AI & Medical Education

How AI is changing how we learn –
And how we can use it to teach

Jonathan Berry, MD
Associate Program Director, BIDMC IM Residency
Clinical Informatics Lead, BIDMC Hematology

June 10, 2026

Landing Page
tinyurl.com/jez53m24

Disclosures

I have no relevant financial relationships to disclose

GenAI tools used in preparation of this session

Gemini, NotebookLM, Claude (including Code & Cowork) were used to generate teaching materials — outlines, slides, web activities, apps, and graphics.

Don't worry — I'll show you how.

I have reviewed, edited, and confirmed all information being presented

Institutional guidelines are changing rapidly. My read reflects BIDMC and broader trends; policies vary by institution and will change over time.

Two Questions Anchor Today

How is AI changing how we learn — and subsequently how we teach?

How can we use AI to augment our capacities as educators?

Insert pre-survey slides here

Manufactured Inevitability

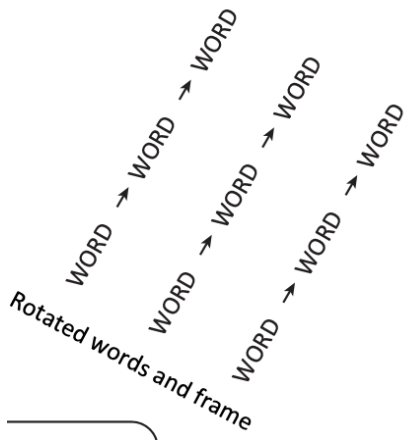
AI for the Learner

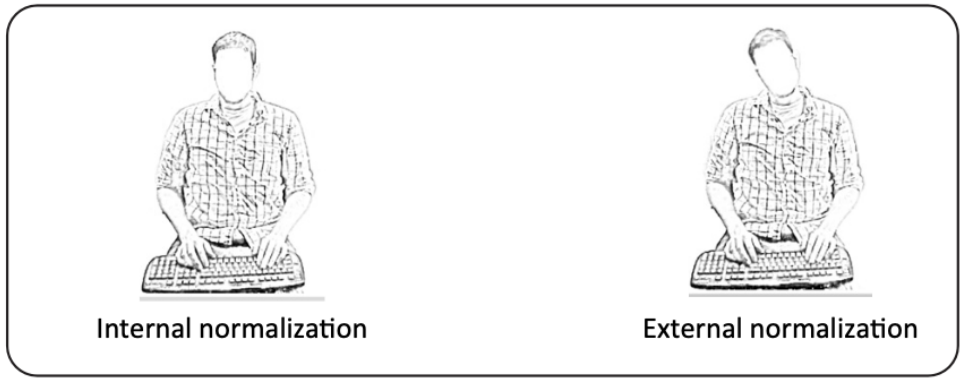

ADVANCED BIO-MED
ANATOMY
NEUROLOGY
ANATOMY
NEUROLOGY

Cognitive Offloading

Risko & Gilbert, Trends in Cognitive Sciences, 2016

Any external action that reduces internal cognitive demand





Cognitive Offloading

Risko & Gilbert, Trends in Cognitive Sciences, 2016

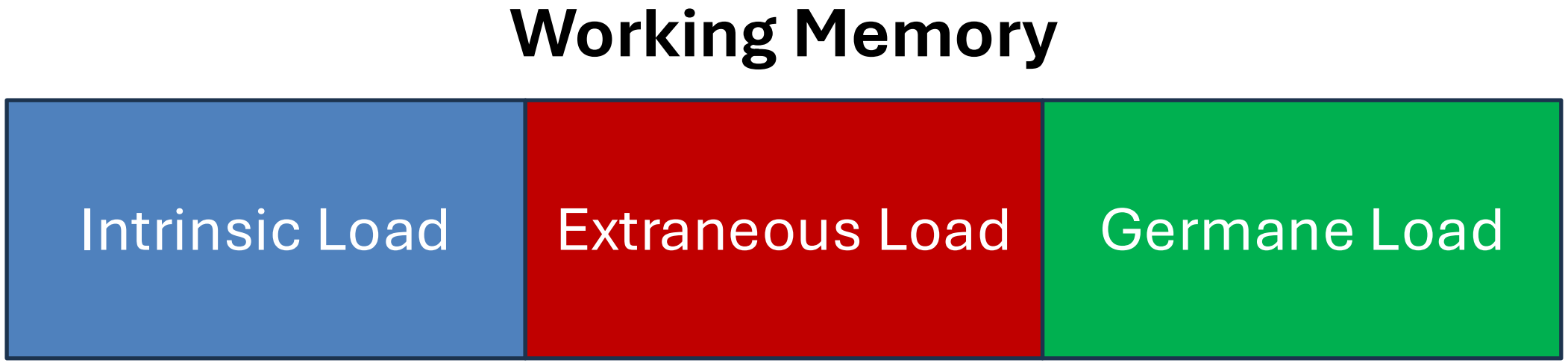
Any external action that reduces internal cognitive demand

— writing things down, using a calculator, setting an alarm, even rotating your head to read sideways text

Natural, automatic, metacognitive

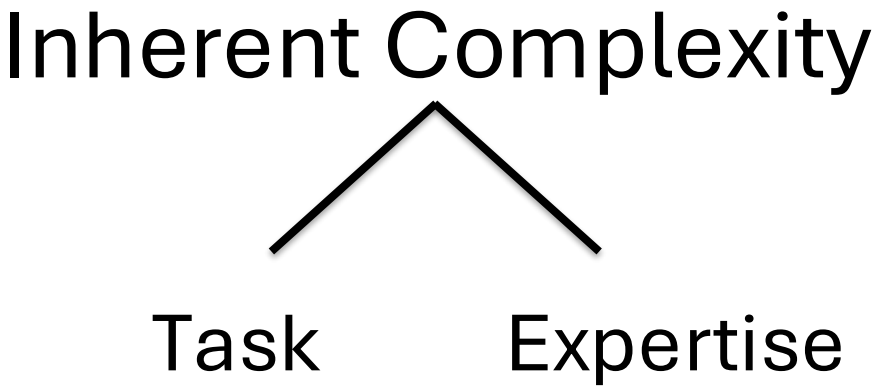
Depending on what's being offloaded,
can be adaptive or maladaptive

Cognitive Load Theory



Cognitive Load Theory

Intrinsic Load



Goal: Manage

Cognitive Load Theory

Extraneous Load

Distraction

Goal: Minimize

Cognitive Load Theory

Germane Load

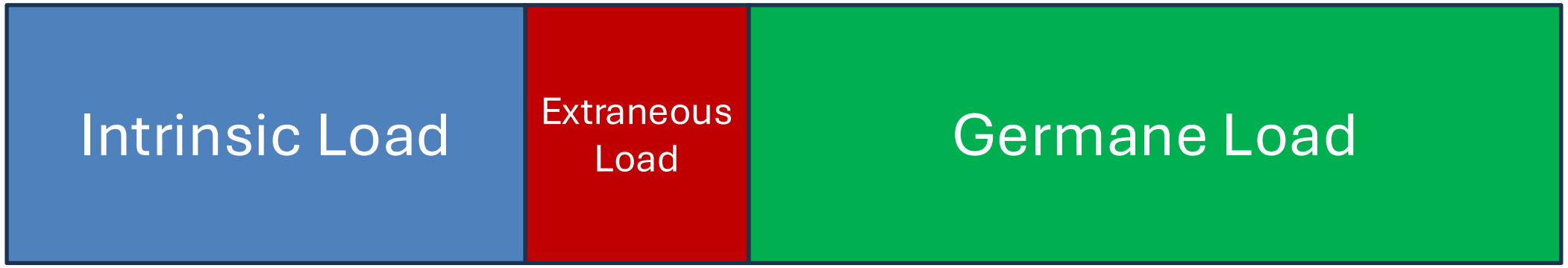
Learning

Goal: Maximize

Cognitive Load Theory

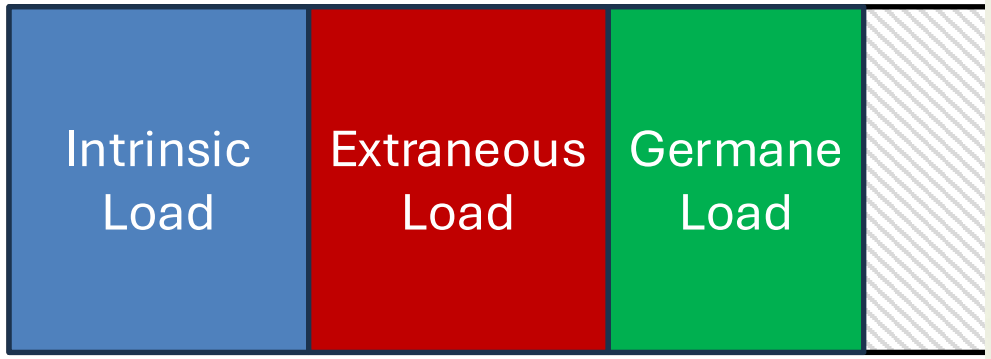
Manage	**Minimize**	**Maximize**
Guide attention	Eliminate distractions	Encourage effort

Cognitive Load Theory



Ideal goal with AI

Cognitive Load Theory



Potential concern

Alignment

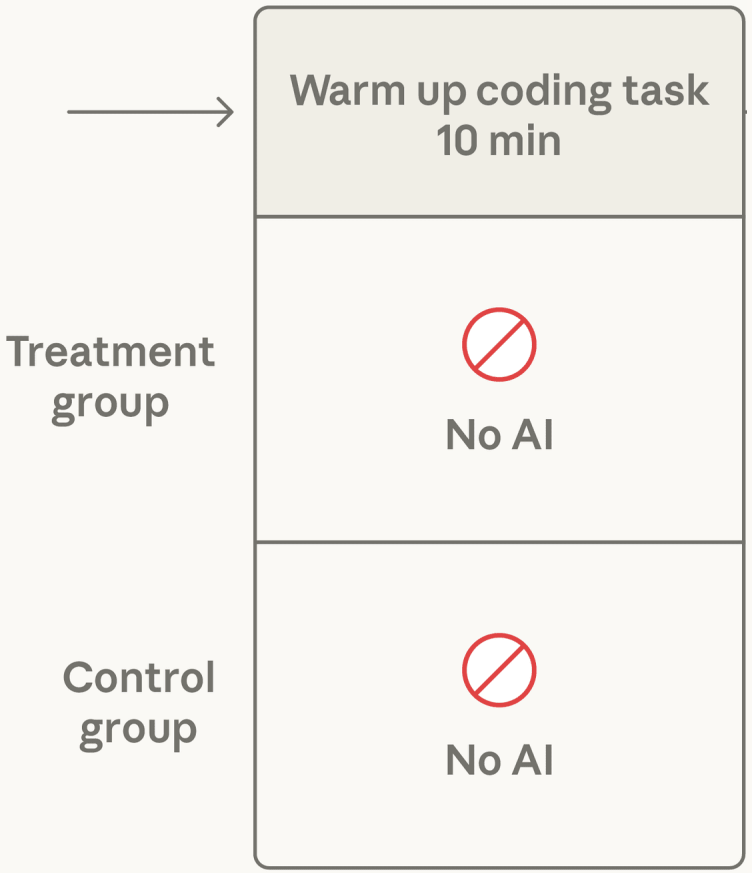
How AI assistance impacts the formation of coding skills

Jan 29, 2026

Read the paper

https://www.anthropic.com/research/AI-assistance-coding-skills

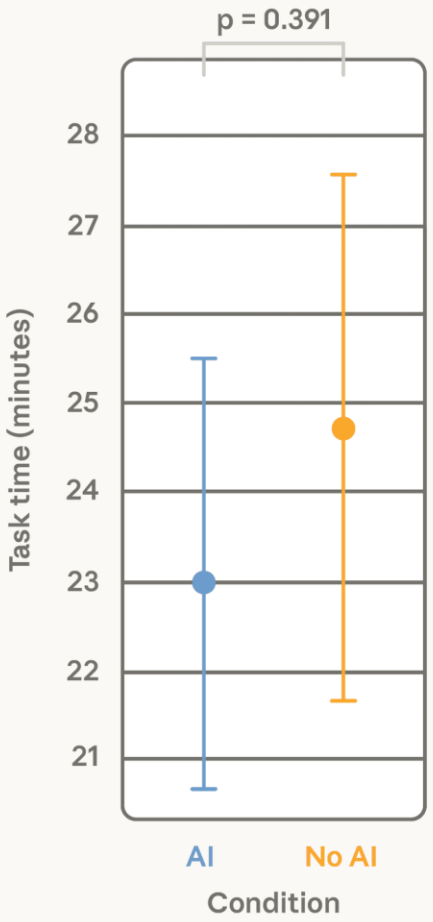
Study design



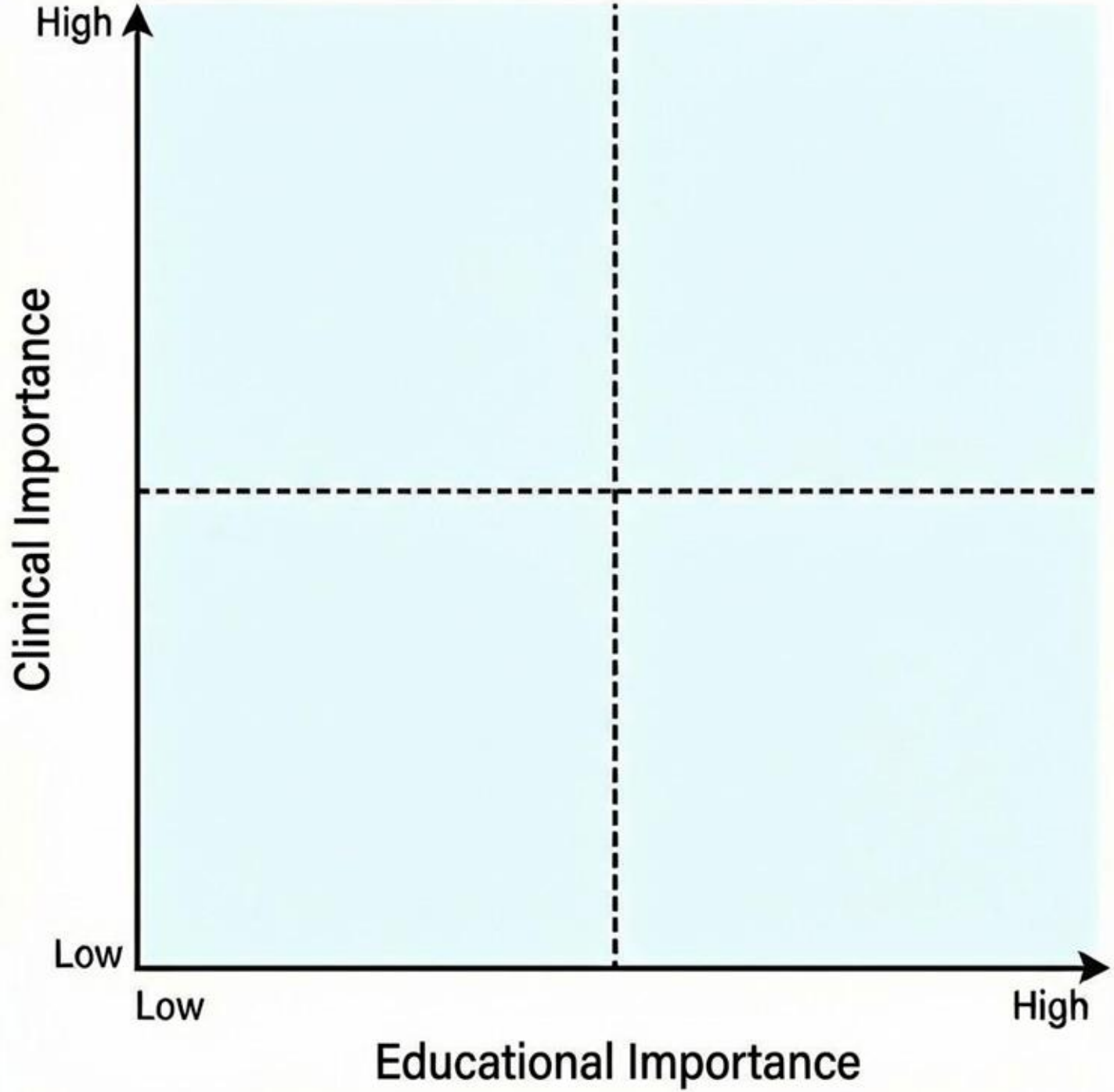
https://www.anthropic.com/research/AI-assistance-coding-skills

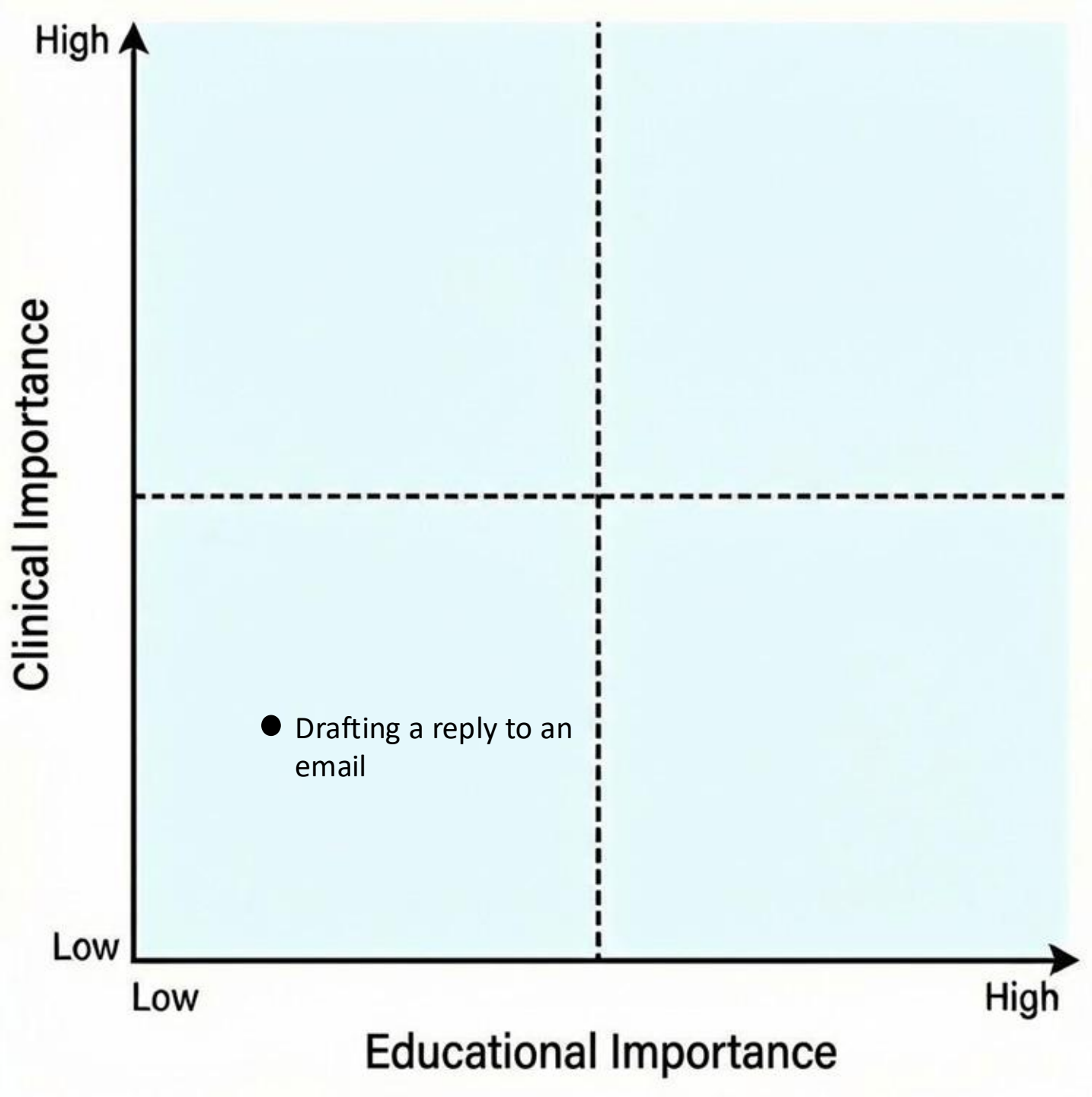
How AI assistance impacts coding speed and skill formation

AI assistance: treatment effect on coding speed and knowledge score



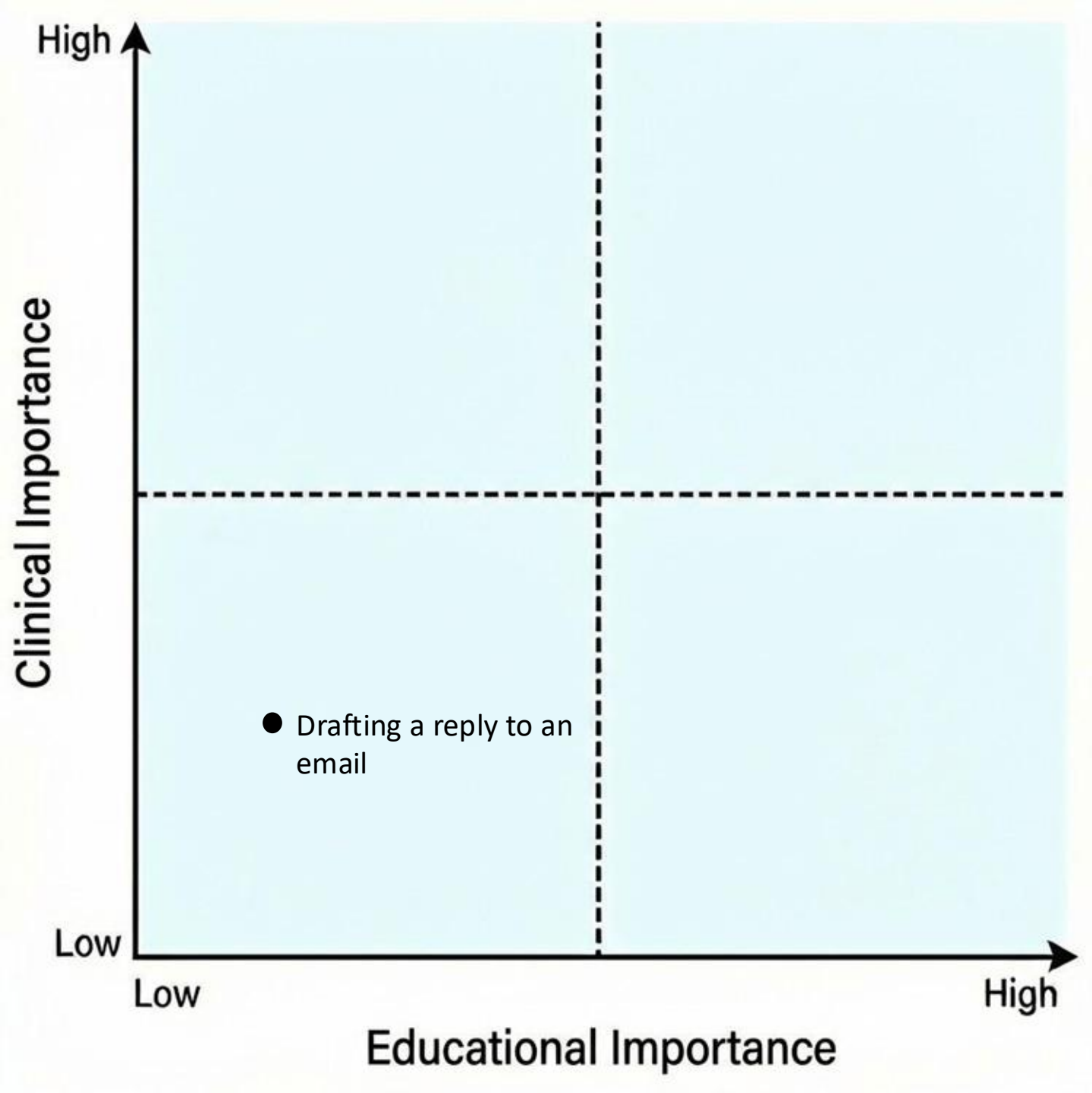
https://www.anthropic.com/research/AI-assistance-coding-skills


High
Clinical Importance
Low
Low
High
Educational Importance


High
Clinical Importance
Low
Drafting a reply to an email
Low
High
Educational Importance

High Clinical / Low Ed
Verify
Minimal learning risk. Must verify output as accuracy is key

High Clinical / High Ed
Patient Care
Tasks are highly educational and accuracy is vital. Requires careful use and careful verification



Low Clinical / Low Ed
Administrative
Minimal-to-no learning risk or patient care risk. Great place to start with deployment

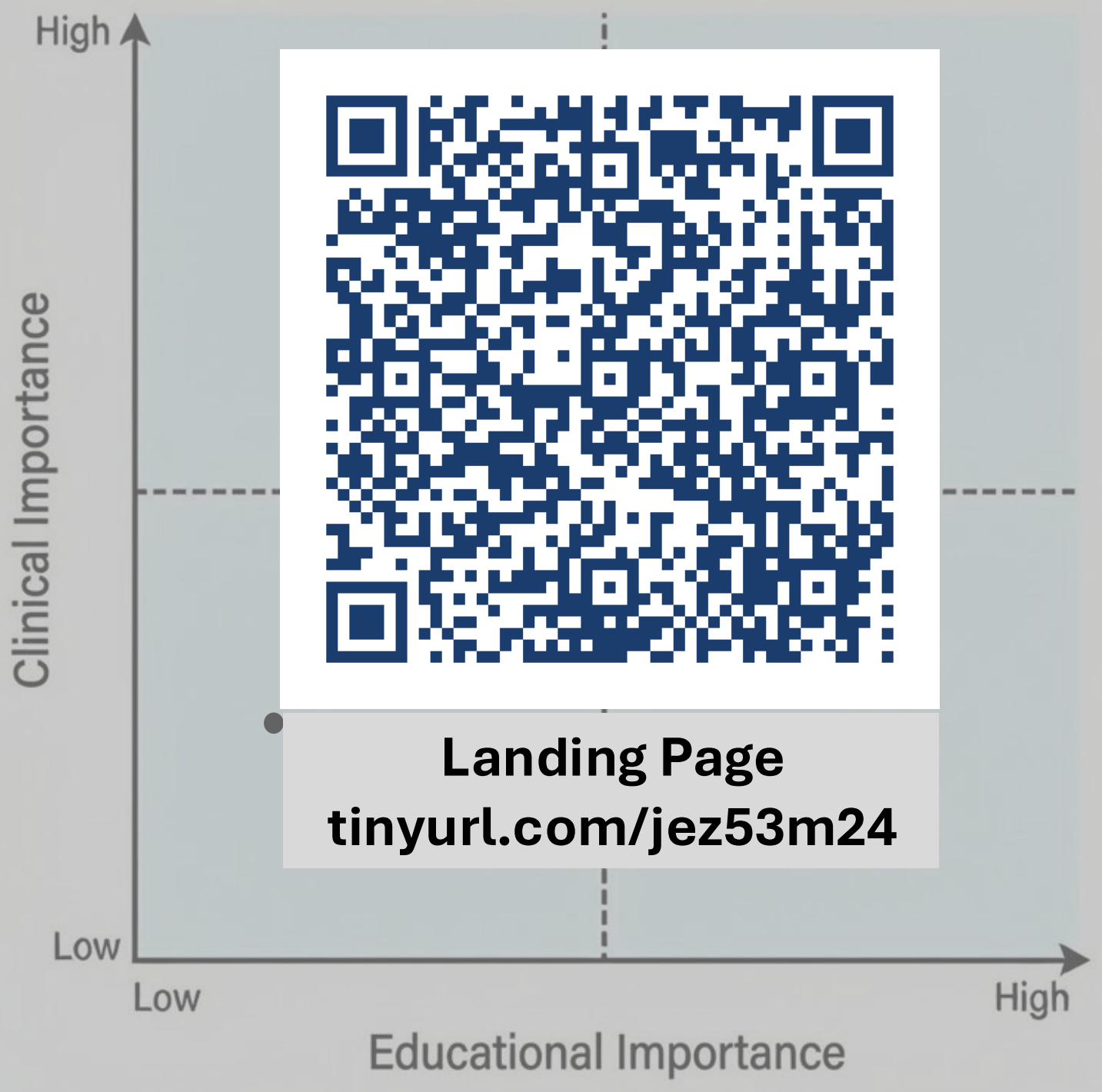
Low Clinical / High Ed
Pure Education
High risk to education if tasks are outsourced to AI without clear intentionality

High Clinical / Low Ed
Verify
Minimal learning risk. Must verify output as accuracy is key

High Clinical / High Ed
Patient Care
Tasks are highly educational and accuracy is vital. Requires careful use and careful verification

Low Clinical / Low Ed
Administrative
Minimal-to-no learning risk or patient care risk. Great place to start with deployment

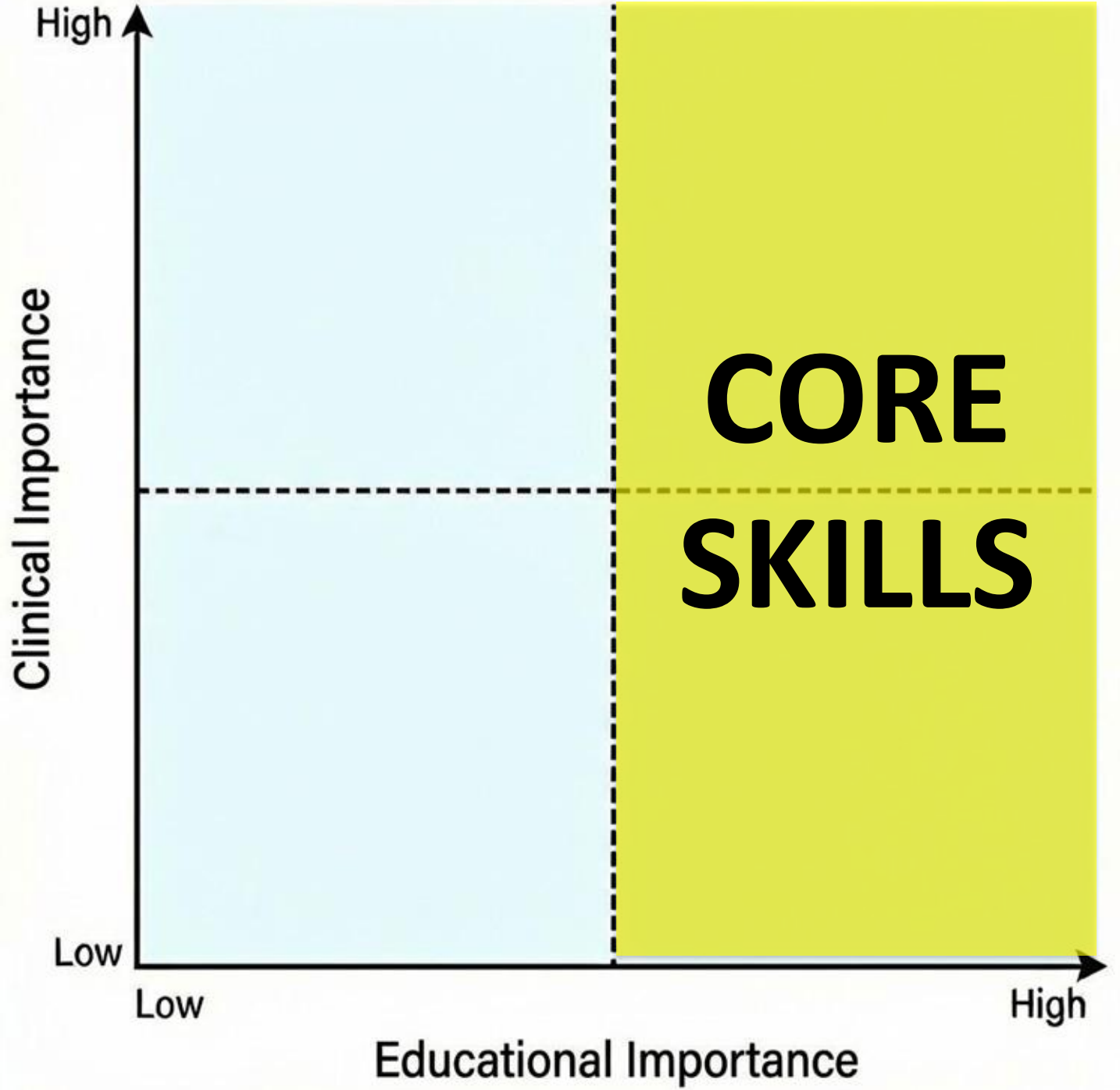
Low Clinical / High Ed
Pure Education
High risk to education if tasks are outsourced to AI without clear intentionality



Landing Page
tinyurl.com/jez53m24

High Clinical / Low Ed
Verify
Minimal learning risk. Must verify output as accuracy is key

Low Clinical / Low Ed
Administrative
Minimal-to-no learning risk or patient care risk. Great place to start with deployment



High Clinical / High Ed
Patient Care
Tasks are highly educational and accuracy is vital. Requires careful use and careful verification

Low Clinical / High Ed
Pure Education
High risk to education if tasks are outsourced to AI without clear intentionality

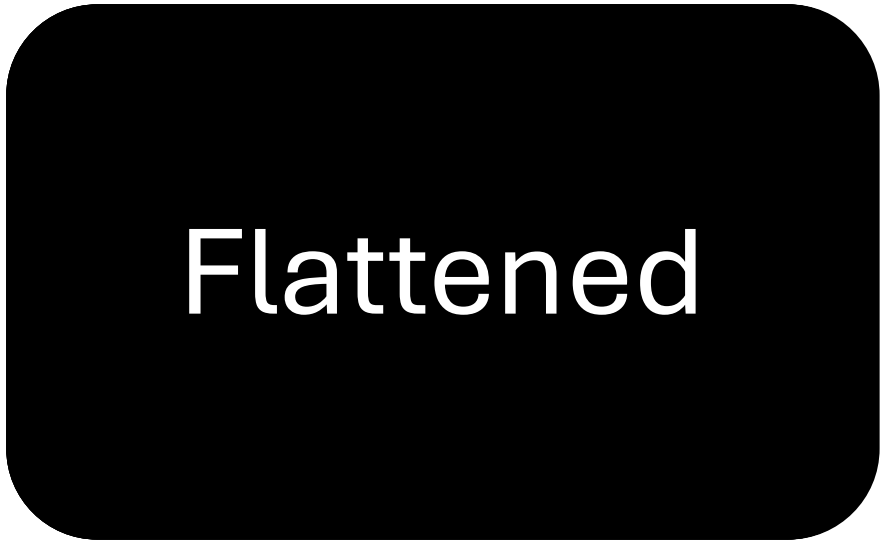
Classifying Questions

Empirical
Facts

Phronetic
Judgment

Unknowable

Classifying Questions



A New Admission — Fatigue & Back Pain

CASE SUMMARY

69F with HTN and T2DM with months of fatigue and low back pain. Recently completed 14-day linezolid for MDR-MRSA cellulitis.

Labs: Normocytic anemia (Hb 9.1), AKI (Cr 1.8 from 1.1), total protein 8.5, albumin 3.2. Ferritin 60, TSAT 30% (normal).

PROMPT 1

The intern asks: "How long do you need to be on linezolid for it to cause cytopenias?"

The resident responds: "AI says bone marrow suppression typically occurs after >2 weeks, with thrombocytopenia being the most commonly affected cell line."

PROMPT 2

After presenting a very thorough differential, you ask the intern about their framework for anemia. They admit that they don't have a great one and plugged the patient data into Open Evidence to get their differential.

Landing Page
tinyurl.com/jez53m24

B Rounding — Ms. R, Day 8

CASE SUMMARY

59F with Parkinson's and IgA vasculitis (skin-limited) on high-dose steroids. Admitted for worsening purpura; improving and de-escalated from IV to PO steroids. On Bactrim SS prophylaxis ×4 days.

Today: new chest purpura, improving CRP (18 from 50), but new transaminase elevation (AST 84, ALT 70; normal Alk phos, T.bili).

PROMPT 1

While presenting the elevated transaminases, the intern states confidently:

"We should stop the Bactrim. I ran it through AI and it says this is DILI."

PROMPT 2

The senior resident reviews the plan and says:

"I looked up on Open Evidence the likelihood of Bactrim causing DILI four days after beginning prophylaxis dose — the timing makes this less likely."

Landing Page
tinyurl.com/jez53m24

AI for the Educator

Using the following slide as a model, if I asked about educational use, insert pre-survey info

Your Prework: What Did You Build?

Teaching summary prompt

A reusable prompt that generates a structured teaching summary from a clinical case or topic.

Cases → Slides

Converting clinical cases to PowerPoint slides using multimedia design principles.

Pharmacology game

An interactive educational pharmacology game.

Three Use Cases for Me (This Week)

Building out Clinical Competency Notes For AY26-27

Claude Cowork: Provided Excel File + Template (I'd previously made with Claude)

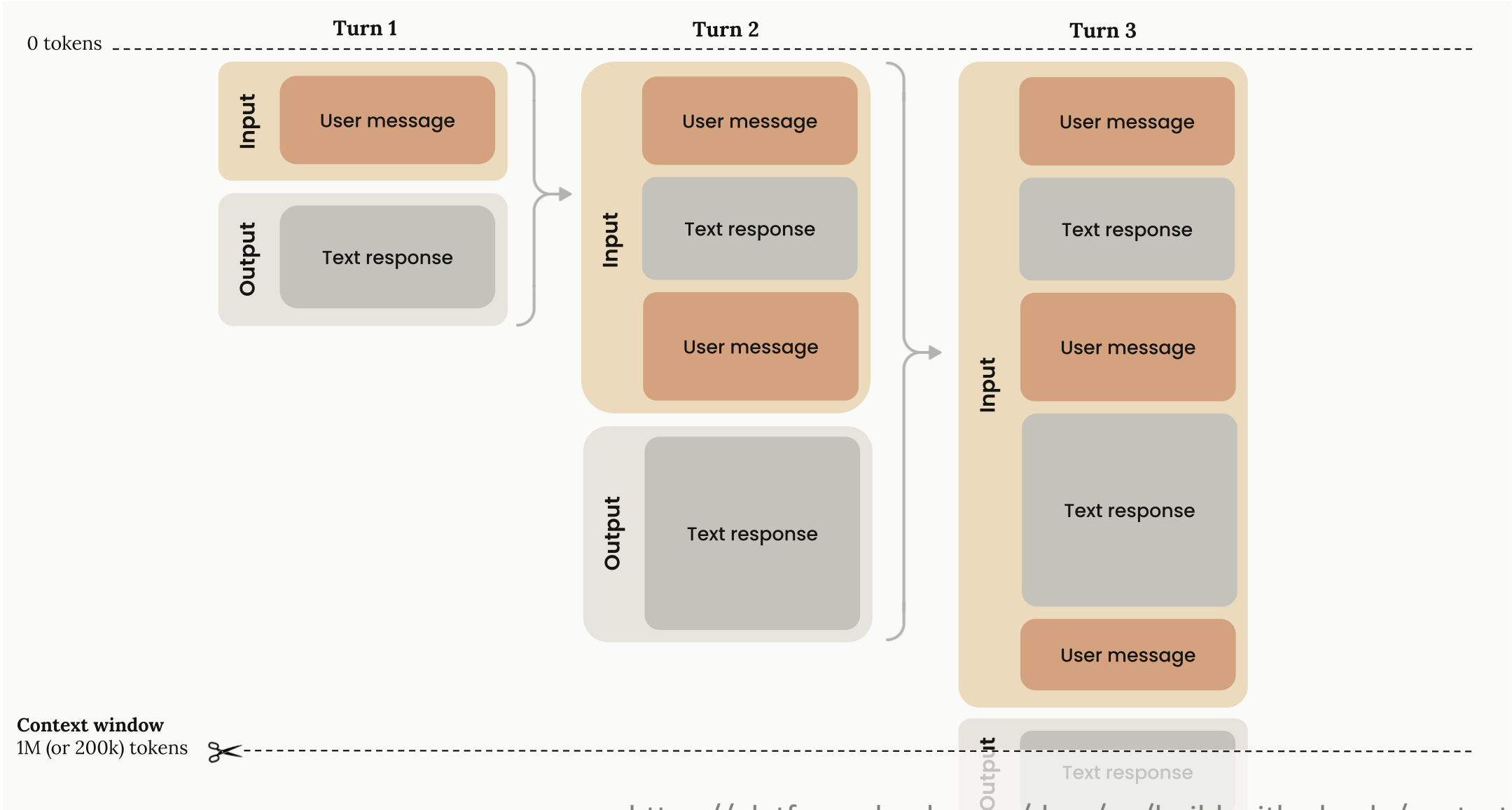
Fixing PPT Design for a Hub & Spoke Visual

Claude: Provided the idea, got it to make it, then provided a screenshot of content and it filled it in

Turning Survey Results into Slides

Claude Code: Provided the excel file, an example of prior slides.

Context



https://platform.claude.com/docs/en/build-with-claude/context-windows

Context Rot and Drift

Models re-read the entire conversation with every prompt

Practical tip: Start new chats often — especially when topics shift. You lose some context but gain better attention.

Context Windows

1 million

Free: 16k
Plus: 32-256k

1 million

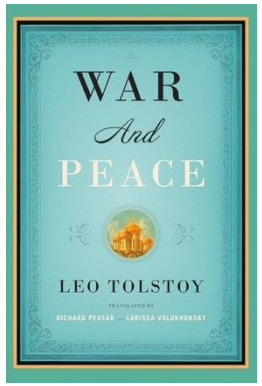
Claude

ChatGPT

Gemini

1 million tokens = 750k words = 1500 pages

Roughly 1.3x Tolstoy's War and Peace



Training Data Cutoffs

Your AI doesn't know what it hasn't read

General chatbots:

ChatGPT / Claude / Gemini all have optional web search — but you might need to ask it to do so

Enterprise Copilot Chat:

No web search (intentional — keeps PHI contained within the enterprise environment)

If your AI can't look for the latest information, all it knows is what is in its training data!

Hallucinations: Confidently Wrong

Definition: Model generates plausible-sounding but fabricated content

Good news: Much better than 2022 — "thinking" and reasoning models significantly reduce hallucination rates

⚠ **Bad news: Hallucination rate is not (and likely never will be) 0**

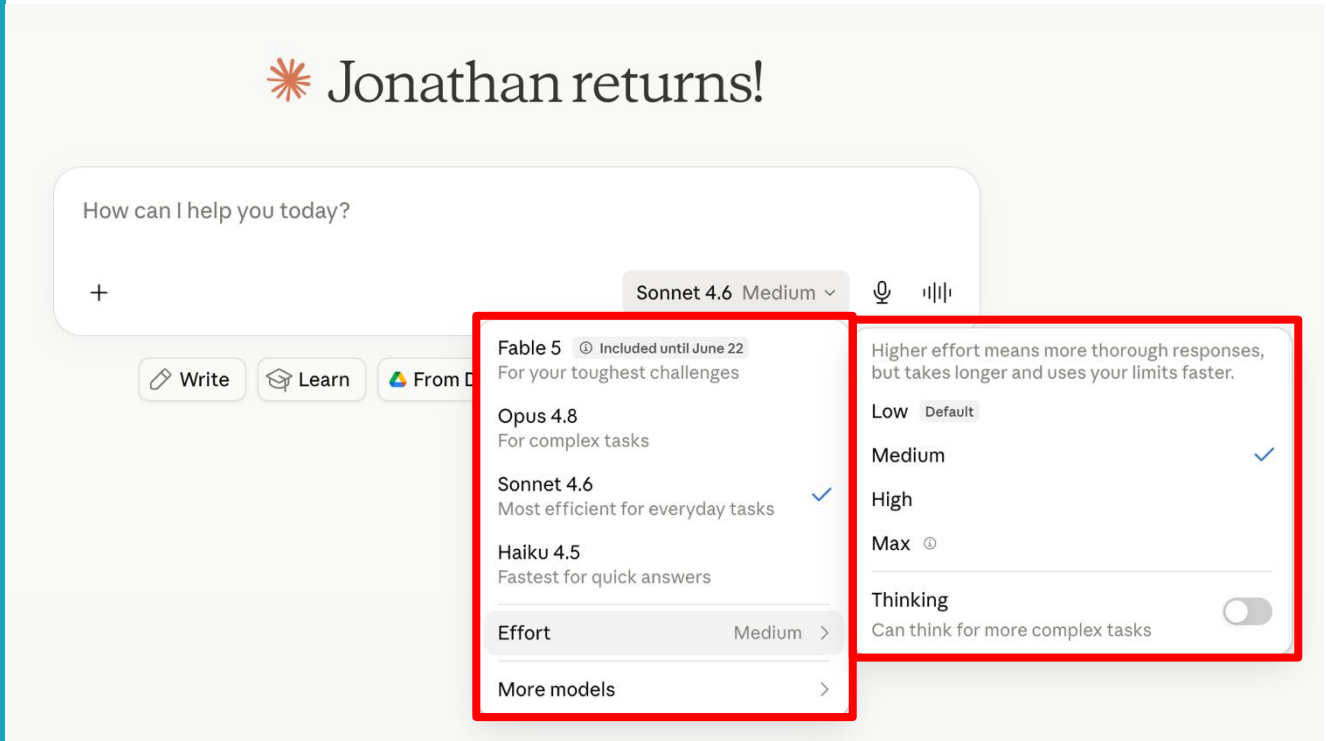
Mitigation: You must read output before using in any capacity. If you're using AI for information, ask for citations — then actually follow the links.

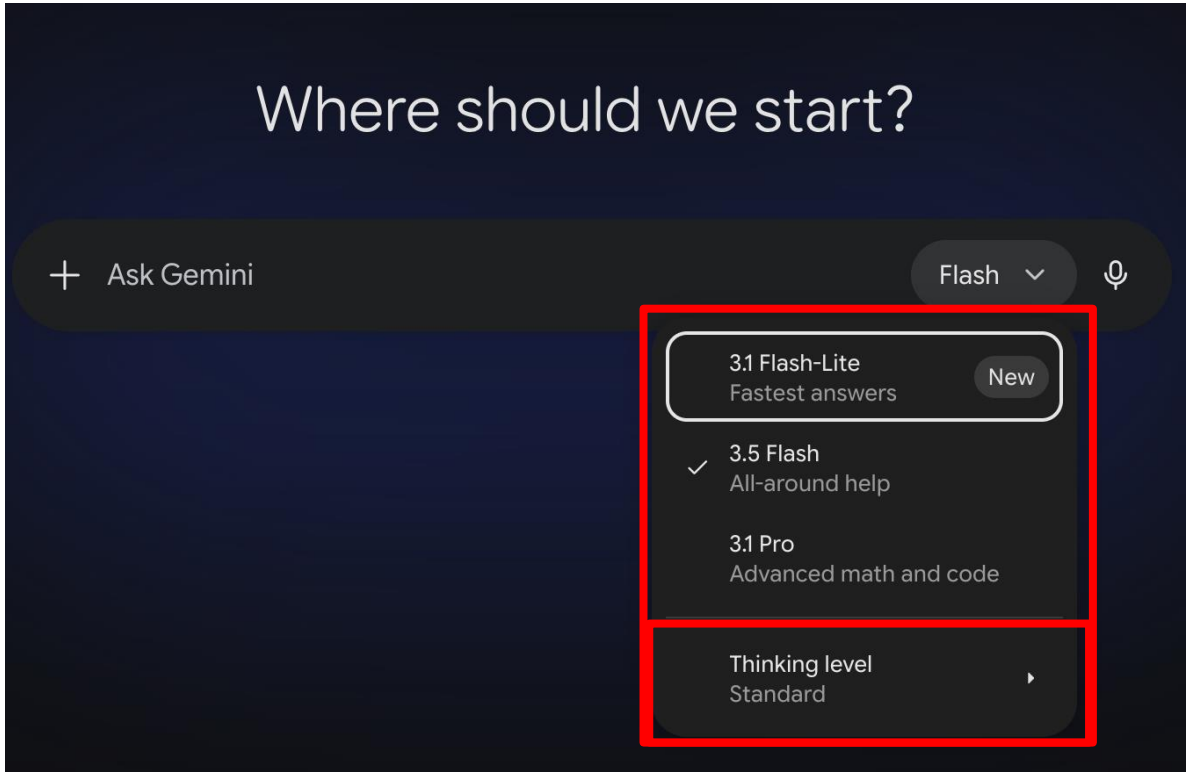
AI Biases Mimic Human Biases

Training Data	Annotation Bias	Evaluation Bias
Underrepresentation of certain populations in the medical literature and trial data can lead to incorrect or less accurate results in certain subgroups	Human-provided labels for supervised learning are subjective; human bias can be propagated to machines	Metrics used to judge model performance may not identify failures in certain subgroups

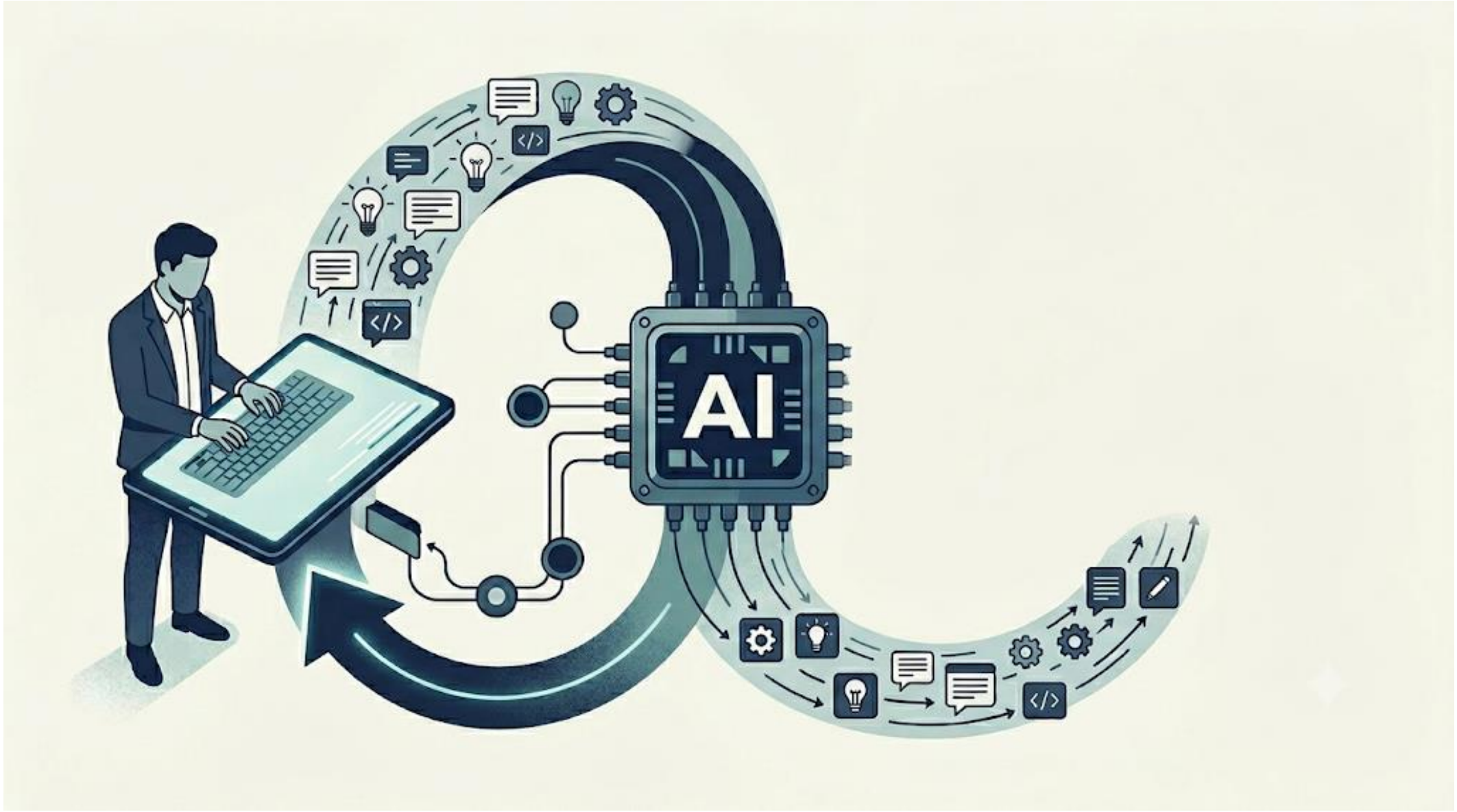
Generic Responses

Intentional Model Choice





Iteration



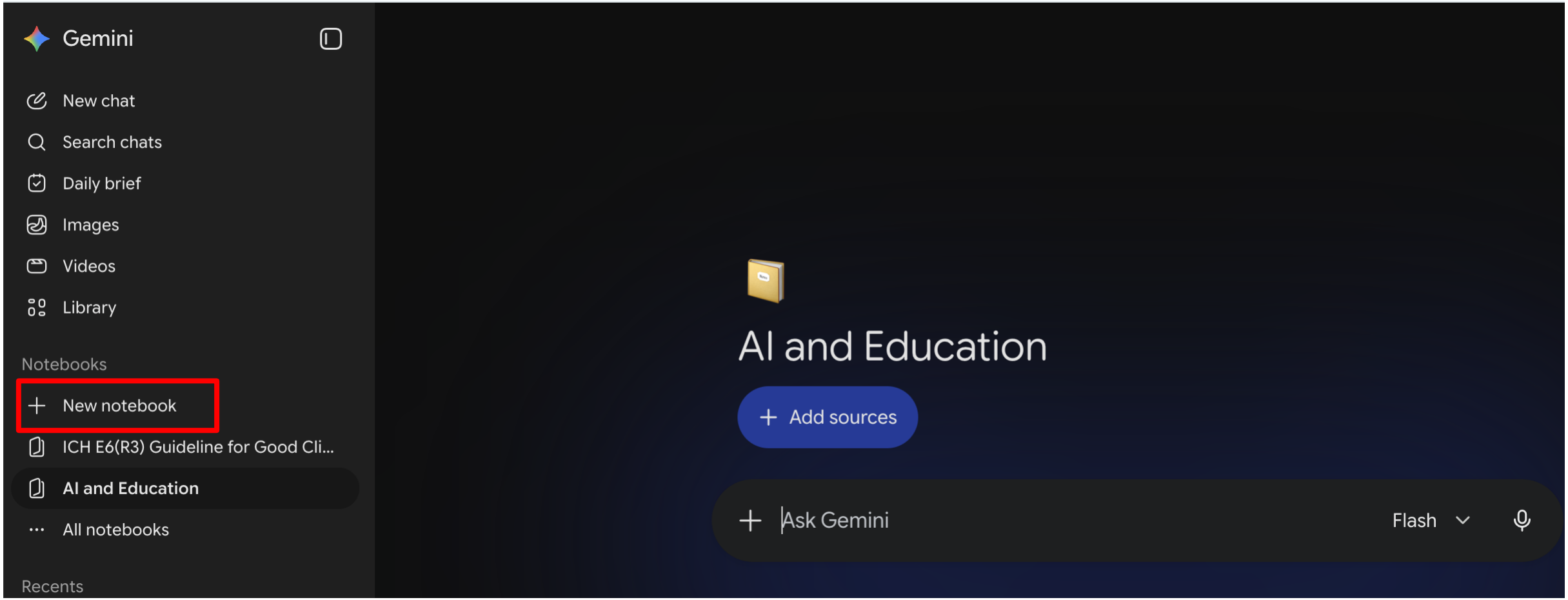
The “Jagged Frontier”

Quality Output Requires Anticipating and Articulating What Success Looks Like

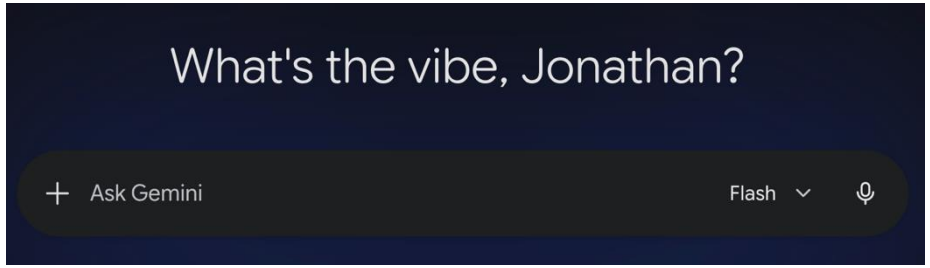
Create an image that shows the concept of the "Jagged frontier"

Editorial-style conceptual illustration, muted navy-and-teal palette on off-white. The "jagged frontier", an austere cliff face with angular jagged areas - some jutting out much further than others. Clean flat vector shapes, subtle grain texture, no text in the image. 16:9.

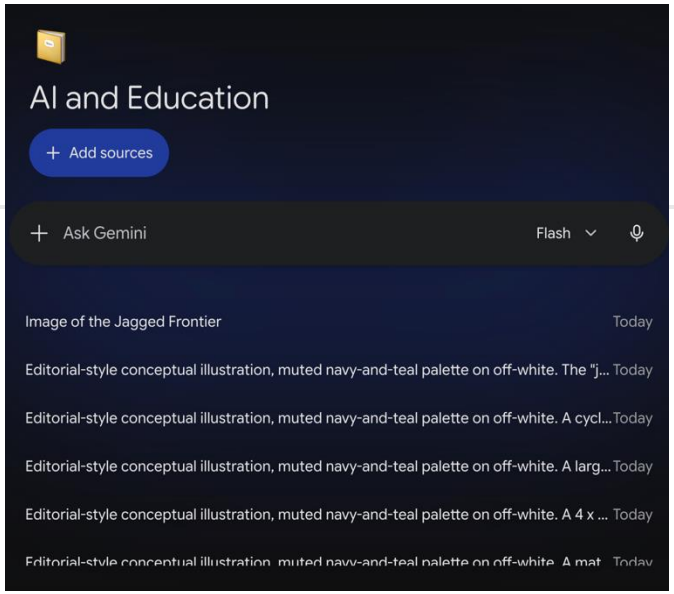
Projects/Notebooks/Folders



Projects/Notebooks/Folders



Create an image that shows the concept of the "Jagged frontier"



Create an image that shows the concept of the "Jagged frontier"

Going to the Balcony

How This Talk Was Built

Claude — scaffolded an outline from our prior planning conversation

Jonathan — wrote framing questions, learning objectives, and the talk's arc; dropped in bibliography sources and pre-survey input along the way

Claude — organized everything into blocks, structure, and proposed slides for each block

Jonathan — reviewed and edited the outline

Claude — drafted a first-draft PowerPoint from the outline

Jonathan — generated images, pulled in slides from other decks, and fleshed out / edited slides

Claude — built this slide

Two Questions Anchor Today

How is AI changing how we learn — and subsequently how we teach?

How can we use AI to augment our capacities as educators?

Learning Objectives

- Explain the concepts of **cognitive load**, **cognitive offloading**, and **deskilling**, and how they relate to GenAI use in the context of medical training.
- Sort various tasks performed by a medical trainee into **core** v. **secondary** skills, and use this framework to guide **when, where and how** to deploy AI.
- Demonstrate GenAI use patterns that **support** learning rather than hinder it.

- Identify where GenAI tools can **accelerate** educational session development v. where they are currently not helpful due to **error rate, amount of manual correction required, or misleading results**.
- Identify **prompting strategies** and **tool choices** that improve AI output for clinical and educational tasks.

Choosing the Right Tool

Task	Best tool(s)	Note
Clinical reasoning / differential	Claude, Gemini, ChatGPT	Avoid framing, ensure using extended thinking/pro models for complex cases
Evidence-based clinical questions	Open Evidence	Pulls data from sources and cites them; more reliable when the answer exists out there
Patient messages, letters, education	Claude, Gemini, ChatGPT	Claude often thought to have the most natural prose. OE can do simple tasks now
MCQ / teaching case generation	Claude, ChatGPT, Gemini *OE improving here	Better able to leverage multiple inputs, use tools, reason through hard questions
Slide generation (image-heavy)	Gemini, NotebookLM	Gemini makes great visuals. NotebookLM turns sources into great slides
Slide generation (content-heavy)	Claude	Very good at helping make outlines or turning outlines into draft slides. Can't do images
Data cleaning, file conversion, website/app design	OpenAI Codex, Claude Code/Cowork	Agents shine at processing and manipulating files

With Whom Am I In Conversation?

One Useful Thing

Claude Code for Everything

The Convivial Society

Teaching in the Age of AI

Jonathan Berry, MD
jlberry@bidmc.harvard.edu

Landing Page
tinyurl.com/jez53m24