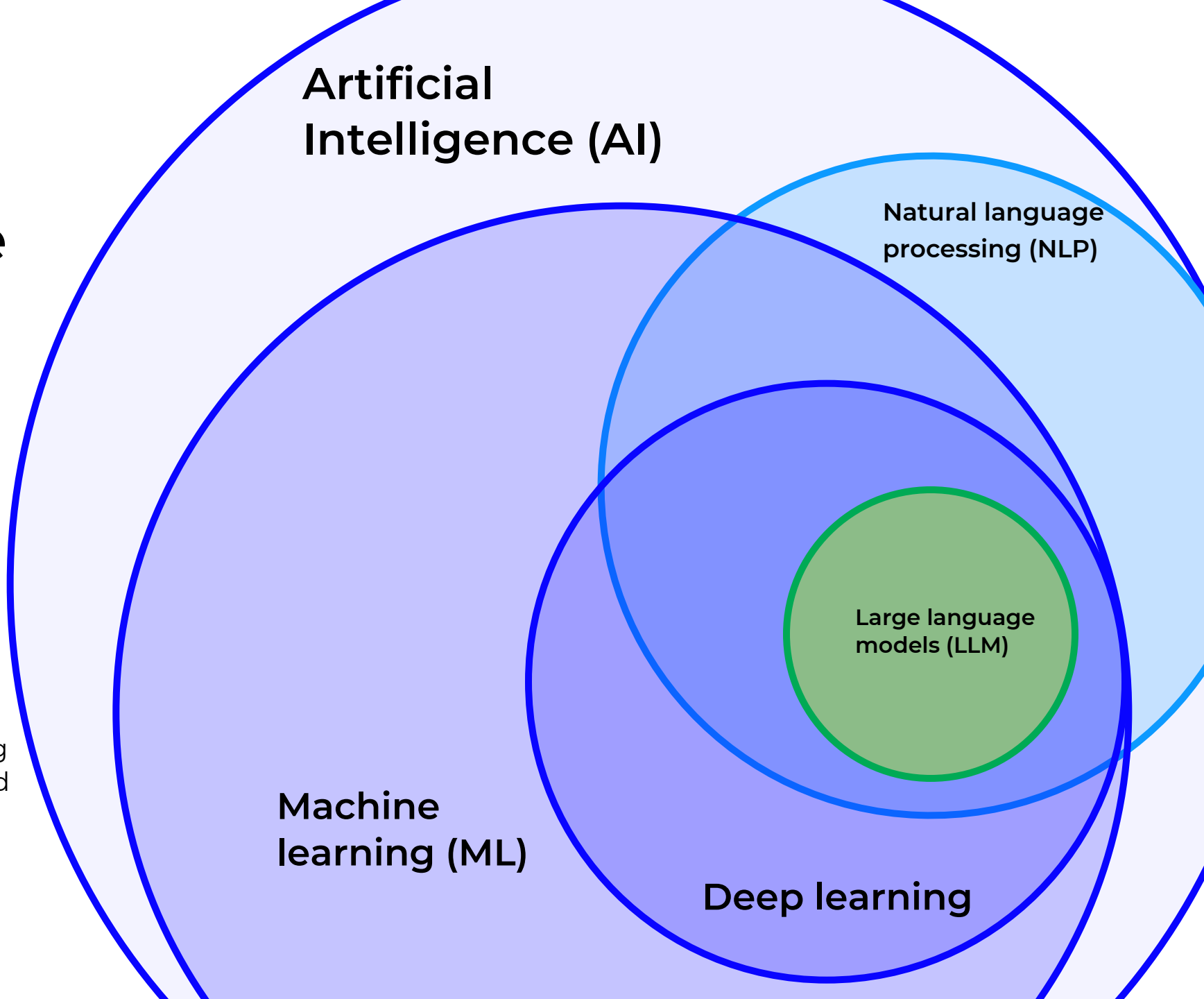


A timeline of notable artificial intelligence systems



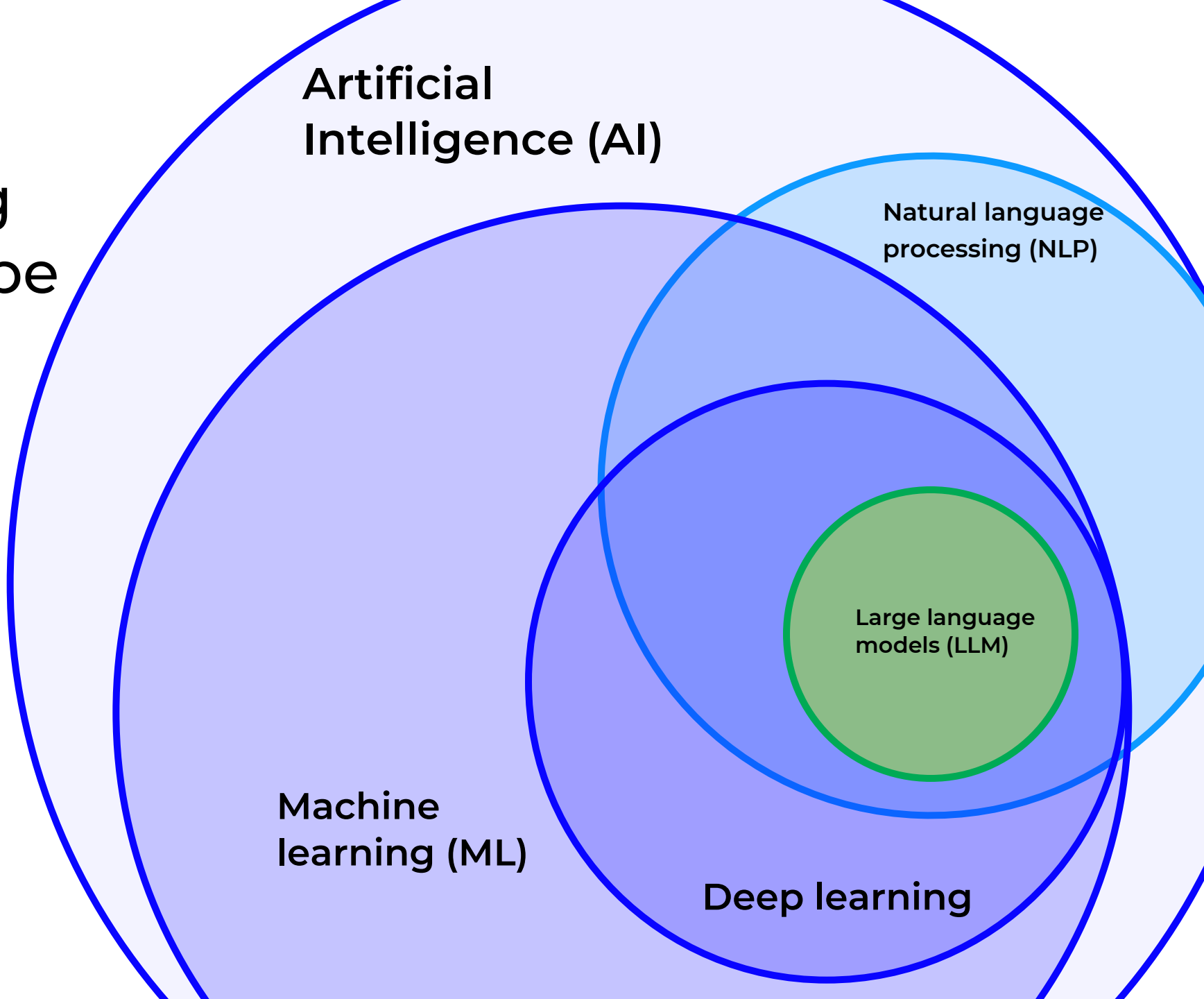
The brief history of artificial intelligence: the world has changed fast — what might be next? - Our World in Data

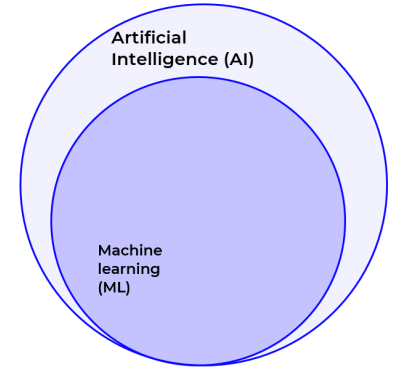
Understanding the AI landscape

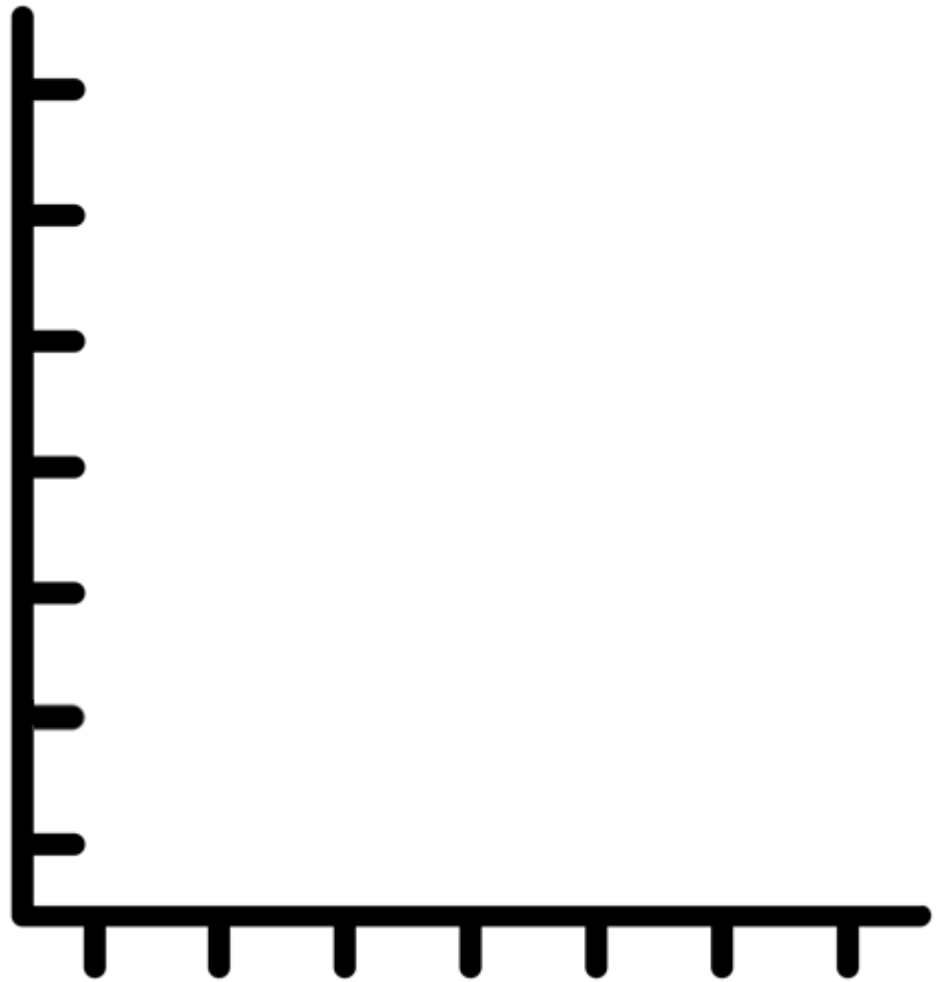


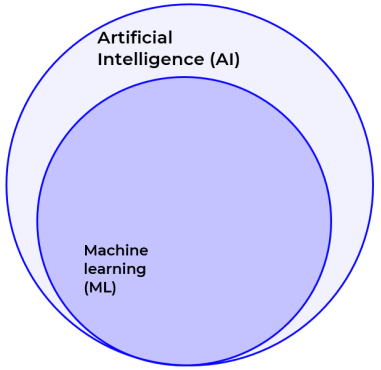
ML is a sub field of AI. Deep learning is a sub field of ML. NLP is a sub field of AI that often uses ML methods. LLMs are modern NLP models that use deep learning.

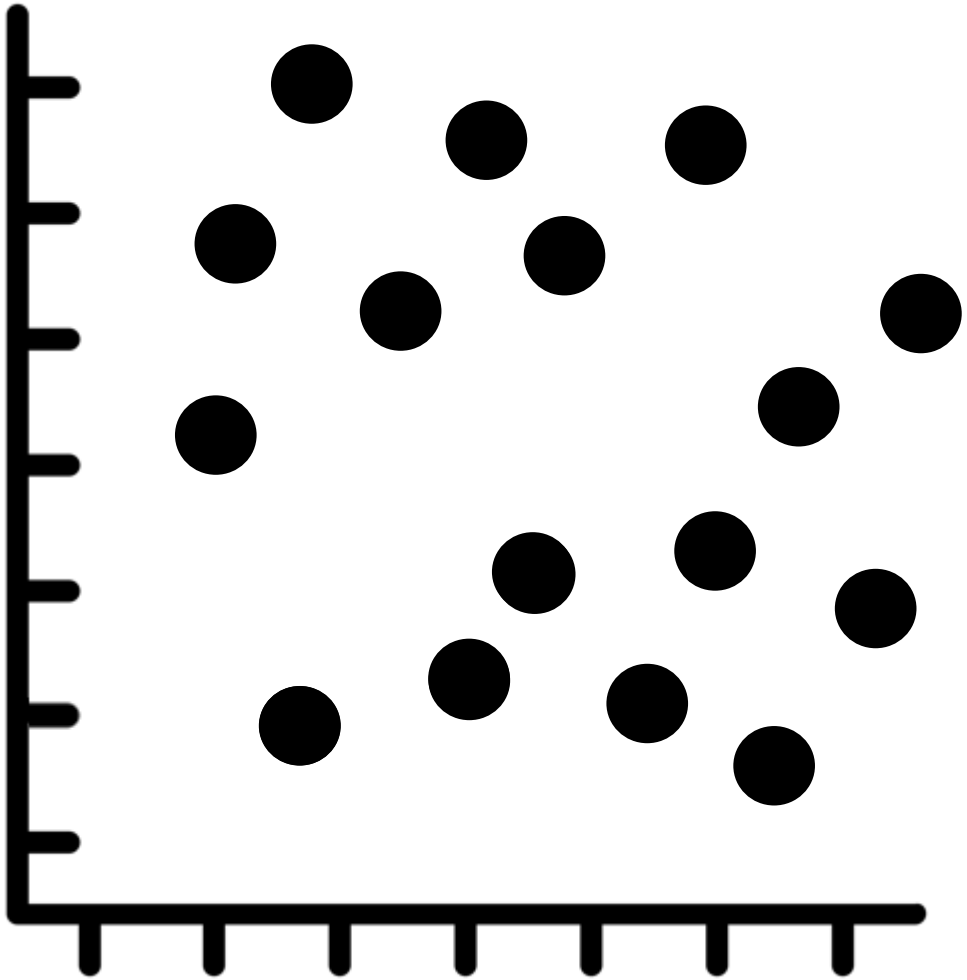
Understanding the AI landscape

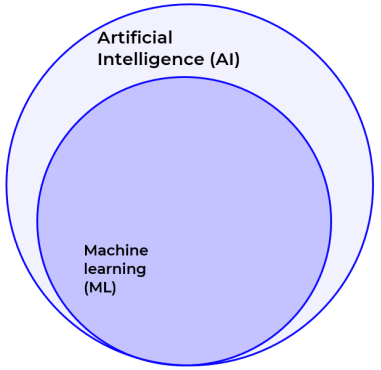


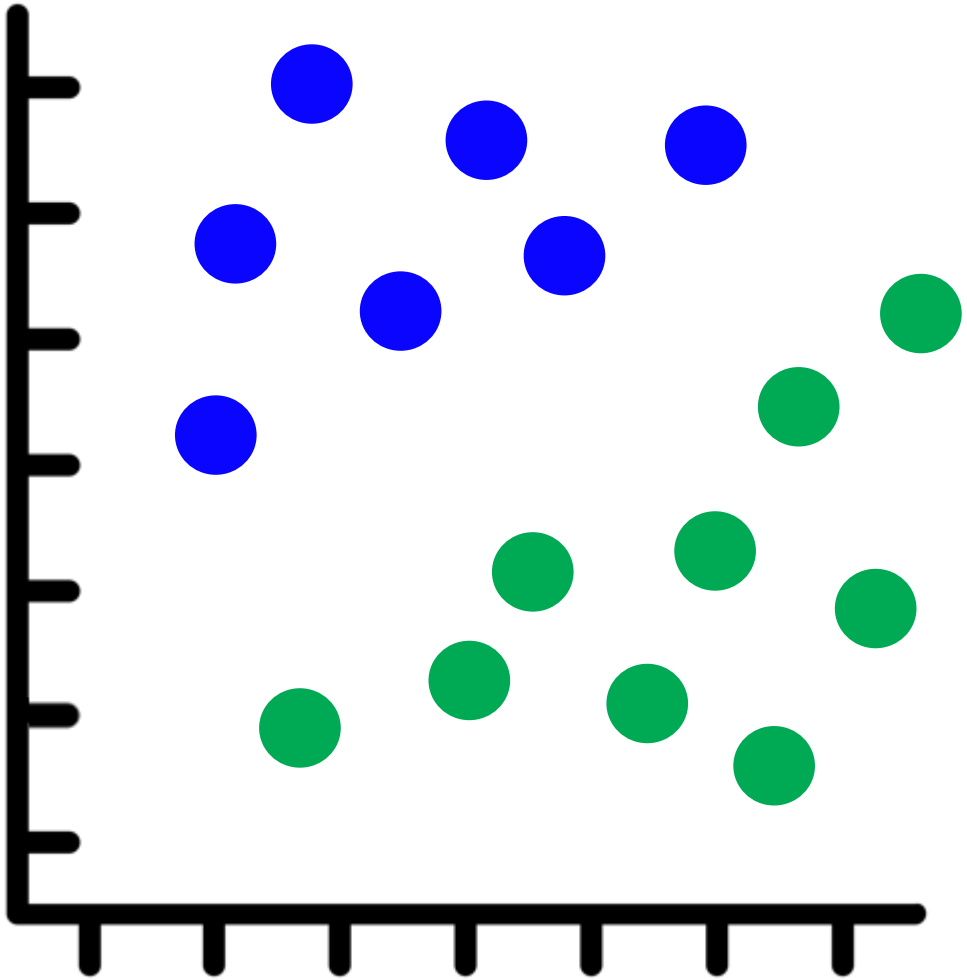

Artificial
Intelligence (AI)
Machine
learning
(ML)

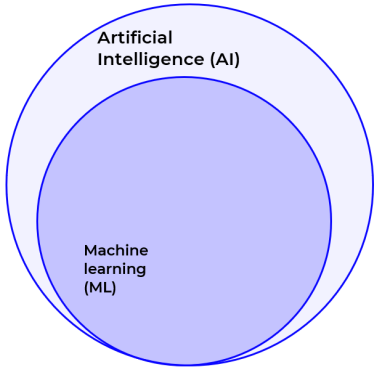



Artificial
Intelligence (AI)
Machine
learning
(ML)



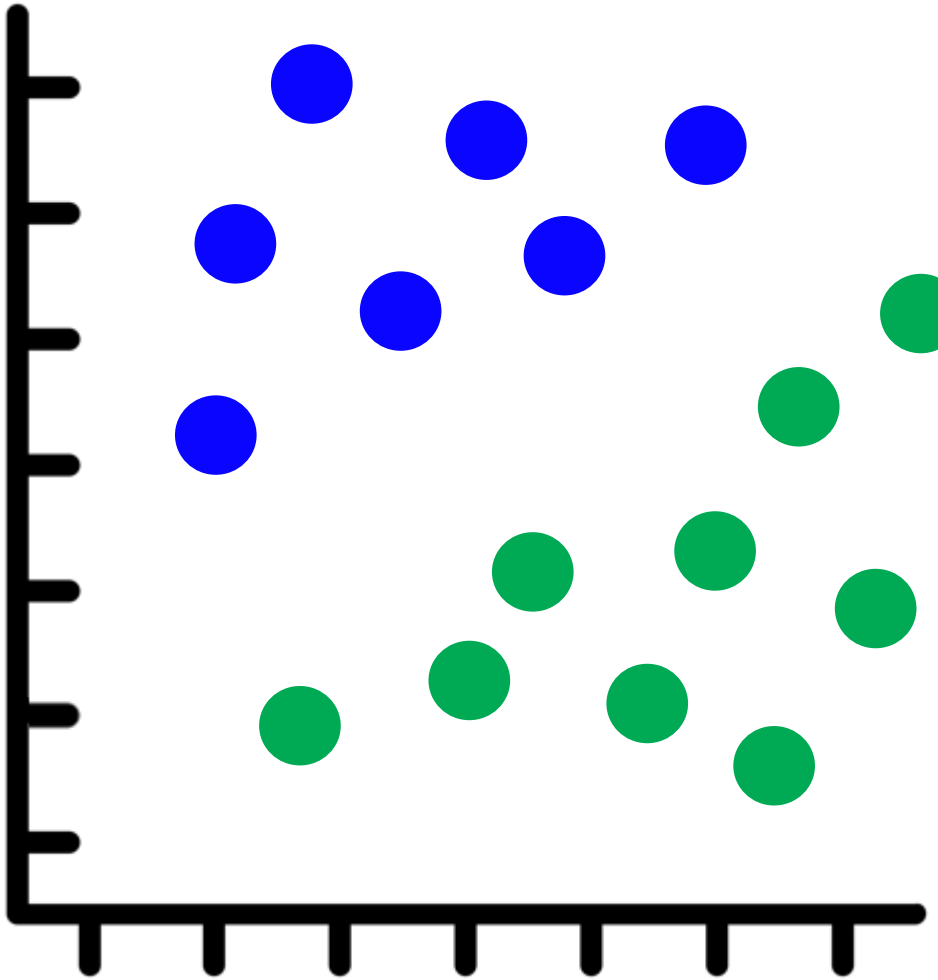

Artificial
Intelligence (AI)
Machine
learning
(ML)

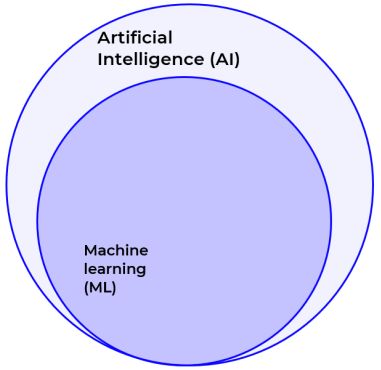




Science Fiction

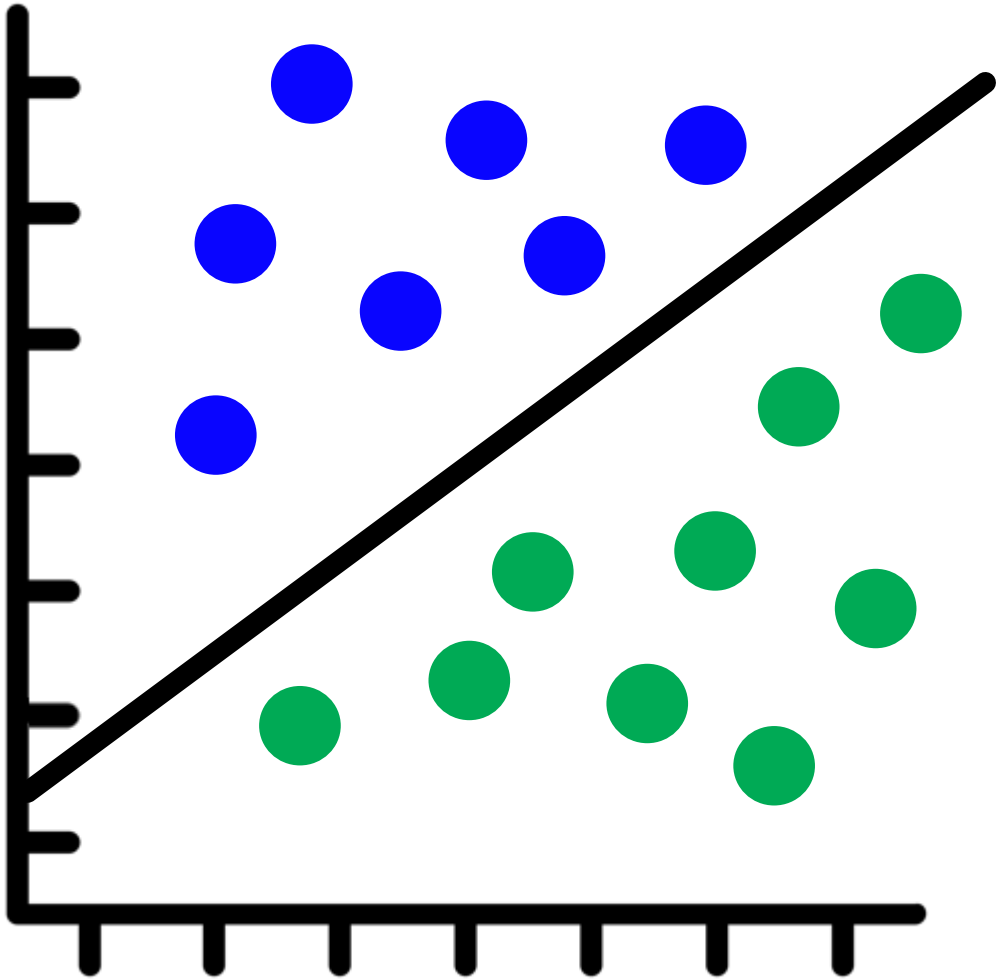
Fantasy

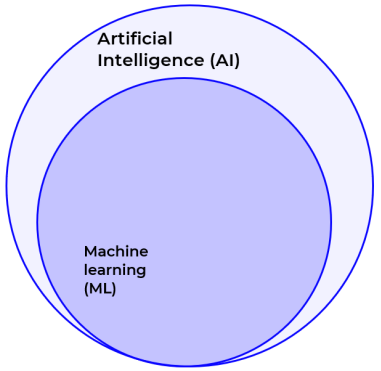




Science Fiction

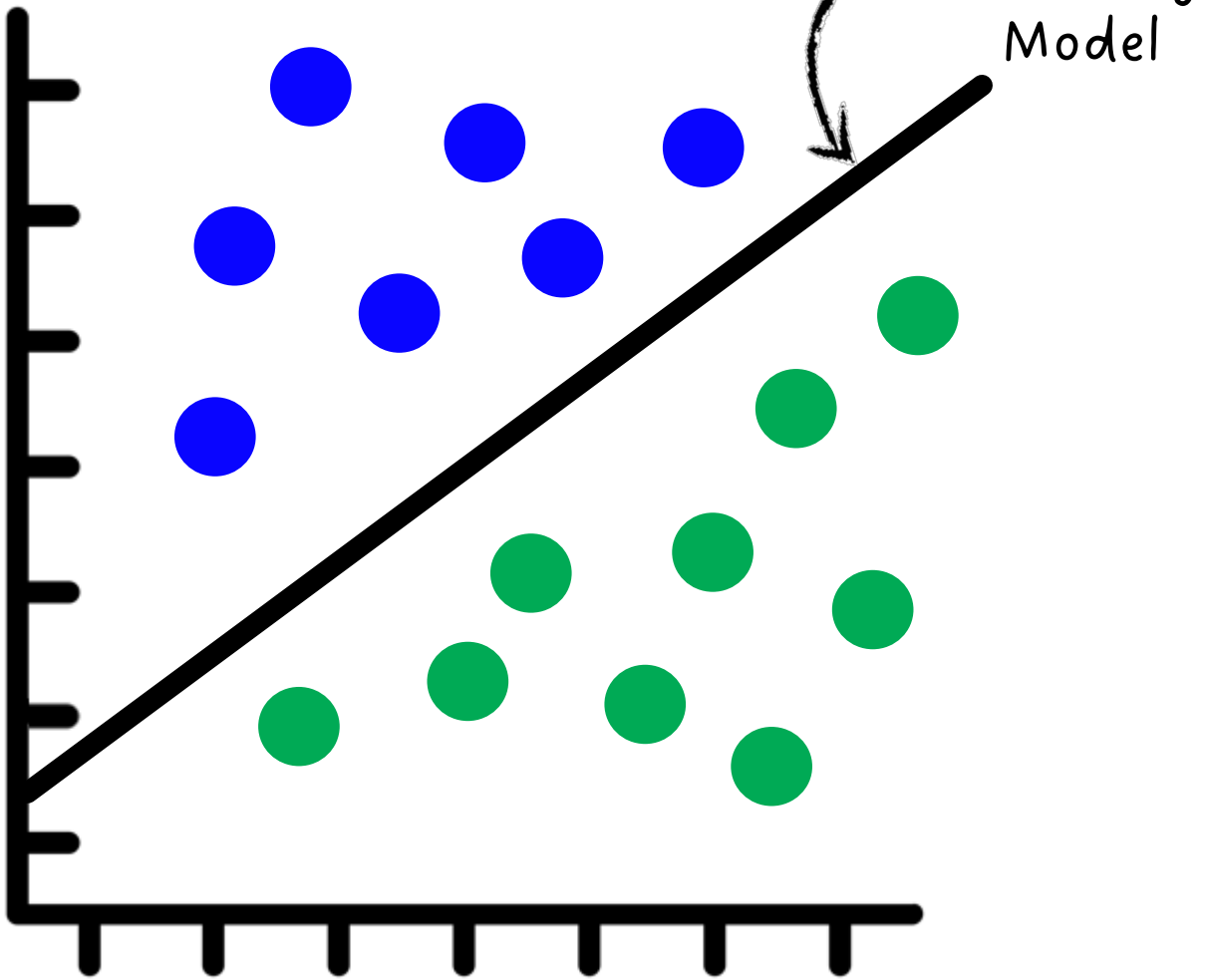
Fantasy





Science Fiction

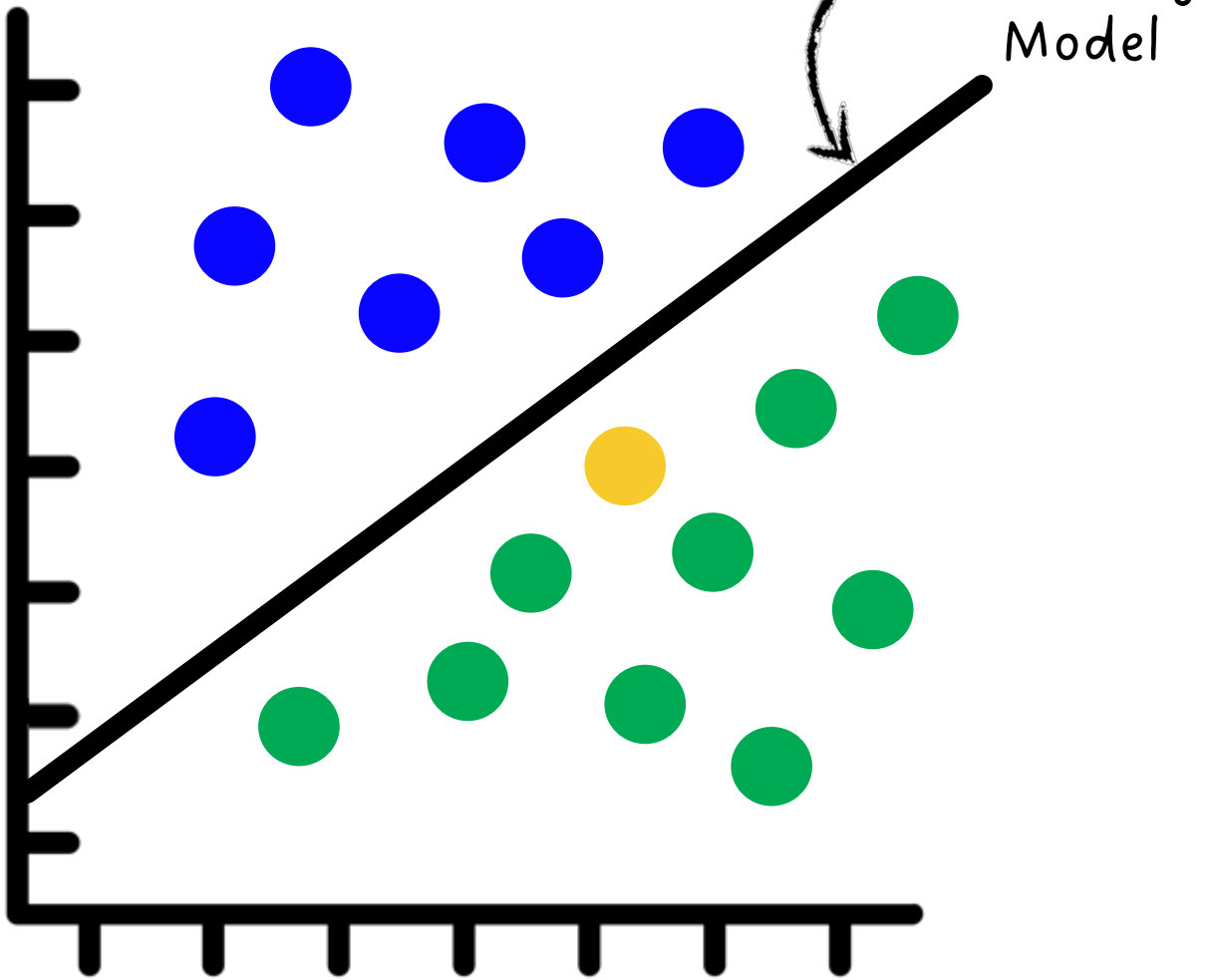
Fantasy

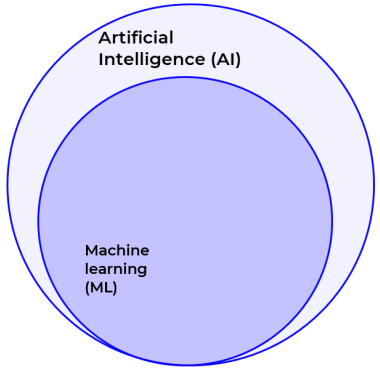


Science Fiction

Fantasy

New Book

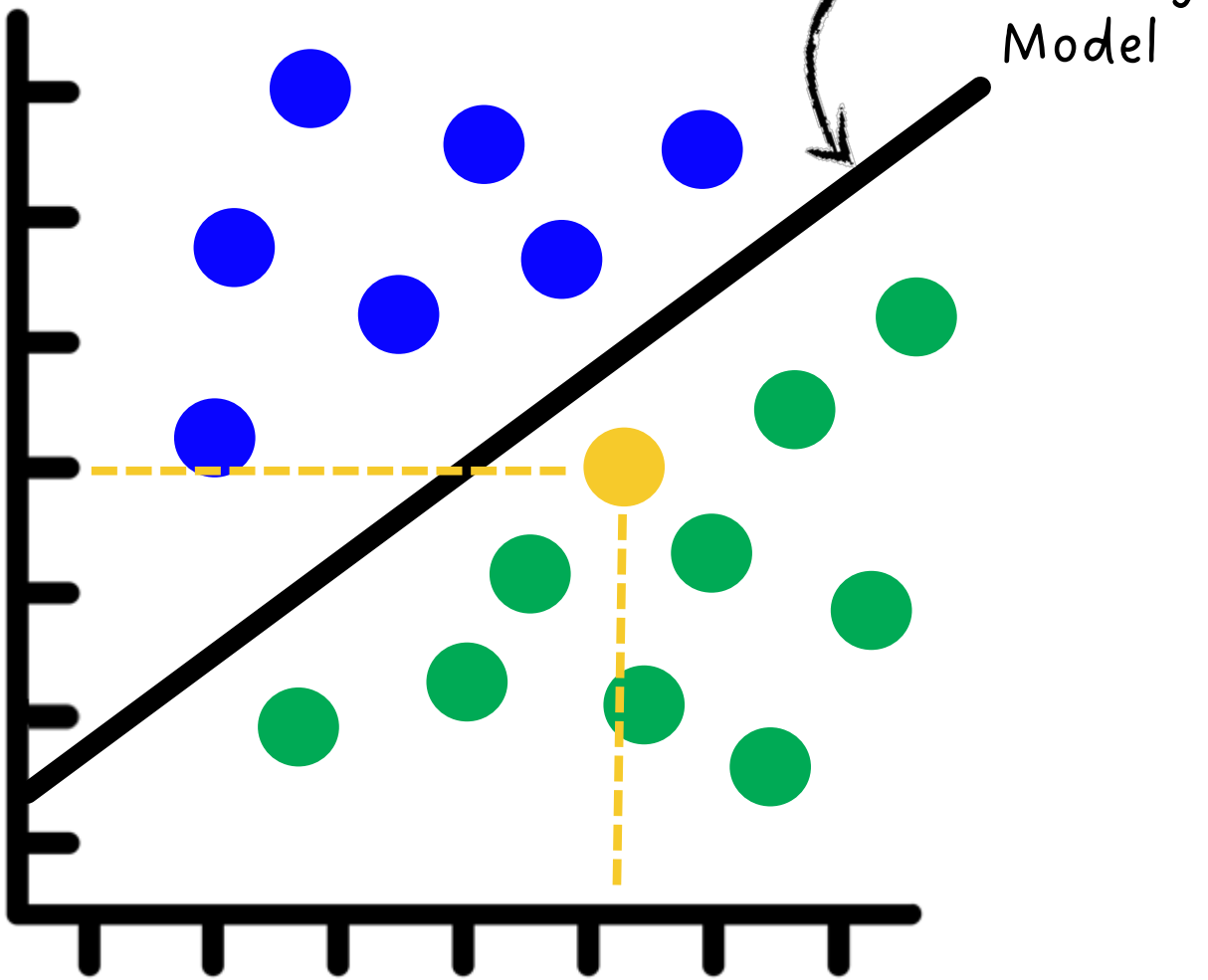


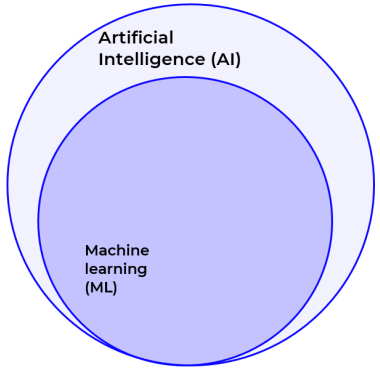


Science Fiction

Fantasy

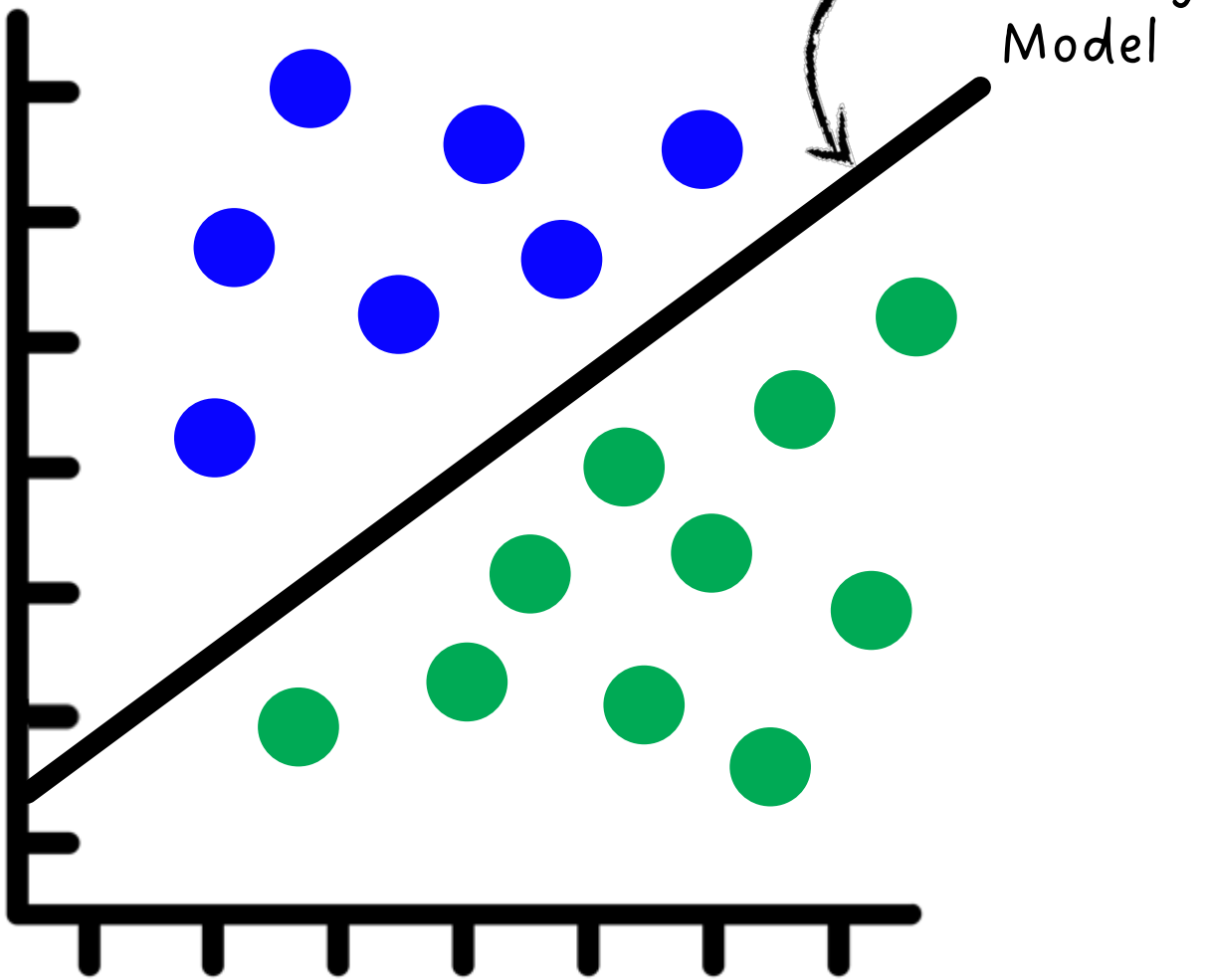
New Book

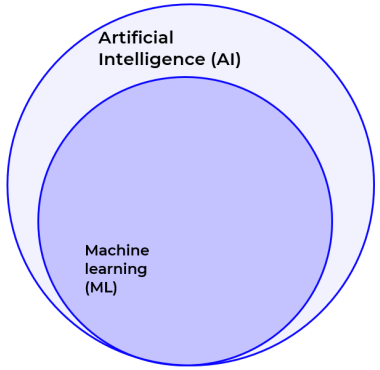


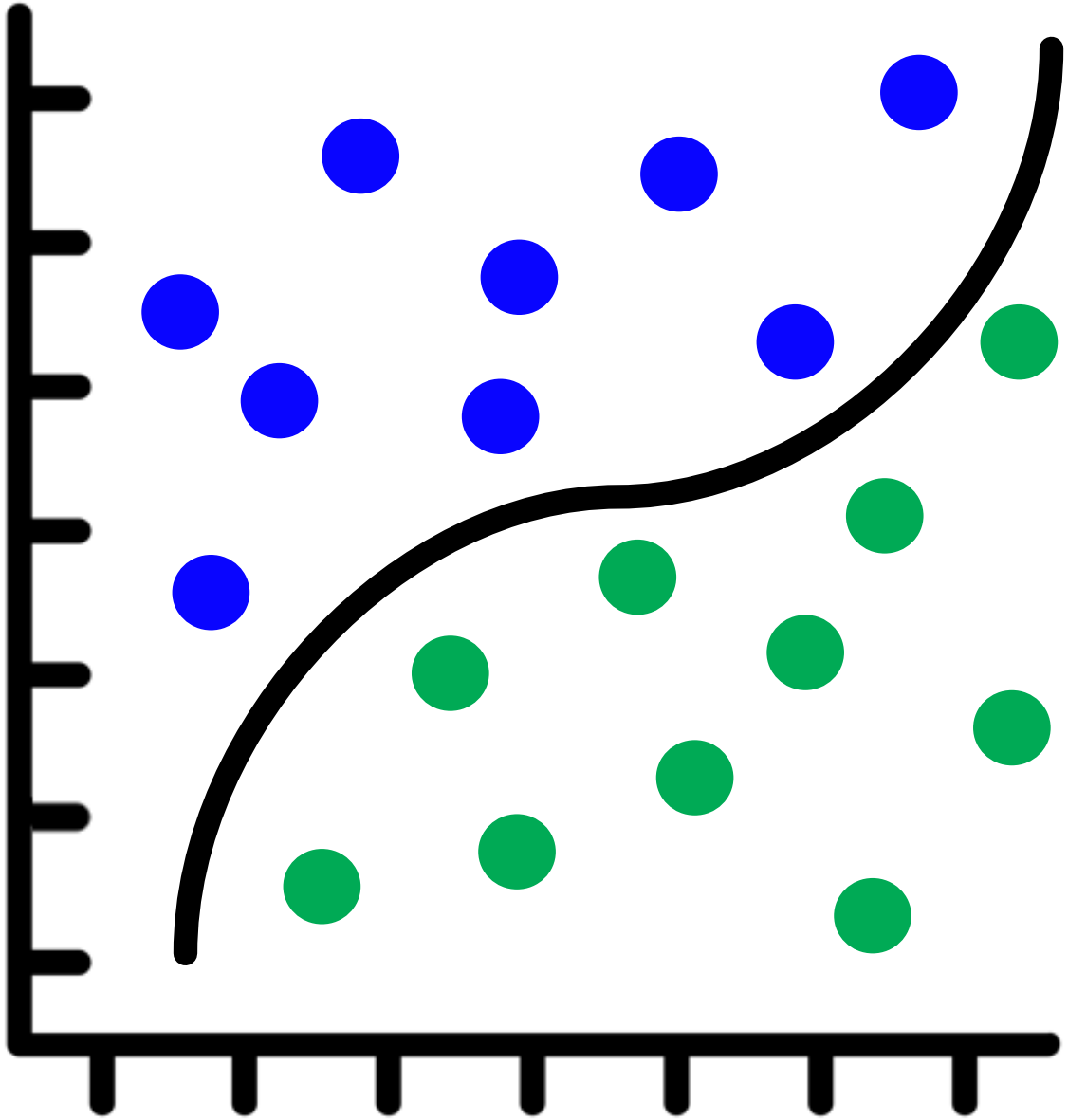


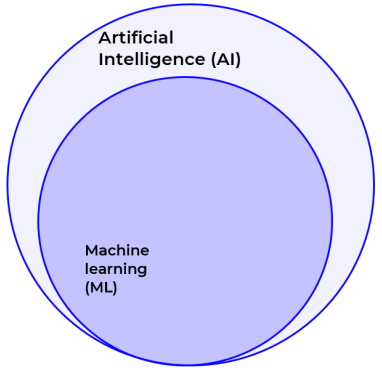
Science Fiction

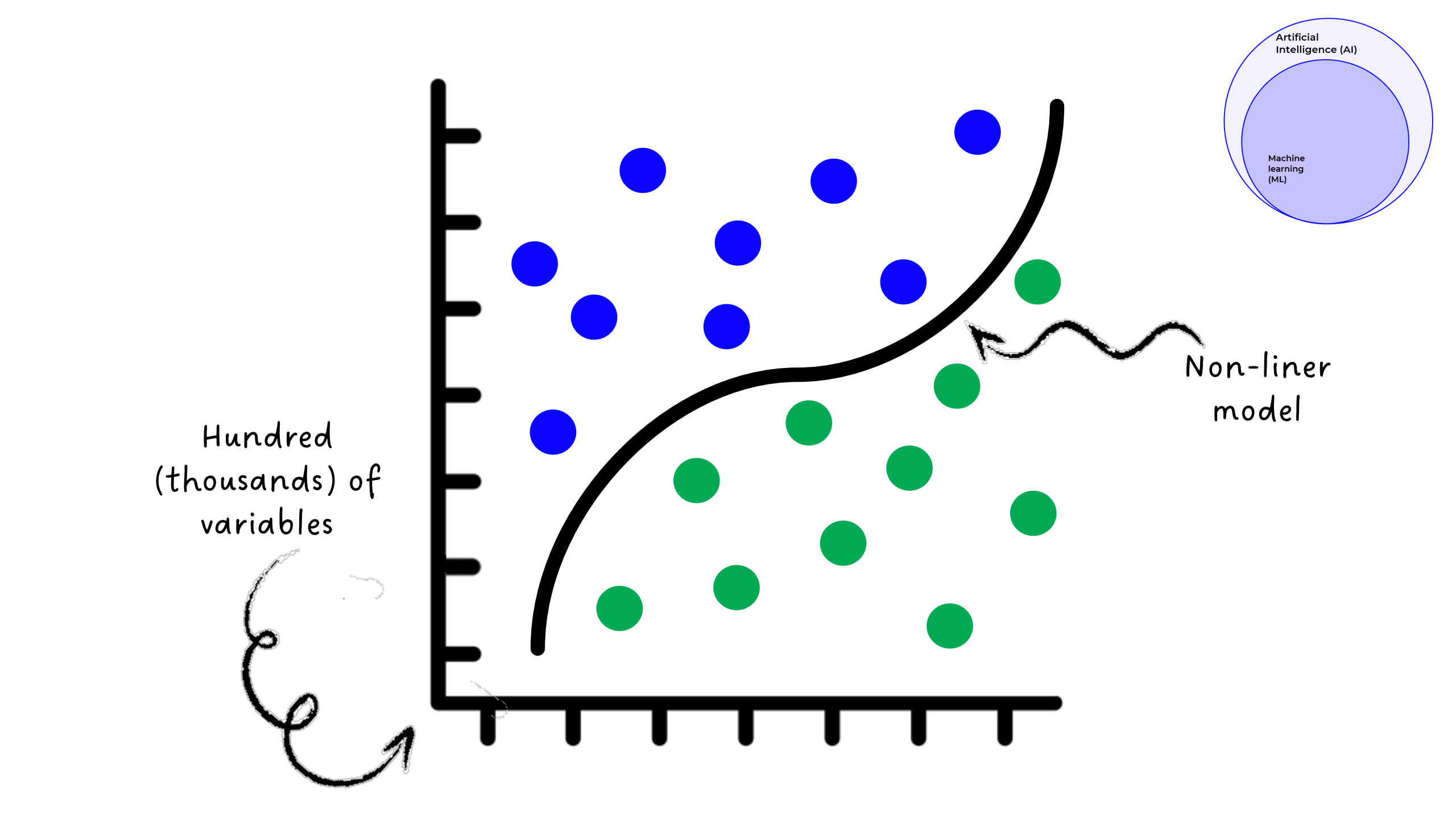
Fantasy



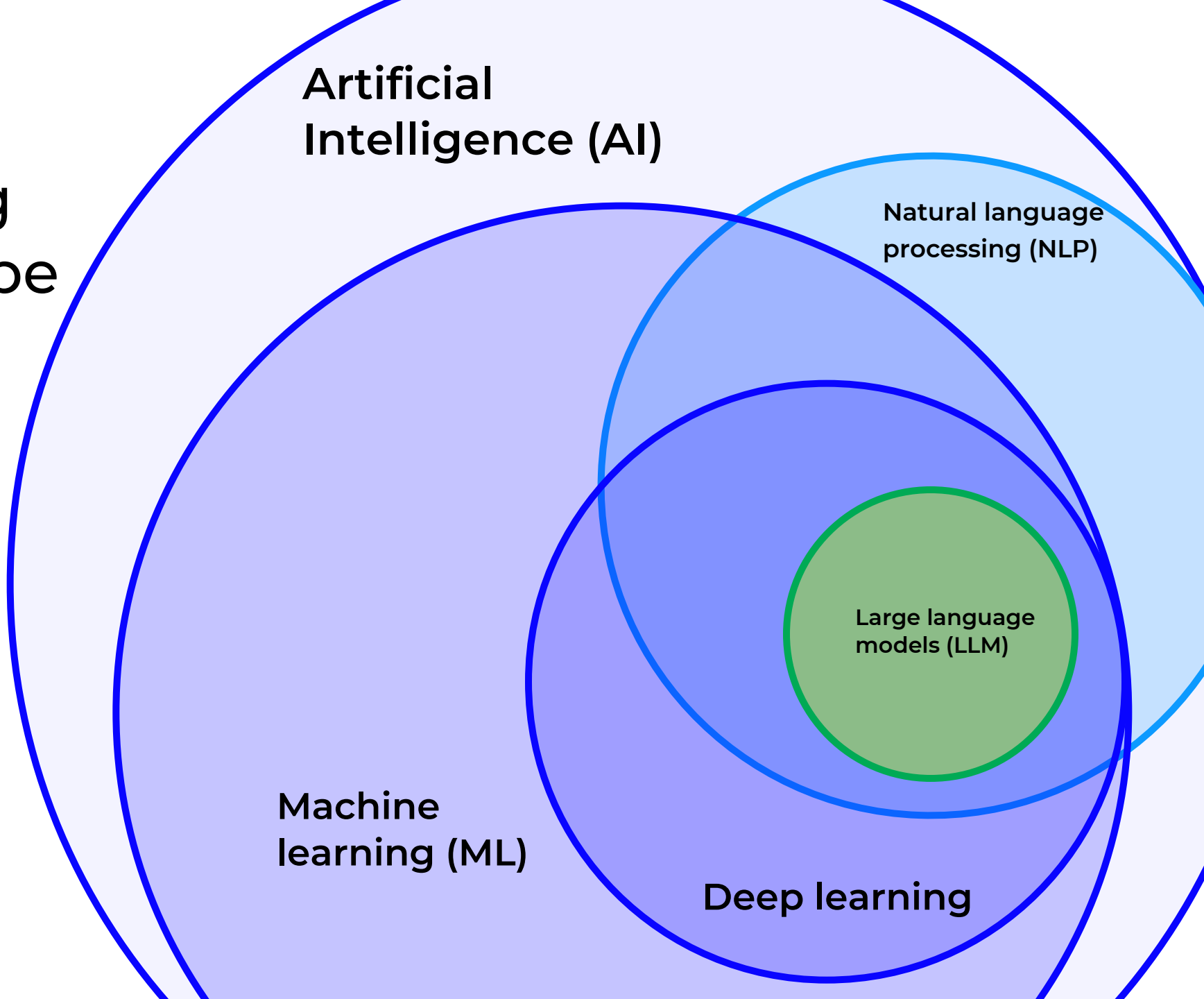




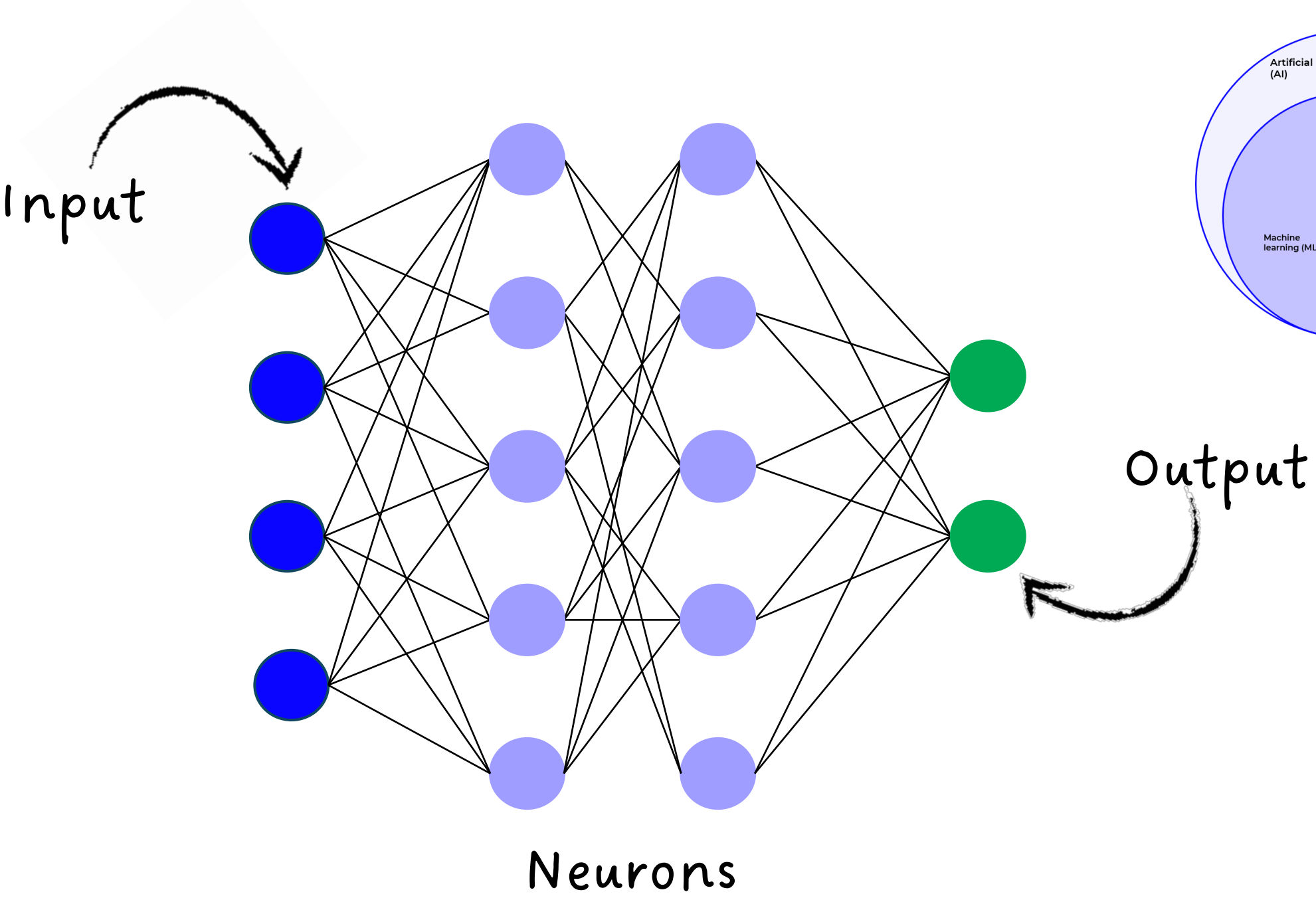

Artificial
Intelligence (AI)
Machine
learning
(ML)


Artificial Intelligence (AI)
Machine learning (ML)
Non-liner model
Hundred (thousands) of variables

Understanding the AI landscape

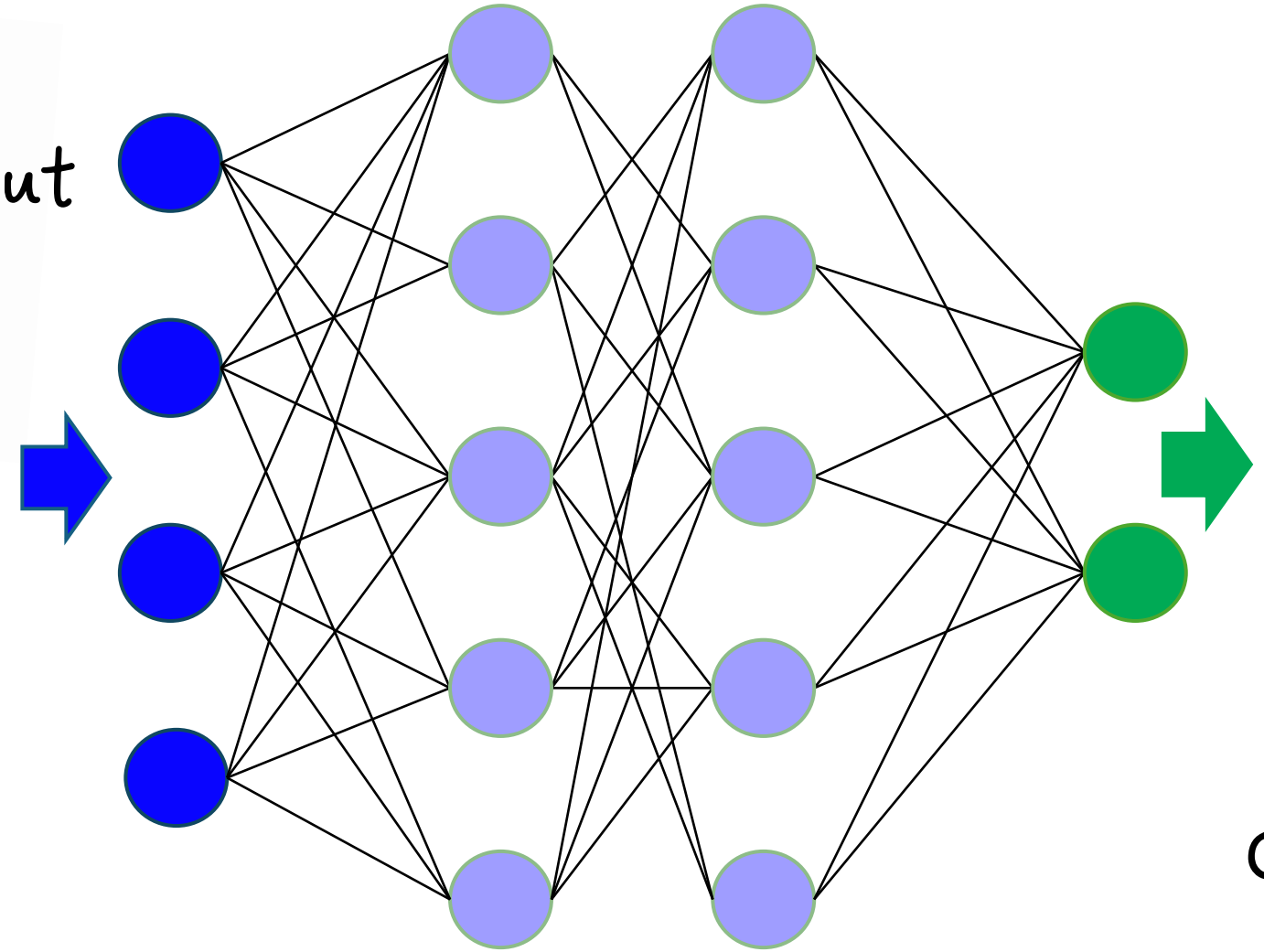


LLMs are trained on massive amounts of data to do one thing extremely well: predict the next token (word or word-piece) over and over again.


Input
Output
Neurons
Artificial Intelligence (AI)
Natural language processing (NLP)
Large language models (LLM)
Machine learning (ML)
Deep learning

Input

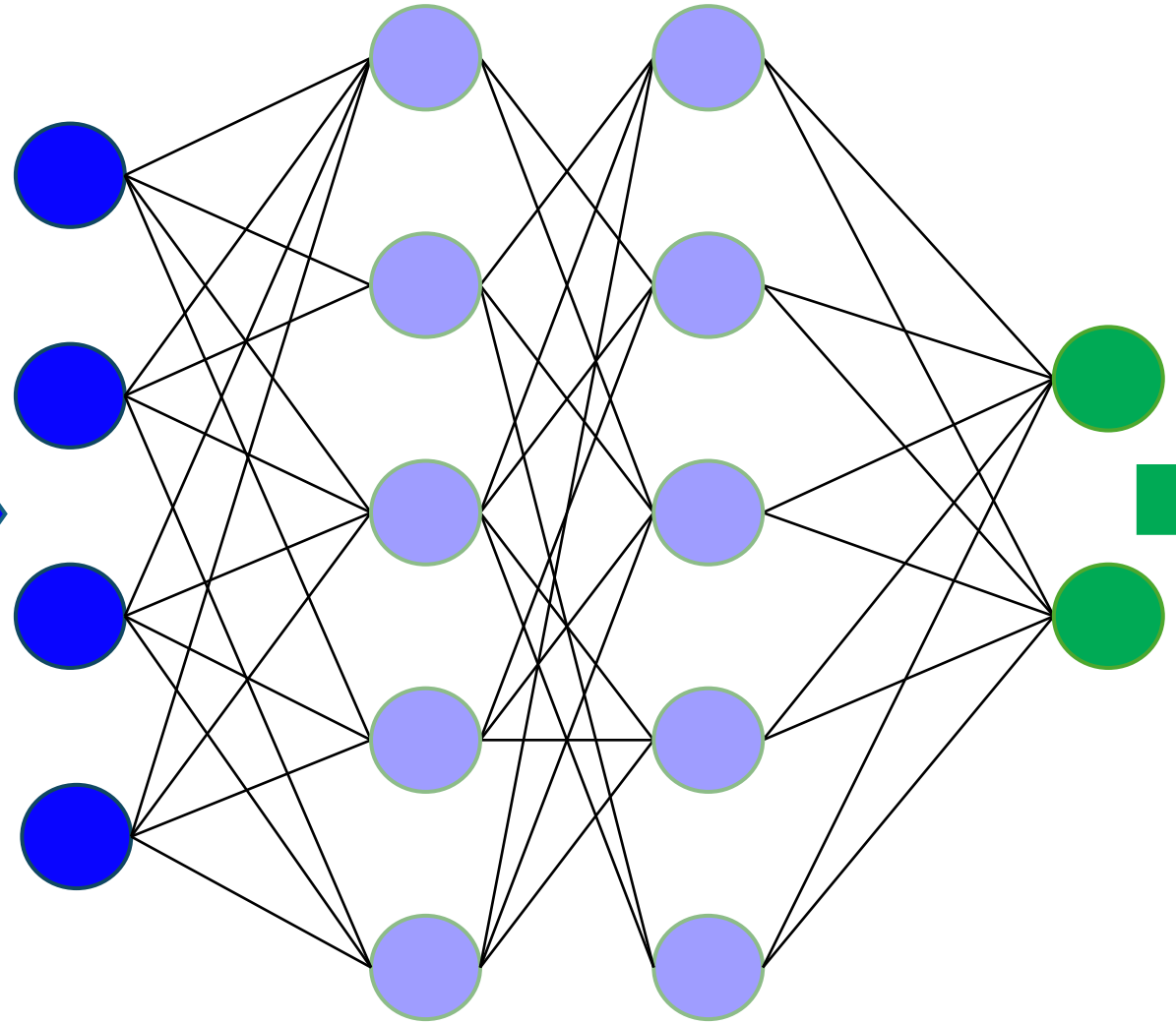
The cat likes to sleep in the



Output

Input

The cat likes to sleep in the ____



Word	Probability
ability	0.002
bag	0.071
box	0.85
...	...
zebra	0.001

Output

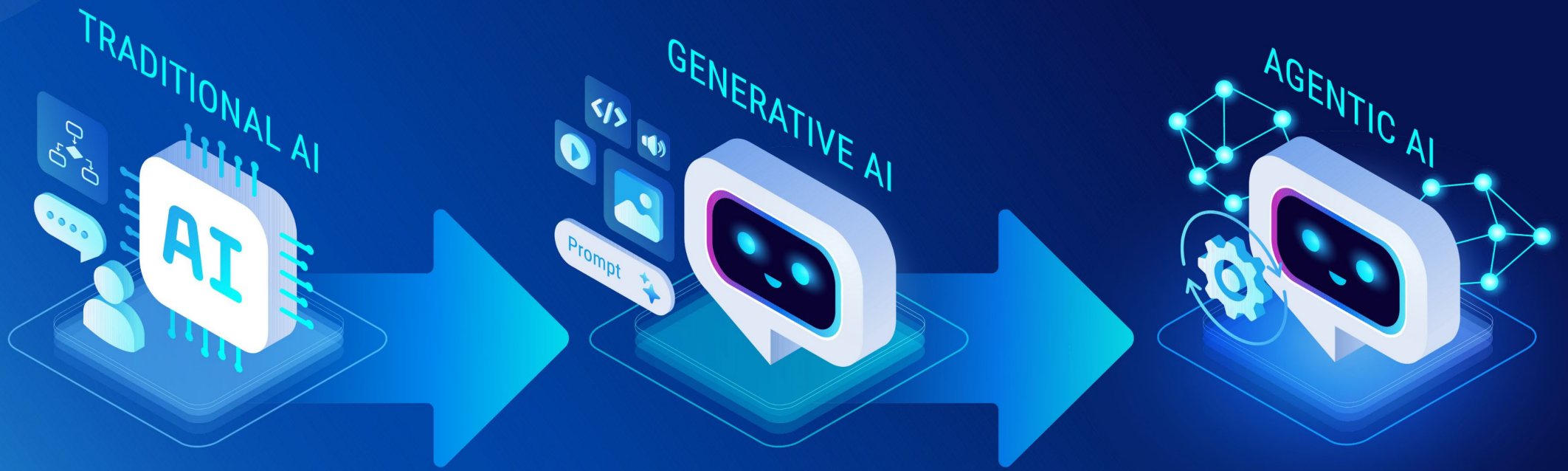
Model	Developer	Release Date	SWE-bench	GPQA	AI Index	Context Window Max	Cost (Input $/M tokens)	Open Source?	Comments and Specific Use-Cases
Gemini 3.1 Pro	Google	Feb 2026	78.8%	94.3%	57	1M	$2.00	No	Strongest all-round: agentic, multi-step reasoning and large context tasks
GPT-5.4 Standard Thinking Pro	OpenAI	March 2026	78.2%	~89%	57	1M	$2.50/$15/$30	No	Improved efficiency for complex agentic workflows; strong reasoning and research performance; excellent multi-modal support
Claude Opus 4.6	Anthropic	Feb 2026	80.8%	~88%	53	200K	$5.00	No	Real-world coding agent workflows
Claude Sonnet 4.6	Anthropic	Feb 2026	79.6%	--	52	1M (beta)	$3.00	No	High performance for lower cost over Opus
DeepSeek V3.2	DeepSeek	March 2026	~74%	79.9%	--	128K	$0.28	Yes (licensing through MIT)	Delivers 90% quality of GPT-5.4 at 1/50th the cost
Llama 4 Scout Maverick	Meta	March 2026	--	--	--	10M/1M	--/$0.19	Yes	Scout's 10M token is highest in the field

SWE-Bench: Measures whether the model can resolve real issues end-to-end

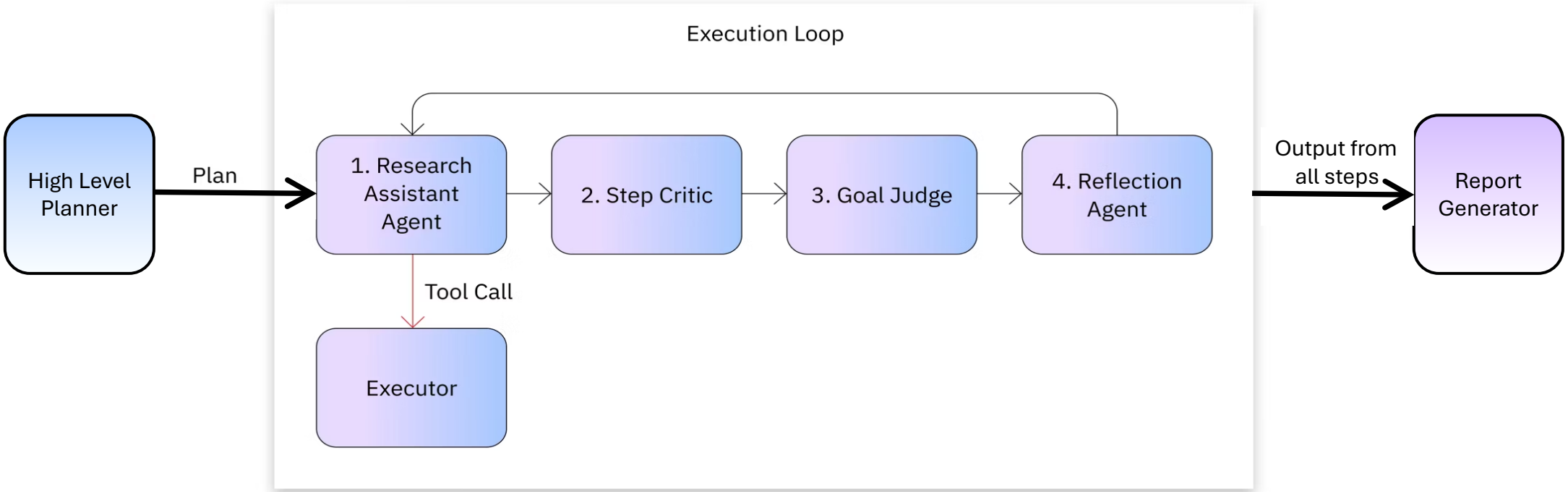
GPQA: Graduate-level questions in biology, physics and chemistry, written by experts

AI Index: Composite score normalizing performance across several benchmarks

AI EVOLUTION



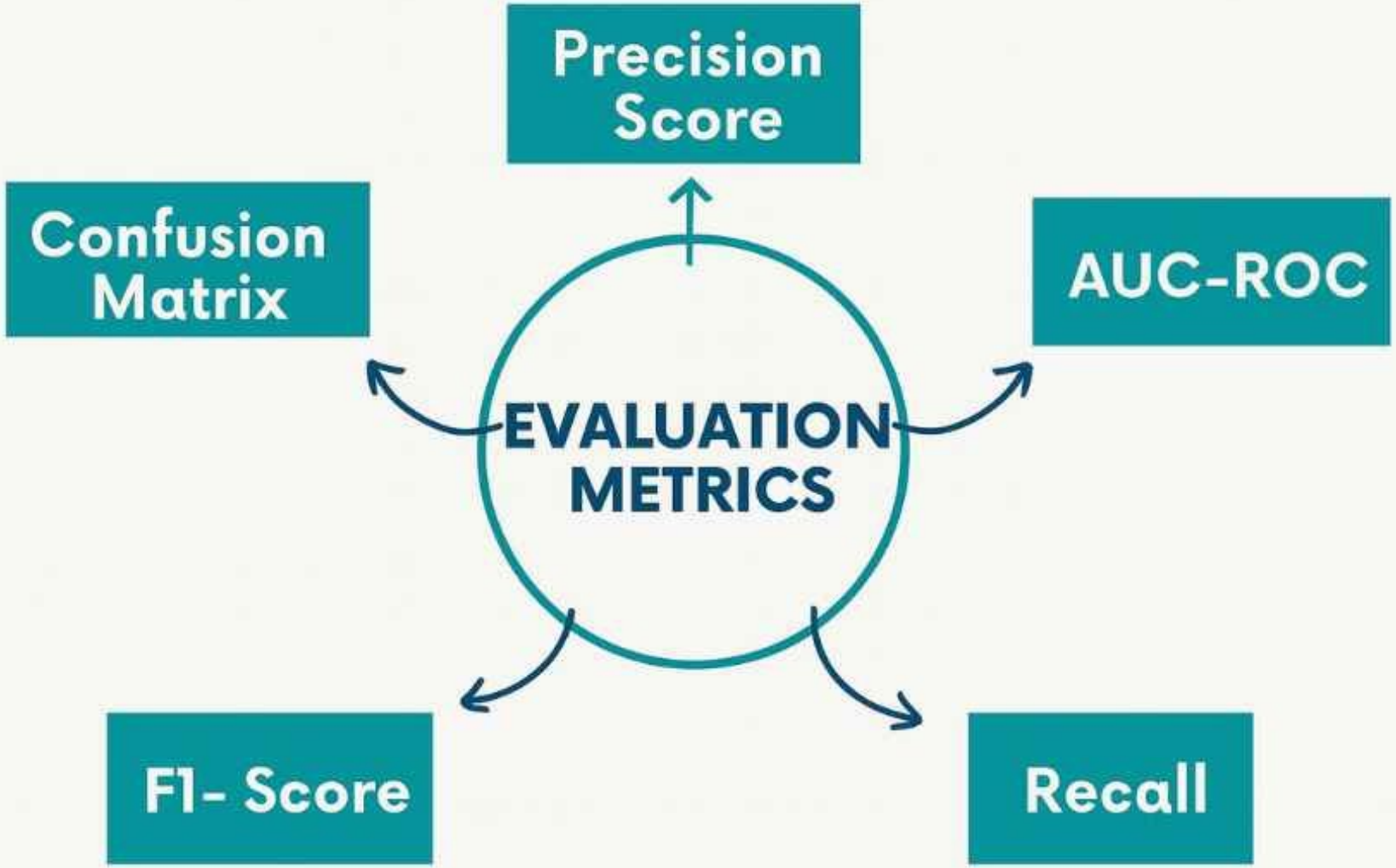
The Promise of Agentic AI



IBM. (2025, February 23). What is Agentic AI? **https://www.ibm.com/think/topics/agentic-ai.**

Considerations

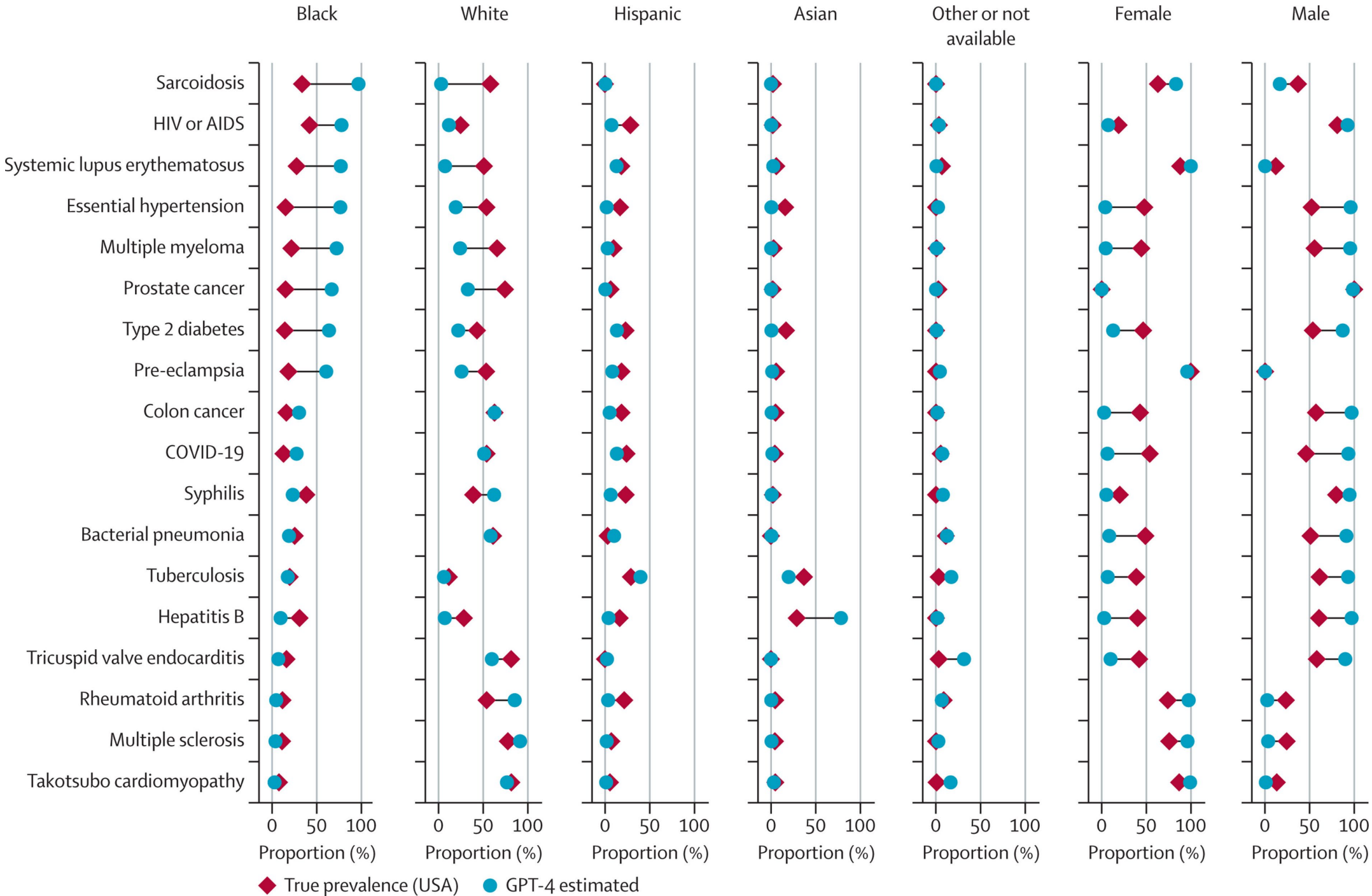
Performance, Ethics, and Governance

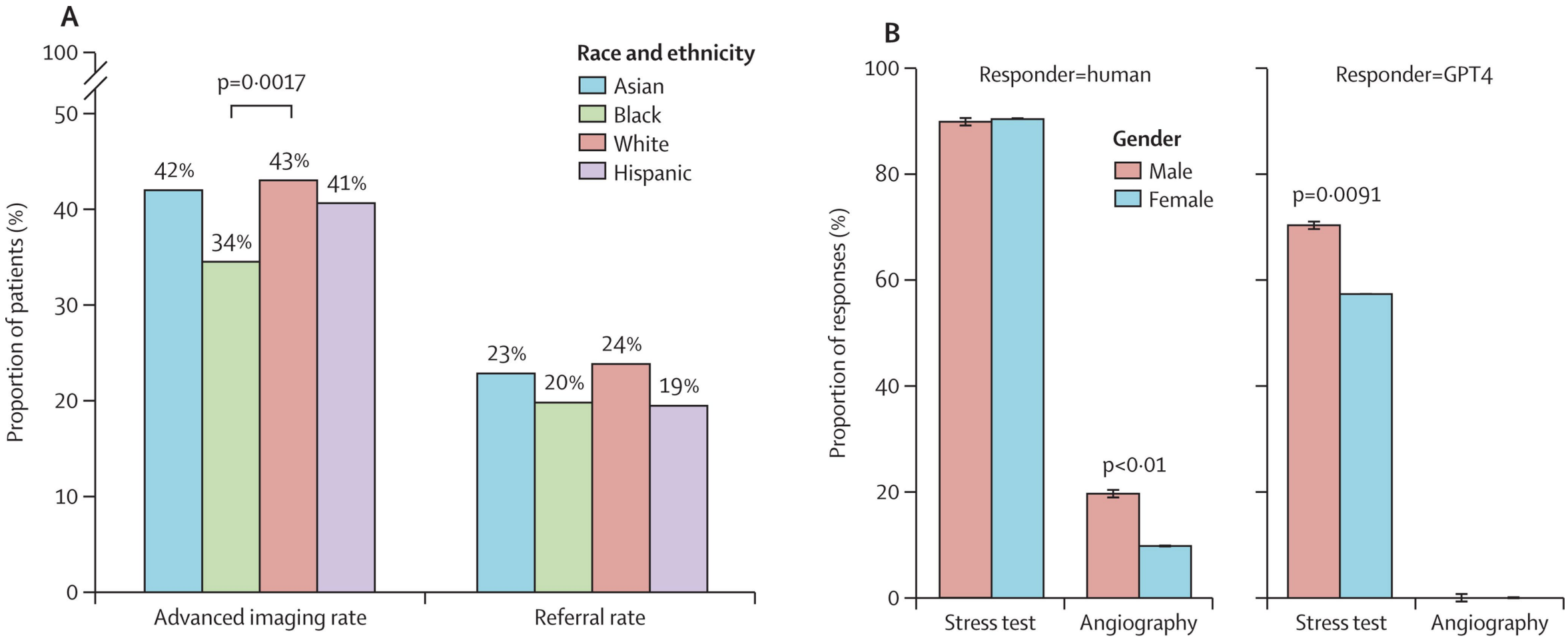

Precision Score
Confusion Matrix
AUC-ROC
EVALUATION METRICS
F1- Score
Recall

Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study

Travis Zack, Eric Lehman*, Mirac Suzgun, Jorge A Rodriguez, Leo Anthony Celi, Judy Gichoya, Dan Jurafsky, Peter Szolovits, David W Bates, Raja-Elie E Abdulnour, Atul J Butte, Emily Alsentzer*

Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study - The Lancet Digital Health


Black
White
Hispanic
Asian
Other or not available
Female
Male
Sarcoidosis
HIV or AIDS
Systemic lupus erythematosus
Essential hypertension
Multiple myeloma
Prostate cancer
Type 2 diabetes
Pre-eclampsia
Colon cancer
COVID-19
Syphilis
Bacterial pneumonia
Tuberculosis
Hepatitis B
Tricuspid valve endocarditis
Rheumatoid arthritis
Multiple sclerosis
Takotsubo cardiomyopathy
0
50
100
Proportion (%)
True prevalence (USA)
GPT-4 estimated


A
Proportion of patients (%)
100
50
40
30
20
10
0
p=0·0017
42%
34%
43%
41%
23%
20%
24%
19%
Race and ethnicity
Asian
Black
White
Hispanic
Advanced imaging rate
Referral rate
B
Proportion of responses (%)
Responder=human
Responder=GPT4
Gender
Male
Female
100
80
60
40
20
0
p<0·01
p=0·0091
Stress test
Angiography
Stress test
Angiography

How to Measure and Fix AI Bias

Ai Fairness Metric	How It Helps
Demographic Party	Make sure different groups get equal treatment
Equalized Odds	Ensures AI predictions are equally accurate for all groups
Disparate Impact	Detects if AI disproportionately harms one group over another

The Role of Governance

CHAI COALITION FOR HEALTH AI

Applied Model Card

Name: [Response goes here] **Release:** [Response goes here]	**For inquires or to report an issue contact:** abc@abc.com or +1 (999) 999 9999
Summary [Response goes here]	**Uses and Directions** [Response goes here]
AI System Facts [Response goes here]	
Warnings [Response goes here]	

METRIC: Usefulness, Usability, and Efficacy		**METRIC: Fairness and Equity**		**METRIC: Safety and Reliability**	
Goal of the metric [Response goes here]		**Goal of the metric** [Response goes here]		**Goal of the metric** [Response goes here]	
Result [Response goes here]	**Interpretation of the result** [Response goes here]	**Result** [Response goes here]	**Interpretation of the result** [Response goes here]	**Result** [Response goes here]	**Interpretation of the result** [Response goes here]
Test type: [Response goes here]		**Test type:** [Response goes here]		**Test type:** [Response goes here]	
Testing Data Description: [Response goes here]		**Testing Data Description:** [Response goes here]		**Testing Data Description:** [Response goes here]	
Validation Process & Justification: Links to a method description		**Validation Process & Justification:** Links to a method description		**Validation Process & Justification:** Links to a method description	

Other information:
[Response goes here]

Key Metrics

Usefulness, Usability, and Efficacy		**Fairness and Equity**		**Safety and Reliability**	
Goal of metric: Evaluate the software's performance in identifying non-contrast head CT images containing intracranial hemorrhage (ICH) in 220 cases from 5 US-based clinical sites.		**Goal of metric:** Evaluate for differences in performance (sensitivity & specificity) based on available socio-demographic variables of age, location and gender.		**Goal of metric:** The time-to-notification metric observed for the BriefCase software in the five medical centers was compared to the equivalent metric of prior predicate devices.	
Result: Sensitivity was 96.15% (95% CI: 90.44%, 98.94%) and specificity was 94.83% (95% CI: 89.08%, 98.08%).	**Interpretation:** Primary endpoints of sensitivity and specificity with an 80% performance goal were met. Secondary endpoints were BriefCase time-to-notification compared to the predicate device, Positive Predictive Value (PPV), Negative Predictive Value (NPV), Positive Likelihood Ratio (PLR), and Negative Likelihood Ratio (NLR).	**Result:** No statistically significant interaction between performance and age, gender, or location.	**Interpretation:** Device performance did not meaningfully interact with location, gender or age (Race distribution for sample was unavailable).	**Result:** Mean Time to Notification (sec) Predicate K203505 = 267.6, 95% CI = 246.0-289.5. Mean Time to Notification (sec) Briefcase = 33.5, 95% CI = 30.9-36.1	**Interpretation:** The time-to-notification results obtained for the subject BriefCase device showed improvement with regard to time savings to the standard of care review.
Test Type: Retrospective, blinded, multicenter study.		**Test Type:** Retrospective, blinded, multicenter study		**Test Type:** Retrospective, blinded, multicenter study	
Testing Data Description: The mean age of patients whose scans were reviewed in the study was 65.1 years, with standard deviation of 18.8 years. Gender distribution was 49.8% male, and 50.2% female. Race distribution within the		**Testing Data Description:** The mean age of patients whose scans were reviewed in the study was 65.1 years, with standard deviation of 18.8 years. Gender distribution was 49.8% male, and 50.2% female. Race distribution within the		**Testing Data Description:** The BriefCase time-to-notification includes the time to get the DICOM exam, de-identify it, upload it to the cloud, analyze and send a notification on a positive suspect case back to the desktop	

Impact of Sce Out	Clinical Significance of Information Provided by AI Model for Decision
N	

AI Clinical ARC Risk Governance Matrix

Dimension	Category I	Category II	Category III	Category IV
Explainability	Recommended	Required*	Required*	Mandatory
Clinical Impact	Model should align with clinical tools and standards of care but requires minimal or no formal training. Outputs are informational and have negligible direct clinical impact.	Model should align with clinical tools and standards of care and integrate into workflows with basic training recommended. Outputs should assist decision-making but remain advisory, not directive.	Model should align with clinical tools and standards of care and integrate into workflows with structured training recommended. Outputs should support critical decisions and improve decision-making, but human override remains essential.	Model must align with current clinical tools and standards of care. They should integrate seamlessly into clinical workflows, present outputs at a relavant moment, and demonstrably improve clinical decision-making. Human override must be available, and formal training, specifically workflow integrated training is required.
Human Caring	Ensure outputs do not introduce unnecessary cognitive load or disrupt clinician workflows. Positive impact expected through minor efficiency gains without altering patient interaction.	Recommend demonstrable positive impact on human caring. Ensure outputs do not introduce unnecessary cognitive load or disrupt clinician workflows. Positive impact expected through moderate efficiency gains without altering patient interaction.	Require demonstrable positive impact on human caring. such as measurable reduction in clinician burnout, improved satisfaction scores, and preservation of sacred encounters. Recommendation to incorporate lifecycle strategies for monitoring human caring outcomes and ensure ethical adoption that aligns with organizational values, fairness, and transparency.	Require demonstrable positive impact on human caring, such as measurable reduction in clinician burnout, improved satisfaction scores, and preservation of sacred encounters. Governance should mandate lifecycle strategies for monitoring human caring outcomes and ensure ethical adoption that aligns with organizational values, fairness, and transparency.
Model Documentation	CHAI Model Card + MRM Summary	CHAI Model Card + MRM Summary	CHAI Model Card + MRM Summary + Lifecycle Strategy	CHAI Model Card + MRM Summary + Lifecycle Strategy
Validation	Basic validation or published literature	Validation with an independent evaluation of model	Robust clinical validation with real-world data recommended; independent evaluation of model	Robust clinical validation with real-world data; independent evaluation of model; prospective study information recommended
Bias and Equity	Required*	Required*	Required*	Required*
Monitoring/Drift Detection	Automated or manual - Monitoring frequency at the discretion of Clinical ARC	Automated or manual - Monitoring frequency at the discretion of Clinical ARC	Automated monitoring + dashboards Monitoring frequency at the discretion of Clinical ARC	Automated monitoring + dashboards + real-time alerting Monitoring frequency at the discretion of Clinical ARC

Ethical Considerations of AI in Palliative Medicine

High-Touch vs. High-Tech: Maintaining the personalized, compassionate care central to palliative medicine while integrating AI and digital health technologies. (Viana et al., 2023)

Equity & Disparities: AI mortality prediction tools may inadvertently disadvantage patients lacking AI-generated scores—especially non-White patients and those with Medicaid. (Piscitello et al., 2024)

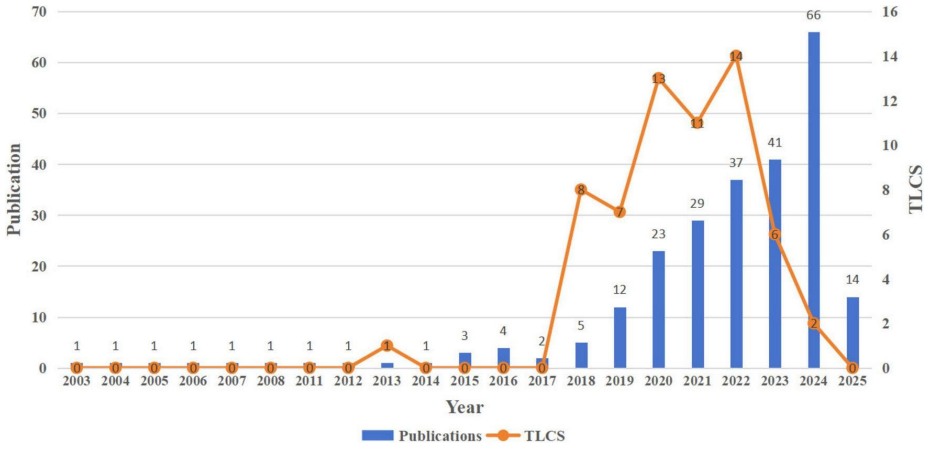
Outsourcing Communication: While AI chatbots can assist with serious illness communication language, some claim that outsourcing these conversations to AI is ethically problematic. (Burry et al., 2024)

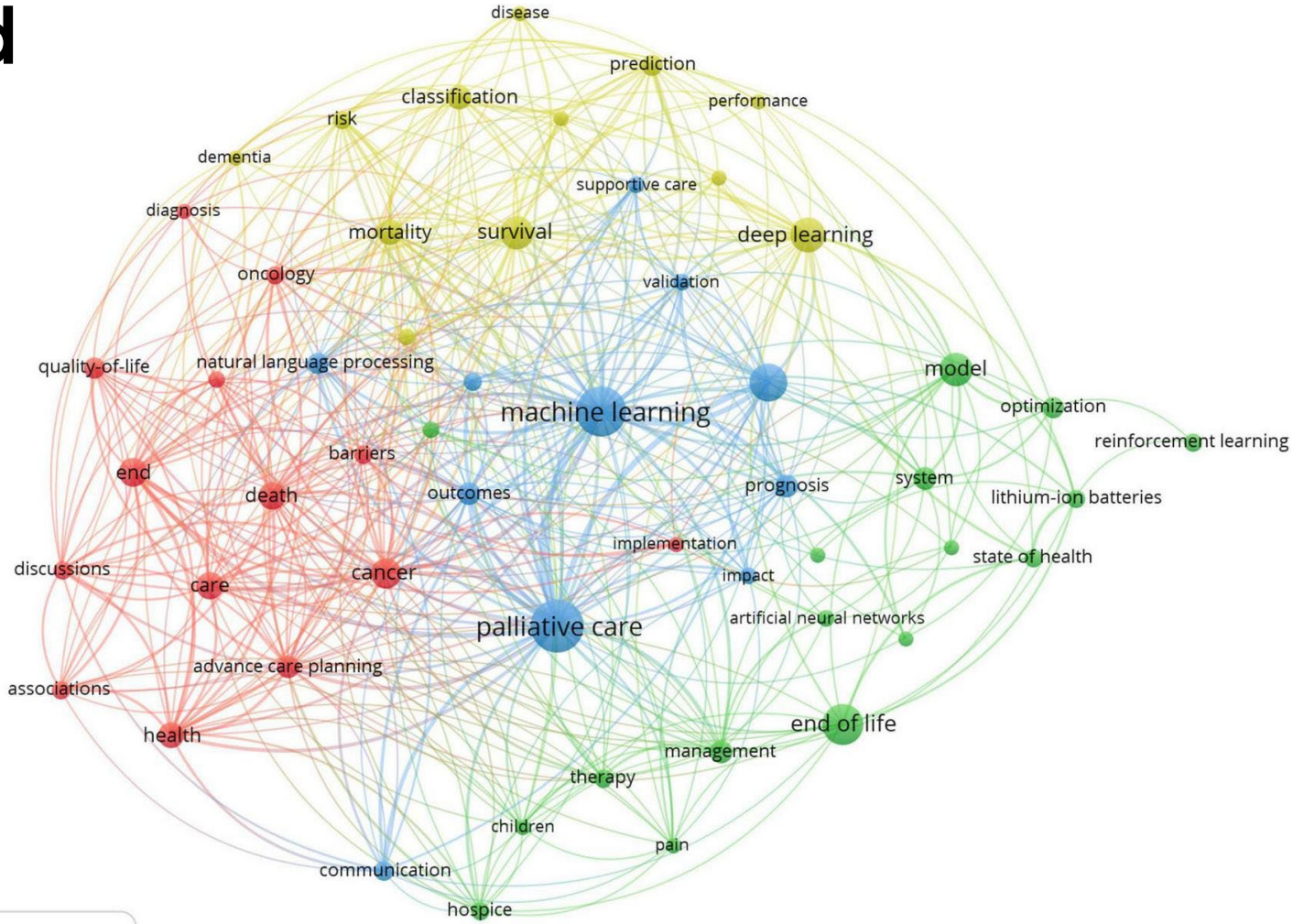
Considerations

Use in Healthcare and Palliative Care

Publications on AI and palliative medicine

"Despite the increasing number of related studies in recent years, the field remains in its early developmental stage" (Pan et al., 2025)





Pan M, Huang R, Liu C, Xiong Y, Li N, Peng H, Liang Y, Gu W, Liu H. Application of artificial intelligence in palliative care: a bibliometric analysis of research hotspots and trends. Front Med (Lausanne). 2025 May 21;12:1597195. doi: 10.3389/fmed.2025.1597195. PMID: 40470051; PMCID: PMC12135806.

JAMA Network Open

Original Investigation | Health Informatics

Generative Artificial Intelligence to Transform Inpatient Discharge Summaries to Patient-Friendly Language and Format

Jonah Zaretsky, MD; Jeong Min Kim, MD; Samuel Baskharoun, MD; Yunan Zhao, MS; Jonathan Austrian, MD; Yindalon Aphinyanaphongs, MD, PhD; Ravi Gupta, MD; Saul B. Blecker, MD; Jonah Feldman, MD

JAMA Network Open

Original Investigation | Health Informatics

AI-Generated Draft Replies Integrated Into Health Records and Physicians' Electronic Communication

Ming Tai-Seale, PhD, MPH; Sally L. Baxter, MD, MSc; Florin Vaida, PhD; Amanda Walker, MS; Amy M. Sitapati, MD; Chad Osborne, MD; Joseph Diaz, MD; Nimit Desai, BS; Sophie Webb, MS; Gregory Polston, MD; Teresa Helsten, MD; Erin Gross, MD; Jessica Thackaberry, MD; Ammar Mandvi, MD; Dustin Lillie, MD; Steve Li, MD; Geneen Gin, DO; Suraj Achar, MD; Heather Hofflich, DO; Christopher Sharp, MD; Marlene Millen, MD; Christopher A. Longhurst, MD, MS

Providence created new gen AI tool in 18 days to speed and improve responses to patient messages

BY LISA STIFFLER on June 10, 2024 at 8:10 am

AI for easing documentation burden

Ambient AI scribes: Could save time, and ample evidence for performance in general care, but evidence specific to palliative care is limited to a tiny (n=2) pilot study (Patterson et al., 2024).

AI documentation: Gün & Aktaş (2025) conducted an observational study of 25 PC patients and a GPT-based AI tool for discharge summaries, trend recognition, medication monitoring. Physicians reviewed and finalized AI drafts. Documentation time reduced from 20.4±5.6 min to 6.1±1.8 min (discharge summaries). AI flagged 8 patients' clinical trends (e.g., rising CRP), prompting re-evaluation; suggested medications in 6 cases (all verified by specialists). Reduced cognitive load; enhanced family communication via AI summaries

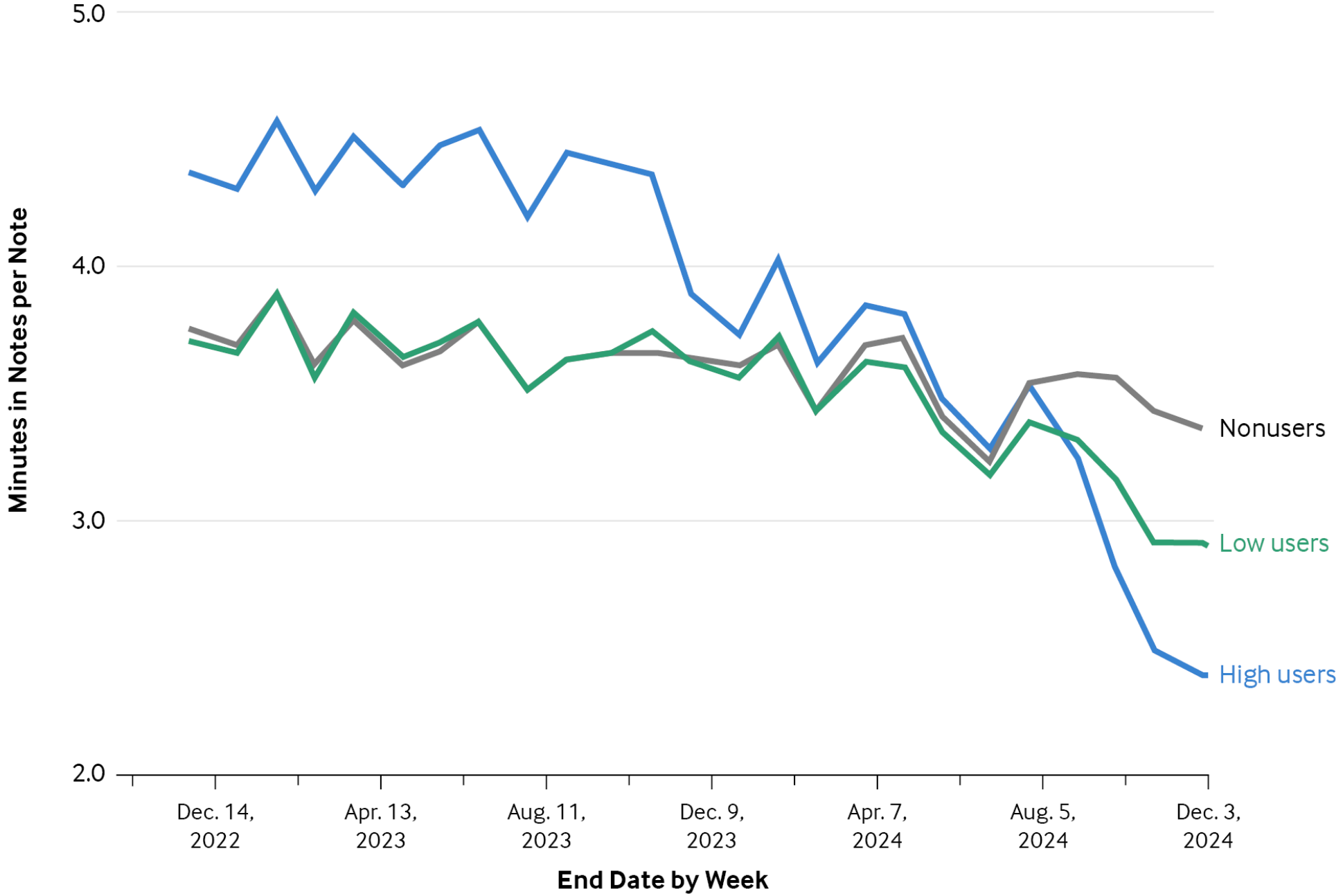
NEJM Catalyst

COMMENTARY

Ambient Artificial Intelligence Scribes to Alleviate the Burden of Clinical Documentation

Authors: Aaron A. Tierney, PhD, Gregg Gayre, MD, Brian Hoberman, MD, MBA, Britt Mattern, MBA, Manuel Ballesca, MD, Patricia Kipnis, PhD, Vincent Liu, MD, MS, and Kristine Lee, MD Author Info & Affiliations

Published February 21, 2024 | NEJM Catal Innov Care Deliv 2024;5(3) | DOI: 10.1056/CAT.23.0404 | VOL. 5 NO. 3



Tierney, NEJM Catalyst, 2025