Universal Imitation Games:
The (Co)End of Generative AI

Sridhar Mahadevan
Adobe Research and University of Massachusetts

Even more category theory

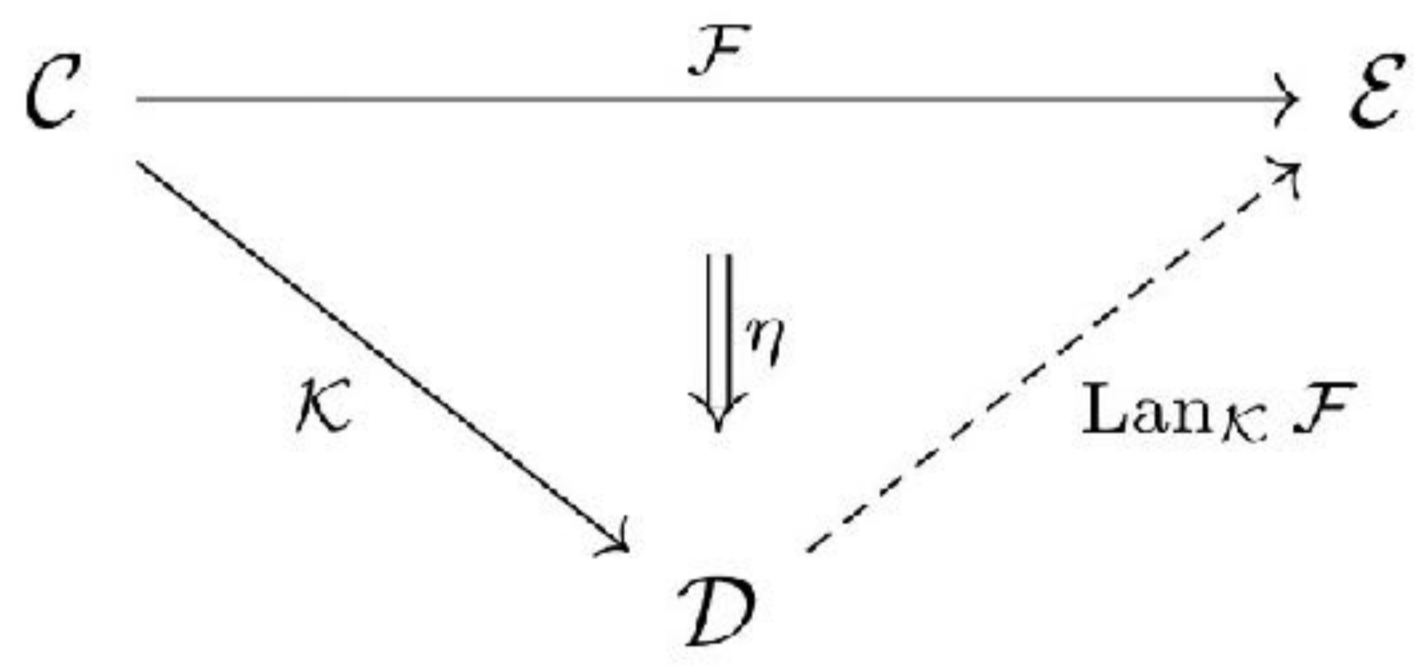
- Continuing the theme of the previous two lectures, I want to give you more examples of how category theory leads to deep insight into familiar problems
- Monads and categorical probability
- Kan extensions
- Lifting diagrams: universal structure for specifying computation
- We will construct universal representers in non-symmetric metric spaces using the powerful Yoneda Lemma
- We will construct novel types of generative AI models based on Yoneda “integral calculus” of coends and ends.

Kan Extensions and Monads

Every concept is a Kan Extension

- Kan extensions are a fundamental universal construction in category theory
- Every other concept can be derived from Kan extensions!
- Foundational result (unlike ML work on extending functions in sets)
    - There are only two canonical ways to extend a functor!

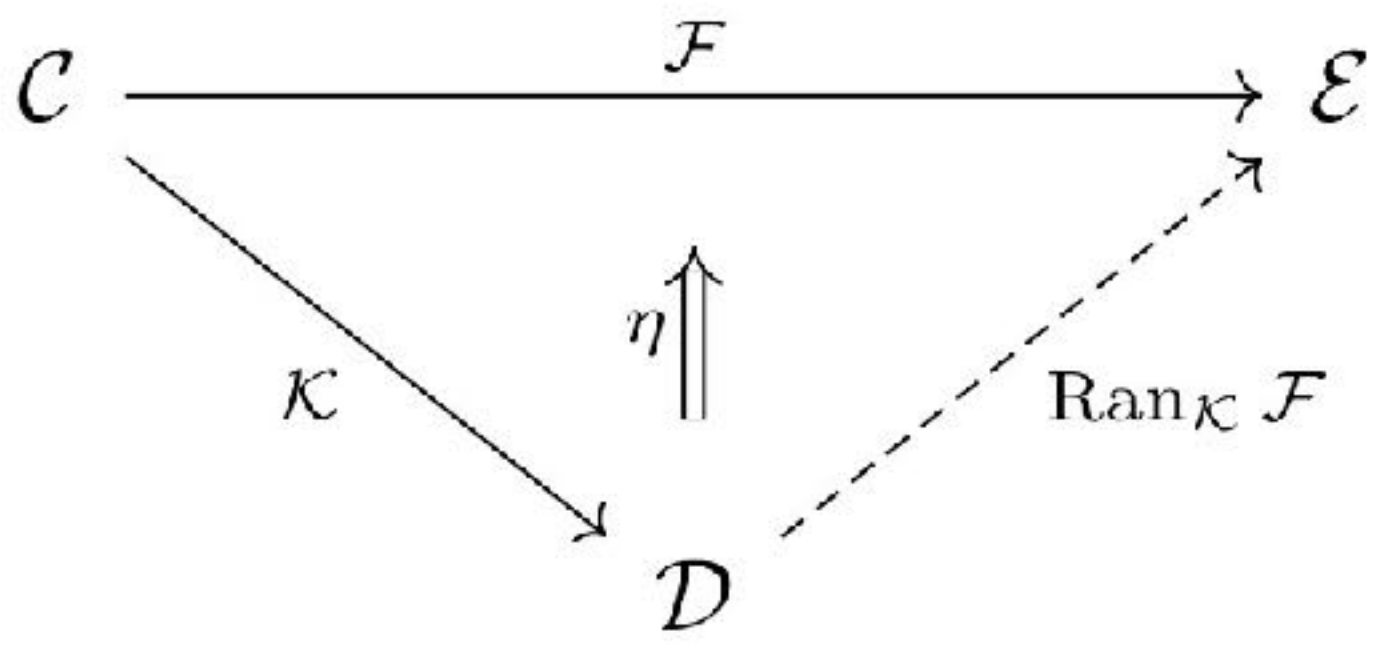
Definition 30. A **left Kan extension** of a functor $\mathcal{F} : \mathcal{C} \to \mathcal{E}$ along another functor $\mathcal{K} : \mathcal{C} \to \mathcal{D}$, is a functor $\mathrm{Lan}_{\mathcal{K}}\mathcal{F} : \mathcal{D} \to \mathcal{E}$ with a natural transformation $\eta : F \Rightarrow \mathrm{Lan}_F \circ K$ such that for any other such pair $(G : \mathcal{D} \to \mathcal{E}, \gamma : F \Rightarrow GK)$, γ factors uniquely through η. In other words, there is a unique natural transformation $\alpha : \mathrm{Lan}_F \Rightarrow G$.



Consider the case when category C is a subcategory of D

Left Kan extensions represent one of only two canonical solutions

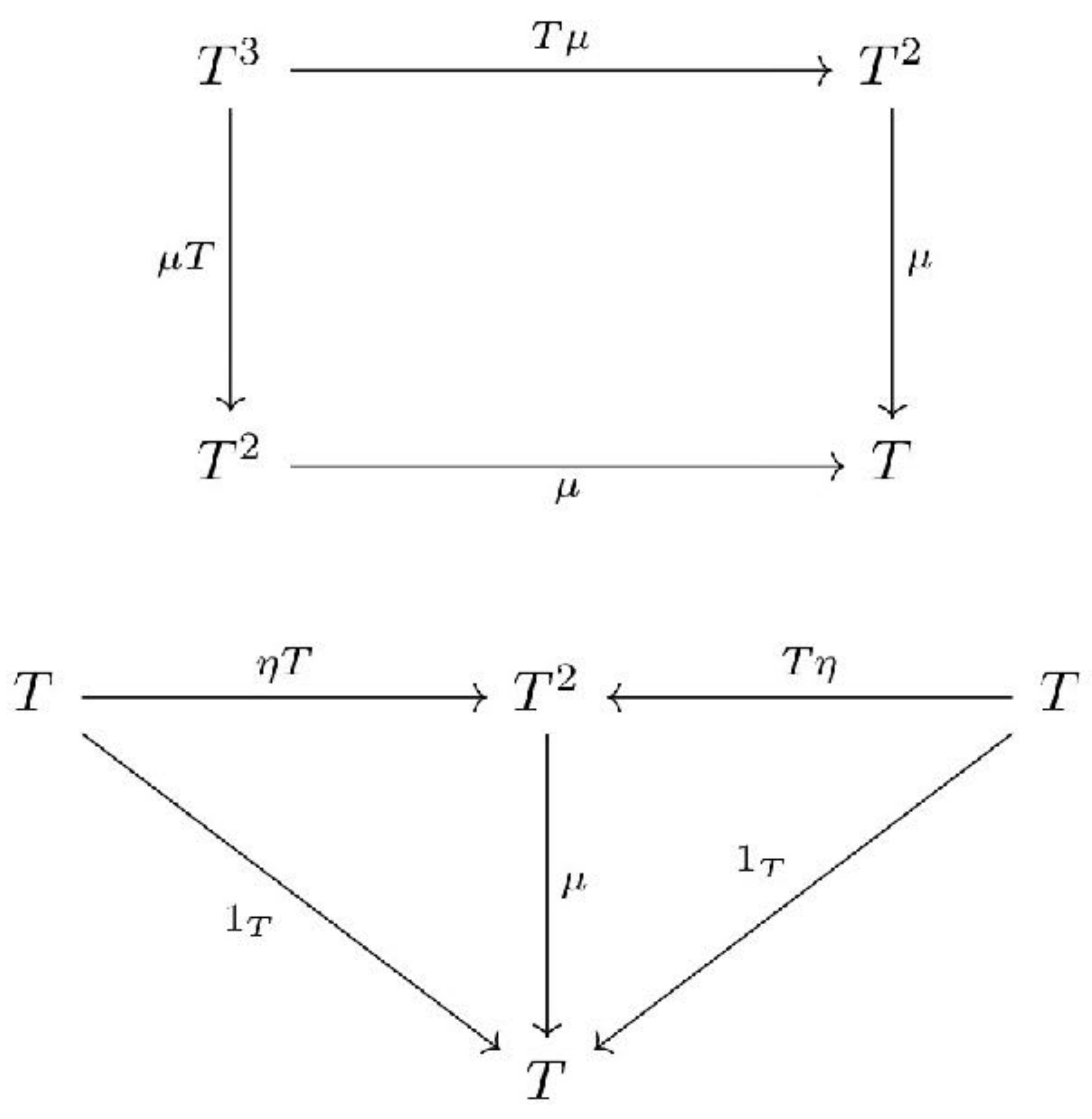
Definition 31. A **right Kan extension** of a functor $\mathcal{F} : \mathcal{C} \to \mathcal{E}$ along another functor $\mathcal{K} : \mathcal{C} \to \mathcal{D}$, is a functor $\eta : \mathrm{Ran}_F \circ K \to F$ with a natural transformation $\eta : \mathrm{Lan}_F \circ K \Rightarrow \mathcal{F}$ such that for any other such pair $(G : \mathcal{D} \to \mathcal{E}, \gamma : GK \Rightarrow F)$, γ factors uniquely through η. In other words, there is a unique natural transformation $\alpha : G \Rightarrow \mathrm{Ran}_F$.



Definition 32. A **monad** on a category C consists of

- An endofunctor $T : C \to C$
- A **unit** natural transformation $\eta : 1_C \Rightarrow T$
- A **multiplication** natural transformation $\mu : T^2 \to T$

such that the following commutative diagram in the category C^C commutes (notice the arrows in this diagram are natural transformations as each object in the diagram is a functor).



Category of directed graphs

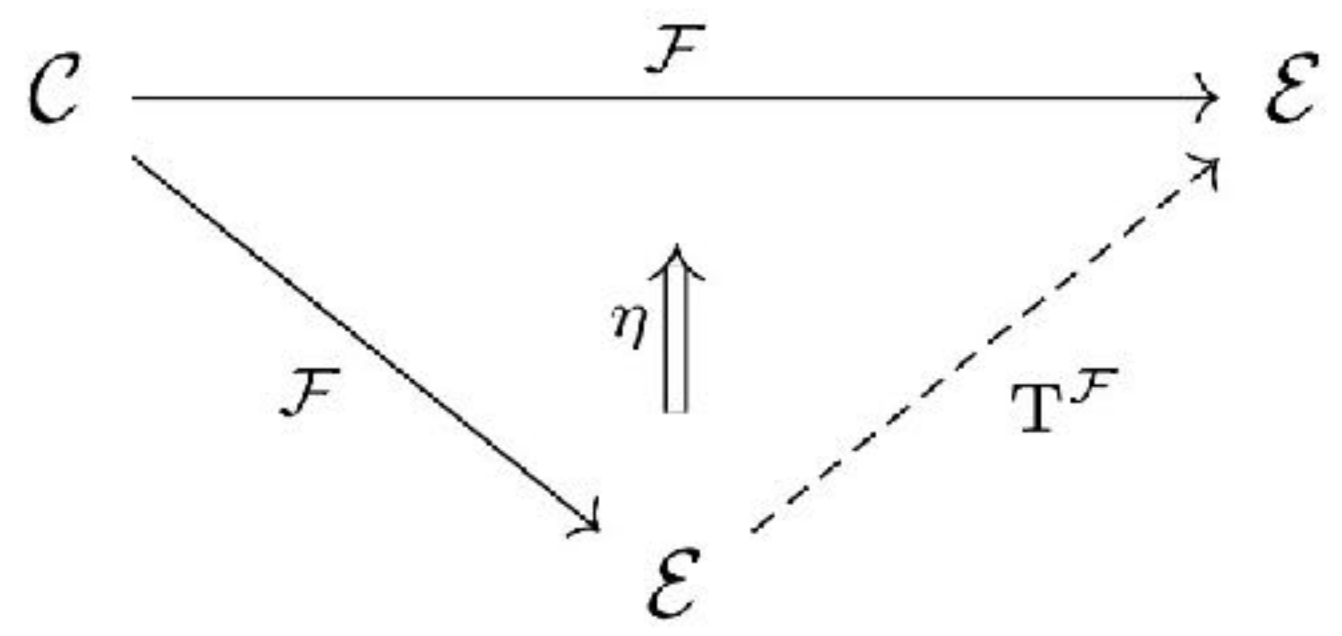
Monad: transitive closure

Category of measurable spaces

Probabilities are monads!

Probabilities are codensity monads

Definition 33. A **codensity monad** $T^{\mathcal{F}}$ of a functor $\mathcal{F}$ is the right Kan extension of $\mathcal{F}$ along itself (if it exists). The codensity monad inherits the university property from the Kan extension.



ELSEVIER

Journal of Pure and Applied Algebra

Volume 220, Issue 3, March 2016, Pages 1229-1251

Codensity and the Giry monad

Tom Avery

Show more

+ Add to Mendeley   Share   Cite

https://doi.org/10.1016/j.jpaa.2015.08.017   Get rights and content

Abstract

The Giry monad on the category of measurable spaces sends a space to a space of all probability measures on it. There is also a finitely additive Giry monad in which probability measures are replaced by finitely additive probability measures. We give a characterisation of both finitely and countably additive probability measures in terms of integration operators giving a new description of the Giry monads. This is then used to show that the Giry monads arise as the codensity monads of forgetful functors from certain categories of convex sets and affine maps to the category of measurable spaces.

Giri monad maps a measurable space

X to the space of all distributions on X

It is a monad since all distributions on

X is also measurable!

Lifting Diagrams

Definition 17. Let $\mathcal{C}$ be a category. A **lifting problem** in $\mathcal{C}$ is a commutative diagram σ in $\mathcal{C}$.

$$\begin{array}{ccc} A & \xrightarrow{\mu} & X \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle p} \\ B & \xrightarrow{\nu} & Y \end{array}$$

Definition 18. Let $\mathcal{C}$ be a category. A **solution to a lifting problem** in $\mathcal{C}$ is a morphism $h : B \to X$ in $\mathcal{C}$ satisfying $p \circ h = \nu$ and $h \circ f = \mu$ as indicated in the diagram below.

$$\begin{array}{ccc} A & \xrightarrow{\mu} & X \\ \downarrow{\scriptstyle f} & \overset{h}{\nearrow} & \downarrow{\scriptstyle p} \\ B & \xrightarrow{\nu} & Y \end{array}$$

The unreasonable power of the lifting property in elementary mathematics

misha gavrilovich*
in memoriam: evgenii shurygin

9 May 2017

instances of human and animal behavior [...] miraculously complicated, [...] they have little, if any, pragmatic (survival/reproduction) value. [...] they are due to internal constraints on possible architectures of unknown to us functional "mental structures".

Gromov, Ergobrain

Abstract

We illustrate the generative power of the lifting property (orthogonality of morphisms in a category) as a means of defining natural elementary mathematical concepts by giving a number of examples in various categories, in particular showing that many standard elementary notions of abstract topology can be defined by applying the lifting property to simple morphisms of finite topological spaces. Examples in topology include the notions of: compact, discrete, connected, and totally disconnected spaces, dense image, induced topology, and separation axioms. Examples in algebra include: finite groups being nilpotent, solvable, torsion-free, p-groups, and prime-to-p groups; injective and projective modules; injective, surjective, and split homomorphisms.

DATABASE QUERIES AND CONSTRAINTS VIA LIFTING PROBLEMS

DAVID I. SPIVAK

ABSTRACT. Previous work has demonstrated that categories are useful and expressive models for databases. In the present paper we build on that model, showing that certain queries and constraints correspond to lifting problems, as found in modern approaches to algebraic topology. In our formulation, each so-called SPARQL graph pattern query corresponds to a category-theoretic lifting problem, whereby the set of solutions to the query is precisely the set of lifts. We interpret constraints within the same formalism and then investigate some basic properties of queries and constraints. In particular, to any database π we can associate a certain derived database $\mathbf{Qry}(\pi)$ of queries on π. As an application, we explain how giving users access to certain parts of $\mathbf{Qry}(\pi)$, rather than direct access to π, improves ones ability to manage the impact of schema evolution.

CONTENTS

1. Introduction 1
2. Elementary theory of categorical databases 9
3. Constraints via lifting conditions 15
4. Queries as lifting problems 25
5. The category of queries on a database 34
6. Future work 42
References 43

FIGURE 3. A topological lifting problem

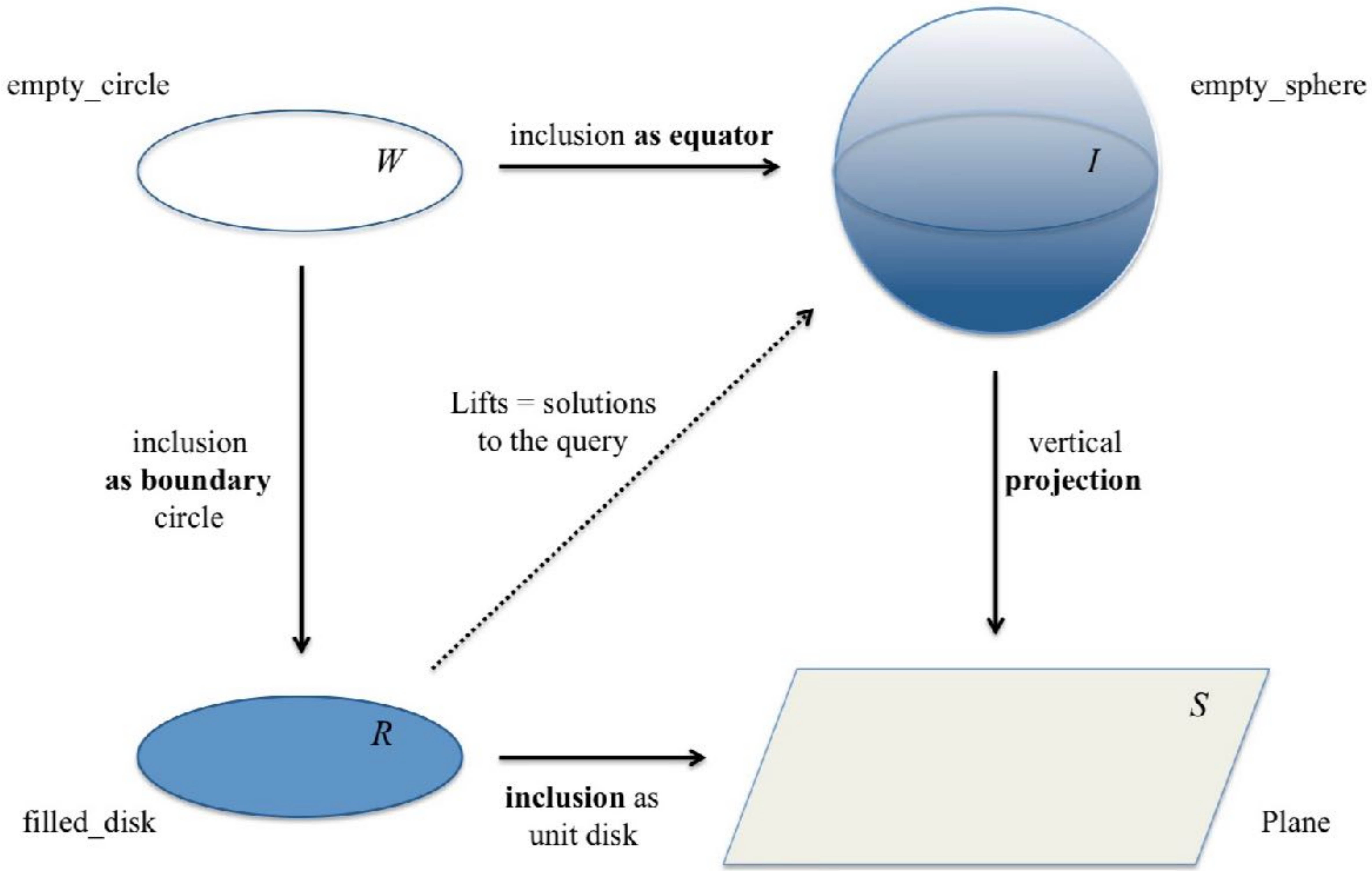


Figure source: Spivak, Database queries and constraints as lifting problems

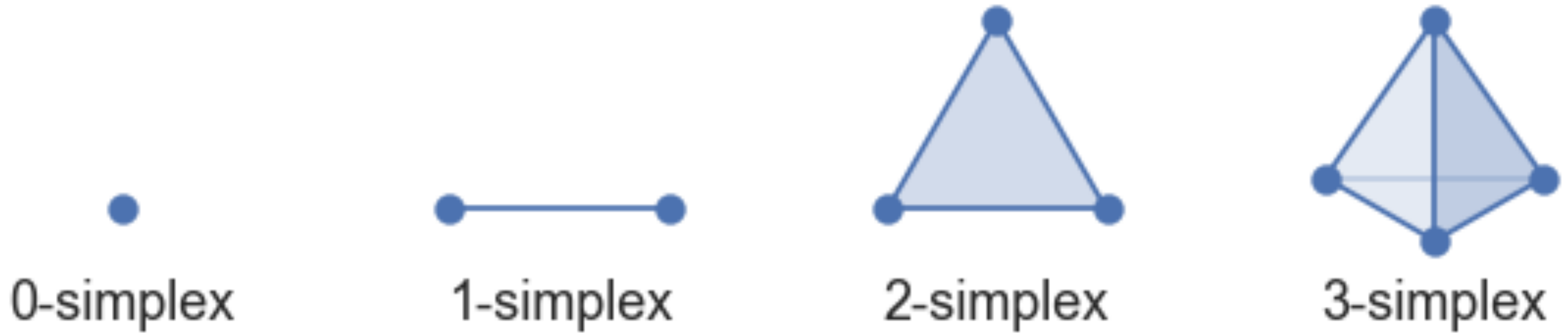
Definition 24. Let $f : X \to S$ be a morphism of simplicial sets. We say f is a **Kan fibration** if, for each $n > 0$, and each $0 \leqslant i \leqslant n$, every lifting problem.

$$\begin{array}{ccc} \Lambda^n_i & \xrightarrow{\sigma_0} & X \\ \downarrow & \overset{\sigma}{\nearrow} & \downarrow f \\ \Delta^n & \xrightarrow{\bar{\sigma}} & S \end{array}$$

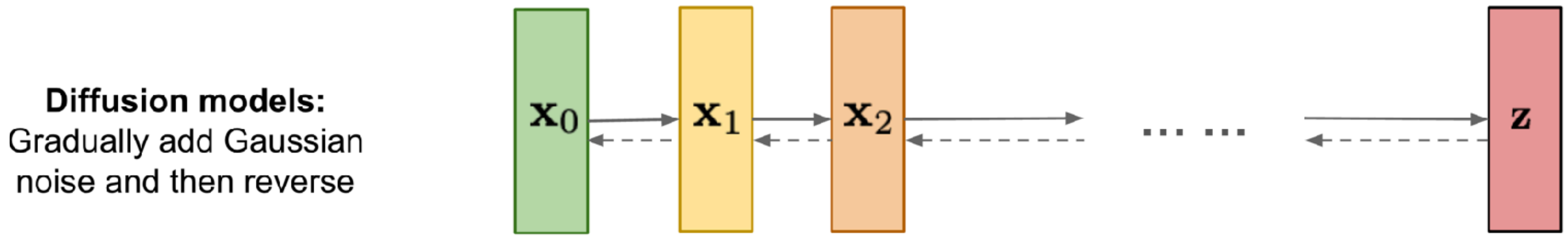
Is solvable!

Do Kan Complexes exist?

- Yes: the simplicial category has a topological realization as a Kan complex
- Each n-simplex is mapped into a topological n-simplex of all n-tuples that sum to 1



Generative AI and Kan Complexes

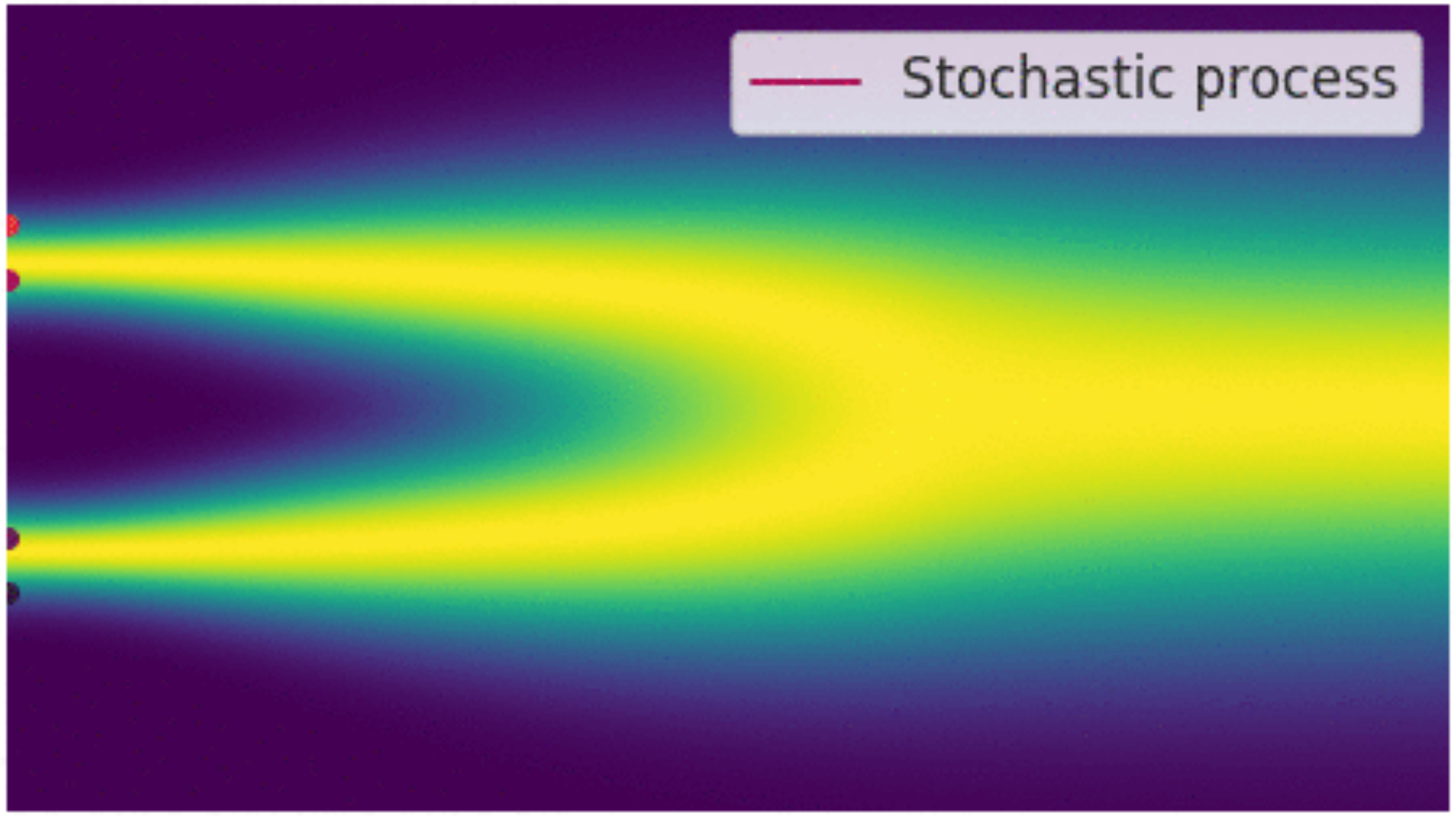


Every morphism

invertible!

Figure Source: https://lilianweng.github.io/posts/2021-07-11-diffusion-models/

Diffusion Process and Kan Complexe

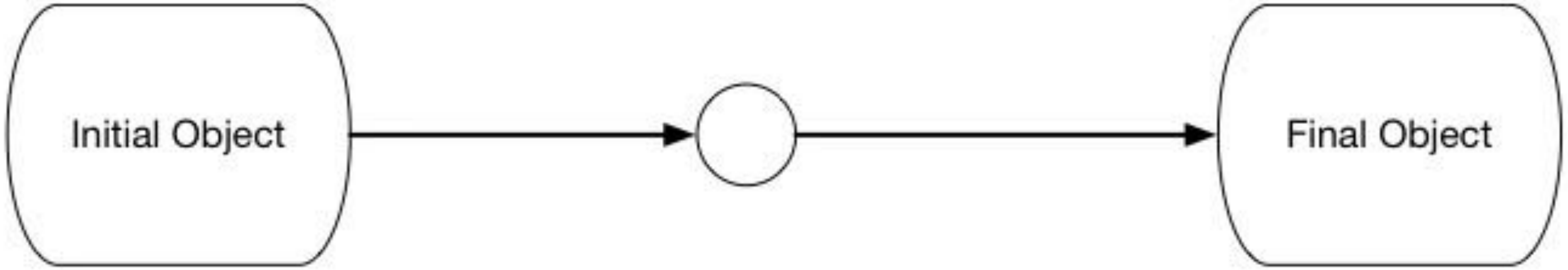


https://yang-song.net/blog/2021/score/

Integral Calculus for Generative AI

Two profound ideas by Yoneda

- Yoneda Lemma (1954):
  - Objects can be characterized by their interactions
  - Yoneda embedding: $C(-,x) : C^{op} \to \mathbf{Set}$
- Co (ends) of bi-functors (1960):
  - Bifunctors $F : C^{op} \times C \to D$
  - Category of (co)wedges defined by dinatural transformations between bifunctors
  - Coends are initial objects in a category of cowedges
  - Ends are final objects in a category of wedges



A covariant functor F is representable iff its category of elements $\int F \simeq \int C(c,-) \simeq c/C$ has an initial object

A contravariant functor F is representable iff its category of elements $\int F \simeq \int C(-,c)$ has a final object

Theoretical Computer Science

ELSEVIER Theoretical Computer Science 193 (1998) 1–51

Fundamental Study

Generalized metric spaces: Completion, topology, and powerdomains via the Yoneda embedding

M.M. Bonsangue[a,*], F. van Breugel[b], J.J.M.M. Rutten[c]

[a] *Rijks Universiteit Leiden, Department of Computer Science, P.O. Box 9512, 2300 RA Leiden, The Netherlands*

[b] *Università di Pisa, Dipartimento di Informatica, Corso Italia 40, 56125 Pisa, Italy*

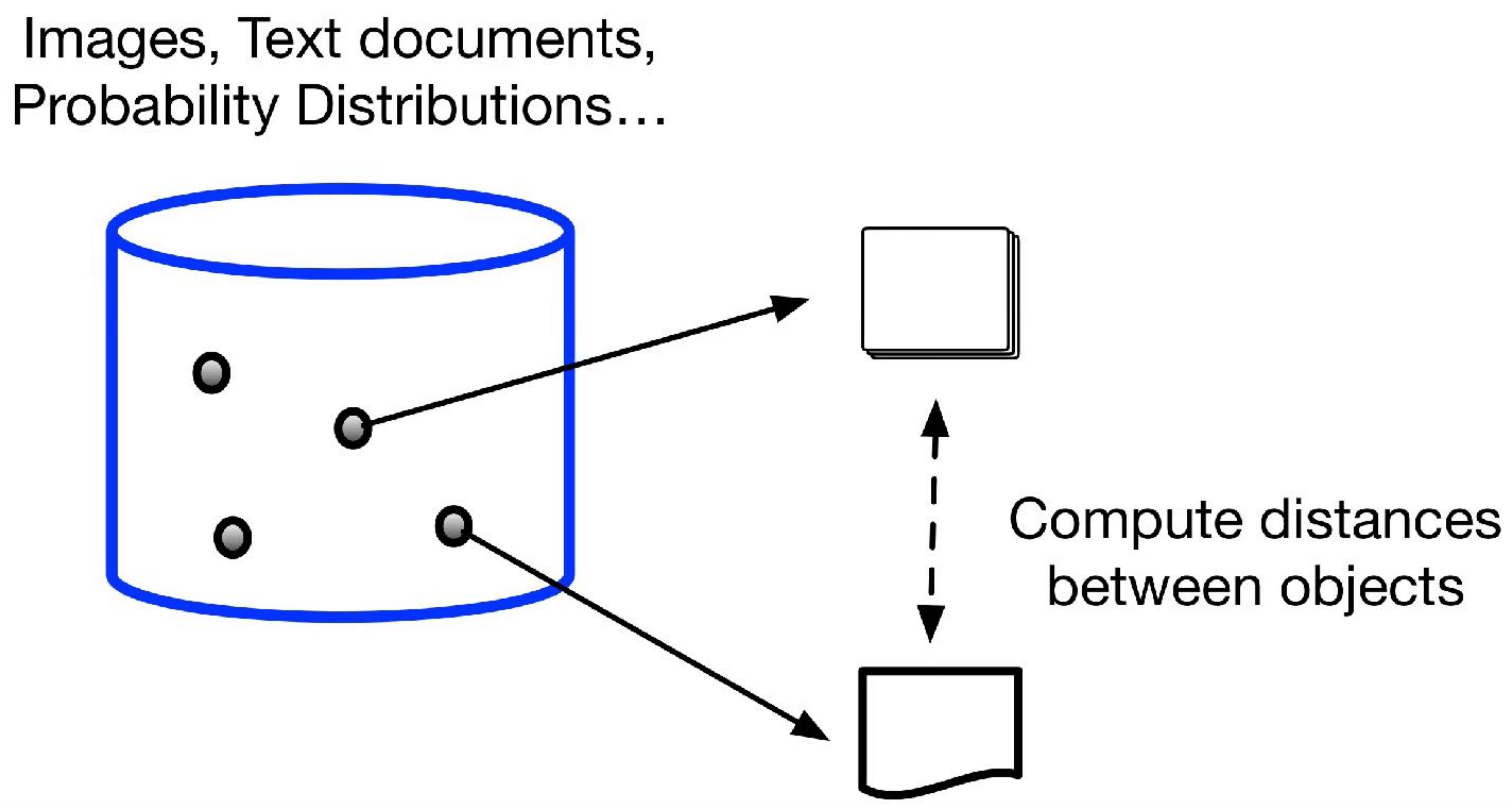
[c] *CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands*

Received June 1996; revised January 1997

Communicated by M. Nivat

Abstract

Generalized metric spaces are a common generalization of preorders and ordinary metric spaces (Lawvere, 1973). Combining Lawvere's (1973) enriched-categorical and Smyth's (1988, 1991) topological view on generalized metric spaces, it is shown how to construct (1) completion, (2) two topologies, and (3) powerdomains for generalized metric spaces. Restricted to the special cases of preorders and ordinary metric spaces, these constructions yield, respectively: (1) chain completion and Cauchy completion; (2) the Alexandroff and the Scott topology, and the ε-ball topology; (3) lower, upper, and convex powerdomains, and the hyperspace of compact subsets. All constructions are formulated in terms of (a metric version of) the Yoneda (1954) embedding.


Images, Text documents,
Probability Distributions…
Compute distances
between objects

Generalized Metric Spaces

- A generalized metric space (gms) is defined as a space X, where
    - $X(x, y) : X \times X \to [0,\infty]$
    - $X(x, x) = 0$
    - Triangle inequality: $X(x, z) \leq X(x, y) + X(y, z)$
    - Note: in a gms, symmetry does not hold, and two objects that are at distance 0 need not be identical

Examples: gms over Preorders

- Let us define a gms over a preordered set $(P, \leq)$
  - Reflexivity: $x \leq x$
  - Transitivity: $x \leq y, y \leq z \Rightarrow x \leq z$
  - The gms is defined as
    - If $x \leq y,$ then $P(x, y) = 0$
    - If $x \nleq y,$ then $P(x, y) = \infty$

Example: gms over strings

- Consider the set of strings $\Sigma*$ over some alphabet Σ
- We can define a gms over the strings $\Sigma*$ as follows:
  - $\Sigma*(x, y) = 0$ if x is a prefix of y
  - $\Sigma*(x, y) = 2^{-n}$ otherwise where n is the longest common prefix

Example: gms over topological spaces

- We can define a gms over the power set $\mathscr{P}(X)$ of all subsets over a metric space as:
  - $\mathscr{P}(X)(V, W) = \inf\left(\epsilon > 0 \,|\, \forall v \in V, \exists w \in W \text{ s.t. } X(v, w) \leq \epsilon\right)$
- This distance is referred to as the non-symmetric Hausdorff distance

Example: gms over distances

- Let us define a gms over the category $[0,\infty]$ of non-negative distances:
  - $[0,\infty](x, y) = 0$ if $x \geq y$
  - $[0,\infty](x, y) = y - x$ if $x < y$
- This category is complete and co-complete, symmetric monoidal, as well as compact and closed
  - Product of two elements is their max (or supremum)
  - Coproducts of two elements is their minimum (or infimum)
  - Monoidal product is defined as addition +

Compact Closed Categories

- Let us define an "internal" Hom functor $[0,\infty](-,-)$ as simply the distance in $[0,\infty]$ as given previously

- The Yoneda embedding $[0,\infty](t,-)$ is ***right adjoint*** to $t+-$ for any $t \in [0,\infty]$

- **Theorem:** For all $r, s, t \in [0,\infty]$,
    - $t + s \geq r$ if and only if $s \geq [0,\infty](t, r)$

Metric Yoneda Lemma for gms

- We can construct “universal representers” in any gms by applying the Yoneda Lemma
- Let X be any gms. For any element $x \in X$
  - $X(-, x) : X^{op} \to [0,\infty] : y \mapsto X(y, x)$
- Let us define a category over gms by using as arrows all non-expansive functions f
  - $Y(f(x), f(y)) \leq c \cdot X(x, y)$
  - Where $c \in (0,1)$

Presheaves in a gms

- For any category $C,$ define its presheaf $\hat{C} = \text{Set}^{C^{op}}$
- In particular, the presheaf for the category of gms is given as
  - $\hat{X} = [0,\infty]^{X^{op}}$
  - Which defines the set of all non-expansive functions from X^{op} to $[0,\infty]$
  - Remarkably, the Yoneda embedding $y \mapsto [0,\infty](y, x)$ is itself a non-expansive mapping, and therefore an element of $\hat{X}$

Metric Yoneda Lemma

- For any non-expansive function $\phi \in \hat{X}$
  - $\hat{X}(X(-,x),\phi) = \phi(x)$
- The Yoneda embedding is an *isometry!*
  - $y(x) = X(-,x)$
  - $X(x,y) = \hat{X}(y(x),y(y)) = \hat{X}(X(-,x),X(-,y))$
- Recall we have made no assumptions about symmetry!

Non-symmetric Attention in LLMs

- Recall that Tranformer modules compute permutation-equivariant maps because attention matrices are symmetric!

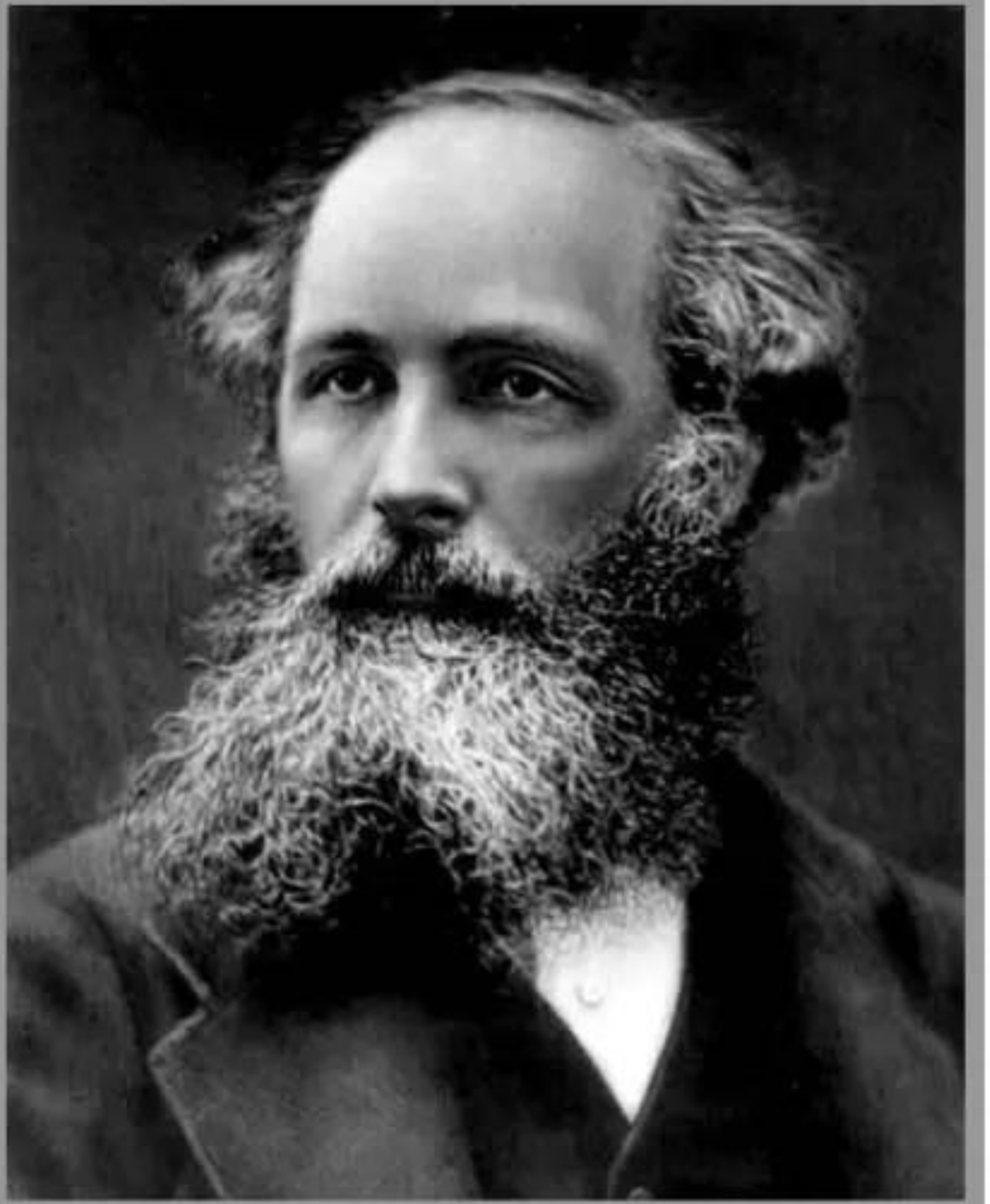
- To fix that problem, a Transformer uses Absolute Positional Encoding

- But, that “fix” causes problems of generalization in long sequences

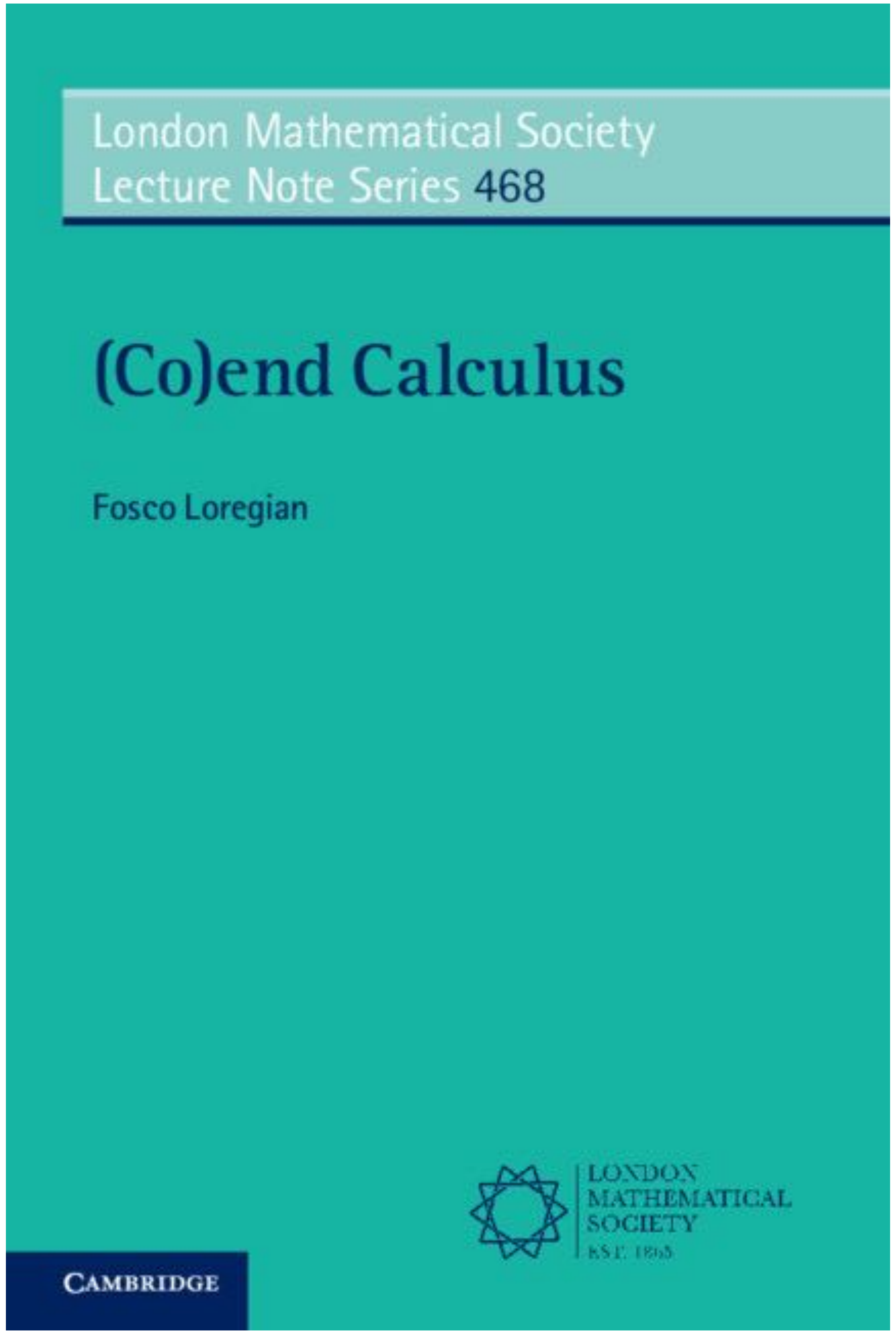
- Conjecture: Yoneda embeddings in a gms may lead to new insights into attention in LLMs

"The true logic of this world lies in the calculus of probabilities"

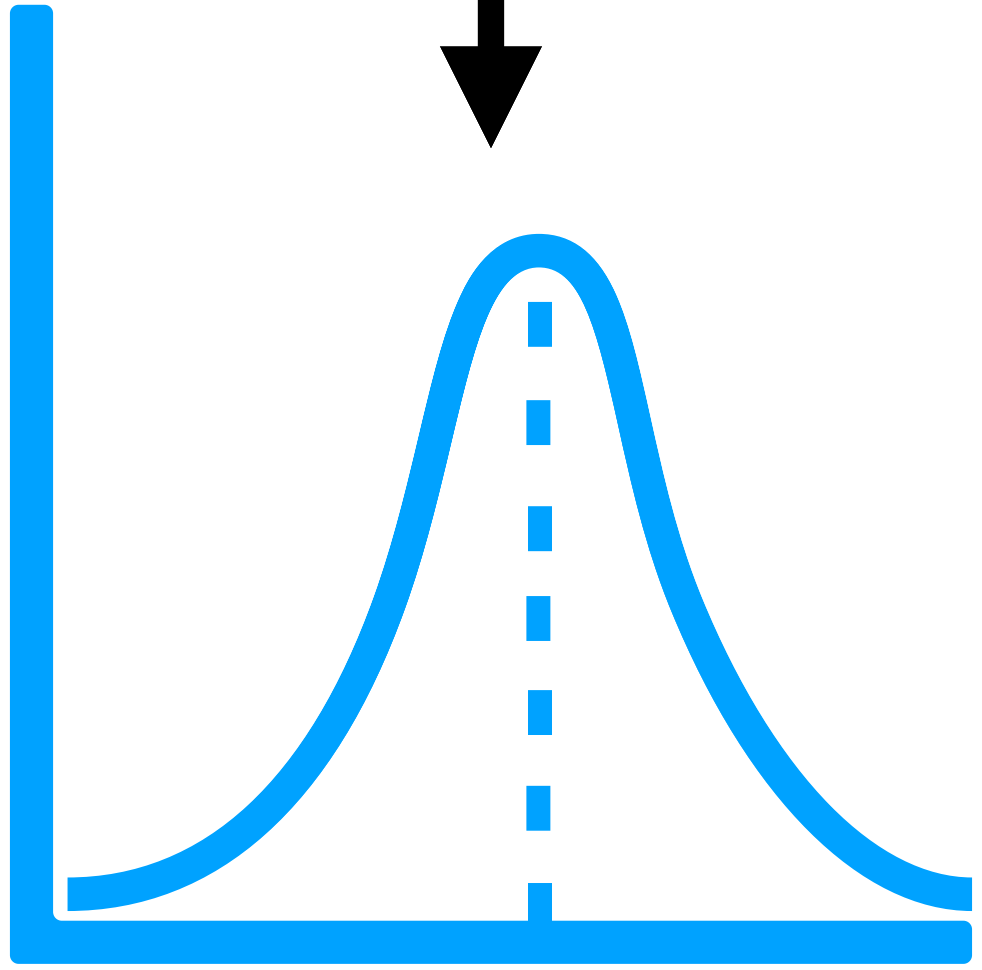
James Clerk Maxwell
Scottish Scientist
1831-1879



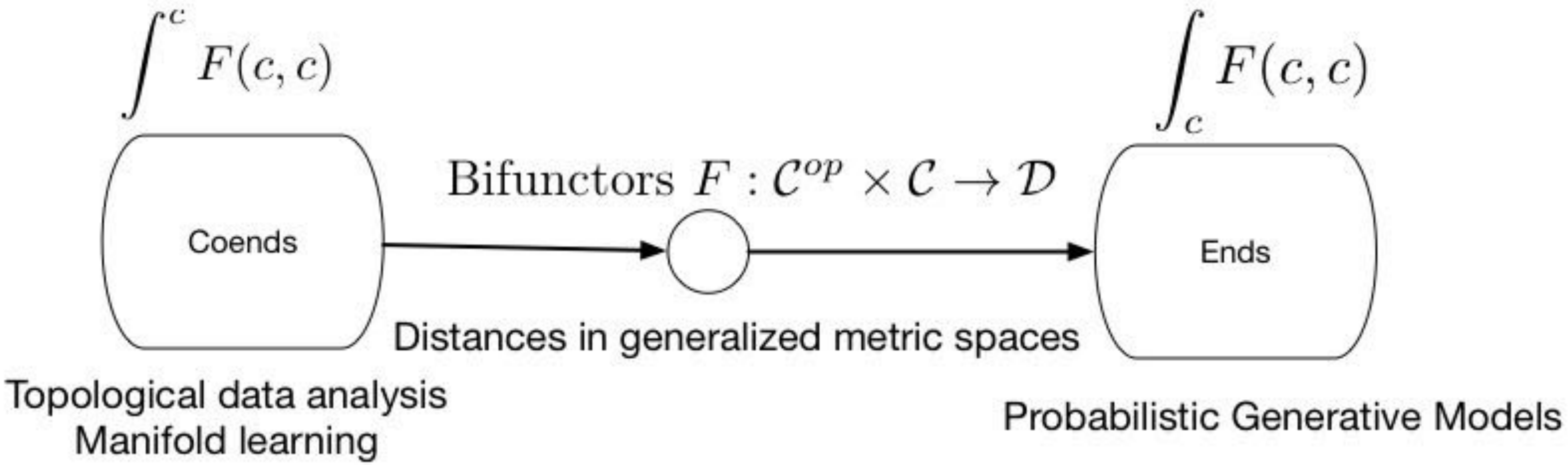
The "true logic" of Generative AI lies in the Calculus of (Co)Ends



$$\int_C F(c, c)$$

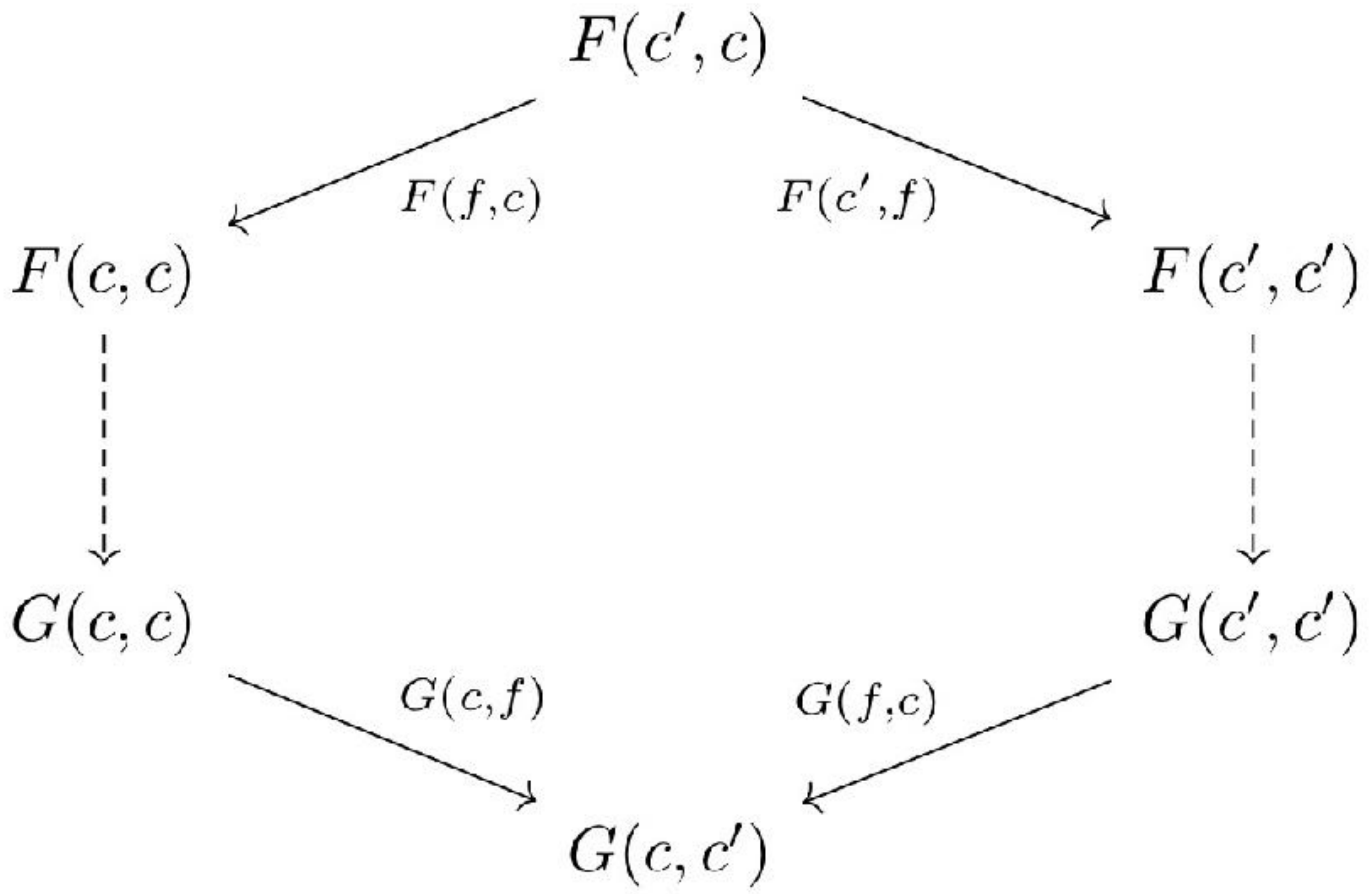


$$\int^c F(c, c)$$

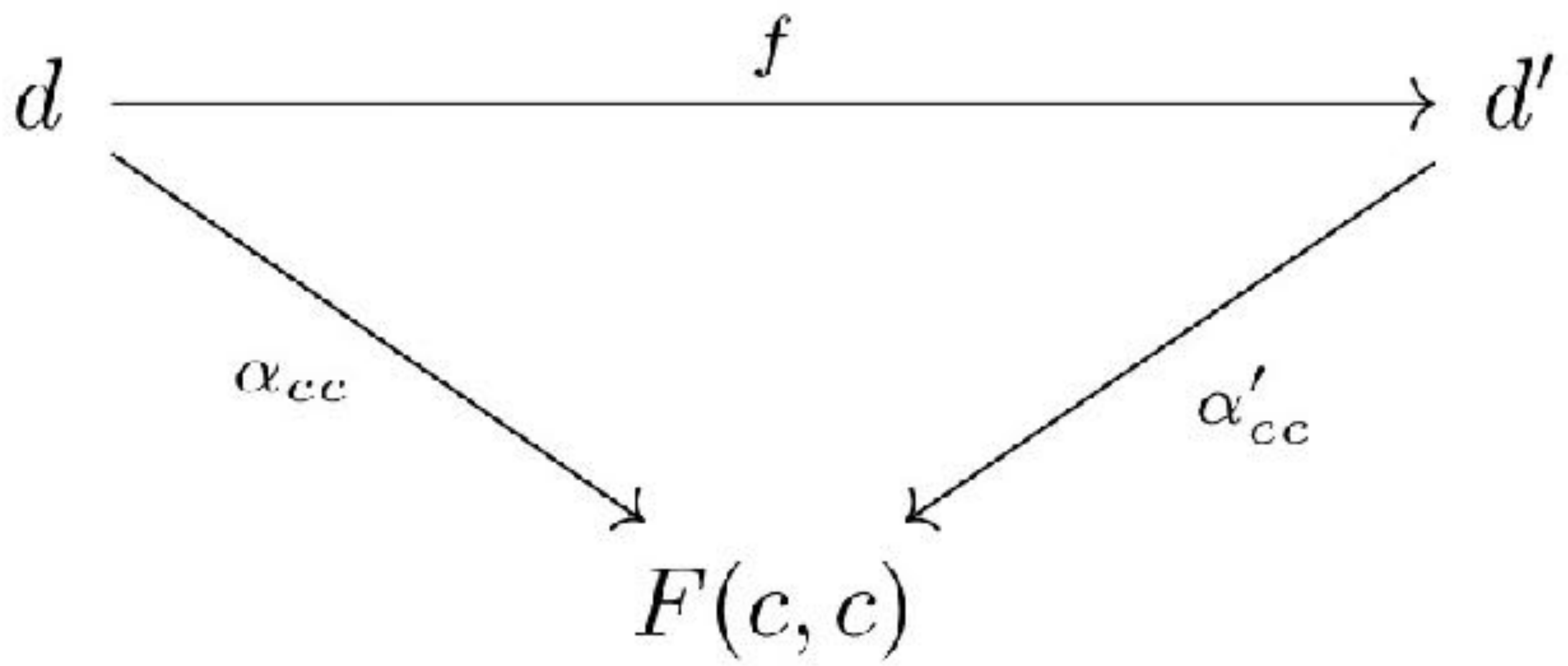


$\int^c F(c,c)$
Coends
Topological data analysis
Manifold learning
Bifunctors $F : \mathcal{C}^{op} \times \mathcal{C} \to \mathcal{D}$
Distances in generalized metric spaces
$\int_c F(c,c)$
Ends
Probabilistic Generative Models

Definition 26. Given a pair of bifunctors $F, G : \mathcal{C}^{op} \times \mathcal{C} \to \mathcal{D}$, a **dinatural transformation** is defined as follows:



Definition 28. Given a fixed bifunctor $F : \mathcal{C}^{op} \times \mathcal{C} \to \mathcal{D}$, we define the **category of wedges** $\mathcal{W}(F)$ where each object is a wedge $\Delta_d \Rightarrow F$ and given a pair of wedges $\Delta_d \Rightarrow F$ and $\Delta'_d \Rightarrow F$, we choose an arrow $f : d \to d'$ that makes the following diagram commute:



Analogously, we can define a **category of cowedges** where each object is defined as a cowedge $F \Rightarrow \Delta_d$.

Definition 29. Given a bifunctor $F : \mathcal{C}^{op} \times \mathcal{C} \to \mathcal{D}$, the **end** of F consists of a terminal wedge $\omega : \underline{\mathbf{end}}(F) \Rightarrow F$. The object $\underline{\mathbf{end}}(F) \in D$ is itself called the end. Dually, the **coend** of F is the initial object in the category of cowedges $F \Rightarrow \underline{\mathbf{coend}}(F)$, where the object $\underline{\mathbf{coend}}(F) \in \mathcal{D}$ is itself called the coend of F.

Definition 65. The **geometric realization** $|X|$ of a simplicial set X is defined as the topological space

$$|X| = \bigsqcup_{n \geqslant 0} X_n \times \Delta^n / \sim$$

where the n-simplex X_n is assumed to have a *discrete* topology (i.e., all subsets of X_n are open sets), and Δ^n denotes the *topological* n-simplex

$$\Delta^n = \{(p_0, \ldots, p_n) \in \mathbb{R}^{n+1} \mid 0 \leqslant p_i \leqslant 1, \sum_i p_i = 1$$

The spaces $\Delta^n, n \geqslant 0$ can be viewed as *cosimplicial* topological spaces with the following degeneracy and face maps:

$$\delta_i(t_0, \ldots, t_n) = (t_0, \ldots, t_{i-1}, 0, t_i, \ldots, t_n) \text{ for } 0 \leqslant i \leqslant n$$

$$\sigma_j(t_0, \ldots, t_n) = (t_0, \ldots, t_j + t_{j+1}, \ldots, t_n) \text{ for } 0 \leqslant i \leqslant n$$

Note that $\delta_i : \mathbb{R}^n \to \mathbb{R}^{n+1}$, whereas $\sigma_j : \mathbb{R}^n \to \mathbb{R}^{n-1}$.

The equivalence relation $\sim$ above that defines the quotient space is given as:

$$(d_i(x), (t_0, \ldots, t_n)) \sim (x, \delta_i(t_0, \ldots, t_n))$$

$$(s_j(x), (t_0, \ldots, t_n)) \sim (x, \sigma_j(t_0, \ldots, t_n))$$

Topological Embeddings as Coends

We now bring in the perspective that topological embeddings can be interpreted as coends as well. Consider the functor

$$F : \Delta^o \times \Delta \to \text{Top}$$

where

$$F([n], [m]) = X_n \times \Delta^m$$

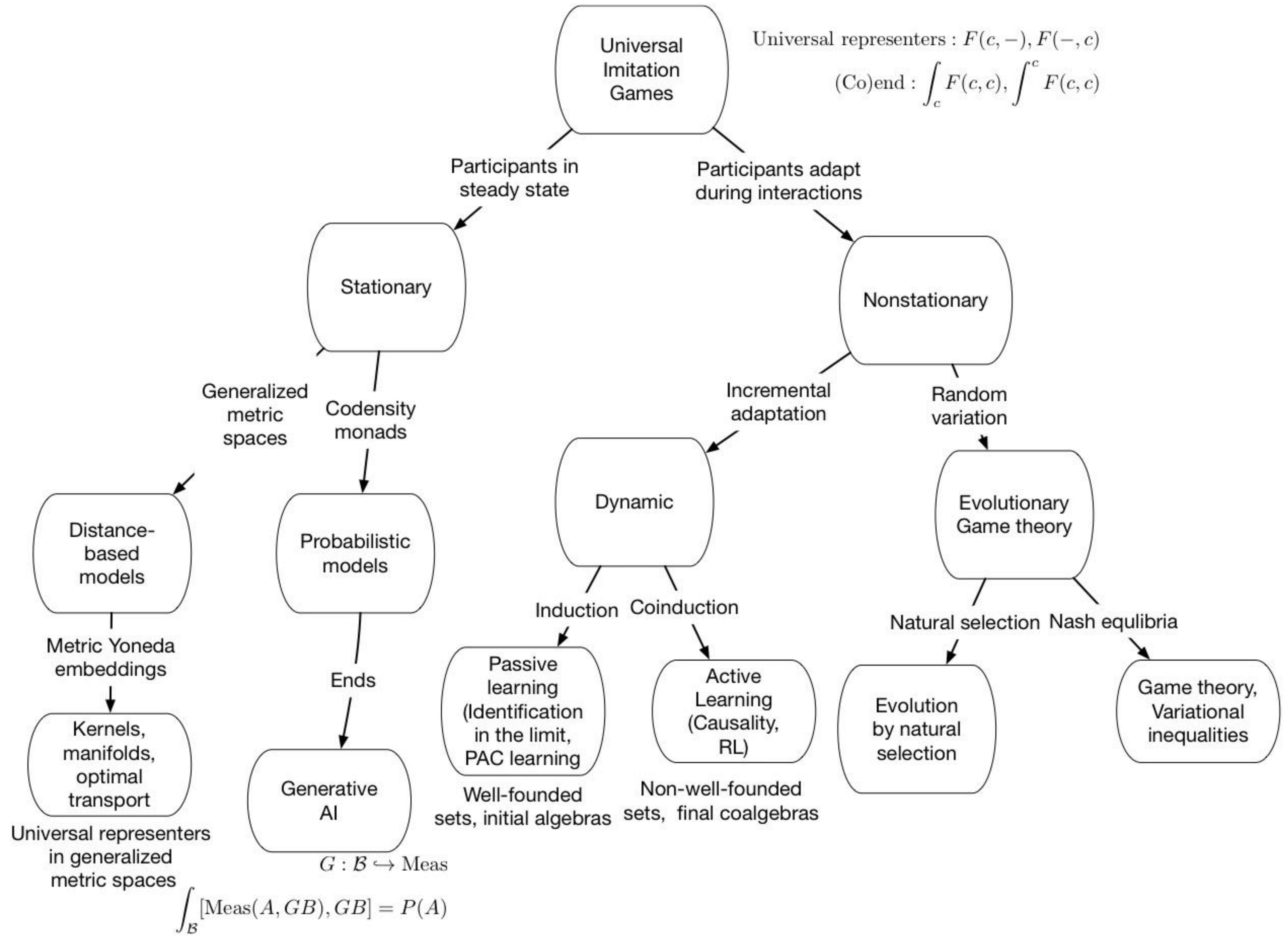
where F acts *contravariantly* as a functor from Δ to $\mathbf{Sets}$ mapping $[n] \mapsto X_n$, and *covariantly* mapping $[m] \mapsto \Delta^m$ as a functor from Δ to the category Top of topological spaces.

The "Geometric" Transformer Model

$$\int^n (\mathrm{Transformer}_\bullet n) \cdot \Delta n$$

Intuition: Construct a simplicial set of of Transformers by composing sequences of length n

Embed them in a Kan complex



Universal Imitation Games
Universal representers : $F(c,-), F(-,c)$
(Co)end : $\int_c F(c,c), \int^c F(c,c)$
Participants in steady state
Participants adapt during interactions
Stationary
Nonstationary
Generalized metric spaces
Codensity monads
Incremental adaptation
Random variation
Distance-based models
Probabilistic models
Dynamic
Evolutionary Game theory
Metric Yoneda embeddings
Ends
Induction
Coinduction
Natural selection
Nash equlibria
Kernels, manifolds, optimal transport
Generative AI
Passive learning (Identification in the limit, PAC learning
Active Learning (Causality, RL)
Evolution by natural selection
Game theory, Variational inequalities
Well-founded sets, initial algebras
Non-well-founded sets, final coalgebras
Universal representers in generalized metric spaces
$G : \mathcal{B} \hookrightarrow \text{Meas}$
$\int_{\mathcal{B}} [\text{Meas}(A, GB), GB] = P(A)$

Summary

- In these three lectures, we constructed a (higher-order) category theory of generative AI, named GAIA

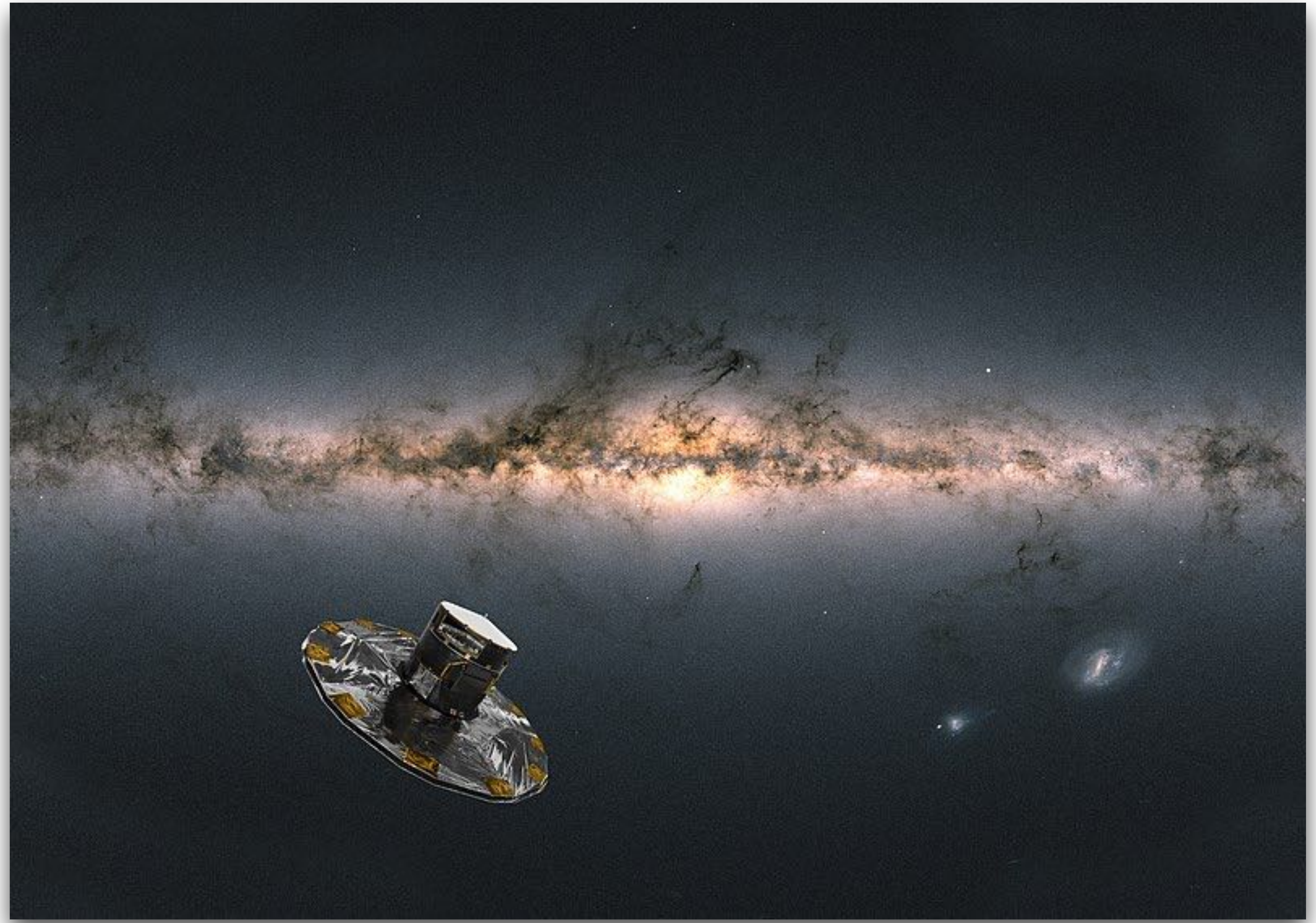
- Our goal was primary theoretical: we want to illustrate how category theory can give deep insight into hard practical problems

- Implementing GAIA is a problem for future work!

- Energy crises are plaguing generative AI — any solution is worth considering!

- Read my book drafts (continually updated) on my UMass web page

GAIA: Generative AI Architecture



Beyond Deep Learning!