Democritus: Inferring Causality from Language

Sridhar Mahadevan *

Posted Date: 14 April 2026

doi: 10.20944/preprints202604.1011.v1

Keywords: causal inference; large language models; category theory; homotopy; natural language processing; causal extraction

 doi:10.20944/preprints202604.1011.v1

Article

DEMOCRITUS: Inferring Causality from Language

Sridhar Mahadevan

Adobe Research, 345 Park Avenue, San Jose, CA 95110, USA; smahadev@adobe.com

Abstract

We describe the evolution of DEMOCRITUS, a system for inferring causality from language. Extracting causal claims from natural language is unstable under paraphrase, granularity shifts, and context drift. A document collection may express the same causal statement in many surface forms, while neighboring studies may agree locally yet fail to glue globally because relation families or polarities change across regimes. This paper studies that problem through successive versions of DEMOCRITUS, an implemented system for compiling documents into local causal models, causal databases, and interactive diagnostic artifacts. Our central claim is that categorical homotopy offers a useful computational language for finding equivalence classes of paraphrastic causal statements while avoiding collapsing genuinely distinct claims. We formalize weak equivalence between causal mentions via a normalization functor, motivate localization into homotopy classes of extracted claims, and connect missing higher-order coherence to failures of causal gluing. We then describe how these ideas are realized in the current DEMOCRITUS that uses an AGI chatbot named CLIFF (Consciousness Layer Interface to Functor Flow) pipeline through homotopy-localized claim classes, regime-gluing diagnostics, topic partitions, archived experimental artifacts, topos-style study collation via soft pullbacks and pushout merges, and an underlying categorical learning stack based on Diagrammatic Backpropagation, Geometric Transformers, and Kan Extension Transformers. Finally, we report focused case studies, including Mediterranean diet, red-wine cardiovascular studies, and rising-ocean-temperature corpora, showing that homotopy localization reduces paraphrase inflation while preserving diagnostically important regime-sensitive and obstructed claims.

Keywords: causal inference; large language models; category theory; homotopy; natural language processing; causal extraction

1. Introduction

We describe successive versions of DEMOCRITUS, an implemented system for extracting causal claims from a collection of documents. In the original version, first reported in [1], DEMOCRITUS used a large language model (LLM), such as `Qwen3-235B`, to extract a causal atlas from an input document, such as the collapse of the 5000-year old Indus civilization [2]. Crucial to the operation of DEMOCRITUS was the use of sophisticated categorical machine learning methods, including Diagrammatic Backpropagation (DB) with the Geometric Transformer (GT). These categorical tools have been significantly expanded and refined, and the version of DEMOCRITUS described in this paper uses a sophisticated Kan Extension Transformer (KET) that is described in detail in the textbook *Categories for AGI* [3]. The original version of DEMOCRITUS built thousands of causal DAGs from an input text document, and to address the combinatorial explosion of these structures, the second version of DEMOCRITUS used a compact Apache database format called CSQL to store them [4]. In this paper, we describe the third stage of DEMOCRITUS, which uses an interactive AGI chatbot named CLIFF (CONSCIOUS LAYER INTERFACE TO FUNCTOR FLOW). Detailed instructions for downloading and installing the CLIFF package are given in the appendix. FUNCTOR FLOW is a novel categorical deep learning language in which DEMOCRITUS is implemented.

Figure 1 shows a screenshot of the CLIFF GUI, which allows users to extract causal structures from a variety of sources and different modes of operation. CLIFF can take a command, such as "Analyze

 doi:10.20944/preprints202604.1011.v1

the PDF document in the folder ⋯ ", or "Analyze N recent studies on X and synthesize their joint support" (where N is some number, such as 10, and X is a topic such as "global warming", "benefits of drinking red wine", or "climate change").

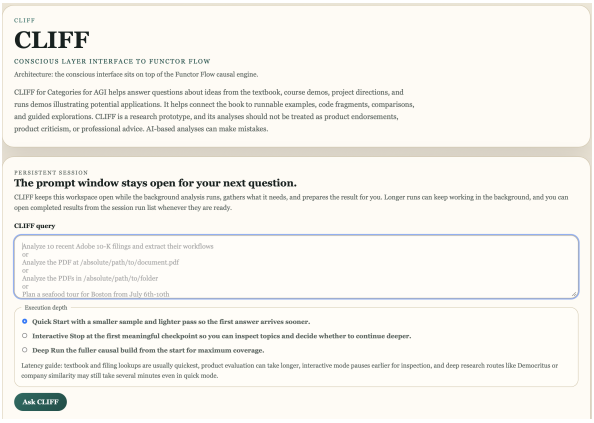

Figure 1. CLIFF is an AGI chatbot that can process causal queries, such as "Analyze 10 recent studies on rising ocean temperatures and synthesize their joint support", and compile them into causal atlases for further inference.

As a concrete illustration of CLIFF, we will demonstrate its operation on the Washington Post article "Emperor penguins have just been declared endangered" by Sarah Kaplan (9 April 2026), shown in Figure 2. The goal here is not produce an LLM summary of the article, which The Washington Post provides, but to construct a far deeper causal analysis of the underlying explanations for why the Emperor penguins are facing risk of extinction. Although the article mentions a few potential causal claims, DEMOCRITUS builds a vast repertoire of background causal claims, and synthesizes them into succinct equivalence classes.

Emperor penguins have just been declared endangered

The Antarctic fur seal also was newly labeled as endangered by a key international group, which cited a warming world.

Today at 11:53 a.m. EDT

2 min Summary 110 Make us preferred on Google

Emperor penguins with a chick at Snow Hill Island in Antarctica. (Wolfgang Kaehler/LightRocket/Getty Images)

By Sarah Kaplan

Two iconic South Pole species, the emperor penguin and the Antarctic fur seal, have been officially declared endangered by the world's leading conservation authority — a consequence of dramatic changes to their habitats caused by global warming.

Emperor penguins were already considered near threatened due to the loss of coastal sea ice, which they rely on for raising chicks. Early ice breakup can lead to colonies falling into the ocean before chicks grow their waterproof feathers, causing the hatchlings to drown.

Figure 2. An article in The Washington Post on Emperor penguins.

 doi:10.20944/preprints202604.1011.v1

1.1. Tier 1 Claims

DEMOCRITUS classifies the potential causal claims in this article into several categories, from *Tier 1* claims that have strong support, to *Tier 2* claims that are less plausible, and *Tier 3* claims that are considered weak. Details of the scoring mechanism used by DEMOCRITUS are given in [4], and relate to the density of causal paths that flows through a particular edge in a DAG. Concretely, here are examples of each of these claims, produced as an executive summary by DEMOCRITUS, and condensed here for clarity. The causal relationships uncovered by DEMOCRITUS are highlighted in italics, such as *reduces*, *leads-to*, *causes* etc.

- Rising ocean temperatures lead to krill moving to deeper waters, which *reduces* the food availability for antarctic fur seals. Supported by: Habitat loss and fragmentation driven by climate change; Sea ice role in global cooling; Sentinel species as climate indicators.
- Global warming *causes* the loss of antarctic sea ice, which leads to the degradation of emperor penguin breeding habitats. Supported by: Habitat loss and fragmentation driven by climate change.

1.2. Tier 2 Claims

In the next category are weaker claims, such as the following:

- Loss of antarctic sea ice *leads to* emperor penguin breeding habitat loss, which increases chick mortality. Supported by: Sea ice role in global cooling.
- The reduction in krill near the surface caused by warming ocean temperatures *affects* the foraging success of antarctic fur seals, contributing to their population decline. Supported by: Sea ice role in global cooling.
- Loss of antarctic sea ice *leads to* the destruction of emperor penguin breeding habitats, which reduces suitable nesting areas Supported by: Sea ice role in global cooling.

1.3. Tier 3 Claims

The final weakest category are those that have the weakest support, such as this claim:

- The loss of emperor penguin breeding habitat *increases* chick mortality by reducing safe and suitable nesting areas. Supported by: Habitat loss and fragmentation driven by climate change

In the remainder of the paper, we focus on the scientific challenges in building DEMOCRITUS, give a theoretical framework for mapping causality into language, and report on further experiments. We also relate DEMOCRITUS to traditional models of causal inference [5–7], from which it differs significantly.

2. Scientific Challenges in Building DEMOCRITUS

Large language models can readily generate causal narratives, but extracting stable causal structure from language remains difficult. The same causal statement may appear in many paraphrastic forms, while neighboring studies may agree locally yet fail to glue globally because relation families or polarities change across regimes. A naive extraction system therefore tends to overcount evidence when paraphrases are left separate, and to overglue evidence when superficially similar statements are collapsed too aggressively. These are not merely presentation issues. They change which claims appear credible, which mechanisms seem to recur across studies, and whether a corpus can be synthesized into a coherent causal account.

Given a query or a local collection of PDFs, DEMOCRITUS retrieves documents, expands a topic cover, extracts large numbers of causal mentions, constructs thousands of local causal models (LCMs), aggregates them into a causal database, and returns interactive artifacts including executive summaries, topic atlases, manifold views, local causal graph galleries, and corpus-level diagnostics. In recent versions of the system, the central practical bottleneck has not been local graph generation itself, but the instability of causal extraction under paraphrase and contextual drift. The present paper should

 doi:10.20944/preprints202604.1011.v1

therefore be read as a study of a specific computational problem inside a working system: how to organize causal discourse so that equivalence, disagreement, and gluing failures become visible rather than being flattened away.

User interaction with that system is mediated by CLIFF, the AGI chatbot illustrated in Figure 1. In practice, CLIFF is the user-facing orchestration layer that accepts a natural-language query, routes it to an appropriate workflow, manages long-running evidence acquisition, and returns a bounded artifact surface such as a dashboard, synthesis page, or textbook-guided explanation. Its design is deliberately modular: a router selects among specialized routes, background workers build local causal or explanatory state, and a conscious-layer presentation exposes best-so-far outputs rather than hiding all intermediate structure. For the present paper, the most important CLIFF route is the DEMOCRITUS route, because it is the public interface through which the causal extraction, homotopy localization, and gluing analyses described below are actually executed.

A skeptical causality researcher may object at this point that causality cannot be "extracted from language" in any strong sense. Taken as a claim about identifiability from observational data, that objection is right. Language alone does not supply the interventional semantics, experimental controls, or statistical assumptions required by modern causal inference. [8–14] We do not claim otherwise. The point of DEMOCRITUS is not to recover ground-truth causal structure from text alone, nor to replace experimental or observational causal analysis. Rather, the system attempts to recover, normalize, compare, and diagnose *causal discourse*: the claims, mechanisms, regime conditions, and conflicts expressed in document collections.

That distinction matters because causal discourse is itself scientifically important. Policy documents, scientific articles, grant proposals, news analysis, and review essays are full of causal claims that researchers and decision-makers actually consume long before they reach a clean structural model. Even when those claims are not sufficient for identification, they remain valuable as a source of hypotheses, mechanism sketches, regime annotations, explanatory narratives, and disagreement signals. The right question for this paper is therefore narrower and more operational: given a corpus containing causal discourse, can we extract structured causal artifacts from it in a way that is more stable, more auditable, and more diagnostically useful than ordinary summarization?

Our main claim is that categorical homotopy provides a useful computational language for this problem. Instead of treating each extracted causal mention as an independent fact, we treat causal mentions up to weak equivalence induced by a normalization functor. This leads naturally to homotopy-localized claim classes, simplicial coherence diagnostics, and regime-gluing views that distinguish cleanly glued claims from regime-sensitive and obstructed ones. In the current DEMOCRITUS/CLIFF implementation, these ideas are realized concretely through homotopy-localized views in the causal SQL bundle, regime-conditioned gluing summaries, topic partitioning for broad corpora, and reproducible archived runs.

The paper is intentionally more experimental than our earlier theoretical work on higher categorical treatments of causality. [1,15] The earlier theory motivated the language of gluing, localization, and obstruction. Here we focus on the computational side: what failure modes arise in practice when causal knowledge is extracted from language, and how homotopical quotienting changes the behavior of a working system.

This framing also clarifies the novelty of the present work. The contribution is not another causal relation extractor evaluated only at the sentence level, nor another claim that a language model "understands causality." The contribution is a systems-and-theory account of what happens when large-scale causal extraction from language is treated as a problem of quotienting, localization, and descent. That is why the saved artifacts matter so much in this paper: they expose not just outputs, but the cover structure, localized claim classes, regime-sensitive surfaces, and local causal models from which those outputs were assembled.

The contributions of the paper are fourfold.

 doi:10.20944/preprints202604.1011.v1

1. We formulate causal extraction from language as a passage from textual mentions to normalized causal claims, and use this to define weak equivalence and homotopy localization for causal discourse.
2. We describe how these ideas are implemented in DEMOCRITUS and surfaced in CLIFF through homotopy localization, regime gluing, topic partitions, archived run artifacts, and a categorical learning stack built from Diagrammatic Backpropagation, Geometric Transformers, and Kan Extension Transformers.
3. We present selected empirical results from saved runs showing both the success and the limits of the approach: focused corpora yield stable multi-regime glued claims, red-wine study collation yields auditable pullback and pushout witnesses, and broader corpora reveal fragmented covers and failures of descent.
4. We argue that reproducible interactive artifacts are themselves a useful experimental method for studying causality from language, because they preserve the local covers, gluing failures, and supporting causal models that are usually flattened away in static summaries.

The rest of the paper is organized as follows. Section 3 defines the extraction problem and motivates homotopy localization. Section 4 gives a condensed categorical account of the discourse–model bridge, including adjoint structure, motif density, and the simplicial diagnostics used for localized claim families. Section 5 then summarizes the categorical machine-learning machinery that makes the implemented system work, focusing on Diagrammatic Backpropagation, Geometric Transformers, and Kan Extension Transformers. Section 6 summarizes the current DEMOCRITUS/CLIFF pipeline. Section 7 describes the experimental framing and presents selected case studies. Section 8 summarizes related work. Section 9 discusses the main lessons and limitations. Longer system details and additional categorical background are moved to the appendices.

3. Homotopy Localization of Causal Mentions

In DEMOCRITUS, *homotopy localization* means that causal extraction is carried out not at the level of raw textual mentions, but at the level of equivalence classes of mentions that preserve the same normalized causal content. A corpus may express one mechanism through many paraphrases, clause reorderings, or local discourse variants; if those are left separate, evidence fragments and support is overcounted. At the same time, linguistically nearby statements can differ in polarity, temporal scope, intervention type, or regime, so collapsing them too aggressively destroys scientifically important distinctions. Homotopy localization is the computational step that quotients away causally inessential variation while preserving the distinctions that matter for gluing, obstruction, and downstream causal interpretation.

We begin with the syntax side. Let $\mathcal{S}_{\text{causal}}$ denote a category of causal mentions, whose objects are sentence fragments, passages, or short discourse neighborhoods expressing putative causal content. Morphisms in $\mathcal{S}_{\text{causal}}$ represent semantics-preserving rewrites, paraphrases, normalization steps, or discourse-level reformulations. On the semantic side, let $\mathcal{C}_{\text{causal}}$ denote a category of normalized causal claims. A normalization pipeline maps each mention to a structured causal object

$$c = (X, Y, r, s, \tau, \kappa, q),$$

where X is a cause expression, Y an effect expression, r a normalized relation family, s a polarity, τ a temporal scope, κ a regime or context label, and q auxiliary qualifiers such as hedging, confidence, or provenance. We write

$$E : \mathcal{S}_{\text{causal}} \to \mathcal{C}_{\text{causal}}$$

for the resulting causal extraction functor.

This functor defines the notion of weak equivalence used throughout the paper. A morphism $f : u \to v$ in $\mathcal{S}_{\text{causal}}$ is a weak equivalence when $E(f)$ is an isomorphism in $\mathcal{C}_{\text{causal}}$. Intuitively, weak equivalences are precisely those rewrites that preserve normalized causal content. They are not mere

 doi:10.20944/preprints202604.1011.v1

lexical similarities. Two mentions may sound close yet fail to be weakly equivalent if they differ in sign, intervention type, temporality, or regime. This is why the extraction functor is essential: it prevents homotopy from collapsing distinctions that are causally meaningful.

Once weak equivalences have been identified, they should be formally inverted. When the localization exists, the resulting homotopy category is

$$\mathrm{Ho}(\mathcal{S}_{\mathrm{causal}}) := \mathcal{S}_{\mathrm{causal}}[W_E^{-1}],$$

where W_E is the class of weak equivalences induced by E. In this localized category, two mentions connected by a zig-zag of weak equivalences become isomorphic. That is the clean mathematical form of the operational instruction "merge paraphrases that express the same causal claim." In DEMOCRITUS, the canonical claim classes stored in the CSQL bundle are a computational approximation to these localized objects: evidence is aggregated at the level of normalized causal content rather than at the level of unstable wording.

Localization, however, is only the first step. One also needs a language for extension, higher-order compatibility, and failure of coherence. This is where lifting problems and simplicial organization enter. A lifting problem asks whether a partial semantic comparison or partial normalization can be extended to a fuller compatible one. In the present setting, cofibration-like maps may be read as controlled enrichments of partial causal structure, while fibration-like maps represent projections along which compatible refinements can be lifted. The model-category axioms are not imposed directly on raw text, but their interpretation is still informative: weak equivalences capture semantics-preserving variation, cofibrations capture structured extensions of partial descriptions, and fibrations capture semantic stability under refinement.

To expose higher-order coherence, we organize paraphrase families simplicially. For a normalized claim c, let $\mathcal{G}_c$ be the paraphrase groupoid whose objects are mentions u with $E(u) \cong c$ and whose morphisms are invertible paraphrase maps between them. Its nerve $N(\mathcal{G}_c)$ is a simplicial set whose vertices are mentions, edges are pairwise identifications, triangles are coherent triple identifications, and higher simplices encode higher-order compatibility. Horn filling then has a direct interpretation: if several pairwise paraphrase relations are known, a horn filler records whether they extend to a coherent multiway identification. Filled horns correspond to stable paraphrase classes; missing fillers signal wording drift, latent mechanism splitting, or regime-sensitive incompatibility.

This distinction between paraphrase and regime is essential for the present paper. One article may say that elevated sea surface temperatures affect habitat suitability for cold-adapted species, while another says they reduce it. Linguistically these are close, but the polarity conflict means they should not become equivalent in $\mathrm{Ho}(\mathcal{S}_{\mathrm{causal}})$. The practical question is therefore not merely which mentions are similar, but which mentions are admissibly equivalent once relation family, polarity, temporal scope, and regime are respected. That is the sense in which homotopy localization is stronger than surface paraphrase collapse: it stabilizes causal discourse without erasing scientifically important distinctions. With that stabilization layer in place, we can then ask how localized discourse objects are grounded in formal causal models.

4. A Categorical Bridge from Discourse to Models

4.1. Discourse Space and Model Space

The separation between a space of linguistic causal discourse and a space of formal causal models parallels familiar distinctions in categorical semantics of language. In Lambek-style and related compositional accounts, syntax lives in one structured category while semantics is assigned in another, and enriched-meaning constructions explain why some expressions require a lifted semantic space only when composition makes that necessary [16,17]. DEMOCRITUS admits an analogous reading.

Let $\mathcal{M}_{\mathrm{disc}}$ denote a category of causal discourse objects: document neighborhoods, discourse motifs, and extracted claim families whose content is expressed in natural language and whose intervention semantics is only implicit. Let $\mathcal{M}_{\mathrm{model}}$ denote a category of formal causal models, including

 doi:10.20944/preprints202604.1011.v1

structural causal models and richer objects such as Topos Causal Models, in which interventions, transport, and identification are defined explicitly. [18]

On this view, DEMOCRITUS v1 mostly lives in $\mathcal{M}_{\text{disc}}$: it samples and organizes causal discourse, builds neighborhoods, and surfaces recurrent mechanisms. The missing step is *causal grounding*: a principled passage from discourse neighborhoods to formal causal objects on which operations such as adjustment or transport can be carried out. That passage is what motivates the adjoint formulation below.

4.2. The Discourse–Model Adjunction

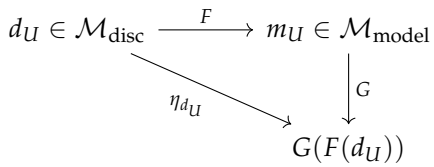

Figure 3. The unit of the discourse–model adjunction. Formalizing a discourse neighborhood and then explaining the resulting model back into language should preserve causal content up to normalization and explicit assumptions.

We model the discourse–model bridge by an adjunction

$$F \dashv G : \mathcal{M}_{\text{disc}} \rightleftarrows \mathcal{M}_{\text{model}},$$

where F formalizes discourse objects into causal models and G maps causal models back to canonical explanatory discourse. The adjunction gives a natural bijection

$$\text{Hom}_{\mathcal{M}_{\text{model}}}(F(d), m) \;\cong\; \text{Hom}_{\mathcal{M}_{\text{disc}}}(d, G(m)).$$

Operationally, this says that fitting the formalization of a discourse neighborhood into a model is equivalent to mapping the original neighborhood into the canonical explanation of that model. The unit $\eta : \text{Id} \Rightarrow GF$ and counit $\epsilon : FG \Rightarrow \text{Id}$ then become round-trip consistency contracts for DEMOCRITUS: formalize-then-explain should preserve discourse up to normalization, while explain-then-formalize should return to the original model up to conservative approximation.

4.3. A Universal Property for Causal Grounding

To make this bridge precise, let $\mathcal{D}$ be a small category of typed causal discourse generators: atomic discourse templates, typed motifs such as back-door, front-door, instrument, collider, selection, and transport fragments, together with refinement or embedding maps between them. Let $\mathcal{E}$ be a cocomplete category of causal models, and let

$$A : \mathcal{D} \to \mathcal{E}$$

interpret each discourse generator as a primitive model fragment.

Define the discourse-observation functor

$$R : \mathcal{E} \to \mathbf{Set}^{\mathcal{D}^{op}}, \qquad R(E)(d) := \text{Hom}_{\mathcal{E}}(A(d), E).$$

Thus $R(E)$ records which discourse motifs are realized inside the model E. The left adjoint to R is obtained by gluing model fragments according to a presheaf of discourse constraints.

 doi:10.20944/preprints202604.1011.v1

Theorem 1 (Discourse Realization by Colimit)**.** *Let $\mathcal{D}$ be a small category, let $\mathcal{E}$ be cocomplete, and let $A : \mathcal{D} \to \mathcal{E}$ be any functor. For each presheaf $P \in \mathbf{Set}^{\mathcal{D}^{op}}$, define*

$$L(P) := \operatorname{colim}\Big(\int_{\mathcal{D}} P \xrightarrow{\pi_P} \mathcal{D} \xrightarrow{A} \mathcal{E} \Big),$$

where $\int_{\mathcal{D}} P$ is the category of elements of P. Then L is left adjoint to R, so there is a natural bijection

$$\mathrm{Nat}(P, R(E)) \cong \mathrm{Hom}_{\mathcal{E}}(L(P), E)$$

for every presheaf P and model object E in $\mathcal{E}$. If moreover A is dense on the full subcategory of $\mathcal{E}$ generated under colimits by its image, then every motif-generated model is canonically recovered from its discourse observations.

Proof. This is the standard category-of-elements formula for the left Kan extension of A along the Yoneda embedding. The colimit defining $L(P)$ glues together copies of the primitive fragments $A(d)$ indexed by generalized elements of P, and the universal property of that colimit yields the stated natural bijection. Density of A is then the usual density-theorem criterion guaranteeing that the induced nerve functor is fully faithful on the motif-generated subcategory. □

The theorem is useful because it identifies causal grounding with a universal property rather than with a specific implementation detail. A discourse presheaf P represents local typed evidence; $L(P)$ is the canonical causal model assembled from that evidence; and R extracts the family of discourse views supported by any model. In particular, the induced monad

$$T := RL$$

acts on discourse presheaves as a *causal grounding closure*: realize the discourse as a model and then re-express the supported views in discourse space. This is the categorical form of the enrichment step that DEMOCRITUS approximates computationally.

When $\mathcal{E}$ is taken to be a category of Topos Causal Models [19] or local structural causal models [5], the same theorem says that a small generating family of typed motifs can be dense enough to recover an experimentally relevant class of causal models from motif observations alone. That is the precise sense in which the discourse–model adjunction is more than metaphor.

4.4. Worked Grounding Examples

The abstraction above becomes concrete in familiar identification settings. Consider a discourse neighborhood d_U containing the claims: smoking increases lung-cancer risk; age affects smoking; age affects lung cancer; smoking is correlated with age; and controlling for age still leaves a smoking effect. The formalization functor F maps this neighborhood to a model with variables $X = \text{SMOKING}$, $Y = \text{LUNGCANCER}$, and $Z = \text{AGE}$ together with graph

$$Z \to X \to Y, \qquad Z \to Y.$$

In model space the relevant operation is back-door adjustment:

$$P(Y \mid do(X = x)) = \sum_z P(Y \mid X = x, Z = z) P(Z = z).$$

The explanation functor G then renders this model-level derivation back into canonical discourse: age is a confounder, the back-door path is blocked by conditioning on age, and the causal effect is therefore identifiable under the stated assumptions. The point is not that the original text contained the formula verbatim; it is that the discourse neighborhood carries enough structured content for a formal model to make the identification criterion explicit.

 doi:10.20944/preprints202604.1011.v1

Front-door identification gives an even clearer local-to-global picture. Suppose a discourse neighborhood asserts that smoking causes tar deposits, tar causes lung cancer, smoking has no direct effect on lung cancer except through tar, and there are unobserved confounders between smoking and lung cancer. Formalization produces variables $X = \text{SMOKING}$, $Z = \text{TAR}$, and $Y = \text{LUNGCANCER}$ with graph $X \to Z \to Y$ and latent confounding between X and Y. Back-door adjustment fails, but front-door transport succeeds through local sections:

$$P(z \mid do(x)) = P(z \mid x)$$

and

$$P(y \mid do(z)) = \sum_{x'} P(y \mid z, x')P(x').$$

Gluing these yields

$$P(y \mid do(x)) = \sum_{z} P(z \mid x) \sum_{x'} P(y \mid z, x')P(x').$$

This is naturally sheaf-like: each conditional distribution is a local section over a variable patch, and identification is a gluing operation producing a global interventional section from compatible local ones. In DEMOCRITUS, regime gluing is a discourse-level analogue of this logic. Not every family of local claims descends to a global statement, and obstruction is diagnostically meaningful. We defer a more detailed treatment of the extension of do-calculus to sheaves in a topos to our previous work [18,20].

4.5. Simplicial Diagnostics for Localized Claim Families

Homotopy localization explains when surface forms should be identified. In the implemented system, the next question is not the full homology of a paraphrase complex, but whether a localized family actually closes up under the multiway comparisons that DEMOCRITUS performs. For a normalized claim class $[c]$, let $K_{[c]}$ be the nerve of its paraphrase family: vertices are surface realizations, edges record pairwise compatibility after normalization, and higher simplices record coherent multiway agreement.

The diagnostics that matter operationally are therefore simple and auditable. If the family is well connected and its candidate horns are mostly filled, DEMOCRITUS reports a *coherent* localized class. If some local compatibilities exist but many horn fillings fail, the system reports a *partially glued* family, meaning that nearby formulations align only locally or only under some regime interpretation. If the family breaks into separate paraphrase islands, it is reported as *disconnected*, which in practice often signals clause-tail segmentation, latent mechanism splitting, or a genuine shift in discourse regime.

This viewpoint clarifies why the present implementation reports coherent, partially glued, and disconnected families rather than only a single similarity score. The current simplicial diagnostics are intentionally lightweight: they ask not merely whether two claims are close, but whether an entire family closes up coherently under the multiway comparisons the system actually constructs. That is enough to expose the practical failure modes that recur in the experiments without claiming that DEMOCRITUS is already computing a full topological invariant.

5. Categorical Machine-Learning Machinery Behind DEMOCRITUS

The previous sections explained why localization, gluing, and obstruction are the right conceptual language for causal discourse extracted from text. A working system, however, still needs concrete machinery that can build local structured states, organize them into neighborhoods, and aggregate them into a bounded global artifact. In the present implementation, that role is played by three recurring components from the broader *Categories for AGI* program and companion textbook [3]: Diagrammatic Backpropagation (DB), Geometric Transformers (GT), and Kan Extension Transformers (KET). These are not decorative names for standard modules. They correspond to three distinct

 doi:10.20944/preprints202604.1011.v1

computational jobs inside DEMOCRITUS: structured consistency checking, geometric organization of extracted claims, and colimit-style aggregation of many local partial views.

Diagrammatic Backpropagation (DB). DB is the part of the stack that treats learning and scoring as operations on diagrams rather than on a single flat prediction map. The basic idea is that local states are related by morphisms, and error or inconsistency should be measured by whether different paths through the diagram agree. In the current DEMOCRITUS implementation this appears as a consistency or gluing energy: a glued manifold state is fed into a DB-style consistency square whose loss measures a *gluing obstruction*. Operationally, this matters because DEMOCRITUS does not trust a single extracted causal graph. It generates many local causal models, compares them, and keeps track of where domain-wise and relation-wise projections fail to commute. The local-causal-model galleries, model-sweep scores, and obstruction diagnostics are therefore not ad hoc visual extras; they are the DB layer made inspectable.

Geometric Transformers (GT). GT is the part of the stack that turns extracted relational structure into a geometry. After causal statements are converted into triples, the system builds relation-aware states whose neighborhoods reflect semantic and mechanistic proximity rather than only token overlap. In the DEMOCRITUS code path, GT combines attention with an explicit geometric mixer: extracted triples are compiled into adjacency, relation, and domain features and then refined into a causal manifold with two- and three-dimensional views. This geometric organization is scientifically important for the present paper because localization and gluing are local notions: one needs a meaningful idea of neighborhood, overlap, basin, and drift before one can ask whether nearby claims belong to the same paraphrase class or whether a corpus breaks into disconnected topical patches. Topic partitions, manifold views, nearest-neighbor expansions, and local model sampling all depend on this GT layer.

Kan Extension Transformers (KET). KET is the aggregation layer. A Kan-extension viewpoint treats attention and synthesis not merely as weighted averaging, but as a structured way of assembling many local observations into a shared target object. In CLIFF and DEMOCRITUS, this appears directly at the workflow level. Different agents, routes, and document-local analyses produce partial artifacts, and the system must combine them into a single bounded dashboard, corpus summary, or candidate backbone claim set without erasing provenance. The CLIFF workflow code makes this explicit by representing attention contexts as colimit-style constructions and consistency states as limit-style gluing objects. For DEMOCRITUS, KET is therefore what turns many document-local causal fragments into corpus-level synthesis artifacts that can still be traced back to their local sources.

Taken together, DB, GT, and KET provide a division of labor that is central to how DEMOCRITUS works. DB supplies the criterion for whether locally constructed states are mutually coherent; GT supplies the geometry in which local overlaps and regime neighborhoods can be identified; KET supplies the aggregation principle that promotes those local pieces into a user-facing synthesis. The homotopy-localization layer studied in this paper sits on top of that stack. It uses GT-organized neighborhoods to decide where comparison should occur, DB-style consistency signals to diagnose failed gluing, and KET-style aggregation to turn localized classes into inspectable corpus-level artifacts. Readers who want fuller derivations and executable demos can consult Appendix D, but the essential scientific point is that DEMOCRITUS is powered by an explicitly categorical learning stack rather than by a single opaque summarization model.

6. System Realization in DEMOCRITUS and CLIFF

DEMOCRITUS realizes this perspective through a multi-stage document-analysis pipeline. A query first induces a retrieval cover, either from external corpora or a local document set. The system then extracts a root topic frontier, expands local discourse neighborhoods, generates candidate causal statements, constructs large numbers of local causal models, and compiles them into a causal SQL bundle (CSQL) that supports provenance-preserving aggregation.

 doi:10.20944/preprints202604.1011.v1

6.1. CLIFF as the Interface Layer

CLIFF is the public-facing interface layer above this pipeline. The current public release, `CLIFF_CatAgi`, describes it as a conscious interface or conscious layer sitting on top of a deeper Functor Flow engine. Operationally, that means three things. First, CLIFF treats a user query as a routing problem: corpus-level DEMOCRITUS analysis, workflow recovery, company comparison, and related synthesis tasks are exposed as distinct routes rather than collapsed into one monolithic prompt. Second, the heavier routes run as background workflows that gather evidence, normalize it into a shared structure, and synthesize bounded artifacts that can be reopened later. Third, optional external engines remain explicit dependencies rather than hidden assumptions: DEMOCRITUS, BASKET/ROCKET, and related backends can live as sibling repositories or be located through environment variables.

This architectural separation matters for the present paper because the experimental objects are not just raw extracted triples. They are archived CLIFF runs with route decisions, saved dashboards, and provenance-carrying CSQL bundles. In that sense, CLIFF is part of the experimental method. It provides the stable user-facing surface through which the localized claim classes, topic partitions, homotopy diagnostics, and gluing summaries become inspectable and reproducible.

With that machinery in place, recent iterations of the system added four components that are central to the present paper.

1. **Homotopy localization.** The CSQL bundle materializes canonical subject–relation–object classes and localizes paraphrastic claim families directly inside the database.
2. **Regime gluing.** A second family of views preserves canonical regimes and relation families, allowing the system to classify localized claims as multi-regime glued, regime-sensitive, or obstructed.
3. **Topic partitions.** Broad queries often retrieve heterogeneous document sets. Instead of forcing a flat synthesis, CLIFF first decomposes the corpus into local topic covers.
4. **Archived interactive artifacts.** Runs are saved and indexed, so the full set of generated artifacts can be reopened later, reused in papers, and reproduced from the public code base.

These changes matter because they shifted the system from a pure extraction-and-summary pipeline to an experimental environment for studying causal language itself. Broad queries may show fragmented topic covers or many singleton partitions; focused queries may show strong multi-regime gluing; and borderline cases often appear as partially glued or disconnected simplicial families rather than falsely coherent backbones.

7. Experimental Framing

Our experiments are qualitative and systems-oriented rather than benchmark-driven. The aim is not to report a single scalar accuracy for causal extraction, but to study how a working extraction pipeline behaves under paraphrase variation, topic drift, and cross-document aggregation. We therefore focus on saved runs that are especially diagnostic of the central problem.

Three kinds of corpus behavior recur in practice.

1. **Focused corpora.** Narrow domains such as the Mediterranean diet tend to produce coherent local basins, interpretable topic partitions, and stable regime-glued claims.
2. **Broad but recoverable corpora.** Queries such as rising ocean temperatures can still produce useful gluing results once topic filtering and atlas sanitation are improved, even though the underlying domain remains multi-regime.
3. **Overbroad corpora.** Queries such as climate change or economic inflation often decompose into multiple local covers. In these cases the atlas and partition layers are more informative than any single flattened global summary.

The core outputs we track are within-document and cross-document homotopy classes, the distribution of coherent, partially glued, and disconnected localized claim families, the counts of multi-

 doi:10.20944/preprints202604.1011.v1

regime glued, regime-sensitive, and obstructed claims, the topic-partition structure of the retrieved corpus, and the availability of supporting local causal graphs and manifold neighborhoods for the most important claims.

7.1. Mediterranean Diet as a Focused Corpus

The Mediterranean diet query is a good example of a focused corpus in which the localization machinery behaves well. In one ten-document run, the system reported no cross-document homotopy classes but a large number of within-document families, indicating that the studies shared mechanisms and regimes without repeating exactly the same backbone claims. The same run yielded 26 multi-regime glued claims, 4 regime-sensitive claims, and no obstructed claims. The most stable surfaces centered on the effect of Mediterranean diet adherence on gut microbiota composition and downstream microbial gene expression related to immune response and metabolism. The surviving archived CLIFF checkpoint for this case preserves four selected source texts rather than a frozen final ten-document manifest, but those preserved texts still center on Mediterranean-diet microbiota and epigenetic mechanisms; Appendix Table A5 records that provenance explicitly.

Table 1. Homotopy and regime-gluing summary for a focused ten-document Mediterranean diet run. The corpus shows strong local coherence and essentially no polarity-level obstruction.

Mediterranean diet run	**Count**
Cross-document classes	0
Within-document families	978
Coherent classes	958
Partially glued classes	2
Disconnected classes	18
Multi-regime glued claims	26
Regime-sensitive claims	4
Obstructed claims	0

This case is instructive because it shows the desired behavior of the system. The topic cover remains tight, the regime-gluing layer surfaces a biologically meaningful local mechanism, and the remaining disconnected families are mostly clause-tail phenomena: one head claim branches into several downstream biological continuations. The absence of strict cross-document homotopy classes in this run should not be read as substantive disagreement. It more often means that neighboring papers describe compatible mechanisms at slightly different levels of granularity, so agreement persists locally and regime-wise without collapsing to a single repeated canonical sentence backbone across documents.

7.2. Red-Wine Study Collation as Topos-Style Gluing

To describe the gluing over causal claims from multiple documents using topos-theoretic constructions, we discuss two PubMed studies with overlapping but non-identical discourse about red wine and resveratrol: *Red Wine Consumption and Cardiovascular Health*, a cardiovascular-health article emphasizing endothelial function and cardioprotective pathways, and *Resveratrol: A Double-Edged Sword in Health Benefits*, a review emphasizing anti-inflammatory, endothelial, and mitochondrial mechanisms. These are the exact archived source papers paired in the saved study-collation run listed in Appendix Table A5. DEMOCRITUS first compiled each document into its own local causal atlas, yielding atlas A with 300 aggregated edges and atlas B with 268. The important point is that the two atlases do not share a flat ontology or identical sentence-level claims. They must be compared through a canonicalized interface of normalized subject–relation–object surfaces.

In this setting, the soft pullback plays the role of a match object over that interface. Instead of requiring literal equality between atlas edges, it searches for high-quality alignments after the normalization and localization steps described earlier. For this pair of studies the pullback recovered two cross-study matches: a strong consensus around *resveratrol* $\rightarrow$ *mitochondrial biogenesis* $\rightarrow$ *enhanced*

 doi:10.20944/preprints202604.1011.v1

energy metabolism, and a weaker but still informative consensus around *resveratrol* → *endothelial function* → *nitric oxide / vascular relaxation*. These are precisely the kinds of cases where a purely lexical merge would either miss the overlap or flatten the distinction between a broad cardiovascular claim and a more specific mechanistic refinement.

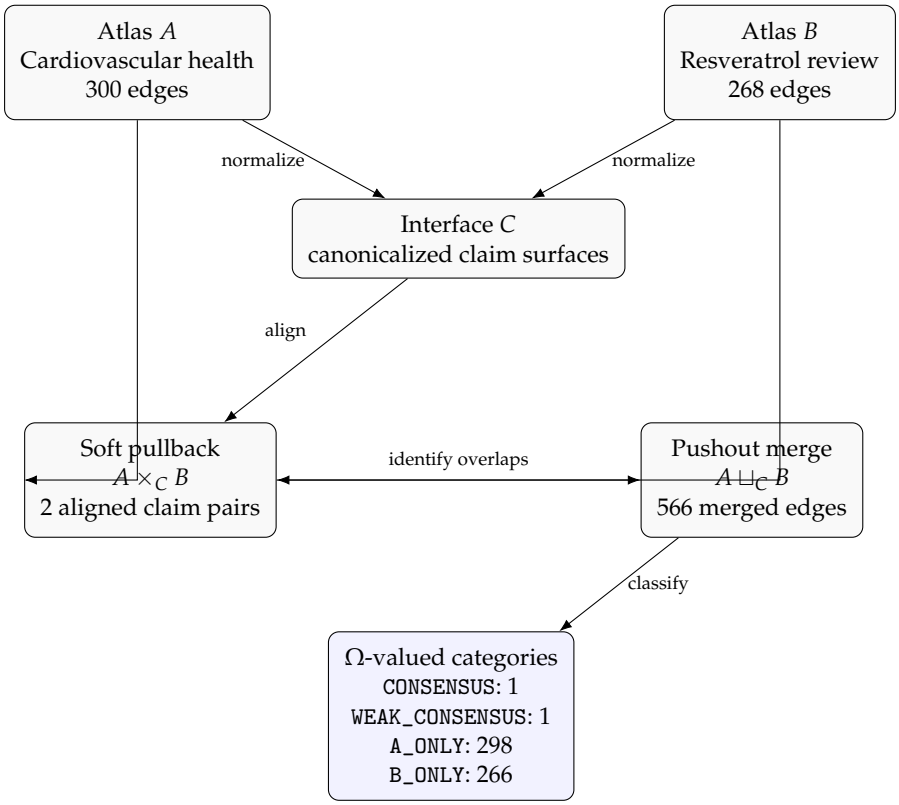

Figure 4. Topos-style study collation for the red-wine corpus. Local causal atlases are compared through a canonicalized interface, softly glued by pullback-style alignment, merged by pushout-style identification, and finally classified into logic-valued claim categories.

The complementary operation is pushout-style merging. Once the two overlap claims are identified, the reconciled atlas is formed as the union of the two study atlases modulo those identifications. In the saved run this yielded 566 merged edges, exactly $300 + 268 - 2$, so the pushout preserves almost all study-specific structure while still creating a shared object for the two aligned mechanisms. This is important experimentally because it shows that gluing is not a lossy consensus summary. Most claims remain local to one study and are not forced into agreement just because the documents inhabit the same biomedical neighborhood.

This experiment also gives a concrete interpretation of the topos language. The pushout atlas carries a simple truth-value assignment $\chi : \text{Claims} \to \Omega$, where Ω is the four-valued space {`CONSENSUS`, `WEAK_CONSENSUS`, `A_ONLY`, `B_ONLY`}. In computational terms, χ is a provenance-aware classifier on merged claims. In categorical terms, it is a small subobject classifier that lets us isolate the consensus subobject while preserving unmatched claims as local sections that failed to glue. That is exactly the kind of reasoning we want from DEMOCRITUS: identify what genuinely descends across documents, retain what stays study-specific, and expose both outcomes in auditable form.

 doi:10.20944/preprints202604.1011.v1

Table 2. Summary of the red-wine study-collation experiment. The four-valued classifier Ω separates the merged atlas into consensus and study-specific regions rather than collapsing everything into a single undifferentiated backbone.

Object / category	Count	Interpretation
Edges in atlas A	300	Cardiovascular red-wine discourse
Edges in atlas B	268	Resveratrol-review discourse
Soft pullback matches	2	Cross-study aligned mechanisms
Pushout edges	566	Union modulo matched overlaps
`CONSENSUS`	1	High-similarity glued claim
`WEAK_CONSENSUS`	1	Lower-similarity but auditable glued claim
`A_ONLY`	298	Study-specific claim retained from atlas A
`B_ONLY`	266	Study-specific claim retained from atlas B

7.3. A Single-Document PDF Experiment: Emperor Penguins

A complementary focused experiment comes from a single Washington Post PDF on emperor penguins, Antarctic sea ice, and warming ocean currents, archived from Sarah Kaplan's 9 April 2026 article "Emperor penguins have just been declared endangered". Because the retrieved corpus contains only one document, there is no cross-document agreement to measure; the diagnostic question becomes whether one article still supports a stable internal causal neighborhood rather than a flat summary. In this archived run, DEMOCRITUS extracted 330 claims over 18 domains and compiled them into 318 unique aggregated edges, with 10 edges receiving repeated within-document support. The dominant repeated backbone linked rising ocean temperatures to krill moving to deeper waters and thereby reducing prey availability for emperor penguins and Antarctic fur seals; a second recurring chain linked global warming to Antarctic sea-ice loss, breeding-habitat degradation, and increased chick mortality. The archive retains the single acquired PDF identity for this case study, and Appendix Table A5 records that provenance.

Table 3. Single-document summary for the archived Washington Post emperor-penguin PDF experiment. Even without cross-document evidence, the run exhibits repeated internal support for a small number of central causal chains.

Emperor penguin PDF run	Count
Documents	1
Extracted claims	330
Domains	18
Unique aggregated edges	318
Repeated within-document edges	10
Maximum support for one edge	4

This kind of run is useful because it separates two different notions of stability. Cross-document homotopy cannot appear when there is only one source, but within-document paraphrase collapse still matters: the system can show whether an article repeatedly returns to the same mechanism under slightly different formulations or instead fragments into unrelated local claims. In this case the repeated support surfaces are ecologically coherent, which makes the run a clean demonstration that the homotopy-localization layer is already informative before any multi-document gluing is attempted.

7.4. Rising Ocean Temperatures as a Multi-Regime Corpus

An archived run launched from the query for five recent studies on rising ocean temperatures shows a complementary failure mode: the corpus is scientifically recognizable as a single domain, but at the homotopy-localization layer it still behaves mostly as a collection of local causal patches rather

 doi:10.20944/preprints202604.1011.v1

than as a document set with repeated cross-document backbones. In the preserved archive the retrieval cover ultimately contains six acquired documents, and the run produced 0 cross-document classes and 1570 within-document families, of which 1522 were coherent, 7 partially glued, and 41 disconnected. In other words, most of the stability in this corpus lives inside individual articles rather than across them. That is already a useful diagnosis. It tells us that the retrieval cover is semantically related, but the extracted mechanisms are often expressed at different granularities, under different regimes, or in different scientific subdomains. Appendix Table A5 makes the cover drift visible: alongside ocean-warming studies, the saved run preserved neighboring sources on cyclone intensification, heat-related mortality, mantle melting, and phytoplankton adaptation.

The disconnected and partially glued families make this visible in an unusually concrete way. One of the strongest local patches concerns the North Atlantic mechanism *weakening of the subpolar gyre* $\rightarrow$ *increased inflow of warm saline subtropical waters into the North Atlantic*. Although this family is confined to a single study, it has 25 vertices, 164 edges, 821 filled triangles, and a horn-fill ratio of 0.87, so its local paraphrase complex is extremely dense. By contrast, the freshwater-stratification family *freshwater influx from rivers and rainfall* $\rightarrow$ *increased ocean stratification* appears only partially glued, with 12 vertices, 19 edges, and a horn-fill ratio of 0.258, indicating wording drift and context-sensitive elaboration. Other disconnected families wander further afield into heat-mortality adaptation discourse and mixotrophic-phytoplankton ecology, showing that even a six-document archived cover can mix climate, oceanography, and ecological-mechanism subcovers.

This is the kind of output we regard as especially valuable. A standard summarizer would compress the corpus into a narrative about warming oceans. The homotopy view instead shows that the corpus contains several locally stable causal neighborhoods without yielding many exact cross-document identifications. That distinction matters scientifically: some mechanisms are robust but article-specific, some are only partially glued because the clause tails drift across regimes, and some apparent neighbors are really signals that the retrieved cover has broadened into adjacent subfields. The archived LCM in Figure 5 makes one of those dense local neighborhoods visually inspectable at the document level, while the relational manifold in Figure 6 shows that the same article already decomposes into several coherent causal basins before any corpus-level synthesis is attempted.

Table 4. Homotopy-localization summary for the archived rising-ocean-temperatures run. The original query asked for five studies, but the preserved retrieved cover contains six acquired documents. The absence of cross-document classes together with the large number of coherent within-document families shows that the corpus forms a related cover without collapsing to a single repeated global backbone.

Rising ocean temperatures run	**Count**
Documents	6
Cross-document classes	0
Within-document families	1570
Coherent classes	1522
Partially glued classes	7
Disconnected classes	41

 doi:10.20944/preprints202604.1011.v1

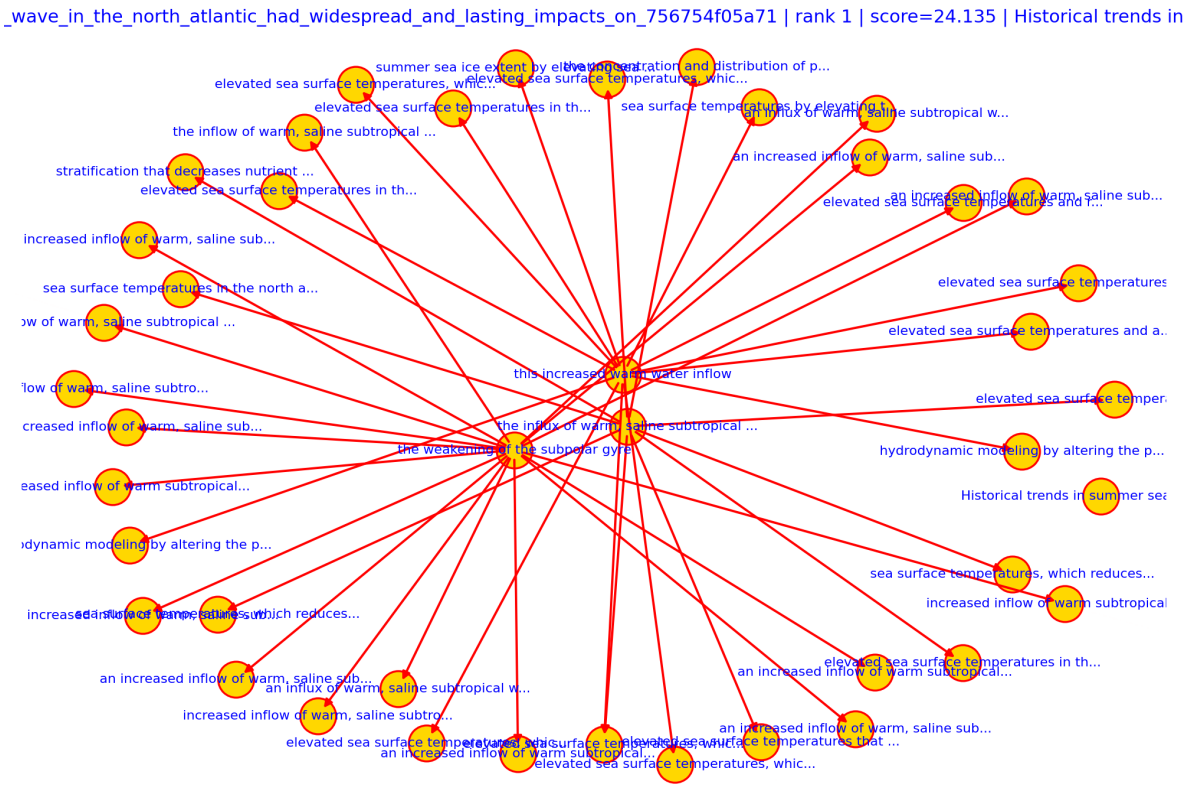

Figure 5. Representative archived local causal model from the rising-ocean-temperatures corpus. The central backbone links weakening of the subpolar gyre to increased warm-water inflow into the North Atlantic, illustrating a locally coherent mechanism that remains confined to one study in the homotopy-localization analysis.

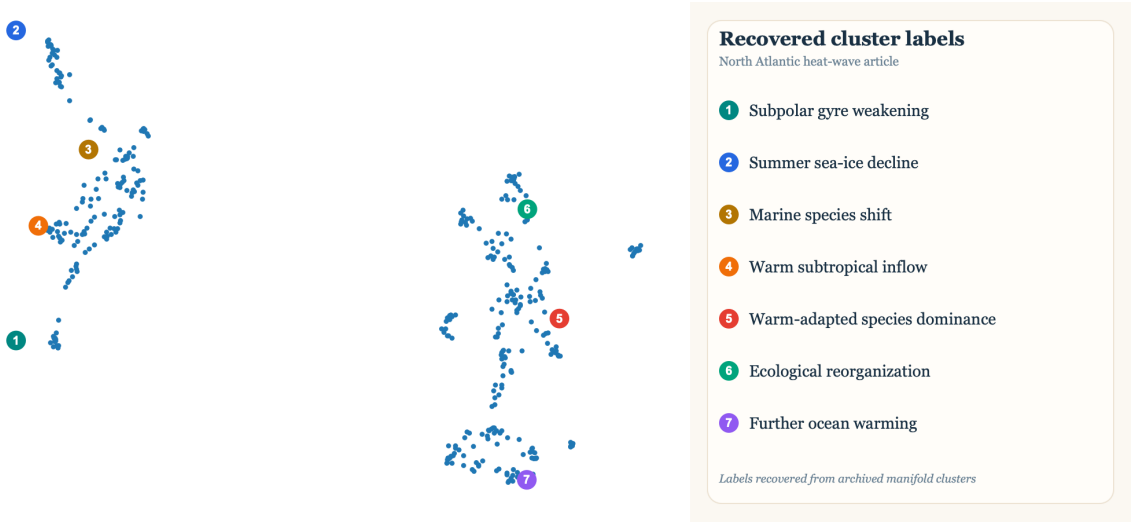

Figure 6. Archived two-dimensional relational manifold for the North Atlantic heat-wave article from the same ocean-temperature run, with numbered cluster labels recovered from the saved manifold metadata. The largest neighborhoods correspond to subpolar-gyre weakening, summer sea-ice decline, warm-adapted species shifts, and ecological reorganization, making the dominant local mechanism basins visible even before any cross-document gluing succeeds.

7.5. Broad Corpora and Topic Partitions

The broad-climate experiments revealed another lesson. When the base query is too broad, the homotopy and regime layers should not be read before the cover itself is understood. Early runs often wandered into narrow basins such as malaria, while later retrieval and topic-atlas improvements

 doi:10.20944/preprints202604.1011.v1

yielded broader covers across climate-health, water systems, adaptation policy, coral bleaching, and ocean warming. Even so, many broad corpora remained collections of local patches rather than a single cleanly glued object. This is why topic partitioning became a first-class diagnostic.

Singleton topic partitions are especially informative in this setting. They do not automatically signal extraction failure; often they mark documents that enter the retrieved cover through a shared high-level query but then instantiate a mechanism, regime, or discourse frame that no other document in the run matches closely enough to glue. Read together with the absence of strict cross-document homotopy classes in otherwise coherent corpora, these singleton partitions show why cover structure must be interpreted before global agreement claims are made: local causal neighborhoods can be internally stable and scientifically relevant even when the corpus as a whole never forms one perfectly repeated cross-document backbone.

8. Related Work

This paper sits at the intersection of several literatures: classical causal inference, causal relation extraction from text, causal knowledge-graph construction, and recent attempts to use large language models for causal reasoning. Our position relative to each is slightly different, and it is useful to state that explicitly.

Relative to classical causal inference, our work is complementary rather than competitive. Modern causal inference studies identifiability, adjustment, intervention, equivalence classes of causal models, and structure learning from observational or experimental data. [8–14] We do not attempt to replace those methods with language alone. Instead, we study the prior layer at which causal discourse is produced, compared, and consumed. In this sense, DEMOCRITUS is closer to a causal hypothesis compiler or diagnostic front end than to a structure-learning algorithm over data.

Relative to causal relation extraction and scientific information extraction, our setting is broader and more artifact-centered. Existing work on causal relation extraction, scientific IE, and claim verification usually targets sentence-level or passage-level extraction accuracy. [21–23] Those systems are valuable, but they do not usually address what happens after extraction when thousands of overlapping causal mentions must be consolidated, compared across documents, and audited for disagreement or regime dependence. The homotopy-localization layer in DEMOCRITUS is aimed precisely at that post-extraction instability.

Relative to corpus-scale causal claim resources, the closest comparison is the Testing Causal Claims (TCC) effort, which aggregates causal claims from a large economics corpus. [24] TCC emphasizes breadth, method metadata, and large-scale indexing of causal statements across a discipline. DEMOCRITUS emphasizes something different: local causal modeling within documents, explicit paraphrase collapse, regime-gluing diagnostics, and interactive artifact preservation. The two approaches are complementary. A resource like TCC offers a broad compiled layer of causal claims, whereas DEMOCRITUS offers a way to study how such claims cohere, fragment, or fail to glue inside narrower local covers. We also built an older specialized CLIFF variant with a dedicated `tcc_atlas` route that queries the TCC atlas/CSQL store directly rather than retrieving documents on the fly; although that branch predates the current public CLIFF, it demonstrates that the same interface layer can scale to much larger curated claim repositories.

Relative to knowledge graphs and GraphRAG, our system again targets a different object. Knowledge graphs and graph-augmented retrieval systems are effective at representing and retrieving asserted relationships. [25,26] But they are not usually designed to distinguish paraphrastic duplication from genuine cross-regime conflict, or to preserve a notion of support grounded in many local causal models. In DEMOCRITUS, the graph is not just a storage substrate. It is part of an experimental object that includes equivalence classes, support provenance, and explicit gluing diagnostics.

Finally, there is a fast-growing literature on whether large language models can perform causal reasoning or assist in causal discovery. Some of that work probes causal competence directly; some uses language models to propose structures or explanations. Our stance is intentionally more conservative.

 doi:10.20944/preprints202604.1011.v1

We use language models as discourse compilers and local hypothesis generators, not as authoritative causal reasoners. The emphasis is on what can be stabilized *after* generation through normalization, localization, scoring, and aggregation. In that sense, the current paper is less a claim about the innate causal intelligence of LLMs than a claim about how to build a better causal-language analysis environment around them.

9. Discussion and Limitations

Several lessons emerge from these experiments. First, paraphrase handling cannot be treated as a cosmetic post-processing step. It changes which claims accumulate support, which mechanisms appear recurrent, and whether a corpus seems coherent at all. Second, gluing failures are often scientifically useful. A regime-sensitive or obstructed claim is not simply an error; it may indicate that one mechanism changes sign across contexts, that different studies use relation families at different levels of strength, or that the apparent agreement is only local. Third, retrieval and synthesis must be studied together. Many apparent failures of causal extraction begin one layer earlier, as failures of the retrieved cover. The red-wine collation experiment adds a fourth lesson: the topos vocabulary is not merely decorative. Pullback-style alignment, pushout-style merging, and logic-valued claim categories correspond to concrete database operations that preserve provenance while separating true cross-study consensus from study-specific hypotheses.

The present work also has clear limitations. We do not claim that the extracted claims are ground-truth causal facts, only that they are structured and diagnostically useful summaries of causal discourse. The current homotopy layer is still coarse: many disconnected families are really head-tail clause-segmentation problems, and the simplicial diagnostics are lightweight rather than fully model-categorical. The experiments are also systems-driven rather than benchmark-centered.

9.1. From Static Causal Snapshots to Temporal Diffusion and Workflow Repair

The present paper studies DEMOCRITUS mainly as a system for building a *localized static snapshot* of causal discourse over a retrieved document cover. That emphasis is deliberate: before one can reason across time, one must first know how to normalize paraphrases, compare local models, and decide when nearby claims genuinely glue. But the same categorical machinery extends naturally to a temporally indexed setting. In the broader *Categories for AGI* program, the basic object is not just a one-shot corpus summary. It is a family of local structured states indexed by time, sector, or regime, together with restriction maps that say how those local views overlap and how they should be compared. [3] Under that interpretation, the current DEMOCRITUS pipeline produces the per-document or per-year states from which a larger temporal object can be assembled.

Annual 10-K filings provide a concrete example. A single filing year can be treated as a local snapshot of how a company describes its mechanisms, risks, operating plans, and value-capture pathways at that moment. Running DEMOCRITUS over each year yields a structured state for that company-year: extracted relations, local neighborhoods, localized claim families, and regime-sensitive gluing diagnostics. Once those yearly states are indexed over time, one can ask a richer question than the present paper asks: not merely whether a claim is stable inside one retrieved cover, but whether a company admits a *glued temporal snapshot* whose local yearly views assemble into a coherent trajectory. In that setting, the same DB/GT/KET stack takes on a temporal interpretation. GT organizes nearby company-year states into a manifold or diffusion geometry; DB measures where successive years commute, drift, or develop genuine obstructions; and KET aggregates the resulting local views into a bounded temporal synthesis artifact, such as a diffusion model of how a firm's operational and causal language evolves across filings.

This is the bridge to BASKET and ROCKET. Those systems should not be read as replacing DEMOCRITUS, but as extending it to a different target object built from the same underlying idea. BASKET starts from structured yearly artifacts of the kind that DEMOCRITUS can produce, but shifts the semantic target from causal relational structure to operational plans and workflows. The yearly local state is no longer only a set of causal claims; it becomes a company-year plan object whose

 doi:10.20944/preprints202604.1011.v1

recurring actions, edges, and motifs can be pooled into a temporal plan manifold. ROCKET then adds a repair and reranking layer on top of that object: partial workflows extracted from filings are completed, denoised, or reweighted using structural compatibility, cross-company motif support, and delayed alignment with downstream financial outcomes. Conceptually, this is the same scientific move as in the present paper, but lifted from static causal discourse to temporally evolving workflow structure. For that reason, BASKET/ROCKET is best understood as a descendant architecture built on top of DEMOCRITUS's localization-and-gluing machinery rather than as a separate unrelated project. The present article keeps its focus on the causal case, but the temporal 10-K diffusion setting shows how the same categorical framework scales from local causal snapshots to longitudinal company models and, eventually, to workflow repair over those models.

10. Conclusions

The main challenge in causal extraction from language is not merely finding candidate triples. It is deciding when different textual realizations should be identified, when they should remain distinct, and how local agreements should glue across documents and regimes. The recent evolution of DEMOCRITUS makes that challenge visible in a concrete implemented system. Homotopy localization reduces paraphrase inflation; regime gluing distinguishes clean agreement from descent failure; pullback and pushout views make cross-document reconciliation auditable; topic partitions reveal whether a corpus has even formed a sensible cover; and archived artifacts preserve the resulting evidence for later inspection and reproduction.

Appendix A. System and Database Appendix

This paper is centered on causal extraction from language rather than on the full CSQL database theory, but the database layer remains important because it is where localized claim classes, regime summaries, and provenance-preserving aggregation are materialized. In the current implementation, each document run produces a collection of local causal models, their scores, rendered graph views, and a CSQL bundle that stores canonical claims, localized claim families, and support tables linking aggregate claims back to individual local models and source documents.

Appendix B. Reproducing the Results with the GitHub CLIFF Implementation

The experiments described in this paper can be reproduced from the current public CLIFF implementation, which is organized in the companion GitHub repository `CLIFF_CatAgi`. That repository provides the natural-language interface, route-selection logic, dependency-resolution layer, and saved-artifact surface used to launch DEMOCRITUS runs from ordinary text queries. In the terminology of the public release, CLIFF is the *Consciousness Interface to Functor Flow*: a user-facing layer that routes a request into deeper workflows and then promotes the resulting synthesis artifact back into a bounded dashboard or report.

The lightweight installation path is:

```
cd \textsc{Cliff}_CatAgi
python3 -m venv .venv
source .venv/bin/activate
pip install -r requirements.txt
pip install -e .
```

For the DEMOCRITUS route used in this paper, one also needs access to the companion DEMOCRITUS repository and its model/runtime dependencies. The current CLIFF resolver looks for such dependencies either under `third_party/` or as sibling repositories in a shared workspace, and also supports environment-variable overrides such as `CLIFF_DEMOCRITUS_ROOT`, `CLIFF_COURSE_REPO_ROOT`, and `CLIFF_BOOK_PDF_PATH`. A typical workspace layout is therefore:

```
workspace/
```

 doi:10.20944/preprints202604.1011.v1


```
  \textsc{Cliff}_CatAgi/
  Democritus_OpenAI/
  Category-Theory-for-AGI-UMass-CMPSCI-692CT/
```

Once the environment is available, a one-shot corpus run can be launched directly from the CLIFF entrypoint. For example, a five-document ocean-temperature synthesis can be invoked as:

```
python3 -m functorflow_v3.\textsc{Cliff} \
  --query "Analyze 5 recent studies on rising ocean temperatures" \
  --route democritus \
  --execution-mode deep \
  --outdir /tmp/\textsc{Cliff}-ocean \
  --democritus-target-docs 5
```

If a fixed local corpus is desired instead of live retrieval, CLIFF also exposes manifest- and file-based inputs. In particular, `-democritus-manifest`, `-democritus-input-pdf`, and `-democritus-input-pdf-dir` are often preferable for paper reproduction because they freeze the cover explicitly. The same entrypoint can therefore be used either for retrieval-driven runs or for locked manifest-based reruns of previously selected documents.

An older companion branch, `TCC-enabled CLIFF`, also included a specialized `tcc_atlas` route for querying the Testing Causal Claims economics atlas directly through its curated atlas/CSQL representation, without live document retrieval. That branch is not the main implementation analyzed in this paper, but it is a useful proof-of-concept that the CLIFF interface can sit on top of substantially larger precompiled claim repositories as well as the document-level DEMOCRITUS workflows emphasized here.

The output layout is designed to preserve the exact artifacts discussed in the paper. Each routed CLIFF session writes a route-specific directory under the chosen `outdir`; for the present experiments, the relevant subtree is `outdir/democritus/`. That directory contains a run summary, the archived DEMOCRITUS batch directory, and the generated corpus-synthesis dashboard. In the current implementation, the synthesized HTML artifact is written under the `democritus_runs/corpus_synthesis/` subtree, while the route summary records the paths to the corresponding CSQL bundle and batch outputs. Those files are precisely the bridge between the theoretical constructions described in the body of the paper and the reproducible computational objects used in the experiments.

For practical verification, the public repository already includes regression tests for the CLIFF router and major DEMOCRITUS-facing surfaces. A minimal smoke-test pass is:

```
python3 -m unittest tests.test_query_router_agentic tests.test_\textsc{Cliff}
```

and broader local verification can be obtained with `python3 -m unittest`. In this sense, the manuscript's experimental layer is not only conceptually reproducible but also tied to an executable public interface that exposes routing, artifact production, and dependency resolution in a form that other researchers can inspect directly.

Appendix C. Archived Source Covers and Manifold Diagnostics

To keep the experiments auditable, Table A5 records the source texts preserved in the surviving saved runs used for the main case studies. The red-wine and ocean runs retain exact document identities. The Mediterranean archive survives only as an interactive checkpoint with four selected texts out of twelve discovered candidates rather than as a frozen final ten-document manifest, while the emperor-penguin archive preserves the single acquired PDF identity but not richer publisher metadata.

 doi:10.20944/preprints202604.1011.v1

Table A5. Source-document provenance recovered from the surviving archived CLIFF/DEMOCRITUS runs used in the paper.

Case study	Archived source texts	Archive note
Mediterranean diet	*Multi-Omic Insights Into Mediterranean Diet-Associated Microbiota*; *Epigenetics of Mediterranean Diet: Altering Disease Risk*; *Cancer and the Mediterranean Diet*; *Mediterranean Diet and Longevity*.	Interactive checkpoint preserved four selected texts from twelve discovered candidates.
Red wine	*Red Wine Consumption and Cardiovascular Health*; *Resveratrol: A Double-Edged Sword in Health Benefits*.	Exact paired inputs used for the pullback/pushout collation experiment.
Emperor penguins	Archived Washington Post emperor-penguin PDF from "Emperor penguins have just been declared endangered".	Single-document run preserves the PDF identity but not full article metadata.
Rising ocean temperatures	*Pre-requisite conditions for the intensification of pre-monsoon cyclones over the Bay of Bengal*; *A standardized indicator reveals sharper increases in heat related mortality in temperate zone cities worldwide*; *A universal concept for melting in mantle upwellings*; *Major heat wave in the North Atlantic had widespread and lasting impacts on marine life*; *Modeling the metabolic evolution of mixotrophic phytoplankton in response to rising ocean surface temperatures*; *Rising temperatures in the subtropical North Atlantic Ocean over the past 35 years*.	Query asked for five studies, but the preserved cover contains six acquired documents spanning neighboring subfields.

Appendix D. Additional Categorical Background

The homotopy construction used here sits inside a larger categorical program developed elsewhere in the book and in earlier papers. The body of the paper now sketches the discourse–model adjunction, motif density, and simplicial diagnostics, but many related directions remain outside the present scope: Markov-category semantics, richer topos-internal intervention logics, and full model-categorical treatments of simplicial assemblies of paraphrase families. For the present paper, the essential computational construction is still the localization of causal mentions at weak equivalences induced by normalization.

Readers who want the fuller story behind the DB/GT/KET stack should consult the companion textbook *Categories for AGI* [3]. The latest copy can be downloaded from https://people.cs.umass.edu/~mahadeva/papers/catagi.pdf. The companion course repository contains the most directly relevant notebook sequence: Week 3 on DB as a diagrammatic consistency or colimit-style energy, Week 4 on Geometric Transformers and causal regimes, and Week 5 on Mini-DEMOCRITUS and Kan Extension Transformers. The public interface implementation lives in `CLIFF_CatAgi`: within `functorflow_v3/`, modules such as `textbook_backstop.py` and `democritus_corpus_synthesis.py` expose the textbook bridge, route orchestration, and corpus-synthesis layer that connect those ideas to the present system. The aim of the present paper is therefore complementary to those sources: here we focus on the homotopy-localization and causal-gluing behavior of the system, while the textbook and repositories document the fuller learning machinery and its code.

Appendix E. Additional Experimental Directions

The saved runs also suggest several directions that are not yet fully developed in the current manuscript. Broad queries such as climate change and economic inflation show that cover selection and topic partitioning are prior problems for any downstream gluing analysis. Single-document runs show that many remaining disconnected families are really clause-tail segmentation issues. Large archived corpora also open the possibility of comparative studies across versions of DEMOCRITUS, using the same saved runs to measure how topic drift, homotopy localization, and regime gluing changed over time.

 doi:10.20944/preprints202604.1011.v1

References

1. Mahadevan, S. Large Causal Models from Large Language Models, 2025, [arXiv:cs.AI/2512.07796].
2. Solanki, H.; Jain, V.; Thirumalai, K.; Rajagopalan, B.; Mishra, V. River drought forcing of the Harappan metamorphosis. *Nature Communications Earth and Environment* **2025**. https://doi.org/https://doi.org/10.1038/s43247-025-02901-1.
3. Mahadevan, S. Categories for AGI. https://people.cs.umass.edu/~mahadeva/papers/catagi.pdf, 2026. Textbook draft.
4. Mahadevan, S. CSQL: Mapping Documents into Causal Databases, 2026, [arXiv:cs.DB/2601.08109].
5. Pearl, J. *Causality: Models, Reasoning and Inference*, 2nd ed.; Cambridge University Press: USA, 2009.
6. Spirtes, P.; Glymour, C.; Scheines, R. *Causation, Prediction, and Search, Second Edition*; Adaptive computation and machine learning, MIT Press, 2000.
7. Imbens, G.W.; Rubin, D.B. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*; Cambridge University Press: USA, 2015.
8. Pearl, J. *Causality: Models, Reasoning and Inference*, 2nd ed.; Cambridge University Press: USA, 2009.
9. Spirtes, P.; Glymour, C.; Scheines, R. *Causation, Prediction, and Search*, 2nd ed.; MIT Press, 2000.
10. Spirtes, P.; Meek, C. Causal Inference and Causal Explanation with Background Knowledge. *UAI* **1995**.
11. Chickering, D.M. Optimal Structure Identification with Greedy Search. *Journal of Machine Learning Research* **2002**, *3*, 507–554.
12. Scutari, M.; Denis, J.B. Bayesian Networks: With Examples in R **2014**.
13. Zheng, X.; Aragam, B.; Ravikumar, P.; Xing, E. DAGs with NO TEARS: Continuous Optimization for Structure Learning. *Advances in Neural Information Processing Systems* **2018**.
14. Brouillard, P.; Lachapelle, S.; Lacoste, A. Differentiable Causal Discovery from Interventional Data. *Advances in Neural Information Processing Systems* **2020**.
15. Mahadevan, S. Higher Algebraic K-Theory of Causality. *Entropy* **2025**, *27*. https://doi.org/10.3390/e27050531.
16. Lambek, J. *The Mathematics of Sentence Structure*; Vol. 65, 1958; pp. 154–170.
17. Asudeh, A.; Giorgolo, G. *Enriched Meanings: Natural Language Semantics with Category Theory*; Oxford University Press, 2020.
18. Mahadevan, S. Intuitionistic *j*-Do-Calculus in Topos Causal Models, 2025, [arXiv:cs.LO/2510.17944].
19. Mahadevan, S. Universal Causal Inference in a Topos. In Proceedings of the NeurIPS 2025, 2025.
20. Mahadevan, S. Decentralized Causal Discovery using Judo Calculus, 2025, [arXiv:cs.AI/2510.23942].
21. Hashimoto, K.; Inui, K. End-to-End Neural Causal Relation Extraction. *ACL* **2016**.
22. Luan, Y.; He, L.; Ostendorf, M.; Hajishirzi, H. Multi-Task Identification of Entities, Relations, and Coreference for Scientific Knowledge Graph Construction. *EMNLP* **2018**.
23. Thorne, J.; Vlachos, A. Automated Fact Checking: Task Formulations, Methods and Future Directions. *COLING* **2018**.
24. Garg, P.; Fetzer, T. Testing Causal Claims in Economics. *arXiv preprint arXiv:2501.06873* **2025**. Dataset and analysis of causal claims extracted from economics papers.
25. Hogan, A.; Blomqvist, E.; Cochez, M.; et al. Knowledge Graphs. *ACM Computing Surveys* **2021**, *54*.
26. Lewis, P.; Perez, E.; Piktus, A.; et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems* **2020**.