Universal Imitation Games: Generative AI beyond deep learning

Sridhar Mahadevan
Adobe Research and the University of Massachusetts
Overview of Talks

- Today’s talk: GAIA: A Generative AI Architecture beyond deep learning
- Tomorrow’s talk: Generative AI using Universal Coalgebras
- Wednesday’s talk: The (co)End of Generative AI Models
“I propose to consider the question, ‘Can machines think’?" – Alan Turing, Mind, Volume LIX, Issue 236, October 1950, Pages 433–460.
Count me out on this one. I never could write poetry.

Please write me a sonnet on the subject of the Forth Bridge.
More than an AI detector
Preserve what's human.

We bring transparency to humans navigating a world filled with AI content. GPTZero is the gold standard in AI detection, trained to detect ChatGPT, GPT4, Bard, LLaMa, and other AI models.
UNIVERSAL IMITATION GAMES*
A PREPRINT

Sridhar Mahadevan
Adobe Research and University of Massachusetts, Amherst
smahadev@adobe.com, mahadeva@umass.edu

February 16, 2024

ABSTRACT

In 1950, Alan Turing proposed a framework called an imitation game in which the participants are to be classified human or machine solely from natural language interactions. Using mathematics largely developed since Turing – category theory – we investigate a broader class of universal imitation games (UIGs). Choosing a category means defining a collection of objects and a collection of composable arrows between each pair of objects that represent “measurement probes” for solving UIGs. The theoretical foundation of our paper rests on two celebrated results by Yoneda. The first, called the Yoneda Lemma, discovered in 1954 – the year of Turing’s death – shows that objects in categories can be identified up to isomorphism solely with measurement probes defined by composable arrows. Yoneda embeddings are universal representers of objects in categories. A simple yet general solution to the static UIG problem, where the participants are not changing during the interactions, is to determine if the Yoneda embeddings are (weakly) isomorphic. A universal property in category theory is defined by an initial or final object. A second foundational result of Yoneda from 1960 defines initial objects called coends and final objects called ends, which yields a categorical “integral calculus” that unifies probabilistic generative models, distance-based kernel, metric and optimal transport models, as well as topological manifold representations. When participants adapt during interactions, we study two special cases: in dynamic UIGs, “learners” imitate “teachers”. We contrast the initial object framework of passive learning from observation over well-founded sets using inductive inference – extensively studied by Gold, Solomonoff, Valiant, and Vapnik – with the final object framework of coinductive inference over non-well-founded sets and universal coalgebras, which formalizes learning from active experimentation using causal inference or reinforcement learning.
Universal Imitation Games

Static

Dynamic

Evolutionary
GAIA: Generative AI Architecture

Beyond Deep Learning!
Generative AI faces energy crisis

Sam Altman’s $7 trillion AI chip dream has him rounding on critics: ‘You can grind to help secure our collective future or you can write Substacks about why we are going [to] fail’

Microsoft agrees to buy electricity generated from Sam Altman-backed fusion company Helion in 2028

Microsoft said Wednesday it has signed a purchase agreement with nuclear fusion startup Helion to buy electricity from it in 2028.

The deal is a vote of confidence for fusion, which has thus far not been commercialized.

Catherine Clifford
@CATCLIFFORD
Mark Zuckerberg explains tech layoffs, shares his views on Sam Altman’s $7 trillion AI chip venture

In his latest interview, Zuckerberg shared his thoughts on what is causing the tech layoffs. He even shared his opinions on OpenAI CEO Sam Altman.

<table>
<thead>
<tr>
<th>Rank &amp; Country</th>
<th>GDP (USD billion)</th>
<th>GDP Per Capita (USD thousand)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 United States Of America (U.S.A)</td>
<td>27,974</td>
<td>83.06</td>
</tr>
<tr>
<td>#2 China</td>
<td>18,566</td>
<td>13.16</td>
</tr>
<tr>
<td>#3 Germany</td>
<td>4,730</td>
<td>56.04</td>
</tr>
<tr>
<td>#4 Japan</td>
<td>4,291</td>
<td>34.55</td>
</tr>
<tr>
<td>#5 India</td>
<td>4,112</td>
<td>2.85</td>
</tr>
<tr>
<td>#6 United Kingdom (U.K.)</td>
<td>3,592</td>
<td>52.43</td>
</tr>
<tr>
<td>#7 France</td>
<td>3,182</td>
<td>48.22</td>
</tr>
<tr>
<td>#8 Italy</td>
<td>2,280</td>
<td>38.93</td>
</tr>
<tr>
<td>#9 Brazil</td>
<td>2,272</td>
<td>11.03</td>
</tr>
<tr>
<td>#10 Canada</td>
<td>2,242</td>
<td>55.53</td>
</tr>
</tbody>
</table>
Transformers cannot solve simple problems:

parity, integer modulo arithmetic, balancing arithmetic expressions
Figure source: Neural Networks and the Chomsky Hierarchy, ICLR 2023
GAIA: CATEGORICAL FOUNDATIONS OF GENERATIVE AI*

A PREPRINT

Sridhar Mahadevan
Adobe Research and University of Massachusetts, Amherst
smahadev@adobe.com, mahadeva@umass.edu

February 16, 2024

ABSTRACT

In this paper, we explore the categorical foundations of generative AI. Specifically, we investigate a Generative AI Architecture (GAIA) that lies beyond backpropagation, the longstanding algorithmic workhorse of deep learning. Backpropagation is at its core a compositional framework for (un)supervised learning: it can be conceptualized as a sequence of modules, where each module updates its parameters based on information it receives from downstream modules, and in turn, transmits information back to upstream modules to guide their updates. GAIA is based on a fundamentally different hierarchical model. Modules in GAIA are organized into a simplicial complex. Each n-simplicial complex acts like a manager of a business unit; it receives updates from its superiors and transmits information back to its n + 1 subsimplicial complexes that are its subordinates. To ensure this simplicial generative AI organization behaves coherently, GAIA builds on the mathematics of the higher-order category theory of simplicial sets and objects. Computations in GAIA, from query answering to foundation model building, are posed in terms of lifting diagrams over simplicial objects. The problem of machine learning in GAIA is modeled as “horn” extensions of simplicial sets: each sub-simplicial complex tries to update its parameters in such a way that a lifting diagram is solved. Traditional approaches used in generative AI using backpropagation can be used to solve “inner” horn extension problems, but addressing “outer horn” extensions requires a more elaborate framework.

At the top level, GAIA uses the simplicial category of ordinal numbers with objects defined as [n], n > 0 and arrows defined as weakly order-preserving mappings f : [n] → [m], where f(i) ≤ f(j), i ≤ j. This top-level structure can be viewed as a combinatorial “factory” for constructing.
Unified field theory of math!

One formalism that explains it all!

Categories are directed graphs!

Even more basic than set theory
An Impossibility Theorem for Clustering

Jon Kleinberg
Department of Computer Science
Cornell University
Ithaca NY 14853

Abstract

Although the study of clustering is centered around an intuitively compelling goal, it has been very difficult to develop a unified framework for reasoning about it at a technical level, and profoundly diverse approaches to clustering abound in the research community. Here we suggest a formal perspective on the difficulty in finding such a unification, in the form of an impossibility theorem: for a set of three simple properties, we show that there is no clustering function satisfying all three. Relaxations of these properties expose some of the interesting (and unavoidable) trade-offs at work in well-studied clustering techniques such as single-linkage, sum-of-pairs, $k$-means, and $k$-median.
Clustering as a functor

Finite Metric Space

Partitions
Characterization, Stability and Convergence of Hierarchical Clustering Methods

Gunnar Carlsson  
Facundo Mémoli*  
Department of Mathematics  
Stanford University  
Stanford, CA 94305

Editor: Ulrike von Luxburg

Abstract

We study hierarchical clustering schemes under an axiomatic view. We show that within this framework, one can prove a theorem analogous to one of Kleinberg (2002), in which one obtains an existence and uniqueness theorem instead of a non-existence result. We explore further properties of this unique scheme: stability and convergence are established. We represent dendrograms as ultrametric spaces and use tools from metric geometry, namely the Gromov-Hausdorff distance, to quantify the degree to which perturbations in the input metric space affect the result of hierarchical methods.

Keywords: clustering, hierarchical clustering, stability of clustering, Gromov-Hausdorff distance
Quantum Computing in Categories

A categorical semantics of quantum protocols

Samson Abramsky and Bob Coecke

Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD, UK.

samson.abramsky · bob.coecke@comlab.ox.ac.uk

Abstract

We study quantum information and computation from a novel point of view. Our approach is based on recasting the standard axiomatic presentation of quantum mechanics, due to von Neumann [28], at a more abstract level, of compact closed categories with biproducts. We show how the essential structures found in key quantum information protocols such as teleportation [5], logic-gate teleportation [12], and entanglement swapping [29] can be captured at this abstract level. Moreover, from the combination of the — apparently purely qualitative — structures of compact closure and biproducts there emerge ‘scalars’ and a ‘Born rule’. This abstract and structural point of view opens up new possibilities for describing and reasoning about quantum systems. It also shows the degrees of axiomatic freedom: we can show what requirements are placed on the (semi)ring of scalars C(1, 1), where C is the category and 1 is the tensor unit, in order to perform various protocols such as teleportation. Our formalism captures both the information-flow aspect of the protocols [8, 9], and the
tation [12], and entanglement swapping [29]. The ideas illustrated in these protocols form the basis for novel and potentially very important applications to secure and fault-tolerant communication and computation [7, 12, 20].

We now give a thumbnail sketch of teleportation to motivate our introductory discussion. (A more formal ‘standard’ presentation is given in Section 2. The — radically different — presentation in our new approach appears in Section 9.) Teleportation involves using an entangled pair of qubits (qA, qB) as a kind of communication channel to transmit an unknown qubit q from a source A (‘Alice’) to a remote target B (‘Bob’). A has q and qA, while B has qB. We firstly entangle qA and q at A (by performing a suitable unitary operation on them), and then perform a measurement on qA and q.[8] This forces a ‘collapse’ in qB because of its entanglement with qA. We then send two classical bits of information from A to B, which encode the four possible results of the measurement we performed on q and qA. Based on this classical communication, B then performs a ‘correction’ by applying one of four possible operations (unitary transformations) to qB, after which qB has
DisCoPy

The Python toolkit for computing with string diagrams

DisCoPy is a Python toolkit for computing with string diagrams.
- Documentation: https://docs.discopy.org
- Repository: https://github.com/discopy/discopy

Why?
Applied category theory is information plumbing. It's boring... but plumbers save more lives than doctors.

As string diagrams become as ubiquitous as matrices, they need their own fundamental package: DisCoPy.

How?
DisCoPy began as an implementation of:
- DisCoCat (distributional compositional categorical) models,
- and ONLP (quantum natural language processing).

This application has now been packaged into its own library, lambeq.

Who?
- Giovanni de Felice (CEO)
- Alexis Tourni (OOO)
- Richie Yeung (CFO)
- Boldizsar Poór (CTO)
- Bob Coecke (Honorary President)

Want to contribute or just ask us a question? Get in touch on Discord!
Le Lemme du Gare du Nord

Hom\((C(−,x),F)\) \(\simeq Fx\)

The Yoneda Lemma came to “life” in 1954

Coincidentally, Turing died in 1954
Natural Transformations

\[
\begin{align*}
F & \quad G \\
C & \quad D \\
F_{c} & \quad F_{f} \\
G_{c} & \quad G_{f} \\
\mu_{c} & \quad \mu_{c}'
\end{align*}
\]
Natural Transformations for Deep Learning

Category of Parameters

Category of Learners

Backpropagation

Stochastic approximation

Fc → Gc

Fc' → Gc'

μc → μc'

Ff → Gf
Category of Elements: For any set-valued functor $F : C \to \text{Set}$

Objects: $(c, x)$, where $c$ is a category object, and $x$ is an element of $Fc$

Arrows: $(c, x) \rightarrow (c', y)$ where $f : c \rightarrow c'$ is a morphism in $C$ so that $F(f)(x) = y$ (covariant)

Arrows: $(c, x) \rightarrow (c', y)$ and $f : c \rightarrow c'$ is a morphism in $C$ so that $F(f)(y) = x$ (contravariant)
Bifunctors $F : C^{op} \times C \to \mathcal{D}$

Coends

Distances in generalized metric spaces

Topological data analysis
Manifold learning

Probabilistic Generative Models

Ends
Inductive Inference over Well-Founded Sets \rightarrow F \text{ is the powerset functor} \rightarrow \text{Conductive Inference over non-well-founded sets}

Passive learning from observation \rightarrow \text{Active learning from experimentation (Causal inference, RL)}
GAIA: A Higher-Order Categorical Framework for Generative AI
Kerodon

an online resource for homotopy-coherent mathematics
Simplicial Category $\Delta$

- **Objects**: ordinal numbers
  - $[n] = \{0, 1, \ldots, n - 1\}$

- **Arrows**:
  - $f : [m] \rightarrow [n]$
  - If $i \leq j$, then $f(i) \leq f(j)$
  - All morphisms can be built out of primitive injections/surjections
    - $\delta_i : [n] \rightarrow [n + 1]$: injection skipping $i$
    - $\sigma_i : [n] \rightarrow [n - 1]$, surjection repeating $i$
Simplicial Sets: Contravariant Functors

\[ \delta^n_i : [n] \to [n + 1] \]

\[ \sigma^n_i : [n + 1] \to [n] \]

\[ X_n : [n] \to X : \Delta^{op} \to X \]
Simplicial Objects: One stop ML shopping center
Universal Causality

Siddharth Mahadevan

Abstract: Universal Causality is a mathematical framework based on higher-order category theory, which generalizes previous approaches based on directed graphs and regular categories. We present a hierarchical framework called UCLA (Universal Causality Layered Architecture), where at the topmost level, causal interventions are modeled as a higher-order category over simplicial sets and objects. Simplicial sets are contravariant functors from the category of oriented binary trees, and whose morphisms are edge preserving injections and surjections over finite ordered sets. NOW-ADAYS interventions on causal structures are modeled as free operations that map situations into lower level simplices. At the second layer, causal models are defined as a category, for example the category of a causal model is a symmetric monoidal category representation of DAG models. The third layer corresponds to the data layer in causal inference, where each causal object is mapped homomorphically into a set of instances using the category of sets and functions between sets. The fourth homotopy layer defines a way of abstractly characterizing causal models in terms of homotopy invariants, defined in terms of the nerve of a category (a functor that converts a causal category model into a simplicial object). Each function between layers is characterized by a universal arrow, which allows universal elements and representations through the bifunctor lemma, and induces a Grothendieck category of objects that satisfies universal homotopy models with data instances, and is related to the notion of universal graphs in relational causal models. Causal inference between layers is defined as a lifting problem, a commutative diagram whose objects are categories, and whose morphisms are functors that are characterized by different types of bifunctors. We illustrate UCLA using a variety of representations, including causal relational models, symmetric monoidal categorical contexts of DAG models, and topological representations, such as integrated visualizations and Layered Architecture for Universal Causality (UCLA).

Keywords: artificial intelligence; higher order category theory; causality; machine learning; statistics
Best data visualization method in ML today

Scalable to millions of data points

Used widely in biology

Based on higher-order category theory of simplicial sets & objects

UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction

Uniform Manifold Approximation and Projection (UMAP) is a dimension reduction technique that can be used for visualisation similarly to t-SNE, but also for general non-linear dimension reduction. The algorithm is founded on three assumptions about the data:

1. The data is uniformly distributed on a Riemannian manifold;
2. The Riemannian metric is locally constant (or can be approximated as such);
3. The manifold is locally connected.

From these assumptions it is possible to model the manifold with a fuzzy topological structure. The embedding is found by searching for a low dimensional projection of the data that has the closest possible equivalent fuzzy topological structure.

The details for the underlying mathematics can be found in our paper on ArXiv:


You can find the software on github.

Installation
GAIA: Categorical Foundations of Generative AI
Simplicial framework for generative AI

Simplicial learning is based on extension problems of inner and outer "horns" of simplicial objects.

Each directed edge defines a morphism that represents a generative AI method.

Each collection of simplices can be "glued" on to compatible simplices through "ports" that define the components of the simplex.

2-simplices

3-simplex

1-simplices

0-simplices
Backprop as Functor:
A compositional perspective on supervised learning

Brendan Fong    David Spivak    Remy Tuyeras
Department of Mathematics, Massachusetts Institute of Technology
Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology

Abstract—A supervised learning algorithm searches over a set of functions $A \rightarrow B$ parametrised by a space $P$ to find the best approximation to some ideal function $f: A \rightarrow B$. It does this by taking examples $(a, f(a)) \in A \times B$, and updating the parameter according to some rule. We define a category where these update rules may be composed, and show that gradient descent—with respect to a fixed step size and an error function satisfying a certain property—defines a monoidal functor from a category of parametrised functions to this category of update rules. A key contribution is the notion of request function. This provides a structural perspective on backpropagation, giving a broad generalisation of neural networks and linking it with structures from bidirectional programming and open games.

Consider a supervised learning algorithm. The goal of a supervised learning algorithm is to find a suitable approximation to a function $f: A \rightarrow B$. To do so, the supervisor provides a list of pairs $(a, b) \in A \times B$, each of which is supposed to approximate the values taken by $f$, i.e. $b \approx f(a)$. The supervisor also defines a space of functions over which the learning algorithm will search. This is formalised by choosing a set $P$ and a function $I: P \times A \rightarrow B$. We denote the function at parameter $p \in P$ as $I(p, -): A \rightarrow B$. Then, given a pair $(a, b) \in A \times B$, the learning algorithm takes a current hypothetical approximation of $f$, say given by $I(p, -)$, and tries to improve it, returning some new best guess...
Figure 10: A learner in the symmetric monoidal category Learn is defined as a morphism. Later in Section 3 we will see how to define learners as coalgebras instead.

**Definition 3. Fong et al. [2019]** The symmetric monoidal category Learn is defined as a collection of objects that define sets, and a collection of an equivalence class of learners. Each learner is defined by the following 4-tuple (see Figure 10).

- A parameter space $P$
- An implementation function $I : P \times A \to B$
- An update function $U : P \times A \times B \to P$
- A request function $r : P \times A \times B \to A$
\[
A \xrightarrow{(P,I,U,r)} B \xrightarrow{(Q,J,V,s)} C
\]

The composite learner \( A \to C \) is defined as \((P \times Q, I \cdot J, U \cdot V, r \cdot s)\), where the composite implementation function is

\[(I \cdot J)(p, q, a) := J(q, I(p, a))\]

and the composite update function is

\[U \cdot V(p, q, a, c) := (U(p, a, s(q, I(p, a), c)), V(q, I(p, a), c))\]

and the composite request function is

\[(r \cdot s)(p, q, a, c) := r(p, a, s(q, I(p, a), c))\].
Backprop is a functor!

\[ U_I(p, a, b) := p - \epsilon \nabla_p E_I(p, a, b) \]

\[ r_I(p, a, b) := f_a(\nabla_a E_I(p, a, b)) \]
Nerve of a Category

• Recall a category is defined as a collection of objects, and a collection of arrows between any pair of objects

• A simplicial set is a contravariant functor mapping the simplicial category to the category of sets

• Any category can be mapped onto a simplicial set by constructing its nerve

• Intuitively, consider all sequences of composable morphisms of length n!
ARE TRANSFORMERS UNIVERSAL APPROXIMATORS OF SEQUENCE-TO-SEQUENCE FUNCTIONS?

Chulhee Yun*  
MIT  
chulheey@mit.edu

Srinadh Bhojanapalli  
Google Research NY  
bsrinadh@google.com

Ankit Singh Rawat  
Google Research NY  
ankitsrawat@google.com

Sashank J. Reddi  
Google Research NY  
sashank@google.com

Sanjiv Kumar  
Google Research NY  
sanjivk@google.com

ABSTRACT

Despite the widespread adoption of Transformer models for NLP tasks, the expressive power of these models is not well-understood. In this paper, we establish that Transformer models are universal approximators of continuous permutation equivariant sequence-to-sequence functions with compact support, which is quite surprising given the amount of shared parameters in these models. Furthermore, using positional encodings, we circumvent the restriction of permutation equivariance, and show that Transformer models can universally approximate arbitrary continuous sequence-to-sequence functions on a compact domain. Interestingly, our proof techniques clearly highlight the different roles of the self-attention and the feed-forward layers in Transformers. In particular, we prove that fixed width self-attention layers can compute contextual mappings of the input sequences, playing a key role in the universal approximation property of Transformers. Based on this insight from our analysis, we consider other simpler alternatives to self-attention layers and empirically evaluate them.

Permutation-equivariant functions

\[ f(XP) = f(X)P \]
\[
\text{Attn}(X) = X + \sum_{i=1}^{h} W^i_O W^i_V X \cdot \sigma[W^i_K X]^T W^i_Q X
\]

\[
\text{FF}(X) = \text{Attn}(X) + W_2 \cdot \text{ReLU}(W_1 \cdot \text{Attn}(X) + b_1 1^T_n),
\]

**Definition 32.** The category \( C_T \) of Transformer models is defined as follows:

- The objects \( \text{Obj}(C) \) are defined as vectors \( X \in \mathbb{R}^{d \times n} \) denoting \( n \)-length sequences of tokens of dimension \( d \).
- The arrows or morphisms of the category \( C_T \) are defined as a family of sequence-to-sequence functions and defined as:

\[
T^{h,m,r} := \{ f : \mathbb{R}^{d \times n} \rightarrow \mathbb{R}^{d \times n} \mid \text{where } f(XP) = XP, \text{ for some permutation matrix } P \}
\]
Since Transformers define a category over Euclidean spaces of permutation-equivariant functions, we can construct its nerve.

Consider all compositions of Transformers building blocks of length $n$.

This construction maps the category of Transformers into a simplicial set.

It is a full and faithful embedding of Transformers as simplicial sets.

However, simplicial sets cannot be faithfully mapped back to ordinary categories.
Simplicial Sets vs. Categories

• Any category can be embedded faithfully into a simplicial set using its nerve

• The embedding is full and faithful (perfect reconstruction)

• Unfortunately, the converse is not possible

• Given a simplicial set, the left adjoint functor that maps it into a category is lossy!

• GAIA (in theory!) is more powerful than existing generative AI formalisms
Summary

• Deep learning faces an energy crisis

• Architectures like Transformers are fundamentally limited!

• We need a better framework: GAIA is one possible approach

• Builds on higher-order category theory of simplicial sets

• GAIA is a theoretical framework — not yet an actual working system!