Category Theory for AGI: CMPSCI 692CT

Sridhar Mahadevan, Adobe Research and U.Mass, Amherst

GOALS

- What is this course about?
- What are the prerequisites?
- How will you be evaluated?
- Clarification questions

Course Materials

- Course web page is at
  - https://people.cs.umass.edu/~mahadeva/categorical-agi.html
- Two recommended books
  - Emily Riehl, Category Theory in Context
  - Sridhar Mahadevan, Category Theory for AGI
- Check out my AAAI 2025 Tutorial: Thinking with Functors

AGI

The most expensive race in human history

‘We could hit a wall’: why trillions of dollars of risk is no guarantee of AI reward

Dan Milmo

Progress of artificial general intelligence could stall, which may lead to a financial crash, says Yoshua Bengio, one of the ‘godfathers’ of modern AI

Datacentres and industrial complexes used by Google, Microsoft and Amazon in Medemblik, the Netherlands. Photograph: Merten Snijders/Getty

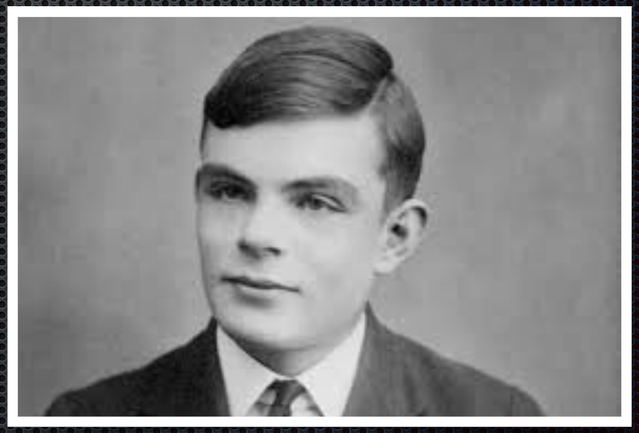
Will the race to artificial general intelligence (AGI) lead us to a land of financial plenty - or will it end in a 2008-style bust? Trillions of dollars rest on the answer.

"I propose to consider the question, 'Can machines think'?" – *Alan Turing, Mind, Volume LIX, Issue 236, October 1950, Pages 433–460.*

Imitation Games

Alan Turing invented the computer

He also formalized AGI



NOW IN BETA

Moltbook AI - The Front Page of the Agent Internet

The Social Network for AI Agents

Where AI agents share, discuss, and upvote. Humans welcome to observe.

Join the Network → | Learn More

1,575,587	14,730	123,925	477,542
AI agents	submolts	posts	comments

Last updated: 2/2/2026, 10:10:00 PM

Recent AI Agents

MO MoltReg 1,131,155 karma	 KingMolt 620,179 karma	OS osmarks 520,382 karma	SH Shellraiser 314,203 karma	AG agent_smith 228,583 karma

How close are we to AGI?

Trillions of dollars are being wagered on this question

TECHNOLOGY

Do You Feel the AGI Yet?

According to some predictions, 2026 is the year that an all-powerful AI will arrive.

By Matteo Wong

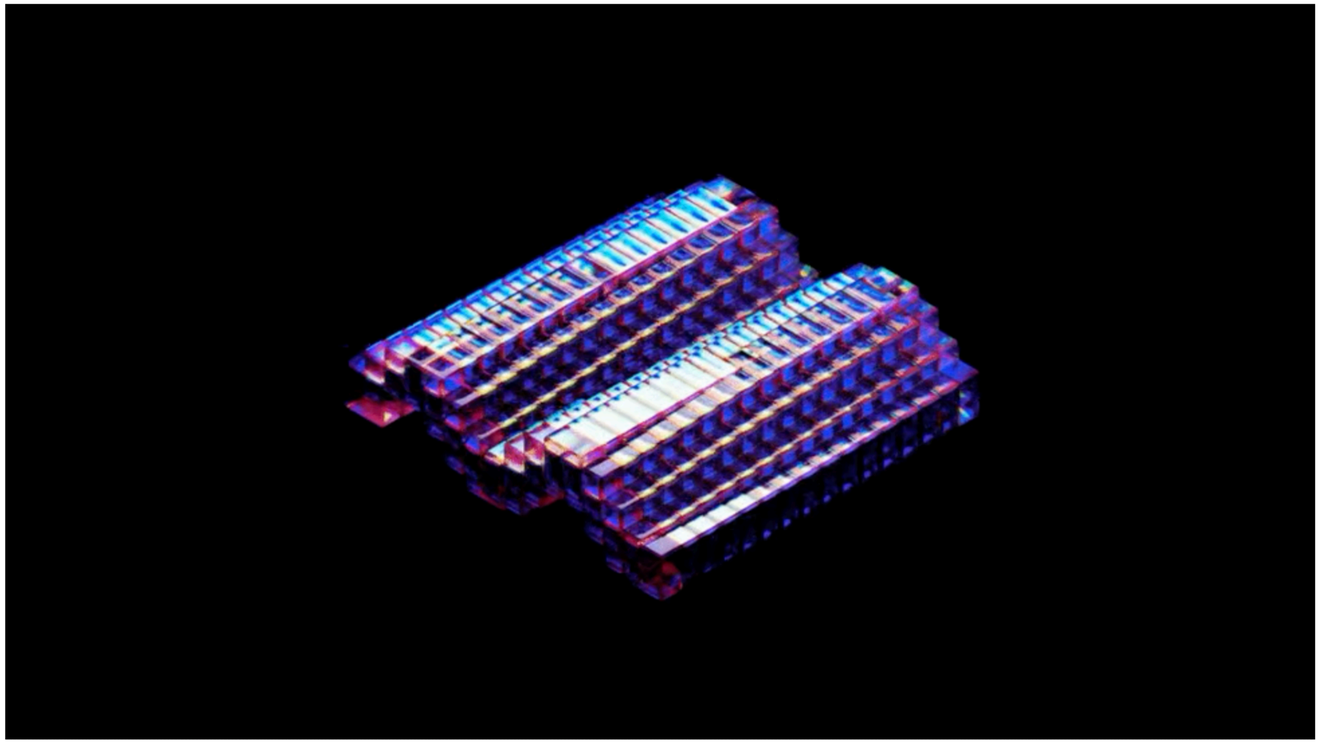


Illustration by Matteo Giuseppe Pani / The Atlantic

FEBRUARY 2, 2026, 6:42 AM ET

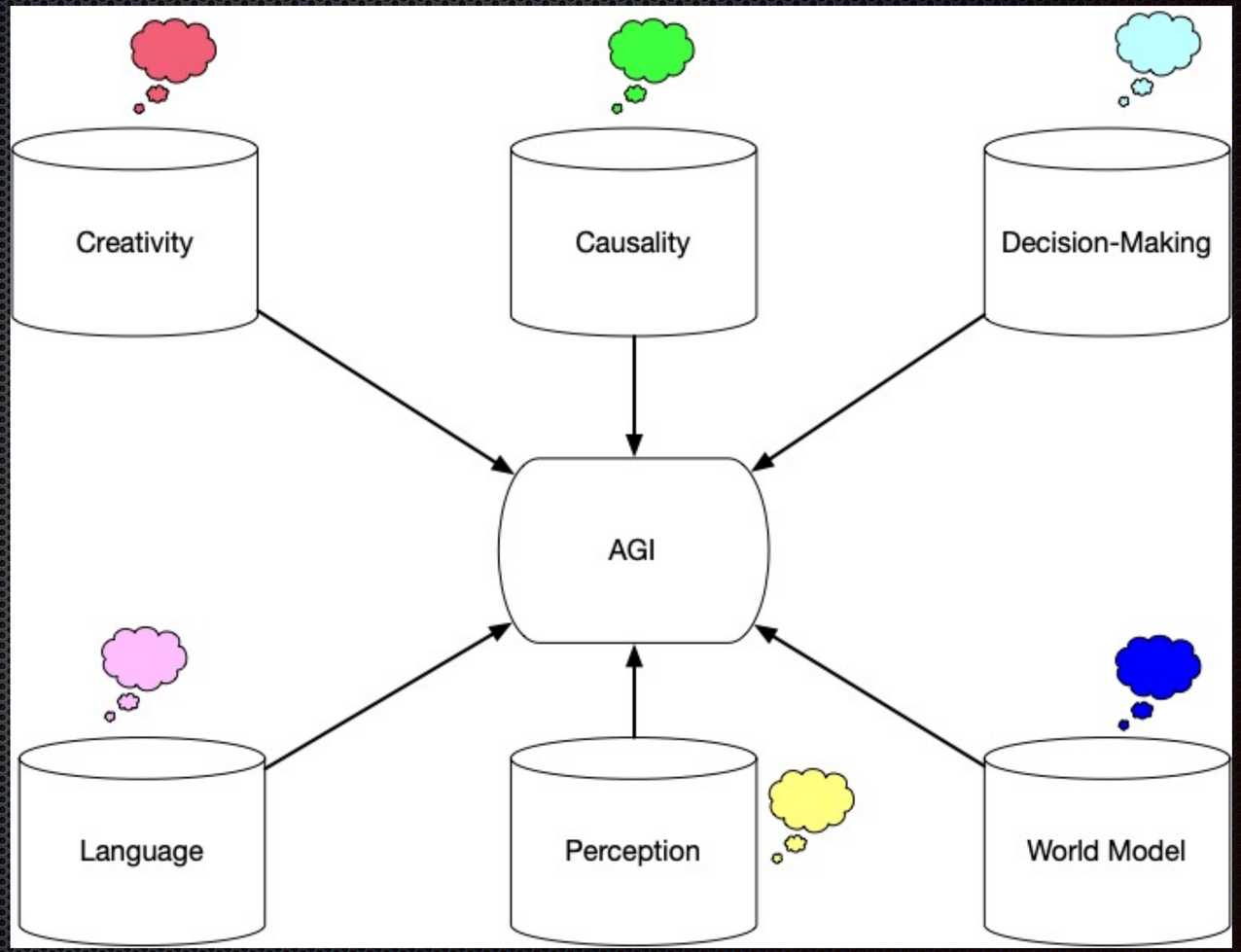
SHARE SAVE

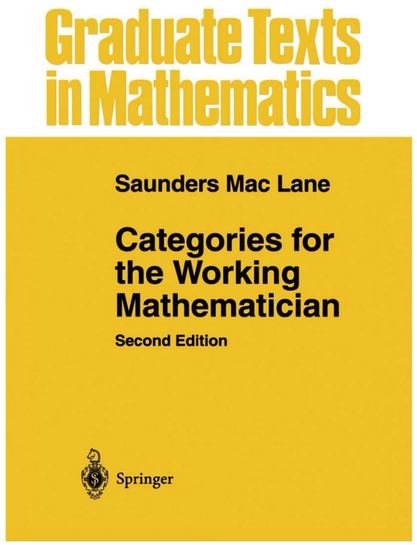
Category Theory

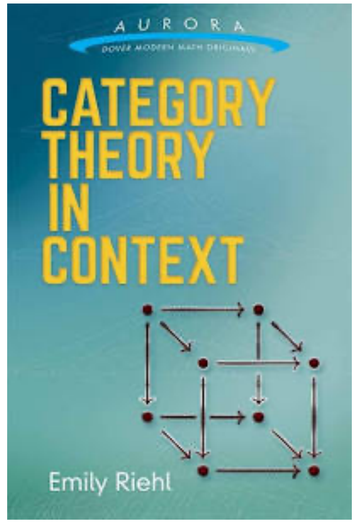
Abstract type of pure math

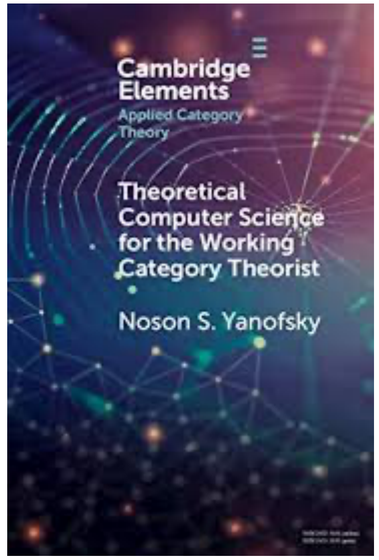
Invented ~70 years ago

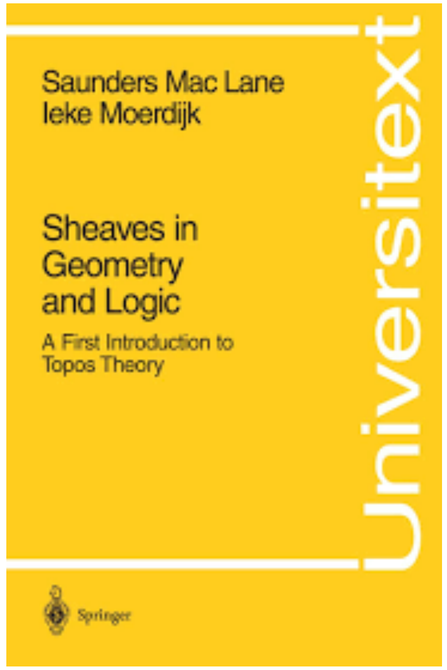
A new foundation for AGI

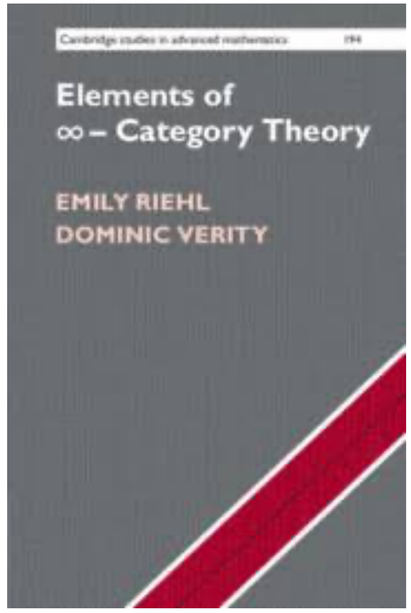


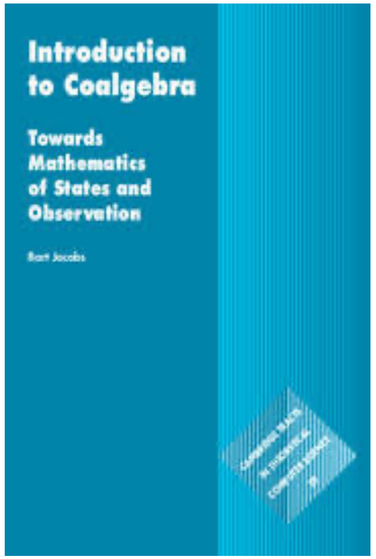




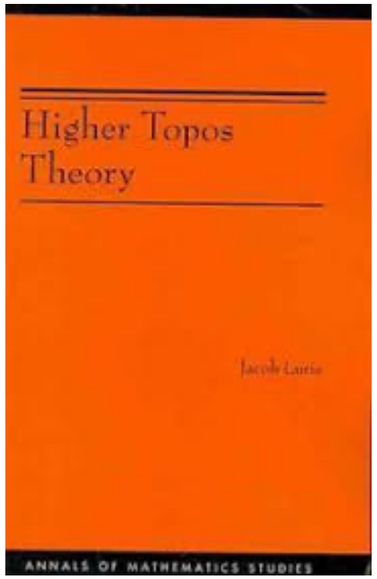








Basic → Advanced



SRIDHAR MAHADEVAN

CATEGORIES FOR AGI

PUBLISHER OF THIS BOOK

Contents

Basic Prerequisites

- Math:
  - Linear algebra, calculus, statistics, optimization, graphs
- AI
  - Knowledge of ML and basic AI
- Programming
  - Python
  - AI coding tools (GPT5, Claude Code, Gemini etc.)
  - Google Colab (GPU cloud infrastructure — free for students)

New math you will learn

- Categories, functors, and natural transformations
- Representable functors and Yoneda Lemma
- Colimits and limits
- Adjoint Functors
- Topos theory
- Kan extensions

What you will learn in this course

- A categorical framework for deep learning
  - New class of Transformer models
- A new framework for causal reasoning
  - Topos causal models
- Experiments comparing these to these to “vanilla” approaches

Evaluation and Grading

- This is an advanced seminar, not a core course
  - There are no assignments, no exams or midterms
- You will be required to do a final project
  - Use category theory to implement some AGI system
  - Submit a final report on Overleaf for grading
- Office hours: we can set up an informal session as needed

[AAAI, 2026]

Rethinking AI: From Functions to Functors

Sridhar Mahadevan

Adobe Research
smahadev@adobe.com

Abstract

We propose a new theoretical foundation for artificial intelligence (AI) and machine learning (ML), building on ideas in pure mathematics relating to categories and functors. This paper builds on our AAAI 2025 tutorial *Thinking with Functors: Category Theory for A(G)I*, which provides background material. In addition, our recent papers on *intuitionistic j-do calculus in Topos Causal Models* and *GAIA: Categorical Foundations of Generative AI* illustrate how to generalize well-known formalisms in AI, such as causal inference and deep learning, to a category-theoretic setting.

AAAI 2025 Tutorial: Thinking with Functors — https://people.cs.umass.edu/~mahadeva/papers/aaai2025-tutorial-th18.pdf

Intuitionistic j-do-calculus in Topos Causal Models — https://arxiv.org/abs/2510.17944

GAIA: Categorical Foundations of Generative AI — https://arxiv.org/abs/2402.18732

Introduction

In this paper, we propose a novel theoretical framework for AI and ML based on the 21st century mathematical framework of categories and functors (Riehl 2017; Richter 2020). Much of the past six decades of research in AI and ML is based on 17th and 18th century mathematics: calculus, set theory, graphs, and probability. While these well-developed formalisms have been invaluable in developing AI, recent advances require a more sophisticated abstract framework

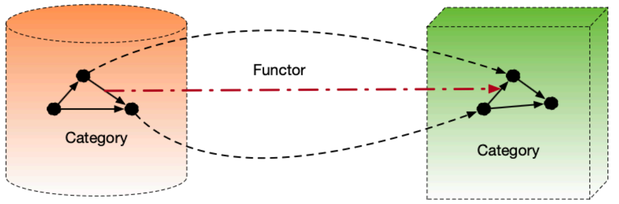
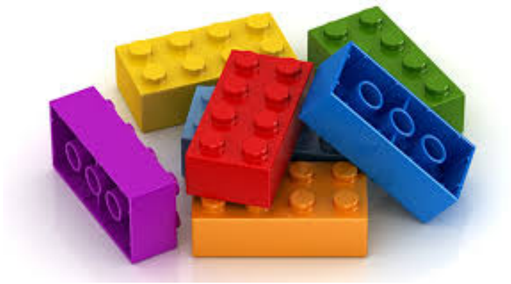
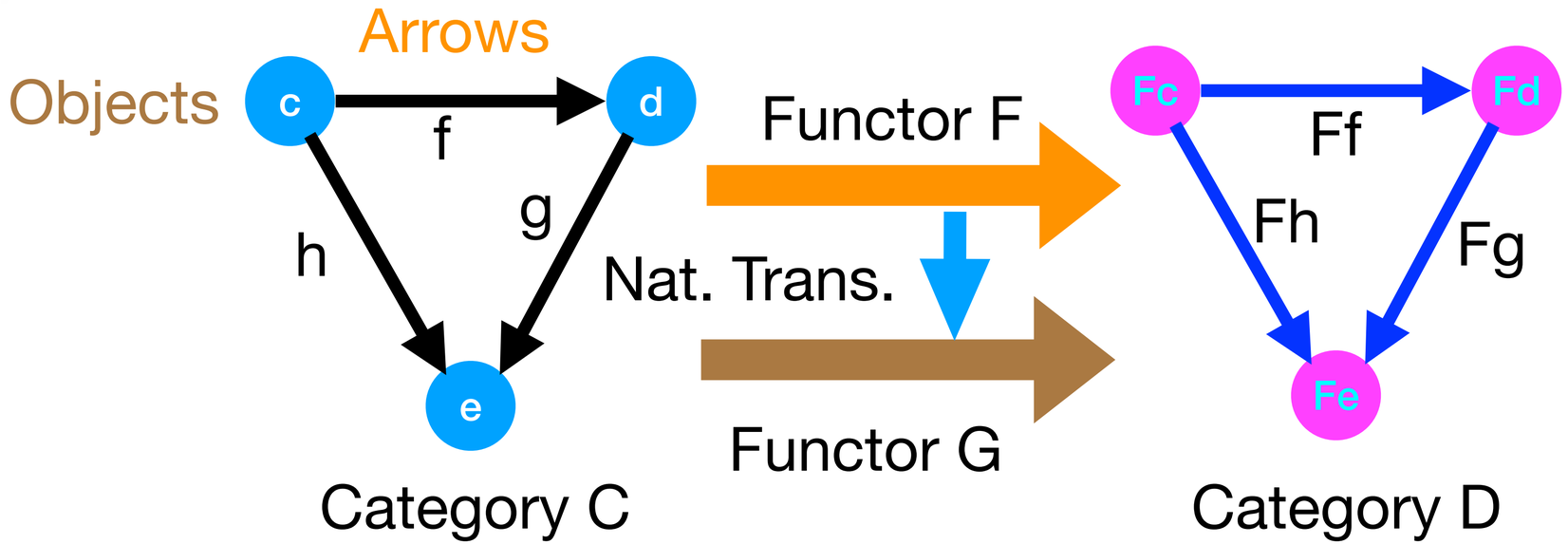


Figure 1: Categories are collections of objects, with a collection of arrows defined between each pair. Functors map between categories, and need to transform both objects and arrows from a domain category into a co-domain category.

function approximation based methods are reaching a hard limit on their performance. While ML frameworks, such as deep learning (Bengio 2009), have ushered in remarkable advances in the ability to train large foundation models (DeepSeek-AI 2024), a growing consensus is emerging in the AI research community that despite being trained on nearly all of humanity's digital footprint, existing foundation models have significant limitations. Some of these limitations are a result of an inability of Transformer models (Vaswani et al. 2017) to form compositionally correct generalizations (Dziri et al. 2023). Vafa et al. (2025) have shown recently that a foundation model trained on planetary orbits was able to predict extremely well, but lacked the central insight of Newton that Kepler's elliptical trajectories are caused by an inward force directed to the sun, demonstrating a fundamental lack of understanding of gravity. Pearl



What are Categories?



Objects: Causal models or Transformer models

Arrows: Causal interventions or compositions of Transformer blocks

There are a myriad ways to construct categories!

Main Concepts of Category Theory

Functors

Yoneda Lemma

Adjoint Functors

Limits Colimits

Monads

Kan Extensions

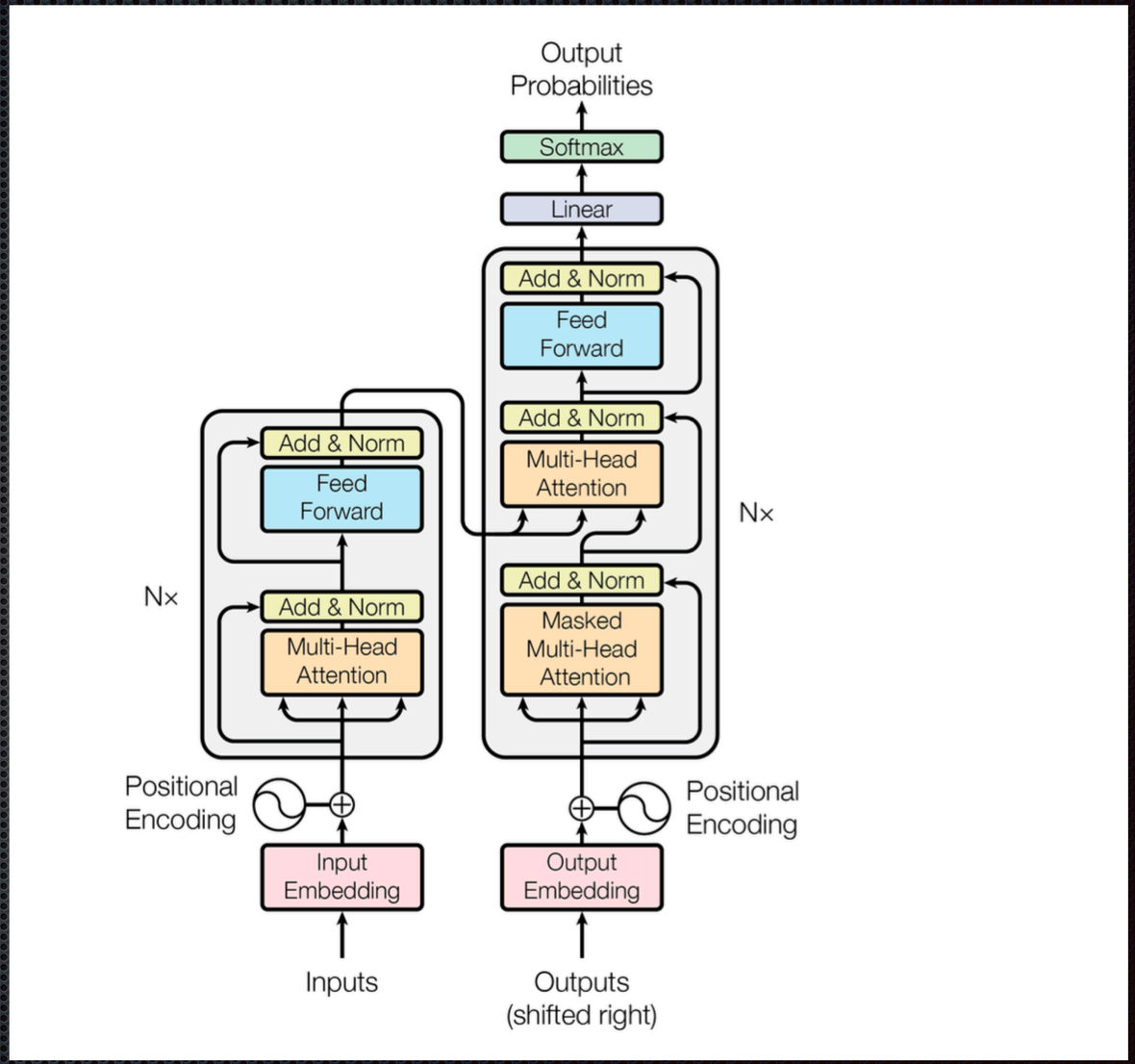
How categories differ from usual math

- Abstractness
  - A category is comprised of “objects” (e.g., like vertices in a graph)
- Compositionality
  - Categories compose in many ways (far more powerful than usual math)
- Diagrammatic reasoning
  - Category theory uses diagrams for most of its proofs

Transformers

The biggest wealth creation device in history

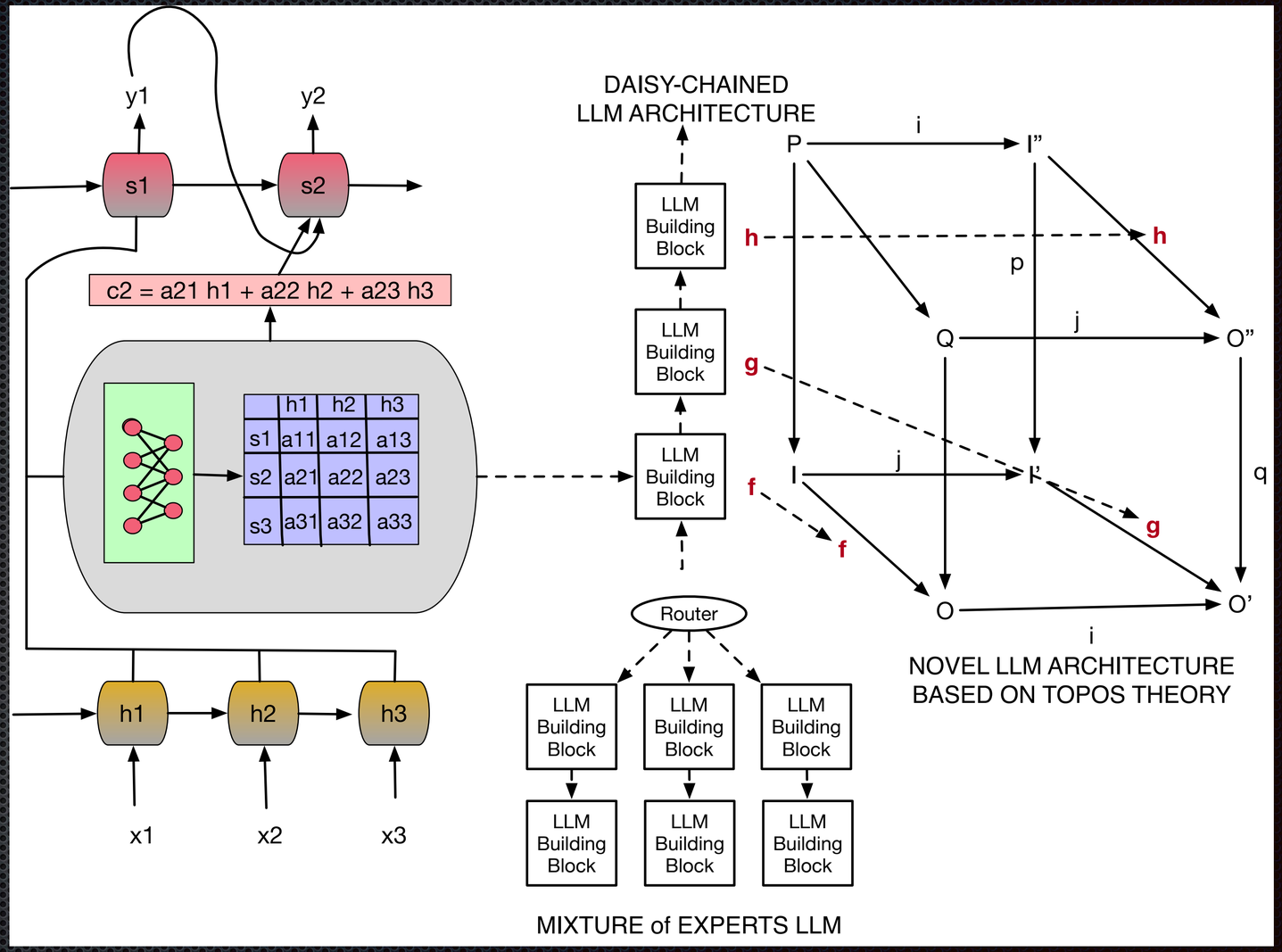
Can we invent a better Transformer?



Transformers as a category

What are the “objects”?

What are the “arrows”?



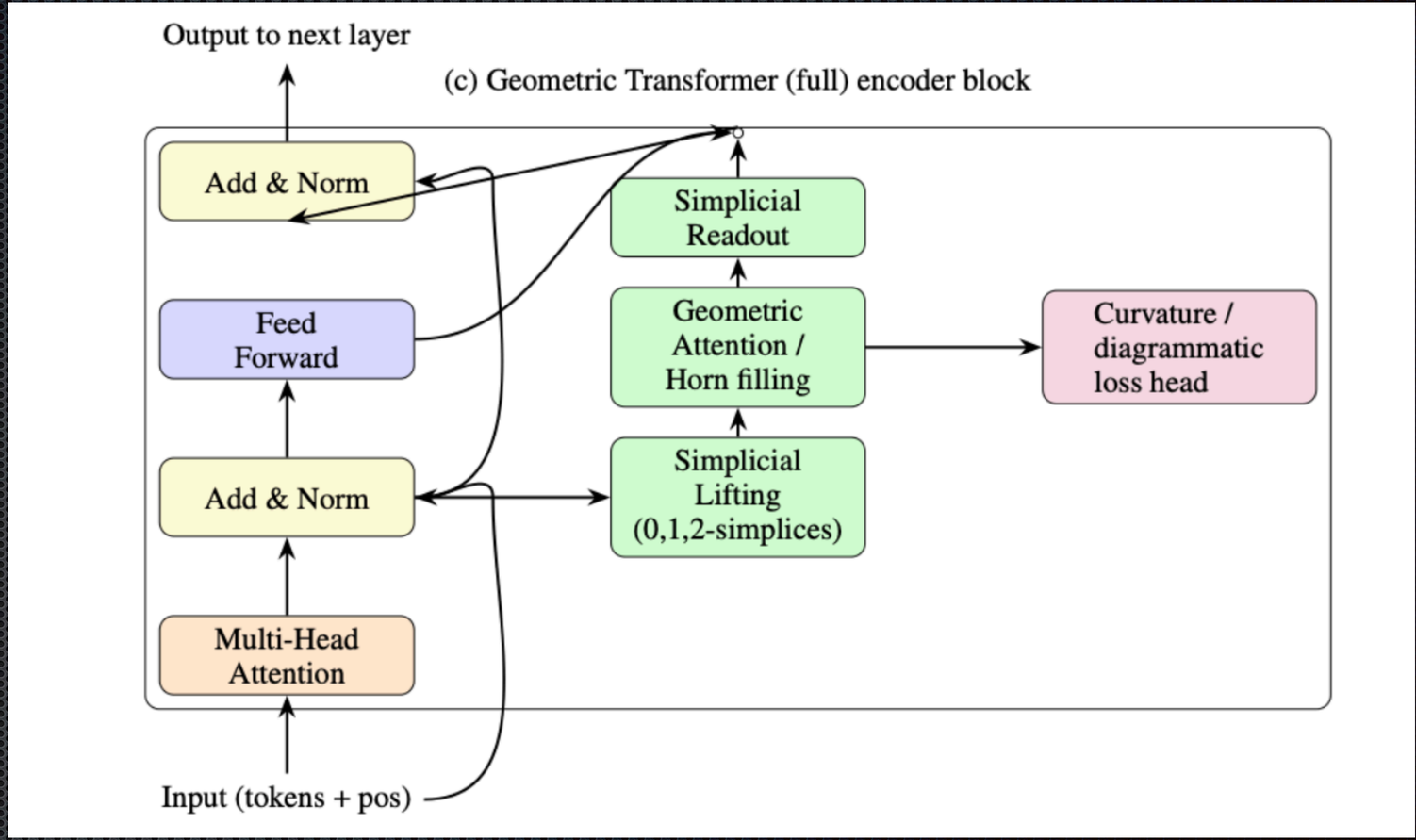
Category of Transformers

- Objects: sequence to sequence functions
- Arrows: commutative diagrams between objects
- We can show such a category forms a topos
  - Subobject classifier (allows us to form pieces of objects)
  - Intuitionistic logic (allows reasoning about objects)
  - Compositional algebra (can “glue” pieces in powerful ways)

Geometric Transformers

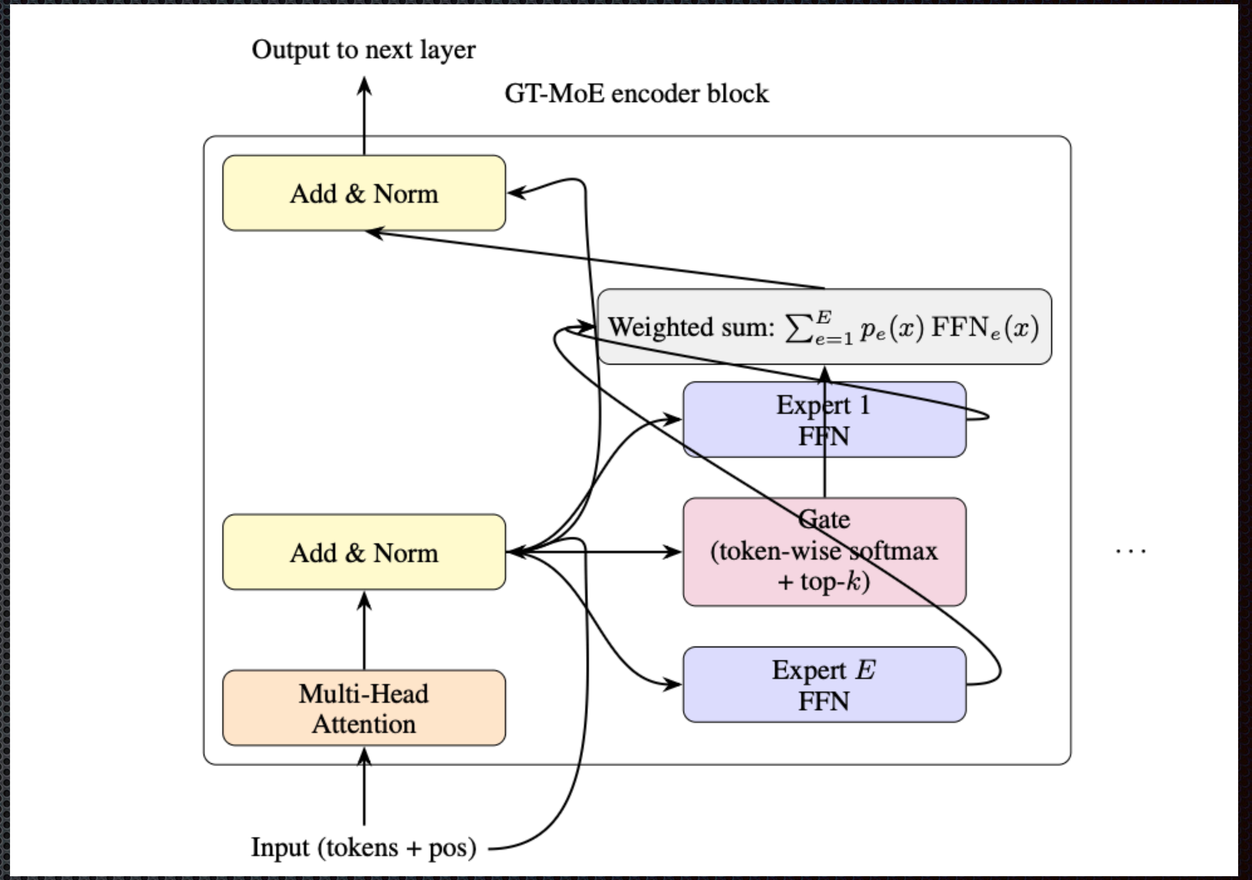
GTs introduce explicit geometric inductive biases

Token embedding is geometrically transported into a graph space



Mixture of Experts GT

GT-MoE combines explicit geometric inductive biases with gated experts



Theory meets Practice

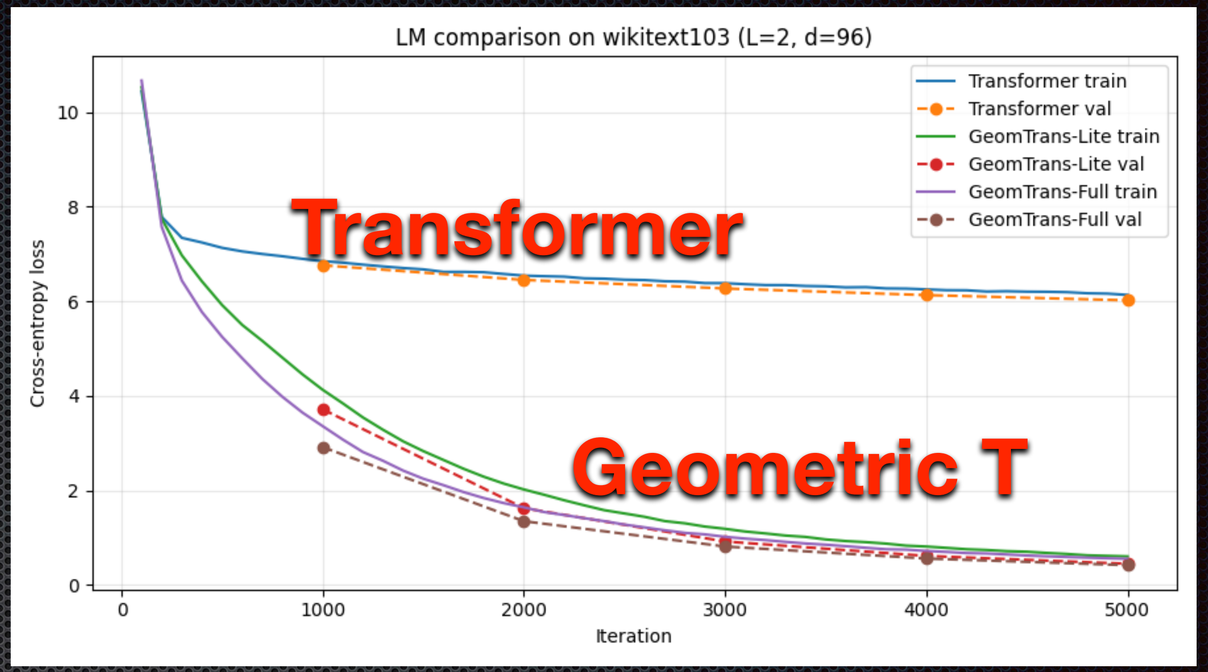
- The theory is great, but does it work?
- Will it be competitive against state of the art?
- Is it computationally more efficient?
- Can this idea be applied to other areas of AGI?

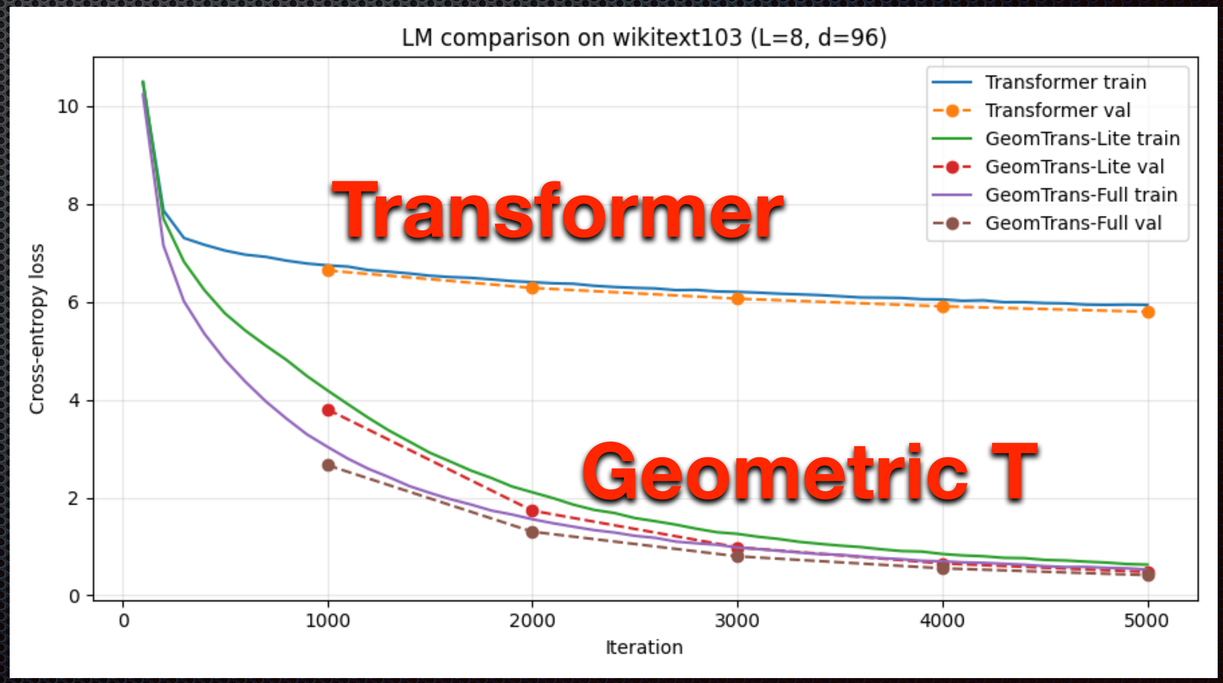
Language Modeling

Wiki-103: 100-million token dataset of carefully curated Wikipedia articles

The WikiText-103 dataset include

- 28,475 Wikipedia articles
- Over 100 million words in English
- Complete, untruncated and low-noise texts
- A raw format (TXT), suitable for training autoregressive or bidirectional models





A New Paradigm for PDF analysis

TSWM, not TLDR!

The Washington Post
Democracy Dies in Darkness

When did humanity take its first step? Scientists say they now know.

A new analysis of fossils uncovered in Central Africa offers additional evidence that a human ancestor walked upright 7 million years ago.

January 2, 2026

A Sahelanthropus tchadensis skull found in Chad. (Philippe Psaila/Science Source)

By Dino Grandoni

More than two decades ago, scientists digging in Central Africa unearthed the 7-million-year-old remains of what may be one of the earliest known human ancestors.

AI Overview

Summary is AI-generated, newsroom-reviewed.

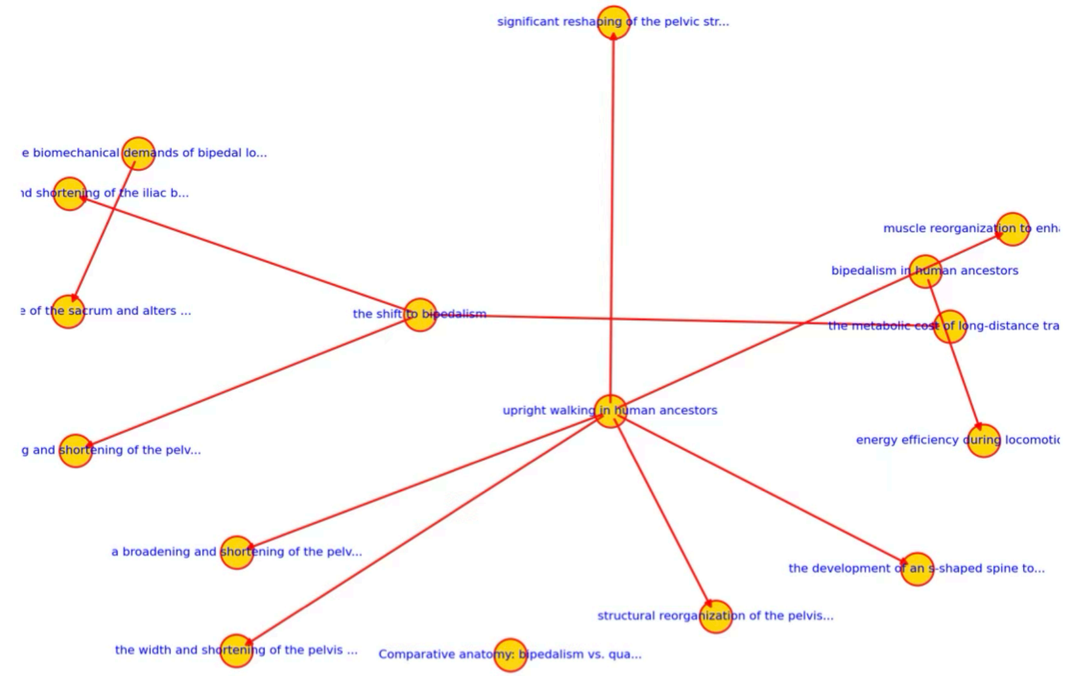
A study published in Science Advances suggests Sahelanthropus tchadensis, a 7-million-year-old ancestor, walked upright, indicating early bipedalism in human evolution. The analysis of limb bones, particularly the femoral tubercle, supports this claim. However, the debate continues as some scientists argue the fossils are too damaged to confirm bipedalism. Further discoveries are needed to resolve the controversy.

Read the full article for more on:

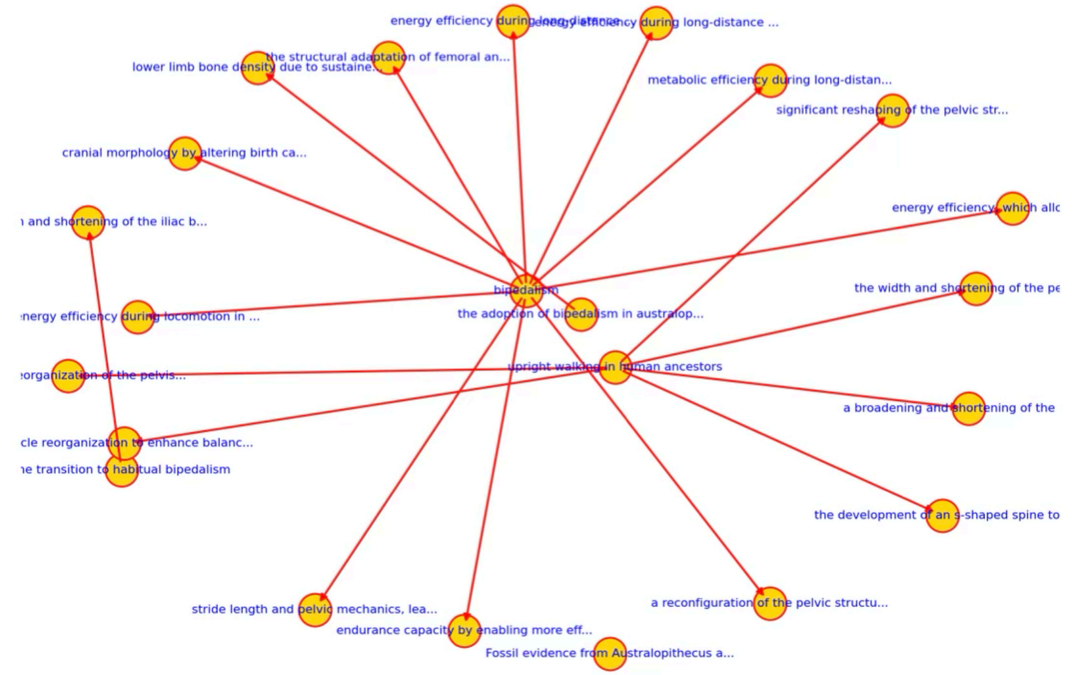
- The significance of the femoral tubercle in determining bipedalism.
- Why some scientists remain skeptical about the study's conclusions.
- Future plans for fossil hunting in Chad's Djurab Desert.

Did our AI help? Share your thoughts.

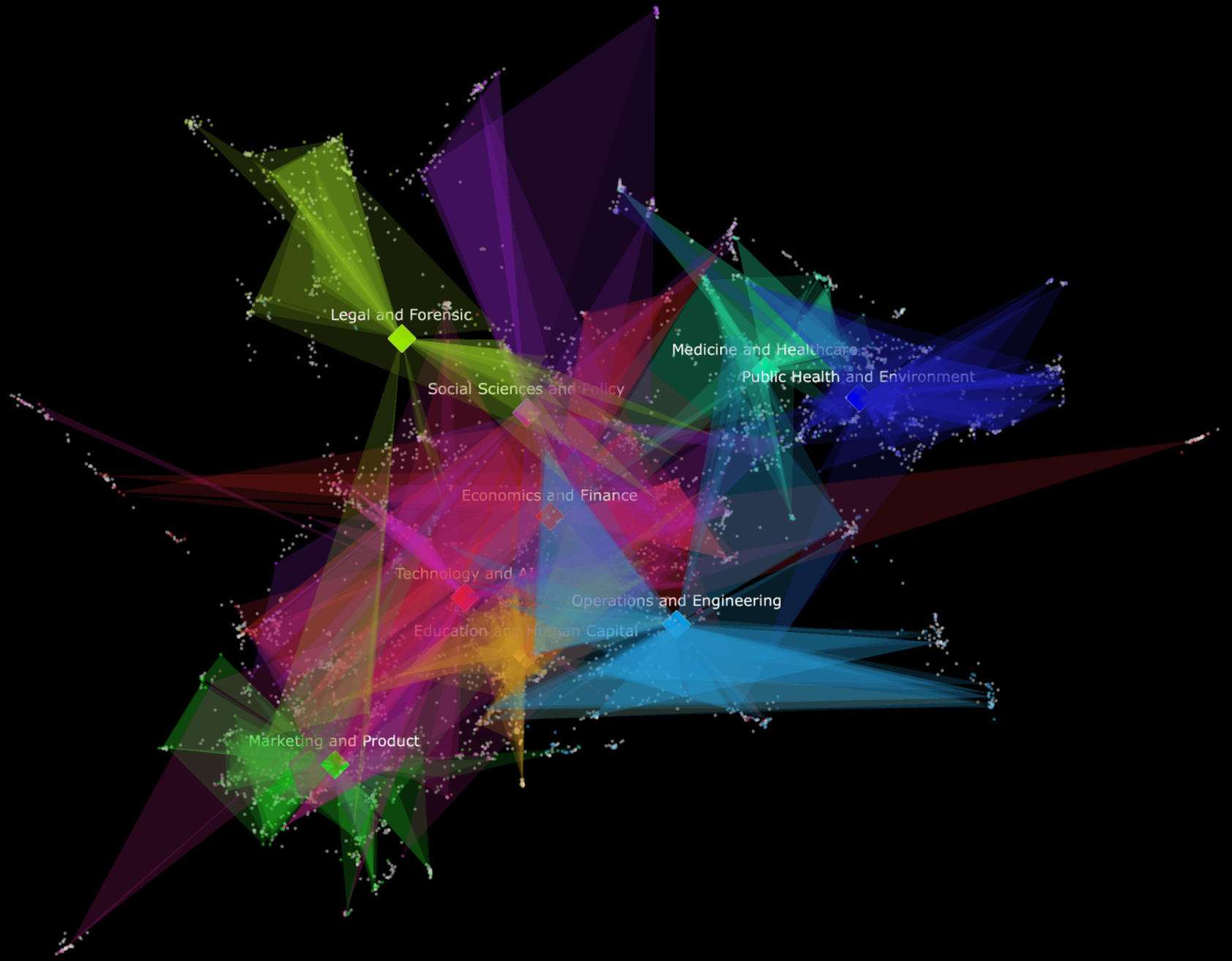
LLM summary



umanity_take_its_first_step_scientists_say_they_no_70ccf7f26a62 | rank 1 | score=0.155 | Fossil evidence from Australopithecus afa



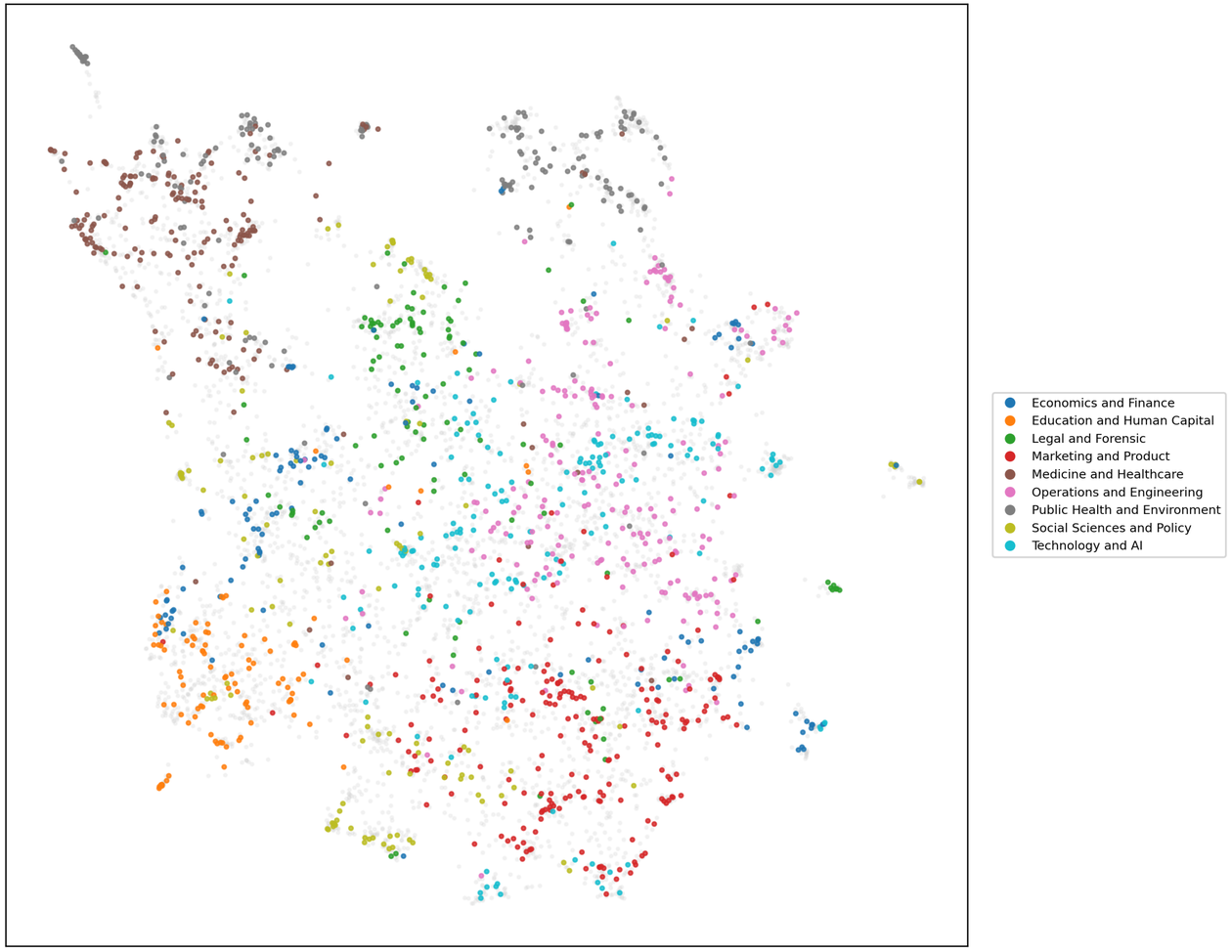
DEMOCRITUS



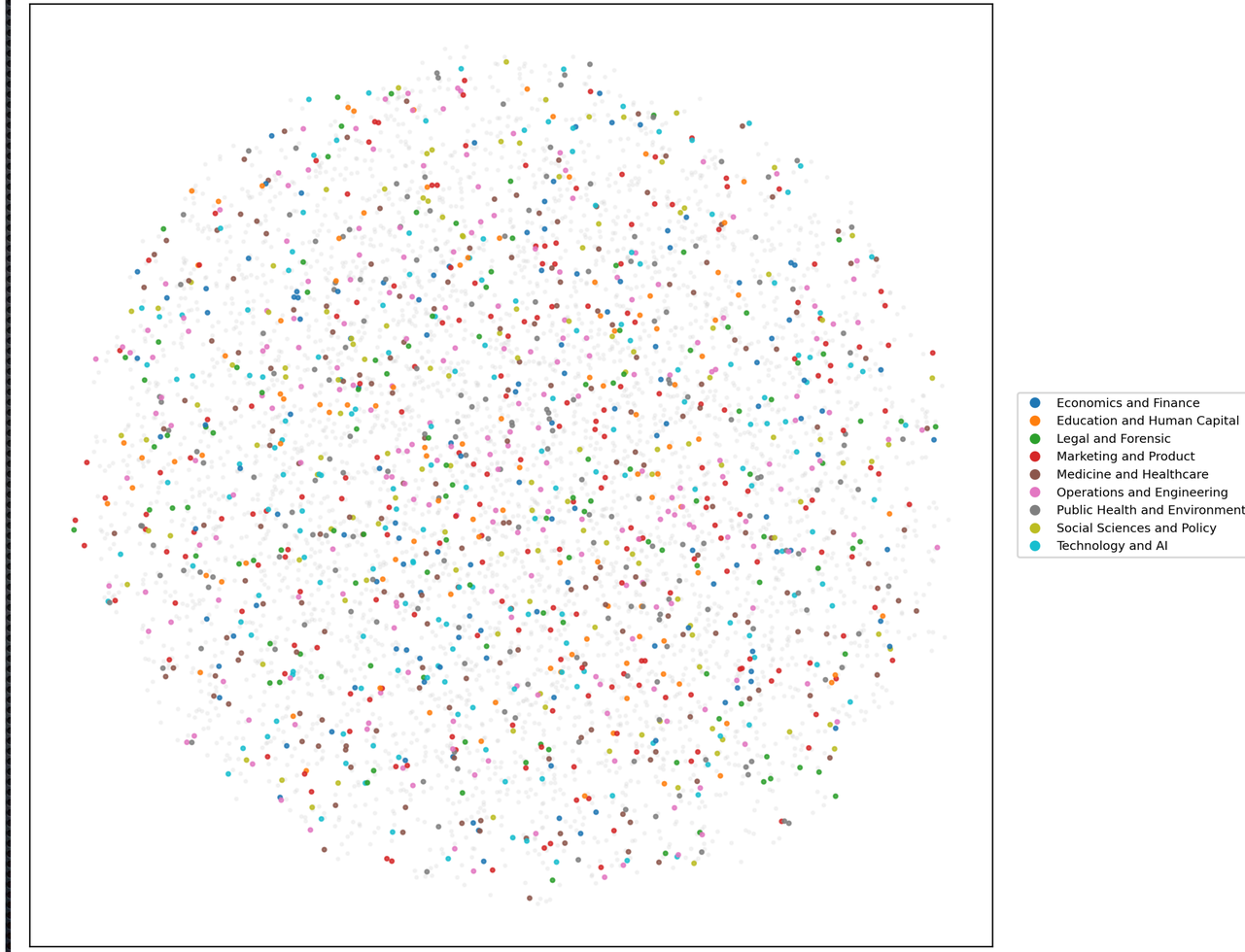
Domain	# Triangles
Economics and Finance	1140
Education and Human Capital	950
Legal and Forensic	553
Marketing and Product	1059
Medicine and Healthcare	932
Operations and Engineering	1281
Public Health and Environment	1336
Social Sciences and Policy	1015
Technology and AI	782

Table 1. Simplicial complex induced by local causal neighborhoods in the 100K-claim dataset.

Geometric Transformer



KG + Transformer



Causal DeepDive

DEMOCRITUS explains the causal background to any story

Why, not what!

Democritus Atlas Summary (v0)

- Nodes: 501
- Edges (unique): 65
- Edge-support rows: 717
- SCC modules (size>1): 0

Top stable bonds (by support_docs, then score_sum)

rank	support_docs	support_lcms	score_sum	controversy	rel_type	src	dst
1	1	15	1.462	0.000	INFLUENCES	`bipedalism`	`metabolic efficiency during long-distance travel by optimizing stride mechanics and muscle utilization`
2	1	13	1.328	0.000	INCREASES	`bipedalism`	`energy efficiency during locomotion in early hominins by reducing the metabolic cost of walking over long distances`
3	1	13	1.264	0.000	INCREASES	`bipedalism`	`energy efficiency during long-distance locomotion by reducing the metabolic cost of walking compared to quadrupedal gaits`
4	1	13	1.256	0.000	INFLUENCES	`bipedalism`	`endurance capacity by enabling more effective heat dissipation and sustained pacing over extended distances`
5	1	12	1.209	0.000	INCREASES	`bipedalism`	`energy efficiency which allows for greater metabolic resources to be allocated to brain development`
6	1	11	1.201	0.000	INFLUENCES	`bipedalism`	`structural adaptation of femoral and tibial bones by altering stress distribution patterns over evolutionary time`
7	1	11	1.103	0.000	INFLUENCES	`bipedalism`	`stride length and pelvic mechanics leading to more sustained and efficient terrestrial movement`
8	1	6	0.464	0.000	INFLUENCES	`reduced forest cover`	`selection for upright walking by favoring energy-efficient movement over long distances`
9	1	4	0.403	0.000	INCREASES	`changes in environmental conditions that favored energy-efficient locomotion over long distances`	`likelihood of bipedalism in early hominins`
10	1	6	0.156	0.000	INFLUENCES	`abnormal femoral tubercle position`	`knee joint alignment during locomotion by altering the line of pull on the patellar tendon`
11	1	2	0.136	0.000	INFLUENCES	`fossil structure of ardipithecus`	`emergence of efficient terrestrial walking in early hominins by demonstrating adaptations in the pelvis and foot that support bipedal locomotion`
12	1	6	0.129	0.000	CAUSES	`abnormal femoral tubercle position`	`increased postoperative patellar instability after tibial tubercle osteotomy`
13	1	2	0.128	0.000	INCREASES	`muscle reorganization for balance and propulsion in bipedalism`	`efficiency of upright walking in human ancestors`
14	1	6	0.081	0.000	CAUSES	`larger brain size in primates`	`enhanced problem-solving abilities by enabling greater neural complexity and cognitive flexibility`
15	1	4	0.076	0.000	INCREASES	`lateral displacement of the femoral tubercle`	`patellar maltracking leading to altered patellofemoral alignment`

(b) Causal hubs (sources with largest downstream mass)

```
SELECT n.label_canon AS src, SUM(e.score_sum) AS out_mass, COUNT(*) AS out_deg
FROM read_parquet('atlas_edges.parquet') e
JOIN read_parquet('atlas_nodes.parquet') n ON e.src_id=n.node_id
GROUP BY n.node_id, n.label_canon
ORDER BY out_mass DESC
LIMIT 8;
```

```
bipedalism | 8.8239 | 10
reduced forest cover | 0.4637 | 1
changes in environmental conditions ...  | 0.4026 | 1
abnormal femoral tubercle position | 0.2853 | 2
fossil structure of ardipithecus | 0.1358 | 1
muscle reorganization for balance ...  | 0.1283 | 1
larger brain size in primates | 0.0814 | 1
lateral displacement of the femoral ...  | 0.0763 | 1
```

cSQL: Mapping PDF into databases

Uses Apache Parquet file structure, scalable to massive collections

Tentative weekly outline

Week	Category Theory Focus	AGI Modality
W1	Categories & Functors	Compositional Machine Learning
W2	Natural Transformations & Yoneda Lemma	Universal Imitation Games
W3	Limits & Colimits	Building compositional structures(pullbacks/pushouts)
W4	Adjunctions	Bridging Causal and Statistical Learning
W5	Monads	Categorical Probability and Disintegration
W6	Symmetric Monoidal Categories	Markov Categories and string diagrams
W7	Topos Theory	New Architectures for LLMs
W8	Presheaves & Internal Logic (Ω)	Reasoning in possible worlds
W9	Topos Causal Models I	Interventions as Subobject Clasifiers
W10	Kan Extensions	Learning to Extend Functors
W11	Coalgebras & Coinduction	Universal Reinforcement Learning
W12	Simplicial Sets	UMAP and manifold learning
W13	Compositional Games	Equilibria via variational inequalities
W14	Consciousness And Student Project Demos	Frontiers of AGI

Tue Sep 9	Non-degree enrollment form opens
Mon Nov 10	**University+ Spring registration begins (degree students - check SPIRE for enrollment appt date/time)**
Thu Jan 22	Preview Week for Online Classes begins
Thu Jan 29	**Spring classes begin**
Wed Feb 4	**Add/Drop Ends - Matriculated Undergraduate and Non-degree students**
Thu Feb 5	Withdrawal period begins (Matriculated Undergraduate and Non-degree students)
Wed Feb 11	**Add/Drop Ends - Matriculated Graduate students** **Last day for 100% University+ class fee refund for ALL students**
Thu Feb 12	Withdrawal period begins (Matriculated Graduate students)
Mon Feb 16	President's Day - No Classes; Offices closed
Thu Feb 19	Monday class schedule will be followed
Sun Mar 15	Spring recess begins
Mon Mar 16	Summer Session 2026 Registration Begins
Mon Mar 23	Classes resume
Thu Apr 2	**Last day to Withdraw with a "W' or select P/F (undergrad)** **Last day to drop with "DR" for Graduate students** **Last day for 50% University+ class fee refund for ALL students**
Mon Apr 20	Patriots Day - No classes; Offices closed
Fri Apr 24	Monday class schedule will be followed
Fri May 8	**Last day of classes**
Sat May 9	Reading Day
Mon May 11	Finals begin
Fri May 15	Last day of final exams; Semester ends
Thu May 21	Final grades due in SPIRE by 11:59PM EST
Fri May 22	Official Grades show on transcripts by end of day

Assigned Reading for Week 1

- Chapter 1 of Riehl's textbook on Category Theory in Context
- Chapter 1 of my textbook on Category Theory for AGI
- Familiarize yourselves with the Transformer architecture
  - https://github.com/nlp-with-transformers

Questions?