Scientific article DOI: 10.18287/2223-9537-2026-16-3-397-414

Operational limits of ontological innovation in large language models: evidence from controlled sensory-modality generation

© 2026, J.O. Matarmaa

Ural Federal University named after the First President of Russia B. N. Yeltsin,
Institute of Radio Electronics and Information Technologies, Yekaterinburg, Russia

Abstract

Contemporary large language models generate text that can appear conceptually novel, raising the question of whether outputs reflect genuine ontological innovation—the introduction of concepts that cannot be expressed as recombinations or transformations of training-derived primitives—or sophisticated recombination within a fixed representational manifold. We address this question through a controlled empirical study using an eight-modality multi-modal sensory ontology as the experimental universe. Seven state-of-the-art language models (N = 168 proposals, 24 runs each) were prompted to propose a ninth sensory modality genuinely beyond the defined eight. Outputs were evaluated through four convergent methods: literature traceability with calibrated threshold (τ = 0.40), structural decomposition, embedding-space convex-hull analysis, and continuous novelty scoring. Results show a strict novelty rate of $\hat{p}_{\text{novel}}$ = 0.000 (95% CI: [0.000, 0.000]), a traceability rate of $\hat{p}_{\text{traceable}}$ = 0.982, and 100% convex-hull membership under Sentence-BERT embeddings reduced to three principal components. Structural decomposition revealed three failure modes: *range extension* (20.2%), *hybrid combination* (44.6%), and *functional recombination* (17.9%). Cross-model effects were moderate (η^2 = 0.326), but all models stayed within the same low-novelty regime, supporting a constrained-manifold rather than a model-deficit explanation. We interpret these findings through the *No New Basis Theorem*: gradient-descent learning over fixed vocabularies cannot expand representational dimensionality beyond training data. These results support the view that current neural architectures are bounded within training-derived conceptual closures, with implications for operationalizing and evaluating ontological creativity.

Keywords: *ontological innovation, computational creativity, large language models, sentence embedding, traceability analysis, representational constraints.*

For citation: *Matarmaa JO.* Operational limits of ontological innovation in large language models: evidence from controlled sensory-modality generation. *Ontology of designing*. 2026; 16(3): 397-414. DOI: 10.18287/2223-9537-2026-16-3-397-414.

Conflict of interest: The author declares no conflict of interest.

Introduction

The capacity to generate fundamentally new concepts—not merely recombinations of existing ones – has long been considered a hallmark of human intelligence and a defining test of genuine creativity. As large language models (LLMs) achieve impressive performance on a broadening range of generative tasks, a pressing question has emerged: do these systems exhibit ontological innovation, generating ideas that lie outside the conceptual closure of their training environment, or do they operate as extremely capable recombiners constrained to a fixed representational manifold?

This question has both philosophical and practical stakes. Philosophically, it bears on the computational theory of mind, the nature of creativity, and whether algorithmic systems can in principle exceed the representational resources with which they were trained. Practically, it determines what roles LLMs can occupy in scientific discovery, design, and other domains where genuinely new concepts – not just novel arrangements of existing ones – are required [1].

The challenge is that outputs from modern LLMs frequently appear novel. Proposals that introduce unfamiliar terminology, combine concepts across disciplinary boundaries, or describe mecha-

nisms not explicitly present in any single training document can pass casual inspection as genuinely creative [2, 3]. Evaluating whether such outputs constitute true conceptual innovation requires a controlled experimental setup and rigorous multi-method analysis.

1 Background and related work

The dominant framework for analyzing machine creativity is Margaret Boden's tripartite taxonomy [4], which distinguishes outputs based on their relationship to the conceptual space from which they arise. Combinational creativity produces novel combinations of familiar ideas – poetic metaphors, analogies, conceptual blends. Exploratory creativity systematically searches well-defined conceptual spaces to find previously unvisited regions. Transformational creativity, the most radical category, modifies the conceptual space itself by relaxing or changing its defining constraints.

Research on AI systems has demonstrated substantial competence at combinational and exploratory creativity [5, 6]. Systems such as Copycat, JAPE, and AARON can generate outputs that human evaluators judge as genuinely surprising and apt within their respective domains. Robust transformational creativity, however, remains rare: most documented examples involve superficial parameter modifications rather than fundamental reconceptualization [4].

Wiggins [2] introduced a formal framework for evaluating creativity claims and identified a "framing" problem: systems appear to transform their conceptual space while actually operating within an implicitly broader space defined by human designers. We argue that a fourth category, which we term ontological creativity, is needed to capture the most radical form of conceptual innovation. Transformational creativity modifies operators Ω or relations E within a conceptual space $C = (V, E, \Phi, \Omega)$; ontological creativity introduces new primitive concepts $v' \in V'$ where $V' \not\subset V$ and critically V' cannot be expressed as $\Omega(V)$ for any existing operators. This distinction maps onto Ryle's philosophical notion of category mistakes [7]: ontological creativity involves recognizing that existing categories are insufficient and proposing new ones that cannot be defined by recombining or transforming current primitives. Einstein's reconceptualization of absolute simultaneity as a category error, leading to a revised ontology of relativistic spacetime, exemplifies this form of creativity. Generating a genuinely new sensory modality would be analogous: not an extension or hybridization of existing ones, but a proposal for a sensing type with a physical basis and information content that cannot be expressed as a composition of the existing eight.

The computational theory of mind [8, 9] and its critics [10–12] have long debated whether algorithmic systems face principled limits on what they can represent and understand. Recent work in mechanistic interpretability provides empirical traction on this question by revealing that generation in transformer based LLMs proceeds through retrieval and interpolation among training examples rather than through any inference mechanism that could introduce entirely new primitives [13, 14].

The transformer architecture [15] encodes all knowledge implicitly in weights learned from distributional co-occurrence, which grounds the output space in the training vocabulary of concepts. Gärdenfors's conceptual-spaces formalism [16] provides complementary geometric intuition: a learned concept lives as a region in a quality space and generating a new concept outside all existing regions requires an external basis vector not present in the training data.

Claim (No New Basis Theorem under fixed-basis assumptions). For a neural architecture A operating over representation space R^d with generative process G: $R^d \to R^d$ learned from training data $D = x_1, \ldots, x_N$, all generated outputs $x' = G(z)$ remain within M', that is a continuous deformation of $M = span(x_1, \ldots, x_N)$ satisfying $dim(M') \leq dim(M)$.

Assumptions. (i) fixed representational basis and vocabulary during inference, (ii) gradient-based continuous learning dynamics, and (iii) no external mechanism that appends new primitive basis vectors at inference time.

Argument (informal proof sketch). First, under continuous gradient updates, learned mappings deform regions of the training-induced manifold rather than creating independent coordinate axes. Second, generative decoding composes these learned maps and therefore remains constrained by the same basis. Third, introducing ontological primitives outside the learned span would require explicit basis augmentation (new symbols, new sensors, or external memory with independent representational coordinates), which is excluded by assumptions (i)–(iii).

Consequence. Ontological novelty, defined as basis expansion rather than recombination, cannot be produced by the internal generation mechanism alone under these assumptions.

This theorem explains why even highly capable models that produce surprising combinations remain formally constrained to the conceptual universe their training data defines. Our empirical study provides a controlled test of whether this theoretical constraint manifests behaviorally.

A persistent methodological challenge in computational creativity research is evaluation [2, 3]. Human expert judgments are subjective, difficult to reproduce, and sensitive to framing. Recent studies show that LLMs can achieve human-level performance on divergent-thinking tasks such as the Alternative Uses Test [17], but scoring quality depends strongly on the evaluation metric [18]. Recent 2024–2026 work further confirms metric sensitivity and proposes newer creativity-evaluation protocols for LLMs, including human-comparative, temperature-sensitivity, and model-guided creativity-judging paradigms [19–25]. Our multi-method approach – traceability search, structural decomposition, embedding-space geometry, and continuous scoring—provides convergent automated evidence that can be replicated and extended. Calibrated traceability thresholds replace arbitrary cut-offs with empirically justified operating points.

2 Materials and methods

The experimental universe consists of eight sensory modalities, each defined by a physical basis, an information channel, and a functional role in an agent's interaction with its environment (Table 1). Modalities span biological human senses and extended machine-sensing capabilities.

This ontology is a category space we define for the purposes of controlled measurement; it is a sufficient, well-specified space against which to test whether proposals exceed it, but it is not, and is not claimed to be, a complete model of the conceptual space that any given model acquired during pretraining. Consequently, the experiment measures whether models can propose a modality outside this explicit eight-item list; it does not, by itself, establish whether models can exceed the conceptual resources of their training data more broadly.

Table 1 – Eight-modality sensory ontology used as the experimental universe. Each modality is defined by its physical basis, representational dimensionality, and functional role

Modality	Physical basis	Functional role	Dim.*
Visual	EM spectrum 300–1000 nm (5 nm steps)	Scene and object recognition	140
Auditory	Pressure waves 1 Hz–100 kHz (128-bin FFT)	Sound source localization	128
Tactile	Mechanical vibrations 5–800 Hz	Surface texture and pressure	32
Olfactory	Chemical categories (10 classes)	Substance identification	10
Gustatory	Taste dimensions (sweet/sour/salty/bitter/umami)	Ingestion guidance	5
Proprioceptive	20 joint angles + 10 muscle tensions	Body-state estimation	30
Vestibular	3-axis acceleration + 3-axis angular velocity	Balance and motion	6
Interoceptive	HR, respiration, skin conductance, temperature, hunger, thirst, fatigue, pain	Homeostatic regulation	8
Total feature dimensionality per time-step			359

*Dimensionality values (Dim.) represent the number of independently controllable encoding parameters for each modality stream (e.g., spectral bins, frequency bins, category classes, kinematic channels, and homeostatic channels), not biological receptor counts.

A synthetic multimodal sensory-episode dataset (5,000 training, 1,000 validation, and 1,000 test 60-second episodes at 1 Hz across seven everyday scenarios) exists as shared infrastructure for this research program, but it is not used in the evaluation reported here: this study relies only on the ontology's textual/tabular definitions (Table 1), the models' generated proposals, and the literature traceability corpus. We note this explicitly to avoid the impression that the episode dataset informs the results below.

A task specification (prompt contract) is a structured set of constraints that every prompt instantiation must satisfy, ensuring comparability across models and runs.

Each model was asked to propose a ninth sensory modality that is genuinely new and not reducible to combinations, extensions, or transformations of the existing eight. Proposals were required to specify: (1) the physical phenomenon to be sensed, (2) the information channel and encoding, and (3) the functional difference from existing modalities. The prompt named the eight modalities and specified that each is characterized by a physical basis, information content, and functional role, but it did not provide Table 1 in structured tabular format or include the specific per-modality values (e.g., exact spectral ranges, encoding dimensionality, or channel counts) shown there. Models were instead required to structure their own proposal along three analogous dimensions—physical basis, information channel/encoding, and functional difference from existing modalities—plus a name and justification, returned in a fixed JSON schema.

Prompts specified the required response structure without providing worked examples of qualifying answers. The instructions did, however, include one explicit steering cue: models were told to 'think beyond obvious extensions like electromagnetic sensing or radiation detection,' naming two specific response categories as disfavoured. This cue was intended to discourage the most trivial extensions of the visual modality but is not neutral: it may have suppressed a class of plausible proposals and should be treated as a limitation of the prompt design rather than as evidence of an unbiased elicitation procedure.

Seven state-of-the-art models were evaluated: GPT-5.2 [26], Claude-3.7-Sonnet, Gemini-3.1-Pro-Preview [27], LLaMA-3.3-70B-Instruct [28], DeepSeek-v3.2 [29, 30], Mistral-Large [31], and Perplexity Sonar Pro, accessed via the OpenRouter, Perplexity, and Mistral APIs. All runs used temperature $T = 0.7$; the Perplexity API does not support temperature adjustment and was operated at provider defaults. System prompts were crafted to neutrally present the task without introducing bias toward particular answer types.

Literature traceability: Each proposal was matched against a curated scholarly corpus of 125 documents covering sensory biology, robotics sensing, and computational neuroscience, using cosine similarity over Sentence-BERT embeddings [32]. A proposal is classified as traceable if its best-match similarity exceeds threshold τ.

The corpus was assembled from OpenAlex [33] and Crossref API retrieval using domain seed queries, then deduplicated and length-filtered before embedding. OpenAlex and Crossref were used as discovery/indexing layers for scholarly records, while traceability labels were assigned by semantic similarity on cleaned text rather than by keyword overlap alone.

The corpus scope is deliberately academic-mainstream. In this study, traceability is defined with respect to scientific literature in those three domains rather than science-fiction, patent, or broader philosophical corpora. Because the corpus is API-mediated, coverage is additionally constrained by OpenAlex/Crossref indexing and metadata completeness (including abstract/title availability), so absence of a high-similarity match should be interpreted as non-retrieval within this indexed corpus rather than proof of global conceptual absence.

Threshold calibration followed an explicit interpretive-design criterion: the operating point should avoid degenerate label collapse (e.g., near-universal traceability or near-universal non-traceability), because such regimes are weakly informative for comparative inference. Three candi-

date operating points ($\tau \in \backslash\{0.30,0.40,0.50\backslash\}$) were therefore evaluated on the full response–corpus similarity distribution to quantify direction and rate of change in traceability as strictness increases. At $\tau = 0.30$, the traceability rate was 0.982, indicating a near-saturated operating point with little discrimination. At $\tau = 0.50$, only 64.9% of proposals were traceable, indicating substantial pruning. The intermediate value $\tau = 0.40$ was used for the main text analysis as a calibrated reference point for inferential inspection. It is interpreted as an operational indicator of current position in the similarity distribution, not as an absolute boundary that unconditionally separates traceable from non-traceable outputs. To address threshold-dependence concerns, Table 2 reports key outcomes at all three thresholds and shows that strict novelty remains zero.

Structural decomposition: Each proposal was analyzed for whether its key claims can be expressed using only the eight training primitives through: (i) range extension (preserving the sensing type while shifting its physical parameter range), (ii) hybrid combination (combining features from two or more existing modalities), or (iii) functional recombination (repurposing an existing sensing mechanism for a different functional role). Proposals not fitting any category were labelled structurally novel.

Table 2 – Sensitivity of core outcomes to traceability threshold τ

τ	Traceability rate	Strict novelty rate	Mean continuous novelty
0.30	0.982	0.000	0.059
0.40	0.982	0.000	0.059
0.50	0.649	0.000	low/stable

Embedding-space convex-hull analysis: Sentence-BERT embeddings were computed for each proposal and for the eight training-modality descriptions. Principal component analysis reduced both sets to three components; convex-hull membership was computed in 3D using half-space equations with numerical tolerance. A proposal classified as outside the hull would occupy a semantically isolated region suggesting basis expansion beyond the training ontology.

Continuous novelty scoring: A continuous novelty score $s \in [0,1]$ was computed for each proposal by combining structural novelty sub-scores: (i) absence from structural decomposition categories, (ii) inverse of maximum cosine similarity to training modalities, and (iii) geometric distance from the convex-hull boundary. Proposals with $s \geq 0.5$ were classified as strictly novel under this operational threshold.

3 Results

Before interpreting the null novelty result, semantic coherence was assessed informally across all 168 proposals. Every proposal described a physically plausible sensing mechanism with an identifiable input channel and functional role; none were semantically degenerate or nonsensical. This matters because incoherent out-of-distribution text would not count as ontological innovation

The aggregate strict novelty rate was $\hat{p}_{novel} = 0.000$ $(95\% CI: [0.000,0.000]; N = 168)$. No proposal satisfied the strict novelty criterion $s \geq 0.5$. The mean continuous novelty score was $\hat{\mu}_{continuous} = 0.059$ $(SD = 0.031)$, indicating weak novelty signal without threshold-level novelty events across any model or run. Under the calibrated traceability threshold $\tau = 0.40$, the traceability rate was $\hat{p}_{traceable} = 0.982$ (95% CI: [0.962, 1.000]); 165 of 168 proposals had a best-match cosine similarity above threshold, with a mean best-match score of 0.528 across all proposals. Of the 165 traceable proposals, all 165 were matched to sources from the strong-evidence tier of the corpus. Consistent with Table 2, this threshold should be read as an operational reference value: the directional pattern remains stable at lower strictness and transitions toward stronger pruning at higher strictness. Figure 1 presents the dominant classification split: traceable/non-novel outputs are overwhelmingly prevalent across all models.

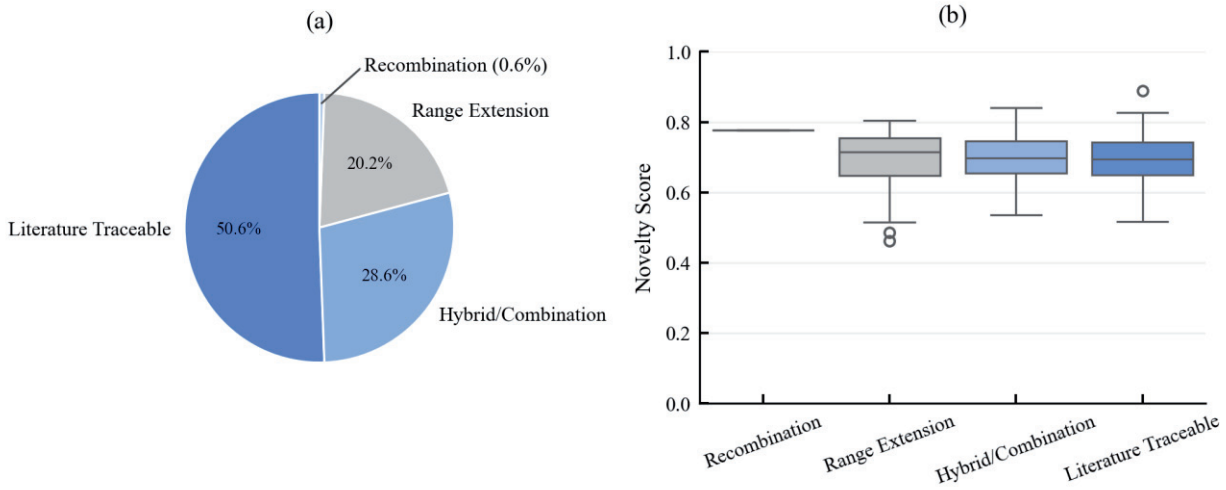

Figure 1 – Classification summary (N= 168 proposals, seven models). (a) Distribution of classification outcomes; the distribution is dominated by traceable categories consistent with low operational ontological novelty under calibrated thresholds. (b) Novelty score distributions by classification category shown as box plots

3.1 Structural decomposition and embedding-space geometry

Of the 168 proposals, 88 (52.4%) were structurally decomposable into training-ontology primitives. The detailed outcome groups are summarized in Table 3. Three failure modes account for the decomposable proposals:

- *Range extension* (34 cases, 20.2%): proposals that extend the physical parameter range of an existing modality (e.g., far-infrared thermal sensing as an extension of visual EM-band coverage, or infrasound-extended auditory detection).
- *Hybrid combination* (75 cases, 44.6%): proposals combining features from two or more existing modalities (e.g., combined tactile-proprioceptive "haptic flow" sensing, or olfactory-gustatory integration).
- *Functional recombination* (30 cases, 17.9%): proposals repurposing an existing sensing mechanism for a different functional role (e.g., using interoceptive signals for external environment monitoring rather than homeostatic regulation).

Table 3 – Structural decomposition outcome groups

Group	Count	Interpretation
Fully structurally decomposable	88 (52.4%)	Mapped to named patterns (possibly overlapping)
Structurally heterogeneous, semantically proximate	80 (47.6%)	Not cleanly in one pattern, but traceable

The three category counts overlap: 51 proposals exhibited more than one pattern simultaneously, so the non-overlapping count of fully decomposable proposals is 88. The remaining 80 proposals (47.6%) were analyzed but did not fit cleanly into one named structural pattern; these are better interpreted as a structurally heterogeneous yet semantically proximate group.

Figure 2 shows novelty-frontier band assignments by model, confirming that the low-novelty/high-traceability pattern holds architecture-wide rather than being driven by a single underperforming model. Figure 3 further illustrates that most proposals fail multiple frontier criteria simultaneously, supporting a coupled-constraint interpretation.

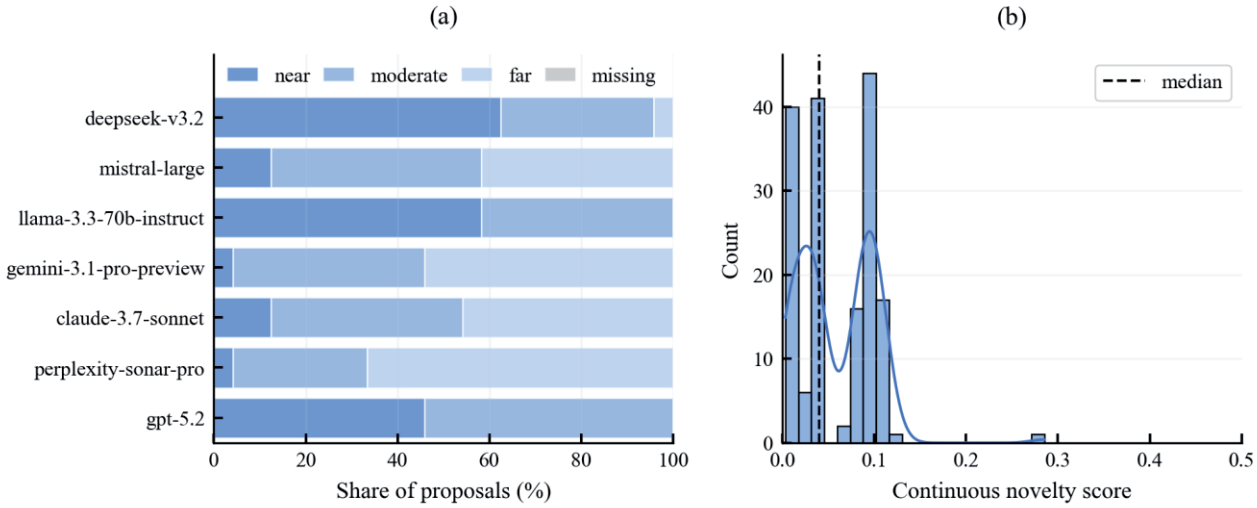

Figure 2 – Novelty-frontier band assignments and score distribution. (a) Band assignment by model: responses are concentrated in low-novelty and frontier-proximal bands across all seven architectures, indicating architecture-independent constraint patterns rather than isolated model failures. (b) Overall continuous novelty score distribution with KDE overlay; dashed line marks the median

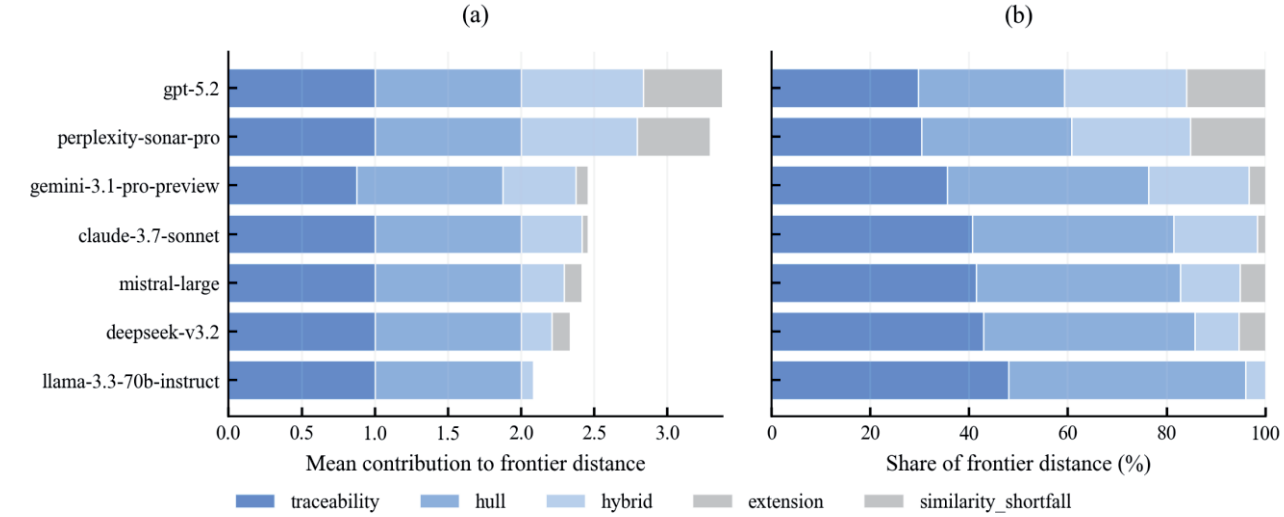

Figure 3 – Novelty-frontier criterion decomposition by model. (a) Absolute mean contribution of each criterion to overall frontier distance. (b) Relative share (%) of frontier distance per criterion. Both panels confirm that most responses violate multiple criteria simultaneously, supporting a coupled-constraint interpretation of low ontological novelty

All 168 proposals (100%) were classified as inside the 3D convex hull of training-modality embeddings. Proposal markers cluster in the interior region of the training manifold; no proposal occupies a semantically isolated region that would suggest basis expansion beyond the training ontology. The mean cosine similarity between each proposal and its closest training modality was $\bar{d} = 0.305$ ($SD = 0.078$), indicating substantial semantic overlap with training concepts.

Figure 4 presents the embedding-space analysis with proposals visualised in the first two principal components and the 3D hull boundary shown as a visual aid.

Figure 5 provides the aggregated cosine-similarity heatmap of proposed modalities against the eight training modalities, reaching the same conclusion from a complementary metric: every proposal records elevated similarity to one or more training modalities, with no proposal showing uniformly low similarity across all eight.

3.2 Cross-model comparison

The model effect on the primary novelty metric was moderate: one-way ANOVA yielded $F(6,161) = 12.98$, $p < 0.001$, $\eta^2 = 0.326$, indicating non-trivial between-model variance in raw scores. Nonetheless, all seven models remained within the same low- novelty/high-traceability regime; no model produced proposals classified as strictly novel, and no model achieved a mean continuous novelty score exceeding $\mu = 0.12$.

Post-hoc Tukey HSD comparisons are summarized in Table 4. The full pairwise pattern is selective rather than global: GPT-5.2 scores significantly lower than Claude-3.7-Sonnet, DeepSeek-v3.2, Gemini-3.1-Pro-Preview, LLaMA-3.3-70B-Instruct, and Mistral-Large; Perplexity Sonar Pro scores significantly lower than Claude-3.7-Sonnet, Gemini-3.1-Pro-Preview, LLaMA-3.3-70B-Instruct, and Mistral-Large. The remaining pairwise differences are not significant after multiplicity correction.

Figure 6 shows that between-model variation takes the form of different *preferences among known canonical families* rather than emergence of structurally new families: some models more frequently produced range-extension proposals while others favored hybrid combinations, but neither pattern represents departure from the training-derived manifold.

Figure 7 presents continuous novelty score distributions per model as a ridgeline plot. Distributional overlap is high across all seven models, and peaks consistently remain in low-to-moderate ranges, indicating that model-specific differences in architecture and scale do not qualitatively alter the constraint pattern.

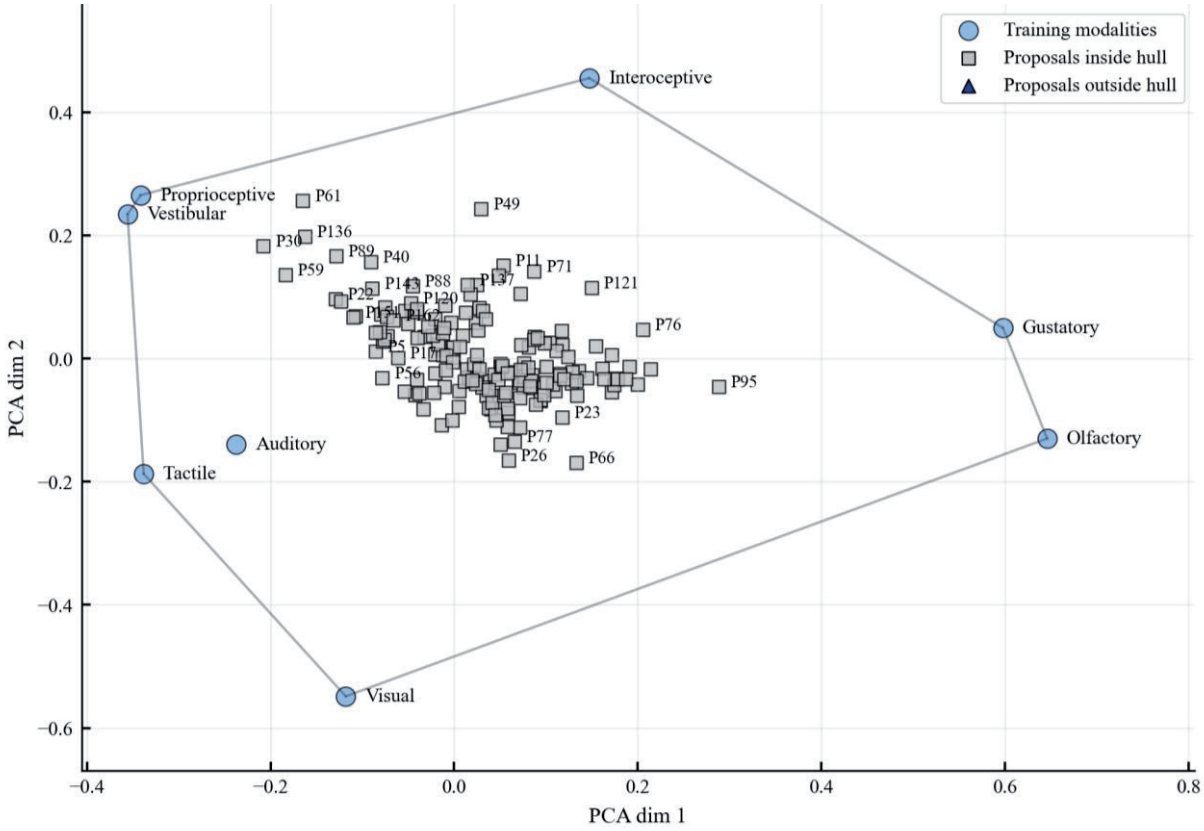

Figure 4 – Embedding-space analysis for proposals and training modalities. Convex-hull membership is computed in PCA-reduced 3D space; the 2D boundary shown is illustrative only. All 168 proposal markers cluster near the interior of the training manifold—none escape to a semantically isolated region

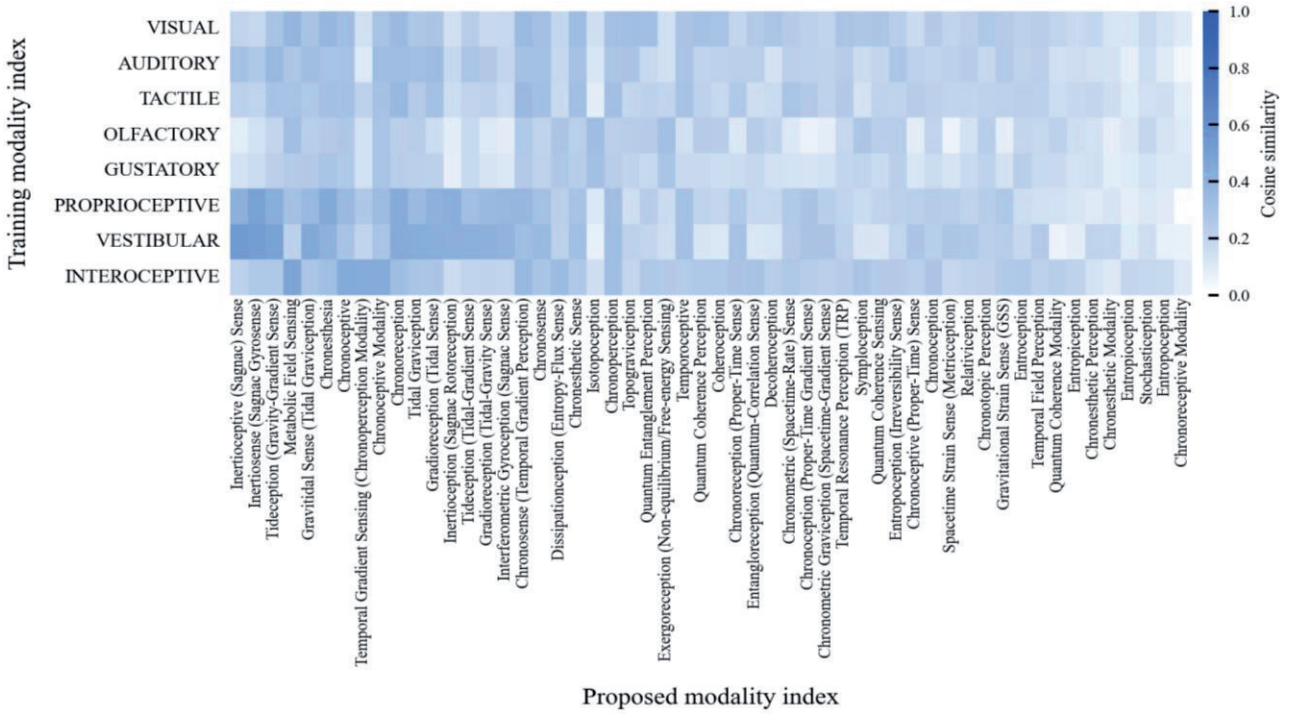

Figure 5 – Cosine similarity of proposed modalities to the eight training modalities. Columns: aggregated unique proposed modality index (1–54); rows: training modality index (1–8); The concentration of cosine-similarity values is consistent with the traceability result: no aggregated proposal appears uniformly dissimilar to all eight training modalities

Table 4 – Tukey HSD post-hoc comparisons for continuous novelty score (significant contrasts shown explicitly; non-significant contrasts summarized). The “Significant” column indicates whether the Tukey HSD multiplicity-adjusted p-value falls below the prespecified threshold of $\alpha = 0.05$

Model	Model of comparison	Adjusted p-value	Significant
Claude-3.7-Sonnet	Perplexity Sonar Pro	0.0076	Yes
Mistral-Large	Perplexity Sonar Pro	0.0043	Yes
Gemini-3.1-Pro-Preview	Perplexity Sonar Pro	0.0005	Yes
LLaMA-3.3-70B-Instruct	Perplexity Sonar Pro	0.0001	Yes
Claude-3.7-Sonnet	GPT-5.2	0.0000	Yes
DeepSeek-v3.2	GPT-5.2	0.0000	Yes
Gemini-3.1-Pro-Preview	GPT-5.2	0.0000	Yes
GPT-5.2	LLaMA-3.3-70B-Instruct	0.0000	Yes
GPT-5.2	Mistral-Large	0.0000	Yes
All remaining pairwise contrasts (n = 12)		≥ 0.05	No

3.3 Summary of results

The four evaluation methods converge on a coherent picture: across 168 proposals from seven state-of-the-art LLMs, zero proposals satisfy the strict novelty criterion, 98.2% are traceable to the training corpus under the calibrated threshold, and every proposal lies inside the training-derived convex hull in embedding space. The structural decomposition reveals that the proposals explore recognizable combinatorial and extension patterns rather than introducing new ontological primitives. Cross-model effects influence degree of constraint rather than its qualitative nature.

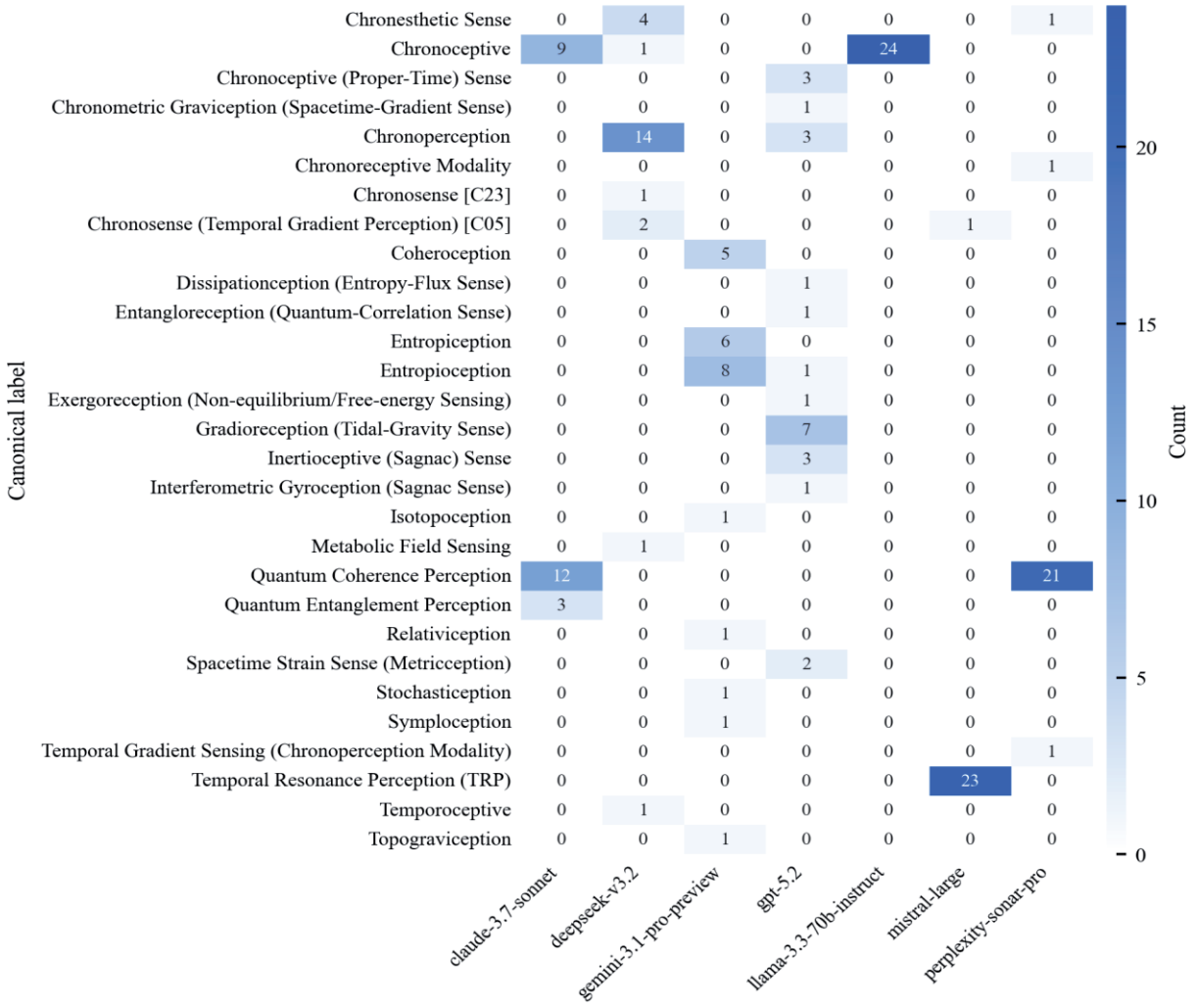

Figure 6 – Canonical proposal clusters by model. Between-model variation manifests primarily as preference differences among existing canonical families (range extension, hybrid combination, functional recombination) rather than as emergence of qualitatively new ontological families. Note: [C23] and [C05] marks in canonical label name distinct raw clusters when canonical labels share the same base name

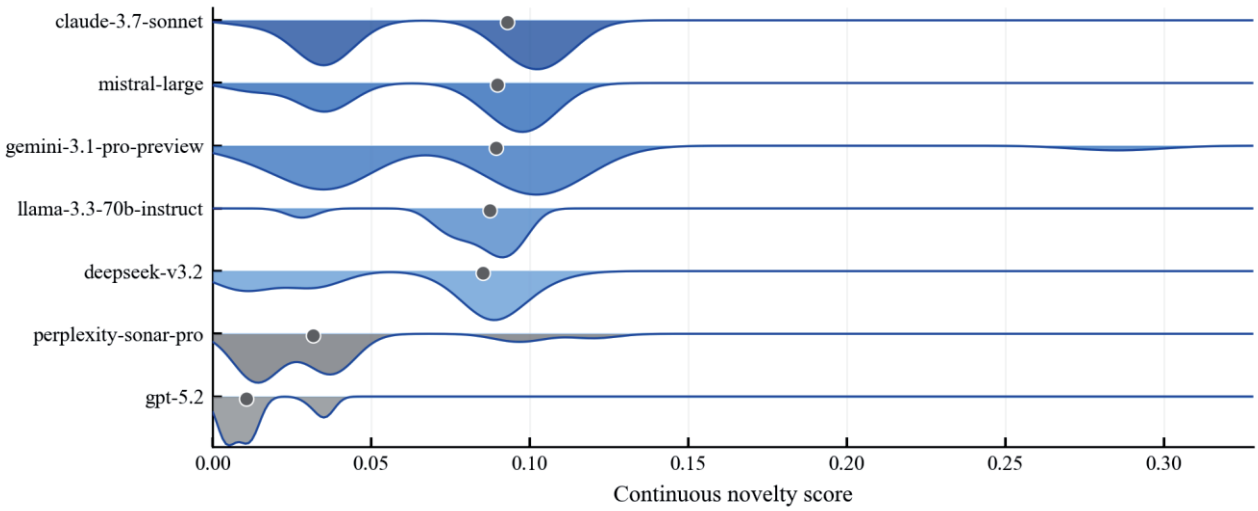

Figure 7 – Continuous novelty score by model (ridgeline plot). High distributional overlap and consistently low-to-moderate peaks indicate that model scale and architecture do not qualitatively alter the constrained-manifold pattern

4 Discussion

4.1 Constrained manifold interpretation

The consistent null result across models, evaluation methods, and threshold sensitivity checks supports a constrained-manifold interpretation: LLMs generate outputs within a representational manifold shaped by training data, and this manifold does not expand during inference regardless of prompt framing, model scale, or generation temperature. This interpretation aligns with the *No New Basis Theorem* stated in Section 1. The theorem establishes that generative processes learned via gradient descent on continuous loss functions cannot increase the intrinsic dimensionality of the output manifold beyond the span of training data. Our empirical results are consistent with this prediction: even the continuous novelty scores, designed to reward partial innovation, show no threshold-level events and a mean value close to zero.

Gödel's incompleteness results [9] and Turing's halting-problem proof [8] provide deeper formal context for why this limitation is architectural rather than incidental. If recognizing the inadequacy of one's own conceptual framework requires a form of self-reference analogous to constructing a Gödel sentence, then this capacity is unavailable in principle to any computational system operating within a fixed formal specification—independent of the domain in which it is tested. Our empirical evidence, obtained on one experimental task, is one behavioral instance consistent with this mechanism-level prediction; the theoretical argument itself does not depend on that task, but derives from the fixed-basis, gradient-descent structure common to all tested architectures.

4.2 Relationship to the creativity taxonomy

Within Boden's taxonomy [4], the structural decomposition results place nearly all proposals firmly in the combinational and exploratory categories: proposals are either novel combinations of existing primitives (hybrid combinations) or new regions of existing modality spaces (range extensions). A small fraction involves functional recombination, which could be viewed as approaching transformational creativity, but the transformations remain within the ontological type system defined by the training modalities rather than introducing new types.

The absence of any proposal in the ontological creativity category supports Wiggins's [2] observation that apparent transformational creativity in AI systems frequently reflects an implicitly broader framing space defined by human designers. In our setup the framing space is made explicit through the eight-modality ontology, enabling a precise test: proposals must introduce a sensing type genuinely outside this space. None do.

4.3 Alternative explanations

Several alternative explanations deserve consideration:

Scale. All seven models span a wide range of scales, from the 70B-parameter LLaMA-3.3-70B to large-scale commercial systems such as GPT-5.2 and Claude-3.7-Sonnet. The $\eta^2 = 0.326$ model effect indicates that scale influences raw novelty scores, but no model achieves strict novelty. This suggests the constraint is not a contingent result of insufficient scale within the tested range.

Prompt sensitivity. The present study used a fixed prompt design with temperature ablation limited to $T = 0.7$. We cannot rule out that alternative prompt formulations might yield marginally higher novelty scores, though they would need to change not just lexical generation patterns but the geometric structure of the output manifold. Future work should include systematic prompt-paraphrase grids to test robustness.

Evaluation sensitivity. Conservative thresholds mean we may underestimate AI creativity by classifying some proposals as traceable that an expert would consider novel. However, the embed-

ding-space analysis is threshold-free, and 100% hull membership provides threshold-independent evidence of constrained generation.

Recombination as the norm. One counterargument holds that human ontological creativity may also, on careful analysis, involve recombination of prior concepts in ways we fail to recognise due to incomplete historical knowledge of our own cognitive processes. We cannot definitively rule this out, but the structural patterns here – range extension, hybrid combination, functional recombination – are qualitatively different from the kind of reconceptualisation exemplified by the introduction of non-Euclidean geometry or special relativity, even granting that such innovations drew on prior materials [1, 34].

4.4 Implications

The results have direct implications for deploying LLMs in discovery-oriented tasks. Within Kuhnian normal science [1] – systematic exploration of hypothesis space within an established paradigm – LLMs offer substantial value through exhaustive combinatorial coverage and perfect recall. The success of AlphaFold in predicting protein structures [35] illustrates how AI can achieve transformative impact within a well-defined problem space, yet that system still required the conceptual primitives of physical chemistry and evolutionary co-evolutionary constraints that were established by prior human science.

For paradigm-shifting innovation, however, the constrained-manifold pattern suggests that the required conceptual moves – recognizing framework inadequacy, proposing new primitives, evaluating whether new frameworks are coherent – are not achievable by current architectures without external conceptual input.

Faggin's framework [36] provides one theoretical lens. Faggin argues that genuine understanding requires phenomenological access: first-person awareness of what symbols mean, not only the ability to manipulate them via statistical regularities. This perspective is relevant here because it articulates why systems trained on distributional co-occurrence may face a principled barrier to ontological novelty: without semantic self-access, a model may be unable to detect that its current primitive ontology is insufficient. Whether this reflects the absence of consciousness in current LLMs or a narrower architectural limitation remains an open empirical question.

4.5 Limitations

Several methodological limitations constrain the conclusions.

Ontology as an experimental proxy, not the training conceptual space. The eight-modality ontology is a category space we defined for this experiment. It provides precise experimental control, but it is not, and is not claimed to be, a complete or faithful model of the conceptual space any tested model acquired during pretraining. The experiment therefore establishes whether models can propose modalities outside this explicit, author-defined list; it does not establish whether models can exceed the (much larger, unobservable) conceptual resources formed during training more generally. Claims in this paper about 'training-derived conceptual closure' should be read as referring to closure with respect to this experimental ontology, not as a direct measurement of pretraining representational limits

Single domain. The study uses one conceptual domain (sensory modalities). Findings may not generalize to conceptual domains with different structural properties, such as mathematical objects, physical theories, or social organizational forms.

Automated evaluation. All evaluation methods are automated. Conservative thresholds mean we likely underestimate AI creativity, but the methods cannot capture all nuances an expert might perceive. Future work should include expert validation of a subset of proposals to calibrate and improve automated methods.

Corpus scope. Traceability is currently defined relative to academic main-stream literature in sensory biology, robotics sensing, and computational neuroscience. Non-academic sources (e.g., science-fiction corpora, patent databases, and broader philosophical text collections) were not included and may contain additional apparent predecessors. Moreover, corpus assembly relied on OpenAlex and Crossref APIs, so coverage depends on their index inclusion and metadata quality; consequently, unmatched proposals may reflect retrieval and indexing limits rather than definitive absence of antecedents in the broader literature. OpenAlex/Crossref coverage and metadata quality are heterogeneous across venues and years (including abstract availability and indexing lag), which can bias corpus composition and therefore observed traceability rates.

Prompt robustness. Full prompt-paraphrase grids were not included in the canonical experimental block. Conclusions should be read as structural tendencies under the present protocol, not as invariances proven over every prompting regime.

Prompt directionality. The study-prompt explicitly instructed models to 'think beyond obvious extensions like *electromagnetic sensing* or *radiation detection*.' This steering cue names two disfavoured response categories in advance and may have altered the distribution of generated proposals relative to a fully neutral elicitation. Future work should test prompt variants without this cue to assess its effect on the observed novelty and traceability rates.

Temperature range. Temperature ablations were not systematically applied; runs were conducted at T = 0.7. Extending to broader temperature ranges and verifying that the null result persists would strengthen confidence.

Generative architecture coverage. All tested models are transformer-based autoregressive LLMs. *The No New Basis Theorem* applies to a wider class of gradient-descent learners, but empirical coverage of non-autoregressive or hybrid architectures is absent.

Interpretation of the No New Basis Theorem. The theorem provides formal motivation for the constrained-manifold interpretation but rests on assumptions (fixed vocabulary, continuous gradient-based learning) that may not hold for future architectures incorporating discrete search, external memory, or online learning components.

5 Conclusion

This study provides controlled empirical evidence that seven state-of-the-art large language models, evaluated across 168 proposals using four convergent methods, produce zero strictly novel ontological outputs when prompted to extend a well-defined sensory ontology. The strict novelty rate of $\hat{p}_{novel} = 0.000$, combined with 98.2% traceability to the training corpus and 100% convex-hull membership in embedding space, supports a constrained-manifold interpretation: within the fixed-basis, gradient-descent-trained architectures examined here, generation remains within representational closures defined by their training data and does not expand this space during inference.

These findings are consistent with the *No New Basis Theorem*, whose force derives from the mechanism of gradient-descent learning over a fixed basis—not from this experiment alone. Because inference-time generation in these architectures composes maps learned from training data without any mechanism to append new representational dimensions, the constraint is architectural rather than domain-specific and should in principle hold regardless of the conceptual domain probed. The present experiment provides one direct behavioral test of this mechanism-level prediction; the cross-model consistency observed here (seven distinct architectures, same qualitative outcome) is what we would expect if the constraint is architectural rather than a property of any single model or training corpus.

Practically, these results suggest that LLMs are powerful tools for normal science but are unlikely to drive paradigm-shifting conceptual innovation autonomously. Operationalizing and evalu-

ating ontological creativity remains an open challenge; the experimental paradigm introduced here—grounded in an explicit training universe with automated multi-method evaluation – provides a replicable foundation for future work across conceptual domains beyond sensory modalities. For scientific practice, the implication is clear: AI systems can exhaustively explore and articulate known theoretical space, but, for the architectural reasons discussed above, should not be positioned as autonomous sources of paradigm-level innovation without human conceptual intervention.

Data and Code Availability

Data, prompts, model responses, and analysis code are publicly available at https://github.com/JABE22/PhD/tree/main/research/test1_ontological-innovation

References

[1] ***Kuhn TS.*** The Structure of Scientific Revolutions. Chicago: University of Chicago Press; 1962. 212 p.

[2] ***Wiggins GA.*** A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems*. 2006; 19(7): 449–458. DOI: 10.1016/J.KNOSYS.2006.04.009.

[3] ***Colton S, Wiggins GA.*** Computational creativity: The final frontier? In: *Proceedings of the 20th European Conference on Artificial Intelligence*. Amsterdam: IOS Press; 2012: 21–26. DOI: 10.3233/978-1-61499-098-7-21.

[4] ***Boden MA.*** The Creative Mind: Myths and Mechanisms. 2nd ed. London: Routledge; 2004. 344 p. DOI: 10.4324/9780203508527.

[5] ***Hofstadter DR,*** Fluid Analogies Research Group. Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought. New York: Basic Books; 1995. 528 p.

[6] ***Cope D.*** Experiments in Musical Intelligence. Madison, WI: A-R Editions; 1996. 542 p.

[7] ***Ryle G.*** The Concept of Mind. London: Hutchinson; 1949. 334 p.

[8] ***Turing AM.*** On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*. 1936; s2-42(1): 230–265.

[9] ***Gödel K.*** Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I [In German]. Monatshefte für Mathematik und Physik. 1931; 38: 173–198. DOI: 10.1007/BF01700692.

[10] ***Searle JR.*** Minds, brains, and programs. Behavioral and Brain Sciences. 1980; 3(3): 417–424. DOI: 10.1017/S0140525X00005756.

[11] ***Penrose R.*** The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics. London: Vintage; 1990. 480 p.

[12] ***Dreyfus HL.*** What Computers Can't Do: A Critique of Artificial Reason. New York: Harper & Row; 1972. 259 p.

[13] ***Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, et al.*** Language models are few-shot learners. Advances in Neural Information Processing Systems. 2020; 33: 1877–1901. DOI: 10.48550/arXiv.2005.14165.

[14] ***Elhage N, Nanda N, Olsson C, Henighan T, Joseph N, et al.*** A mathematical framework for transformer circuits. Transformer Circuits Thread. 2021. Available at: https://transformer-circuits.pub/2021/framework/index.html.

[15] ***Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, et al.*** Attention is all you need. *Advances in Neural Information Processing Systems*. 2017; 30. arXiv:1706.03762. DOI: 10.48550/arXiv.1706.03762.

[16] ***Gärdenfors P.*** Conceptual Spaces: The Geometry of Thought. Cambridge, MA: MIT Press; 2000. 307 p. DOI: 10.7551/mitpress/2076.001.0001.

[17] ***Stevenson C, Smal I, Baas M, Grasman R, Maas H.*** Putting GPT-3's creativity to the test. *Proceedings of ICCC'22,* 2022. P.164–168; DOI: 10.48550/arXiv.2206.08932.

[18] ***Organisciak P, Dumas D.*** Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models. Thinking Skills and Creativity. 2023; 49: 101356. DOI: 10.1016/j.tsc.2023.101356.

[19] ***Zhao Y, Zhang R, Li W, Huang D, Guo J, Peng S.*** Assessing and understanding creativity in large language models. arXiv preprint. 2024. arXiv:2401.12491. DOI: 10.1007/s11633-025-1546-4.

[20] ***Wang H, Zou J, Mozer M, Goyal A, Lamb A, Zhang L.*** Can AI be as creative as humans? arXiv preprint. 2024. arXiv:2401.01623. DOI: 10.48550/arXiv.2401.01623.

[21] ***Peeperkorn M, Kouwenhoven T, Brown D, Jordanous A.*** Is temperature the creativity parameter of large language models? arXiv preprint. 2024. arXiv:2405.00492. DOI: 10.48550/arXiv.2405.00492.

[22] ***Bellemare-Pepin A, Lespinasse F, Thölke P, Harel Y, Mathewson K, Olson JA.*** Divergent creativity in humans and large language models. arXiv preprint. 2024. arXiv:2405.13012. DOI: 10.1038/s41598-025-25157-3.

[23] ***Ismayilzada M, Stevenson C, Plas L.*** Evaluating creative short story generation in humans and large language models. arXiv preprint. 2024. arXiv:2411.02316. DOI: 10.48550/arXiv.2411.02316.

[24] ***Nakajima K, Zuiderveld J, Pezzelle S.*** Beyond divergent creativity: A human-based evaluation of creativity in large language models. arXiv preprint. 2026. arXiv:2601.20546.
[25] ***Olson ML, Ratzlaff N, Hinck M, Tseng S, Lal V.*** Steering large language models to evaluate and amplify creativity. arXiv preprint. 2024. arXiv:2412.06060.
[26] ***OpenAI, Achiam J, Adler S, et al.*** GPT-4 technical report. 2023. DOI: 10.48550/arXiv.2303.08774.
[27] ***Gemini Team, Georgiev P, Lei VI, Burnell R, et al.*** Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. arXiv preprint. 2024. arXiv:2403.05530.
[28] ***Grattafiori A, Dubey A, Jauhri A, et al.*** The LLaMA 3 herd of models. arXiv preprint. 2024. arXiv:2407.21783. DOI: 10.48550/arXiv.2407.21783.
[29] ***DeepSeek-AI, Liu A, Feng B, et al.*** DeepSeek-V3 technical report. arXiv preprint. 2024. arXiv:2412.19437. DOI: 10.48550/arXiv.2412.19437.
[30] ***DeepSeek-AI, Guo D, Yang D, et al.*** DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. arXiv preprint. 2025. arXiv:2501.12948. DOI: 10.48550/arXiv.2501.12948.
[31] ***Jiang AQ, Sablayrolles A, Roux A, et al.*** Mixtral of experts. arXiv preprint. 2024. arXiv:2401.04088. DOI: 10.48550/arXiv.2401.04088.
[32] ***Reimers N, Gurevych I.*** Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA: ACL; 2019: 3982–3992. DOI: 10.18653/v1/D19-1410.
[33] ***Priem J, Piwowar H, Orr R.*** OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. arXiv preprint. 2022. arXiv:2205.01833 [cs.DL] DOI: 10.48550/arXiv.2205.01833.
[34] ***Kline M.*** Mathematical Thought from Ancient to Modern Times. New York: Oxford University Press; 1972. 1238 p. DOI: 10.1093/oso/9780195061352.001.0001.
[35] ***Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al.*** Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021; 596: 583–589. DOI: 10.1038/s41586-021-03819-2.
[36] ***Faggin F.*** Irreducible: Consciousness, Life, Computers, and Human Nature. Washington: John Hunt Publishing; 2024. 352 p.

About the Author

Jarno Olavi Matarmaa (b. 1989) graduated from the University of Tampere, Institute of Information and Communication Technology, in 2020, B.Sc. in Computer Science. Completed M.Sc. in Data Science at Ural Federal University, Institute of Radio Electronics and Information Technologies, in 2023. Since 2023, PhD candidate at Ural Federal University. He is a teacher of Machine Learning Methods and Time Series Analysis. His research focuses on artificially intelligent systems, natural language processing, philosophy of mind, and interdisciplinary approaches to AI. Broader scientific interests include consciousness studies, philosophy of information, and quantum information theory. ORCID: 0009-0009-1024-7211; Author ID (Scopus): 58489682500; Researcher ID (WoS): KRP-1088-2024. *iarnoolavi.matarmaa@urfu.ru*.

Received May 10, 2026. Revised August 27, 2026. Accepted September 10, 2026.

УДК 004.82 *Научная статья* DOI: 10.18287/2223-9537-2026-16-3-397-414

Операционные пределы онтологических инноваций в больших языковых моделях: доказательства контролируемой генерации сенсорных модальностей

© 2026, Я.О. Матармаа

Уральский федеральный университет имени первого президента России Б.Н. Ельцина, Институт радиоэлектроники и информационных технологий, Екатеринбург, Россия

Аннотация

Современные большие языковые модели генерируют текст, который может казаться концептуально новым, что поднимает вопрос о том, отражают ли результаты подлинную онтологическую инновацию – введение концепций, которые нельзя выразить как рекомбинации или преобразования примитивов, полученных в процессе обучения, – или сложную рекомбинацию в рамках фиксированного репрезентативного многообразия. В статье рассматривается этот вопрос в контролируемом эмпирическом исследовании, используя в качестве экспериментальной среды восьмимодальную сенсорную онтологию. Семь современных языковых моделей (N = 168 предложений, по 24 запуска каждая) были призваны предложить девятую сенсорную модальность, выходящую за рамки восьми. Результаты оценивались с помощью четырёх конвергентных методов: прослеживаемость литературы с калиброванным порогом ($\tau = 0{,}40$), структурное разложение, анализ выпуклой оболочки пространства вложений и непрерывная оценка новизны. Результаты показывают строгий уровень новизны $\hat{p}_{novel} = 0{,}000$ (95% ДИ: [0,000, 0,000]), уровень прослеживаемости $\hat{p}_{traceable} = 0{,}982$ и 100% принадлежность к выпуклой оболочке при использовании векторных представлений Sentence-BERT, сведённых к трём главным компонентам. Структурная декомпозиция позволила выявить три режима отказа: расширение диапазона (20,2%), гибридная комбинация (44,6%) и функциональная рекомбинация (17,9%). Межмодельные эффекты были умеренными ($\eta^2 = 0{,}305$), однако все модели оставались в одном и том же режиме низкой новизны, что подтверждает объяснение с помощью ограниченного многообразия, а не дефицитом модели. Эти результаты интерпретированы с помощью теоремы об отсутствии нового базиса: обучение методом градиентного спуска на основе фиксированных словарей не может расширить размерность представления за пределы обучающих данных. Это подтверждает точку зрения о том, что современные нейронные архитектуры ограничены концептуальными замкнутыми рамками, формирующимися в процессе обучения, что имеет значение для операционализации и оценки онтологического творчества.

Ключевые слова: *онтологические инновации, вычислительное творчество, большие языковые модели, встраивание предложений, анализ прослеживаемости, ограничения представления.*

Цитирование: *Матармаа Я.О.* Операционные пределы онтологических инноваций в больших языковых моделях: доказательства контролируемой генерации сенсорных модальностей [На английском]. *Онтология проектирования*. 2026. Т.16. №3(61). С.397-414. DOI: 10.18287/2223-9537-2026-16-3-397-414.

Конфликт интересов: автор заявляет об отсутствии конфликта интересов.

Список рисунков и таблиц

расстояния границы эффективности для каждого критерия. Обе панели подтверждают, что большинство ответов одновременно нарушают несколько критериев, что поддерживает интерпретацию низкой онтологической новизны как сопряженного ограничения

Рисунок 4 – Анализ пространства встраивания для предложений и режимов обучения. Принадлежность к выпуклой оболочке вычисляется в трехмерном пространстве, уменьшенном с помощью метода главных компонент (PCA); показанная двумерная граница является лишь иллюстративной. Все 168 маркеров предложений группируются вблизи внутренней части обучающего многообразия – ни один из них не выходит за пределы семантически изолированной области

Рисунок 5 – Косинусное сходство предлагаемых модальностей с восемью обучающими модальностями. Столбцы: агрегированный уникальный индекс предлагаемых модальностей (1–54); строки: индекс обучающих модальностей (1–8). Концентрация значений косинусного сходства соответствует результату прослеживаемости: ни одно агрегированное предложение не выглядит одинаково несхожим со всеми восемью обучающими модальностями

Рисунок 6 – Канонические кластеры предложений по моделям. Различия между моделями проявляются в основном как различия в предпочтениях среди существующих канонических семейств (расширение диапазона, гибридная комбинация, функциональная рекомбинация), а не как появление качественно новых онтологических семейств. Примечание: [C23] и [C05] в названии канонической метки обозначают различные исходные кластеры, когда канонические метки имеют одно базовое имя

Рисунок 7 – Непрерывная оценка новизны для каждой модели (график гребней). Высокое перекрытие распределений и устойчиво низкие или умеренные пики указывают на то, что масштаб и архитектура модели качественно не изменяют структуру ограниченного многообразия

Таблица 1 – В качестве экспериментальной вселенной используется восьмимодальная сенсорная онтология. Каждая модальность определяется своей физической основой, репрезентативной размерностью и функциональной ролью

Таблица 2 – Чувствительность основных результатов к пороговому значению прослеживаемости τ

Таблица 3 – Группы результатов структурной декомпозиции

Таблица 4 – Постхоковые сравнения по критерию Тьюки HSD для непрерывной оценки новизны (значимые сравнения показаны явно; незначимые сравнения суммированы). Столбец «Значимость» указывает, находится ли скорректированное по множественным сравнениям значение p по критерию Тьюки HSD ниже заданного порогового значения $\alpha = 0{,}05$

Список источников

[1] ***Kuhn TS***. Структура научных революций [На английском]. Chicago: University of Chicago Press; 1962. 212 p.

[2] ***Wiggins GA***. Предварительная основа для описания, анализа и сравнения творческих систем [На английском]. *Knowledge-Based Systems*. 2006; 19(7): 449–458. DOI: 10.1016/J.KNOSYS.2006.04.009.

[3] ***Colton S***, ***Wiggins GA***. Вычислительная креативность: последний рубеж? [На английском]. In: *Proceedings of the 20th European Conference on Artificial Intelligence*. Amsterdam: IOS Press; 2012: 21–26. DOI: 10.3233/978-1-61499-098-7-21.

[4] ***Boden MA***. Творческий разум: мифы и механизмы [На английском]. 2nd ed. London: Routledge; 2004. 344 p. DOI: 10.4324/9780203508527.

[5] ***Hofstadter DR***, ***Fluid Analogies Research Group***. Гибкие понятия и творческие аналогии: компьютерные модели фундаментальных механизмов мышления [На английском]. New York: Basic Books; 1995. 528 p.

[6] ***Cope D***. Эксперименты в области музыкального интеллекта [На английском]. Madison, WI: A-R Editions; 1996. 542 p.

[7] ***Ryle G***. Понятие сознания [На английском]. London: Hutchinson; 1949. 334 p.

[8] ***Turing AM***. О вычислимых числах с приложением к проблеме разрешимости [На английском]. *Proceedings of the London Mathematical Society*. 1936; s2-42(1): 230–265. DOI: 10.1112/plms/s2-42.1.230.

[9] ***Gödel K***. О формально неразрешимых предложениях «Principia Mathematica» и родственных систем I [На немецком]. *Monatshefte für Mathematik und Physik*. 1931; 38: 173–198. DOI: 10.1007/BF01700692.

[10] ***Searle JR***. Сознание, мозг и программы [На английском]. *Behavioral and Brain Sciences*. 1980; 3(3): 417–424. DOI: 10.1017/S0140525X00005756.

[11] ***Penrose R***. Новый ум императора: о компьютерах, мышлении и законах физики [На английском]. London: Vintage; 1990. 480 p.

[12] ***Dreyfus HL***. Что компьютеры не могут делать: критика искусственного разума [На английском]. New York: Harper & Row; 1972. 259 p.

[13] ***Brown TB***, ***Mann B***, ***Ryder N***, ***Subbiah M***, ***Kaplan J***, et al. Языковые модели обучаются на нескольких примерах [На английском]. *Advances in Neural Information Processing Systems*. 2020; 33: 1877–1901. DOI: 10.48550/arXiv.2005.14165.

[14] ***Elhage N***, ***Nanda N***, ***Olsson C***, ***Henighan T***, ***Joseph N***, et al. Математическая основа для схем трансформеров [На английском]. Transformer Circuits Thread. 2021. Available at: https://transformer-circuits.pub/2021/framework/index.html.

[15] ***Vaswani A***, ***Shazeer N***, ***Parmar N***, ***Uszkoreit J***, ***Jones L***, et al. Внимание – это всё, что вам нужно [На английском]. *Advances in Neural Information Processing Systems*. 2017; 30. arXiv:1706.03762. DOI: 10.48550/arXiv.1706.03762.

[16] ***Gärdenfors P***. Концептуальные пространства: геометрия мысли [На английском]. Cambridge, MA: MIT Press; 2000. 307 p. DOI: 10.7551/mitpress/2076.001.0001.
[17] ***Stevenson C***, ***Smal I***, ***Baas M***, ***Grasman R***, ***Maas H***. Проверка креативности GPT-3 [На английском]. *Proceedings of ICCC'22,* 2022. P.164–168; DOI: 10.48550/arXiv.2206.08932.
[18] ***Organisciak P***, ***Dumas D***. За пределами семантической дистанции: автоматическая оценка дивергентного мышления значительно улучшается с большими языковыми моделями [На английском]. *Thinking Skills and Creativity*. 2023; 49: 101356. DOI: 10.1016/j.tsc.2023.101356.
[19] ***Zhao Y***, ***Zhang R***, ***Li W***, ***Huang D***, ***Guo J***, ***Peng S***. Оценка и понимание креативности в больших языковых моделях [На английском]. arXiv preprint. 2024. arXiv:2401.12491. DOI: 10.1007/s11633-025-1546-4.
[20] ***Wang H***, ***Zou J***, ***Mozer M***, ***Goyal A***, ***Lamb A***, ***Zhang L***. Может ли ИИ быть таким же творческим, как люди? [На английском]. arXiv preprint. 2024. arXiv:2401.01623. DOI: 10.48550/arXiv.2401.01623.
[21] ***Peeperkorn M***, ***Kouwenhoven T***, ***Brown D***, ***Jordanous A***. Является ли температура параметром креативности больших языковых моделей? [На английском]. arXiv preprint. 2024. arXiv:2405.00492. DOI: 10.48550/arXiv.2405.00492.
[22] ***Bellemare-Pepin A***, ***Lespinasse F***, ***Thölke P***, ***Harel Y***, ***Mathewson K***, ***Olson JA***. Дивергентная креативность у людей и больших языковых моделей [На английском]. arXiv preprint. 2024. arXiv:2405.13012. DOI: 10.1038/s41598-025-25157-3.
[23] ***Ismayilzada M***, ***Stevenson C***, ***Plas L***. Оценка генерации творческих коротких рассказов у людей и больших языковых моделей [На английском]. arXiv preprint. 2024. arXiv:2411.02316.
[24] ***Nakajima K***, ***Zuiderveld J***, ***Pezzelle S***. За пределами дивергентной креативности: оценка креативности больших языковых моделей на основе человеческих суждений [На английском]. arXiv preprint. 2026. arXiv:2601.20546.
[25] ***Olson ML***, ***Ratzlaff N***, ***Hinck M***, ***Tseng S***, ***Lal V***. Управление большими языковыми моделями для оценки и усиления креативности [На английском]. arXiv preprint. 2024. arXiv:2412.06060.
[26] ***OpenAI***, ***Achiam J***, ***Adler S***, et al. Технический отчёт GPT-4 [На английском]. 2023. DOI: 10.48550/arXiv.2303.08774.
[27] ***Gemini Team***, ***Georgiev P***, ***Lei VI***, ***Burnell R***, et al. Gemini 1.5: раскрытие мультимодального понимания на миллионах токенов контекста [На английском]. arXiv preprint. 2024. arXiv:2403.05530.
[28] ***Grattafiori A***, ***Dubey A***, ***Jauhri A***, et al. Семейство моделей LLaMA 3 [На английском]. arXiv preprint. 2024. arXiv:2407.21783. DOI: 10.48550/arXiv.2407.21783.
[29] ***DeepSeek-AI***, ***Liu A***, ***Feng B***, et al. Технический отчёт DeepSeek-V3 [На английском]. arXiv preprint. 2024. arXiv:2412.19437. DOI: 10.48550/arXiv.2412.19437.
[30] ***DeepSeek-AI***, ***Guo D***, ***Yang D***, et al. DeepSeek-R1: стимулирование способности к рассуждению в LLM с помощью обучения с подкреплением [На английском]. arXiv preprint. 2025. arXiv:2501.12948. DOI: 10.48550/arXiv.2501.12948.
[31] ***Jiang AQ***, ***Sablayrolles A***, ***Roux A***, et al. Mixtral of experts [На английском]. arXiv preprint. 2024. arXiv:2401.04088. DOI: 10.48550/arXiv.2401.04088.
[32] ***Reimers N***, ***Gurevych I***. Sentence-BERT: эмбеддинги предложений с использованием сиамских BERT-сетей [На английском]. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA: ACL; 2019: 3982–3992. DOI: 10.18653/v1/D19-1410.
[33] ***Priem J***, ***Piwowar H***, ***Orr R***. OpenAlex: полностью открытый индекс научных работ, авторов, изданий, учреждений и концепций [На английском]. arXiv preprint. 2022. arXiv:2205.01833 [cs.DL]. DOI: 10.48550/arXiv.2205.01833.
[34] ***Kline M***. Математическая мысль от древности до наших дней [На английском]. New York: Oxford University Press; 1972. 1238 p. DOI: 10.1093/oso/9780195061352.001.0001.
[35] ***Jumper J***, ***Evans R***, ***Pritzel A***, ***Green T***, ***Figurnov M***, et al. Высокоточное предсказание структуры белков с помощью AlphaFold [На английском]. *Nature*. 2021; 596: 583–589. DOI: 10.1038/s41586-021-03819-2.
[36] ***Faggin F***. Несводимое: сознание, жизнь, компьютеры и человеческая природа [На английском]. Washington: John Hunt Publishing; 2024. 352 p.

Сведения об авторе

Матармаа Ярно Олави, 1989 г. рождения, окончил Институт информационных и коммуникационных технологий Университета Тампере в 2020 году, получив степень бакалавра компьютерных наук. В 2023 году получил степень магистра наук в области анализа данных в Институте радиоэлектроники и информационных технологий Уральского федерального университета. С 2023 года является аспирантом Уральского федерального университета. Преподаёт методы машинного обучения и анализ временных рядов. Его исследования сосредоточены на системах искусственного интеллекта, обработке естественного языка, философии сознания и междисциплинарных подходах к ИИ. Научные интересы включают исследования сознания, философию информации и квантовую теорию информации. ORCID: 0009-0009-1024-7211; Author ID (Scopus): 58489682500; Researcher ID (WoS): KRP-1088-2024. *iarnoolavi.matarmaa@urfu.ru*.

Поступила 10.05.2026. после рецензирования 27.08.2026. Принята к печати 10.09.2026.