Structure Complexity Entropy: A Principle for Structural Intelligence in AI

Luan Zhirong

luanzhirong@xaut.edu.cn

Xi'an University of Technology https://orcid.org/0000-0001-8260-0584

Shaobo Zheng

Xi'an University of Technology

Yujun Lai

Xi'an University of Technology

Haoran Ding

Xi'an Jiaotong University

Lujuan Dang

Xi'an Jiaotong University

Lei Xing

Xi'an Jiaotong University

Kaizhong Zheng

Xi'an Jiaotong University

Liangjun Chen

Xi'an Jiaotong University

Xinhu Zheng

The Hong Kong University of Science and Techonology (Guangzhou)

Badong Chen

Xi'an Jiaotong University

Article

Keywords:

Posted Date: January 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8609404/v1

Additional Declarations: There is **NO** Competing Interest.

Structure Complexity Entropy: A Principle for Structural Intelligence in AI

Abstract

Human intelligence excels at decomposing complex information into meaningful hierarchical structures, yet this capability remains a fundamental challenge for artificial intelligence. Current approaches, from manual curation to automated clustering, lack a principled way to quantify and autonomously discover such structures, often resulting in rigid binaries or semantically incoherent groupings. Here, we introduce Structural Complexity Entropy (SCEntropy), a novel metric that quantifies the internal disorder of an information set by measuring the heterogeneity of all pairwise relationships. SCEntropy provides AI with a foundational principle: a low value signifies a coherent concept, whereas a high value signals the need for decomposition. Leveraging this, we develop SCEntropy-driven Hierarchical Clustering (SHC), an algorithm that uses a single complexity threshold to autonomously construct multi-branch, semantically coherent hierarchies from the bottom up. We validate this principle across two core domains. In visual concept discovery, SHC not only recovers known taxonomies in image datasets like CIFAR-100 but also discovers semantically coherent and human-aligned super-categories, demonstrating an ability for autonomous knowledge structuring. Conversely, in natural language generation, we show that SCEntropy-derived hierarchies serve as a scaffold for coherent reasoning. By structurally constraining large language models, we enhance thematic focus and logical flow in multi-turn dialogues, mitigating semantic drift. Our work establishes SCEntropy as a universal framework for structural machine intelligence, enabling AI to not only discover how the world is organized but

also to discipline its own internal processes, paving the way for more autonomous and interpretable systems.

Main

The ability to perceive, construct, and manipulate internal structures of knowledge is a cornerstone of human intelligence[1-4]. This "structural intelligence" manifests in two complementary ways: through inductive abstraction, where we form coherent concepts and hierarchies from disparate observations; and through structured execution, where these internal models guide coherent thought, language, and action[5, 6]. While artificial intelligence (AI) has made staggering strides in pattern recognition, it lacks a fundamental capacity to autonomously reason about and with information structures, remaining reliant on human-defined architectures or brittle, flat data representations[7, 8].

Current methods for instilling structure into AI fall short of this vision. Manual curation approaches embed human insight but cannot scale[9-12], while automated clustering algorithms are fundamentally constrained. The dominant paradigm, exemplified by agglomerative hierarchical clustering[10], is bound to a binary-tree architecture, producing rigid, pairwise merging sequences that fail to capture the natural, multi-branch organization of real-world concepts[13-17]. Although valuable explorations beyond the binary tree exist, including graph-theoretic and reciprocal nearest neighbor methods[18, 19], they largely remain within a paradigm of local, similarity-driven sequential decision-making. These methods face a common, deeper challenge: the absence of a principled, quantifiable standard that directly measures 'conceptual cohesion' independently of pairwise similarity [20-22]. This paradigm-level gap means that even non-binary structures are built without a global "structural complexity" budget, resulting in intermediate clusters whose semantic coherence is not guaranteed and cannot be intrinsically assessed. This structural poverty is not merely a limitation in data organization but a fundamental constraint that restricts AI's ability to discover new knowledge inductively and to maintain coherent reasoning deductively.

A primary obstacle has been the absence of a universal measure for structural complexity itself[23]. How can an AI system know if a set of ideas forms a coherent whole, or if its own internal organization is structurally unsound? Existing information-theoretic measures, such as structural entropy[24, 25], are often designed for analyzing given graph partitions, not for quantifying the intrinsic cohesiveness of a set of elements or for guiding the real-time construction of semantic hierarchies[26].

We hypothesize that this gap stems from a missing computational principle that mirrors a core operation of human cognition: the efficient management of cognitive load through the formation of coherent conceptual chunks[27]. The human mind intuitively groups elements not merely by pairwise similarity, but by the overall consistency of their relational structure, avoiding sets with high internal dissonance[28]. In essence, a cognitively coherent 'concept' is a set of elements with low structural complexity. To formalize this principle, we introduce Structural Complexity Entropy (SCEntropy), a novel metric that quantifies this disorder in the relational fabric of any information set. The innovation of SCEntropy lies in proposing a meta-principle that establishes "structural complexity" itself as a measurable and constrainable entity. It operationalizes conceptual coherence. This thereby provides a unified principle to quantify structural cohesion, directly addressing the long-standing absence of a universal measure.

Leveraging this principle, we develop SCEntropy-driven Hierarchical Clustering (SHC). This algorithm represents a paradigm shift in hierarchy construction. It uses SCEntropy as a global constraint during the building process, rather than a post-hoc evaluation. This allows SHC to autonomously answer a question previously intractable for AI: "What is the richest, meaningful structure we can build without exceeding a manageable level of complexity?" The result is the autonomous discovery of multi-branch, semantically coherent hierarchies from the bottom up, without predefined schemas.

We demonstrate that this structural awareness is a general-purpose capability. Our results across two core domains of AI validate the universality of the SCEntropy principle. In visual concept discovery, SHC acts as an engine for knowledge induction, not only recovering known taxonomies but also discovering novel, valid super-categories on benchmarks like CIFAR-100. Conversely, in natural language generation, we show a novel application where SCEntropy-derived hierarchies serve as a cognitive scaffold. By explicitly organizing the outputs of large language models, we enforce thematic coherence and logical flow in multi-turn dialogues, proving that structural awareness can directly tame and guide complex generative processes. Our work thus paves the way for AI that truly comprehends, rather than merely processes, the complex structure of information.

By providing a unified principle to measure and manage structural complexity, we enable AI systems to not only discover how the world is organized but also to discipline their own internal processes accordingly. This dual advance in structural discovery and structural control opens the door to AI that is more autonomous, interpretable, and truly intelligent in its handling of complexity.

The Mathematical Foundation of Structural Coherence

The hierarchical decomposition of complex information hinges on a fundamental property[29-32], which is the consistency of relationships among its constituent elements[33]. Human cognitive analysis instinctively evaluates this relational fabric, isolating independent elements and grouping those with strong, consistent interdependencies into coherent subsets[34, 35]. The complexity of the resulting structure is thereby governed not merely by the number of elements or subsets, but by the heterogeneity in the relationships between them[36].

Conventional metrics for assessing information organization, such as pairwise similarity, primarily capture the strength of individual interactions[37]. However, they fall short of characterizing the global structural pattern because they do not quantify the variability across all pairwise relationships[38]. Conventional methods that rely on a fixed similarity threshold cannot distinguish between a set with uniformly high similarity (low complexity, requiring no decomposition) and one with a disordered mix of strong and weak connections (high complexity, necessitating partitioning). This critical distinction between the intensity and the consistency of relationships is not merely overlooked by conventional methods relying on a fixed similarity threshold. It remains fundamentally unmeasured by existing similarity-based approaches, which consequently fail to guide semantically-aware decomposition.

We therefore introduce SCEntropy to directly measure the disorder inherent in the relational structure of an information set. SCEntropy quantifies structural complexity by assessing the degree of inconsistency among all pairwise similarities. Formally, it evaluates the aggregate absolute differences between every pair of similarity measurements, capturing the overall relational heterogeneity. A low SCEntropy value signifies a structurally coherent set with uniform relationships, whereas a high value indicates significant internal disorder, guiding an AI system on when and how to decompose the set for more efficient processing. The formal definition is as follows.

Definition 1: Structural Complex Entropy

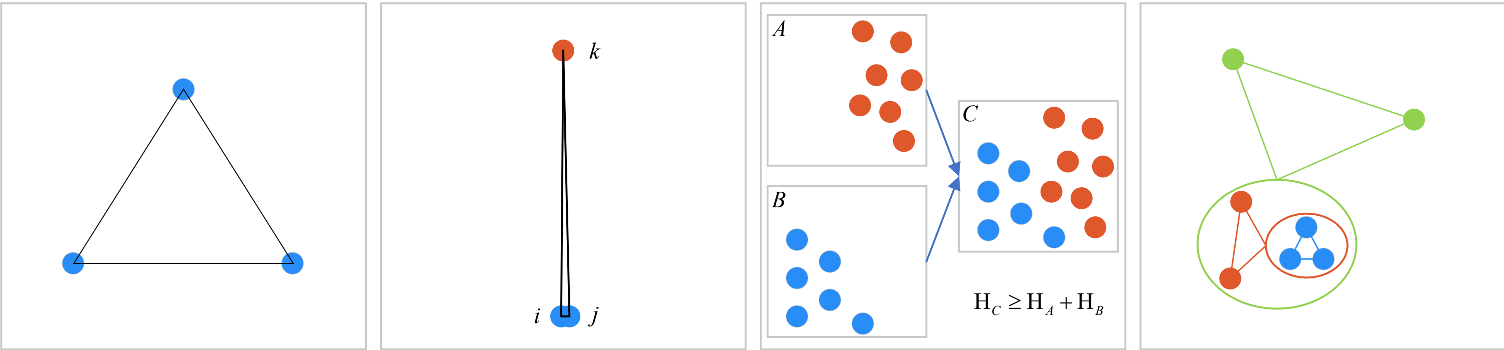
Given an information set consisting of n elements, SCEntropy is defined as:

$$
\begin{aligned}
H = & -\sum_{1\le i<j\le n}\ \sum_{1\le k<l\le n,(i,j)<(k,l)} \log\left(1-\left|S_{ij}-S_{kl}\right|\right) \\
s.t.\quad & 0\le S_{ij}\le 1,\ 0\le S_{kl}\le 1,\ \forall i,j,k,l, \\
& S_{ij}=S_{ji},\ S_{kl}=S_{lk},\forall i,j,k,l, \\
& S_{ii}=0,\ \forall i
\end{aligned}
\tag{1}
$$

where S_{ij} denotes the normalized similarity measurement between the i-th and j-th elements, and $|S_{ij}-S_{kl}|$ represents the absolute value of the difference between S_{ij} and S_{kl}. The constraint $(i,j)<(k,l)$ specifies the enumeration order to avoid redundant calculations. The term "element" is used in a broad sense to refer to any entity that can be embedded in vector space, such as a single data point, an image, an image category, a word, or a sentence. The exact definition depends on the object being analyzed.

SCEntropy measures the unpredictability or dissonance within a system of relationships. Consider a social context: a cohesive team (low SCEntropy) is one where the strength of all pairwise relationships is fairly uniform—everyone knows each other to a similar degree. A disordered group (high SCEntropy) contains a mix of strong cliques and distant acquaintances, making the relational landscape unpredictable. Similarly, in conceptual space, a low SCEntropy set, like {dog, cat, rabbit}, has uniformly high pairwise similarities. A high SCEntropy set, like {dog, car, loyalty}, contains wildly varying pairwise relationships, signaling a lack of a unifying theme. Thus, SCEntropy does not merely sum similarity strengths, it quantifies the consistency of the relational fabric, providing a direct gauge for whether a set of elements "belongs together" in a semantically coherent structure.

Having established that SCEntropy measures relational consistency, we now present its key mathematical properties. These properties govern how SCEntropy behaves and how it informs hierarchical decomposition.



(a) **Structural Homogeneity** (b) **Structural Heterogeneity** (c) **Complexity Reduction through Partitioning** (d) **Unified Criterion for Structural Discovery**

Fig.1 | Schematic illustrations of SCEntropy properties.

Property 1: Structural Homogeneity.

Zero SCEntropy arises if and only if all pairwise similarities are identical, indicating a perfectly homogeneous and coherent structure. In this state of maximum internal consistency, as visualized in Fig. 1(a), the set behaves as a unified whole. No meaningful decomposition is possible or necessary, as dissecting it would not yield semantically distinct or computationally simpler subproblems. This property provides a principled criterion for identifying atomic, indecomposable concepts.

Property 2: Structural Heterogeneity.

SCEntropy increases with the heterogeneity of pairwise similarities, reaching its maximum for a given set size when the relationships are maximally disordered. This occurs when the set contains a mix of highly similar (near-identical) and highly dissimilar (near-orthogonal) element pairs, as illustrated in Fig. 1(b). A high SCEntropy value acts as a direct signal that decomposition is warranted. It guides the system to isolate coherent subgroups (e.g., elements with near-zero distance) from independent elements, thereby reducing global structural complexity.

Property 3: Complexity Reduction through Partitioning.

The SCEntropy of a composite set is greater than or equal to the sum of the SCEntropies

of its constituent partitions. Formally, for a dataset $C = A \cup B$, let datasets A, B and C have sizes $n_A = |A|$, $n_B = |B|$ and $n_C = |C|$, respectively. The SCEntroy of dataset C is given by:

$$H_C = H_A + H_B + H_{A \cap B} \tag{2}$$

where

$$H_{A \cap B} = -\sum_{1 \leq i < j \leq n_C} \sum_{1 \leq k < l \leq n_C, (i,j) < (k,l)} \log\left(1 - \left|S_{ij} - S_{kl}\right|\right) \tag{3}$$
$$s.t. \quad \left(\left\{C_i, C_j, C_k, C_l\right\} \cap A \neq \phi\right) \wedge \left(\left\{C_i, C_j, C_k, C_l\right\} \cap B \neq \phi\right).$$

The constraint implies that the intersections of $\{C_i, C_j, C_k, C_l\}$ with both datasets A and B are non-empty. In other words, for $\{C_i, C_j, C_k, C_l\}$, at least one element must belong to dataset A, and at least one element must belong to dataset B, simultaneously. $H_{A \cap B} \geq 0$, quantifies the structural disorder between subsets A and B. Equality, $H_C = H_A + H_B$ is achieved if and only if $H_{A \cap B} = 0$. This condition signifies that the combined set $A \cup B$ is characterized by a perfectly uniform distribution of pairwise similarities. This non-negativity guarantees that $H_C \geq H_A + H_B$, formally proving that any partitioning of the data reduces the overall structural complexity, as shown in fig 1(c). This mathematically grounds the intuition behind hierarchical decomposition.

Property 4: Unified Criterion for Structural Discovery.

SCEntropy provides a unified, global criterion for structural analysis, replacing the multiple level-specific thresholds required by conventional methods. As illustrated in Fig. 1(d), SCEntropy intrinsically captures inconsistencies in the relational distribution, allowing a single threshold on structural complexity to guide the entire multi-level hierarchy construction. In contrast, methods like agglomerative clustering lack such a unified principle and rely on distinct, ad-hoc similarity thresholds at each dendrogram level. This makes SCEntropy-driven clustering more robust and semantically grounded.

A New Clustering Paradigm: Construction with Structural Complexity Constraints

Theoretically, minimizing SCEntropy provides a direct path for hierarchical decomposition. However, pursuing this minimization to its logical conclusion leads to a binary tree where every internal node contains only two elements. While structurally minimal, this overly fine-grained partitioning is semantically impoverished: it shatters the data into an excessive number of leaf nodes and fails to produce the rich, multi-element intermediate concepts that are essential for human-like reasoning and interpretability. The resulting hierarchy, though perfectly minimized in complexity, loses its value as a meaningful representation of the data's intrinsic semantic structure.

To bridge the gap between theoretical minimalism and semantic utility, we introduce the SCEntropy-driven Hierarchical Clustering (SHC) algorithm. The core innovation of SHC lies in its treatment of SCEntropy not as an objective function to be minimized, but as a dynamic constraint to regulate structural complexity during the hierarchy construction process. This shift in perspective is crucial. Instead of asking "what is the absolute simplest structure?", SHC asks "what is the richest meaningful structure we can build without exceeding a manageable level of complexity?" This allows the algorithm to autonomously construct multi-branch, semantically coherent hierarchies with a single, user-specified complexity threshold

The SHC algorithm proceeds through a bottom-up, iterative merging process, detailed in the following steps:

1. **Input Representation:** The algorithm begins with a dataset $C = \{C_i | 0 \leq i \leq n_c\}$, where each element C_i is represented by its feature vectors.

2. **Complexity Threshold Specification**: A single threshold, H_{th}, is set to define the

maximum allowable structural complexity for any cluster. This threshold is the sole parameter governing the granularity of the resulting hierarchy.

3. **Iterative Cluster Formation:**

a. **Seed Merging**: Identify the two closest clusters C_i and C_j and merge them to form a new cluster C_{new}.

b. **Constrained Candidate Selection**: From the remaining elements, identify a candidate set $\{C_k\}$, where each candidate C_k is closer to some element in C_{new} than it is to any element not in C_{new}. This ensures the candidate is locally closest to the growing cluster.

c. **Optimal Merging based on Structural Cohesion**: From the candidate set for each candidate. Select and merge the candidate C_k that results in the **lowest SCEntropy** for the new union. This step prioritizes the merger that best preserves or improves the structural consistency of the cluster.

d. **Complexity Check and Cluster Finalization**:

if $H(C_{new}) \leq H_{th}$, the cluster can continue growing. Return to step (b) to select a new candidate.

If $H(C_{new}) > H_{th}$, the cluster has reached its maximum allowable complexity. Reverse the last merge. the **centroid of** C_{new} is calculated based on all its finalized constituent elements as the feature of C_{new}. The algorithm then returns to step (a) to begin forming a new cluster.

4. **Termination**: The algorithm terminates when no further merging is possible without violating the H_{th} constraint, typically when only two clusters remain or no three clusters can be merged without exceeding the threshold.

SHC represents a fundamental departure from classic hierarchical clustering paradigms, such as agglomerative clustering, in two critical aspects:

Unified Criterion vs. Multiple Thresholds: Traditional methods first construct a full binary tree (dendrogram) and then perform post-hoc cuts at different similarity levels to extract a flat clustering. To obtain a multi-level hierarchy, one must define a distinct threshold for each desired level. In stark contrast, SHC employs a single, unified criterion—the SCEntropy threshold H_{th}—which governs the entire, multi-level hierarchy construction from the outset. This ensures global structural consistency, as the same principle of manageable complexity applies at every level of the tree.

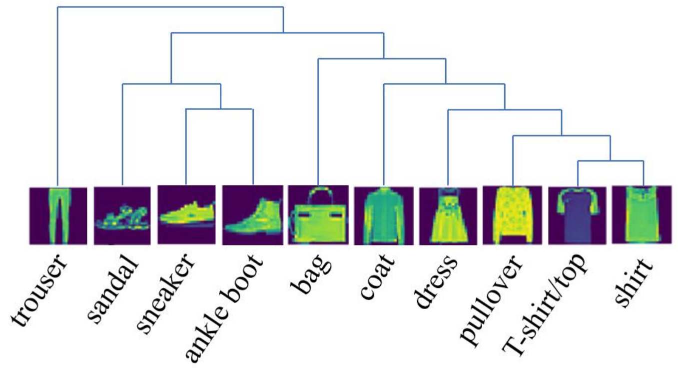
Construction-Time vs. Post-Construction Cutting: Perhaps the most significant algorithmic innovation is that SHC performs complexity-based "cuts" during the tree construction, not after it. This "construction-time cutting" allows for continuous centroid updates and enables the formation of multi-branch nodes. A traditional algorithm, once a merge is done, never revisits it. SHC, by contrast, dynamically decides when a cluster is "full" based on its structural complexity, then stops its growth and starts a new one. This results in a more natural and adaptive fit to the data's intrinsic, multi-branch semantic organization.

In summary, SHC leverages the principle of structural complexity not as a final arbiter of simplicity, but as a guiding principle for building rich, useful, and computationally tractable knowledge hierarchies. This makes it a powerful tool for enabling structural reasoning in AI systems.

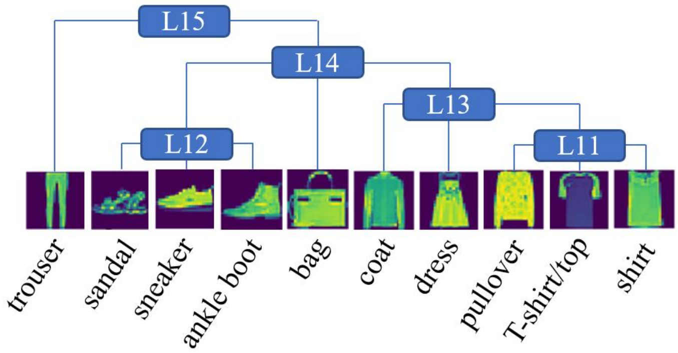
Results

Autonomous Emergence of Multi-Branch Hierarchies

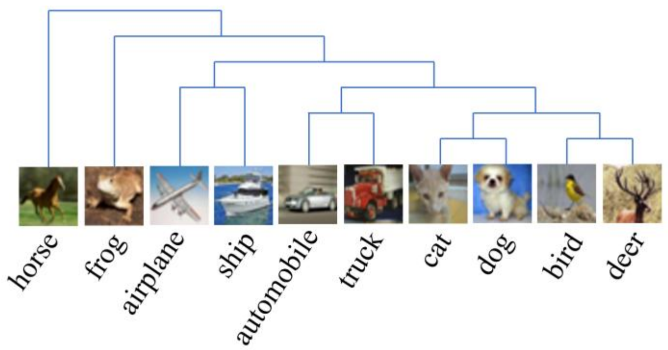
A cornerstone of human intelligence is the ability to organize concepts into flexible, multi-branch hierarchies. We first demonstrate that our SHC algorithm autonomously uncovers such human-aligned structures from high-dimensional image data, transcending the limitations of traditional clustering paradigms. As shown in Fig.2.



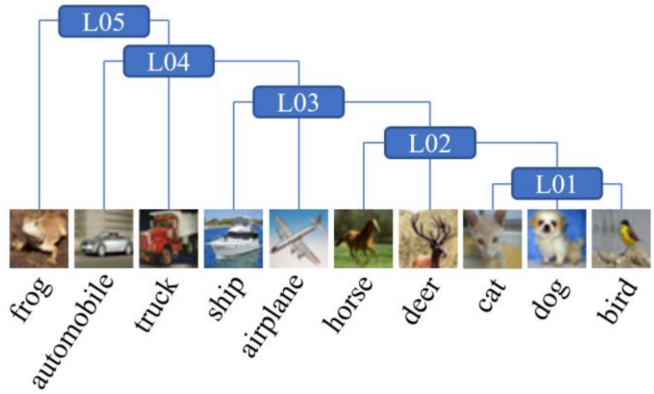
(a) Agglomerative clustering result on FashionMNIST



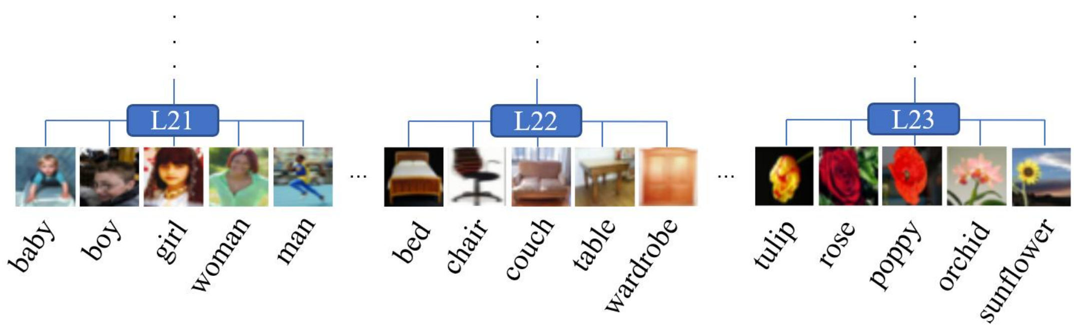
(b) SHC result on FashionMNIST



(c) Agglomerative clustering result on CIFAR-10



(d) SHC result on CIFAR-10



(e) Partial result of SHC on CIFAR-100

Fig.2| Hierarchical concept discovery results by SHC and agglomerative clustering. SHC uncovers hierarchical structures that align with human cognitive organization, producing multi-branching semantic groupings without predefined class numbers. This approach enhances interpretability and reflects natural conceptual relationships in the data.

On FashionMNIST[40], SHC moves beyond pixel-level similarity to uncover functional groupings (Fig. 2(b)). It intuitively coalesces upper-body garments (T-shirts, pullovers, shirts) under a common branch, while distinctly separating footwear (sneakers, sandals, ankle boots) and trousers, forming a taxonomy that aligns with human cognition. This stands in stark contrast to the rigid, binary-tree architecture produced by agglomerative clustering on the same dataset (Fig. 2(a)). This trend of forming more semantically coherent conceptual groupings is consistent on CIFAR-10[41] (Fig. 2(c) and Fig. 2(d)), where SHC's multi-branch structure, for instance, aggregates a broader set of animals under a unified super-category compared to the sequential, pairwise merging of agglomerative clustering.

Most notably, on the more complex CIFAR-100 dataset[41], SHC demonstrates a remarkable ability to autonomously reconstruct the dataset's underlying semantic taxonomy. As revealed in our concept discovery results (Fig. 2(e)), SHC does not merely group visually similar items, it tends to recover cohesive high-level concepts that align with the human-defined superclasses. Specifically, the algorithm successfully formed a cluster that is an exact match for the “people”, “household furniture” and “flowers” superclasses, cleanly grouping its five constituent subclasses.

The hierarchies produced by SHC are inherently adaptive and data-driven. This structural flexibility directly enhances interpretability, providing transparent, multi-level conceptual groupings. Moreover, since clusters are formed without manually tuned distance thresholds or predefined taxonomies, SHC demonstrates a significant leap in autonomy for structural discovery, paving the way for the quantitative and advanced semantic analyses that follow.

SCEntropy as a Portable and Robust Clustering Principle

Traditional clustering metrics are inadequate for evaluating hierarchies, as they penalize semantically coherent superclasses, such as a cluster that correctly combines "trees" and "landscapes" into a "geographic environments" concept. We therefore introduce the Hierarchical

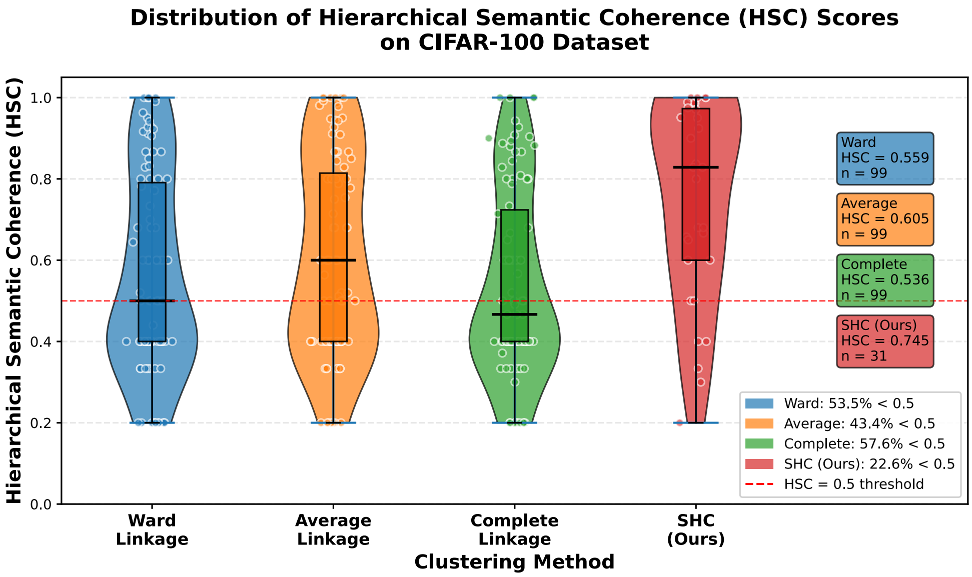
Semantic Coherence (HSC) score, a novel metric that quantifies the semantic completeness of a cluster by its completeness in capturing entire superclasses. For any given cluster, let k_i be the total number of subclasses in superclass i, and m_i be the number of those subclasses present in the cluster. Let $\mathrm{n} = \sum_i m_i$ be the total number of subclasses in the cluster. Let K be the number of superclasses. The HSC for the cluster is defined as:

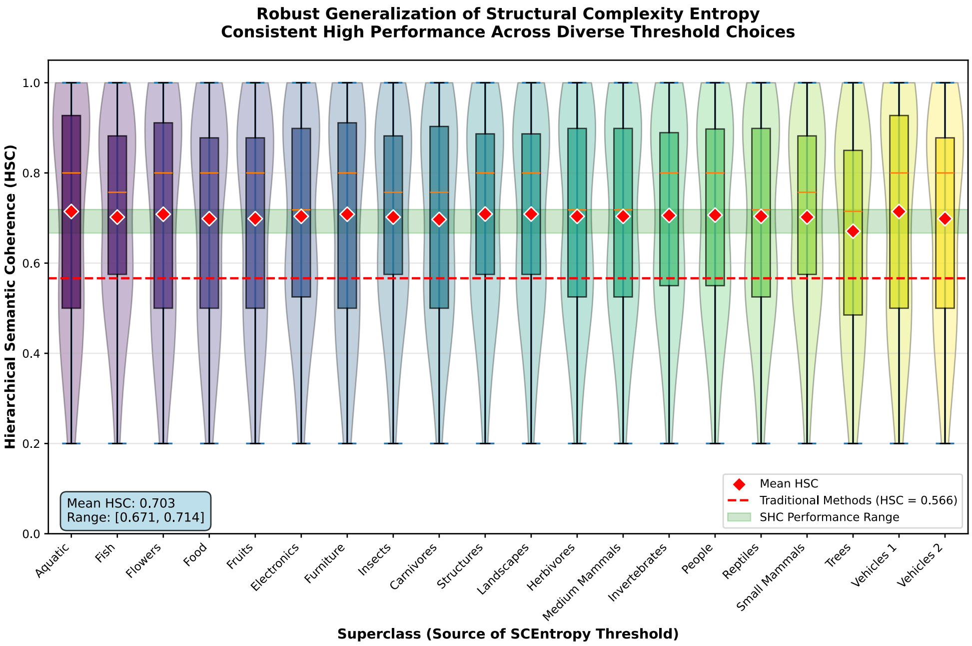
$$HSC = \frac{1}{n}\sum_{i=1}^{K}\frac{m_i^2}{k_i}. \tag{4}$$

Unlike purity or adjusted Rand index that penalize any combination of multiple superclasses, HSC explicitly rewards clusters that capture complete semantic concepts. The mathematical formulation ensures that a cluster containing entire superclasses achieves high scores, while fragmented clusters are penalized. This aligns with the cognitive science insight that human conceptual organization favors coherent, self-contained units. For a detailed introduction to HSC, refer to Supplementary Information Section 5.

Using this measure, our SHC algorithm demonstrates a fundamental advance by producing hierarchies of unprecedented semantic quality. On CIFAR-100 (Fig. 3a), SHC generates a compact hierarchy of only 31 intermediate conceptual clusters, which is threefold fewer than the 99 nodes produced by conventional agglomerative methods. Critically, these clusters exhibit significantly higher semantic coherence, with a mean HSC of 0.745 compared to ~0.566 for traditional approaches (Ward: 0.559; Average: 0.605; Complete: 0.536). The distribution reveals a more profound distinction: merely 22.6% of SHC clusters fall below the HSC=0.5, whereas approximately half of traditional clusters (e.g., 53.5% for Ward) demonstrate poor semantic integrity. This evidence underscores that traditional clustering, constrained by its binary merging, produces semantically ambiguous nodes. In contrast, SHC's complexity entropy principle enables the autonomous discovery of coherent, human-aligned conceptual hierarchies.

(a) SHC produces semantically coherent hierarchies with fewer conceptual clusters





(b) SHC provides robust structural discovery across diverse conceptual domains

Fig. 3 | Quantitative evaluation of structural complexity entropy principles in visual concept discovery.

This capability to automatically form compact, meaningful hierarchies without predefined cluster counts or similarity thresholds represents a significant step toward AI systems that can intrinsically understand and organize complex information structures.

The transferability of SCEntropy across diverse conceptual domains provides compelling evidence for its role as a fundamental principle of structural organization. As demonstrated in Fig. 3 (b), when SCEntropy values derived from individual CIFAR-100 superclasses were employed as global thresholds for organizing the entire conceptual space, the resulting hierarchies consistently maintained high semantic coherence that ranges from 0.671 to 0.714.

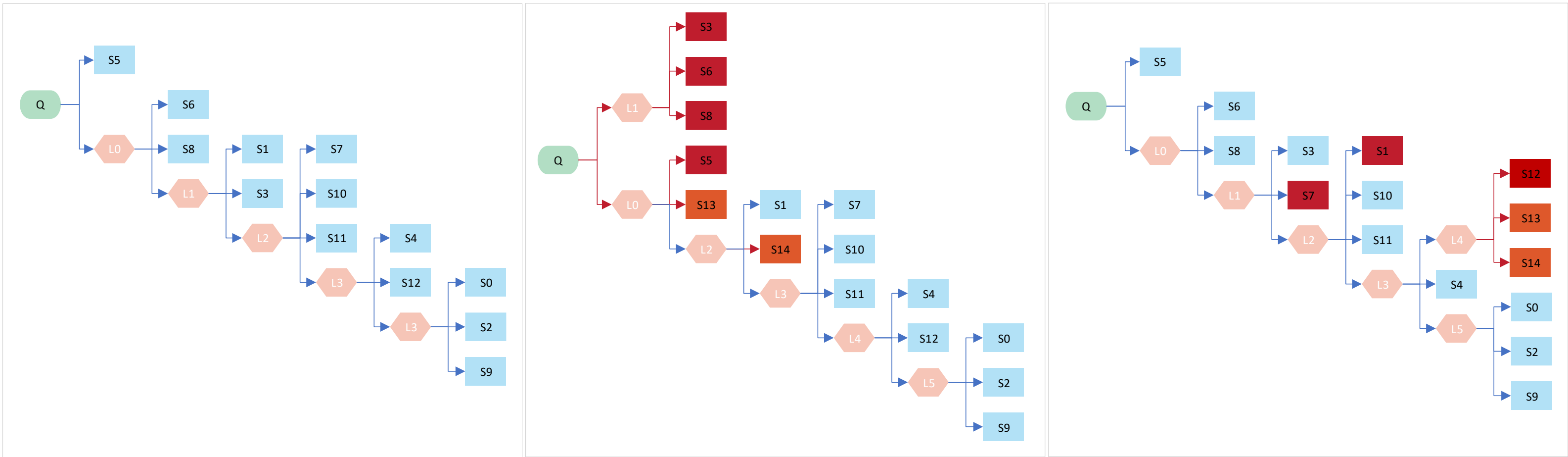
This remarkable consistency, which was achieved across 20 distinct conceptual domains from "people" to "household electrical devices", demonstrates that SCEntropy captures a universal property of structural organization rather than dataset-specific patterns. The method's robustness to threshold selection underscores its practical utility in real-world applications where optimal parameters are unknown. The consistent superiority over traditional methods demonstrates that SCEntropy-driven structural discovery transcends parameter tuning to embody a principled approach to conceptual organization. This capability to autonomously derive effective organizational criteria from the data's intrinsic structure represents a significant advance toward AI systems that can develop human-aligned conceptual frameworks without explicit supervision. Most importantly, the finding that a known concept's SCEntropy can guide the organization of unknown data reveals a concrete mechanism for conceptual induction. Here, SCEntropy serves not merely as a clustering parameter, but as a cognitive scaffold. This mechanism enables machines to leverage mastered concepts as a stable foundation for discovering and structuring novel, semantically coherent categories in truly unsupervised settings.

Structuring Cognition: Scaffolding for Coherent Dialogue

Beyond its applications in visual concept organization, we demonstrate that SCEntropy-derived hierarchies serve as fundamental cognitive scaffolds for structuring language generation. This approach addresses a core limitation in current large language models (LLMs): their inherent lack of explicit structural reasoning capabilities[41]. By treating sentences as structural elements and organizing them through SHC, we provide LLMs with explicit representations of conceptual relationships that mirror the hierarchical nature of human thought processes.

Our method reveals that latent hierarchical structures naturally emerge from LLM outputs, and that these structures provide a principled framework for constraining content generation. When new content is required to extend specified nodes within the SHC-derived hierarchy while preserving surrounding structural relationships, the generated text maintains thematic coherence and logical flow across multiple dialogue turns. This structural guidance proves particularly crucial for complex, multi-faceted queries where maintaining coherent reasoning chains is challenging.

As illustrated in Fig. 4, the efficacy of this approach becomes evident when examining content expansion under structural constraints. For the problem of "*How to balance commercial development and cultural authenticity in the digital preservation of cultural heritage?*", the response of the LLM consists of 13 sentences, with each sentence being treated as a separate element. SHC organizes the initial response into a coherent hierarchical structure (Fig. 4(a)). When expanding specific content—here, sentence S12—unconstrained generation (Fig. 4(b)) produces semantically divergent outputs (S13, S14) that disrupt the original conceptual organization, even altering relationships at the root level. This structural disintegration directly induces thematic drift, undermining the logical coherence and argumentative focus of the narrative. In stark contrast, SHC-constrained generation (Fig. 4(c)) produces tightly aligned expansions that preserve the structural integrity of the original response, with only minimal, locally-contained adjustments. This demonstrates that structural complexity entropy provides not merely a technical constraint, but a cognitive framework that guides LLMs toward human-like conceptual organization.



(a) Constructing the content structure for the initial response to the question. 13 sentences (S0-S12) were generated to address it. Using SHC, the content structure of these 13 sentences was organized, with labels L0-L4 representing the intermediate stages of this process.

(b) Content structure formed by newly generated content without structural constraints (baseline), combined with the original content. Expanding the 13th sentence (S12) resulted in sentences S13 and S14, leading to a revised structure. The expanded content diverges from S12, indicating a lack of coherence.

(c) The content structure formed by the newly generated content under structural constraints, combined with the original content. Sentence S12 was expanded into S13 and S14. The content of S13 and S14 is highly relevant to S12, and the original content structure remains intact without disruption.

Fig.4 | Comparison of the content structure generated under the tree-structured constraints based on SHC

Mitigating Semantic Drift: A Quantitative Evaluation

To quantitatively evaluate the efficacy of structural constraints in mitigating semantic drift, we conducted systematic experiments on 100 complex questions spanning diverse domains. We employed a challenging content expansion task: for each initial LLM response, the sentence with the highest perplexity was identified as the target for expansion, thereby focusing the evaluation on the most semantically ambiguous and drift-prone components. The newly generated content was evaluated against human-authored reference expansions using a comprehensive suite of metrics: BERTScore[38] (assessing semantic alignment via Precision, Recall, and F1), METEOR[40] (evaluating semantic faithfulness and fluency), and ROUGE[39] (measuring lexical and phrasal overlap via ROUGE-1, -2, and -L, each with Precision, Recall, and F1).

As definitively summarized in Table 1, the application of SHC-derived structural constraints yielded substantial and consistent improvements across all evaluation dimensions, confirming its role in disciplining generative processes.

The gains in semantic alignment, as measured by BERTScore, were robust and balanced. Precision increased from 0.7504 to 0.7783 (~3.7% improvement), indicating enhanced alignment of the generated content with reference semantics. Recall saw a more pronounced rise from 0.7540 to 0.8126 (~7.8% improvement), reflecting a superior coverage of relevant conceptual information. Consequently, the overall BERTScore F1 improved from 0.7519 to 0.7946 (~5.7% improvement), demonstrating a coherent advancement in semantic fidelity.

The most dramatic improvement was observed in the METEOR score, which surged from 0.1585 to 0.2972—an increase of 87.5%. This leap signifies an exceptional gain in semantic faithfulness and fluency preservation, moving the generated text much closer to human-like language quality. This suggests that structural constraints directly target and ameliorate the incoherent and often tenuous phrasing characteristic of unconstrained LLM expansions.

A nuanced yet highly informative pattern emerged from the ROUGE metrics. ROUGE-1 F1 score improved from 0.2515 to 0.4814 (a 91.4% increase), demonstrating a major enhancement in unigram-level lexical coverage and relevance. The constrained model's Recall (0.5756) substantially outstripped its Precision (0.4273) on this metric, indicating a tendency to generate more comprehensive content that captures relevant ideas, even at the cost of some lexical precision—a trade-off favorable for content expansion.

Most strikingly, the ROUGE-2 F1 score exhibited an extraordinary near-nine-fold increase, from a baseline of 0.0257 to 0.2429 under constraints. This 845% improvement is a pivotal finding, as it specifically measures bigram overlap. It provides strong quantitative evidence that SHC constraints crucially enhance the model's capacity for maintaining phrasal coherence and structured expression, forcing it to build new content that logically connects with existing ideas at a local, syntactic level. Similarly, ROUGE-L F1 improved from 0.2079 to 0.4287 (~106% increase), confirming better long-sequence agreement.

This comprehensive quantitative profile elucidates the mechanism of SHC's effectiveness. The method does not merely induce lexical matching (which would inflate ROUGE Precision alone) but orchestrates a structured conceptual expansion. The exceptional gains in METEOR and ROUGE-2, in particular, underscore that the SCEntropy principle guides the LLM to extend discourse along semantically faithful and phrasally coherent pathways, effectively instituting a form of "cognitive chunking" that curtails meandering and preserves thematic integrity. These results establish that structural complexity entropy provides not only a theoretical framework but also a practical, empirically validated mechanism for instilling structural discipline into generative AI systems.

Table 1. Performance comparison of generated content with and without SHC-based structural constraints

Method	BERTScore			METEOR
	Precision	**Recall**	**F1-score**	
Unconstrained	0.7504	0.7540	0.7519	0.1585
Constraint	0.7783	0.8126	0.7946	0.2972

Method	ROUGE-1			ROUGE-2			ROUGE-L		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
Unconstrained	0.2481	0.2597	0.2515	0.0251	0.0270	0.0257	0.2046	0.2152	0.2079
Constraint	0.4273	0.5756	0.4814	0.2117	0.2999	0.2429	0.3805	0.5137	0.4287

Discussion

Our work establishes SCEntropy as a foundational principle for structural machine intelligence, addressing a fundamental gap in AI's capacity to autonomously reason about information organization. The consistent performance of SCEntropy across visual concept discovery and language generation demonstrates that this metric captures a universal property of structural coherence that transcends specific domains or data modalities.

The quantitative results reveal several profound implications for AI systems. In In visual concept discovery, SHC produces semantically coherent hierarchies with significantly fewer nodes than traditional methods (31 vs. 99 in CIFAR-100) while achieving superior semantic coherence (HSC: 0.745 vs. ~0.566), demonstrating that SCEntropy offers a more natural principle for conceptual organization than purely similarity-driven merging. This aligns with the cognitive science insight that human categorization follows structural coherence. The transferability of SCEntropy thresholds across domains further confirms its generalizability. A key demonstration of SHC's autonomy is its discovery of novel, semantically coherent super-categories. These are evaluated through their high HSC scores and human-aligned interpretability, validated through our proposed framework, as external ground truth is unavailable for novel knowledge structures. This highlights SHC's capacity for autonomous conceptual abstraction.

Beyond these quantitative advances, the SCEntropy principle offers a mechanistic, computational account for the cognitive naturalness of certain groupings. We posit that the drive to minimize relational dissonance, as formalized by SCEntropy, mirrors a fundamental cognitive operation aimed at reducing mental effort. A cluster with low SCEntropy constitutes a stable 'conceptual chunk' that can be efficiently processed as a single unit, whereas a high SCEntropy set corresponds to a cognitively taxing assembly of disparate elements. Thus, our framework establishes a quantifiable link between information-theoretic structure and the subjective experience of conceptual coherence, potentially informing future studies on the neural correlates of concept formation in biological intelligence.

Most profoundly, our results establish that the transferability of SCEntropy constitutes a form of conceptual bootstrapping. This phenomenon occurs when the signature of a single known concept guides the organization of an entire conceptual space. This provides a compelling case that SCEntropy captures a fundamental "structural grammar" for knowledge. While the current work demonstrates this principle within the constrained ontology of benchmark datasets, it provides a pathway for AI systems to move beyond organizing predefined categories, and toward genuine open-world conceptual discovery: the identification and definition of truly novel categories in entirely unknown data.

In natural language generation, our results reveal how explicit structural awareness can directly discipline and enhance generative processes. The quantitative evaluation demonstrates that SHC-derived constraints do not merely nudge lexical output but induce a systematic shift toward higher-quality discourse. The 87.5% improvement in METEOR score signifies a leap in semantic faithfulness and fluency, directly countering the vague or inconsistent phrasing typical of unconstrained expansions. Crucially, the extraordinary near-nine-fold increase in ROUGE-2 F1-score (from 0.0257 to 0.2429) is particularly telling: it demonstrates a dramatic gain in phrasal coherence, indicating that the model learns to build new content that connects logically at a local syntactic level. This pattern of “constrained creativity”, where thematic consistency is maintained alongside appropriate lexical variation, successfully mitigates the semantic drift that plagues multi-turn LLM generation. It provides a concrete mechanism by which explicit hierarchical representations can scaffold coherent long-form reasoning.

Our findings collectively suggest that a key limitation of current AI systems in advanced reasoning may stem not from a deficit in pattern recognition, but from a lack of principled mechanisms for structural self-organization and self-discipline. By introducing SCEntropy as a universal measure of structural disorder and SHC as an algorithm for complexity-constrained building, we provide AI with the tools to autonomously develop and utilize hierarchical representations that reflect the intrinsic architecture of information. This dual advance in structural discovery (inductive organization) and structural control (deductive guidance) enables a more holistic form of machine intelligence—one that does not merely process data but comprehends and manipulates the relational fabric that gives data meaning.

Considerations and Future Directions

Several considerations and limitations warrant attention, pointing to fruitful avenues for future research.

First, we address the choice of baseline comparisons. In this work, we selected classical agglomerative clustering methods as our primary baselines. This decision was strategic: these algorithms are well-understood, and their fundamental operating principle, which is greedy, pairwise merging based on local similarity, stands in direct, clear contrast to SHC's principle of global structural complexity management for multi-branch growth. This contrast allows for a purer evaluation of the intrinsic value of the SCEntropy principle itself.

Furthermore, this principled approach distinguishes SCEntropy from other contemporary paradigms for structural discovery. For instance, neuro-symbolic methods[12] often integrate neural networks with symbolic reasoning and knowledge graphs, requiring pre-defined logical rules or schema. In contrast, SCEntropy provides a foundation for structural discovery that is purely based on information-theoretic measurement, emerging directly from the data's relational fabric without requiring symbolic representations. Similarly, while deep clustering methods[42] excel at learning powerful representations, their hierarchical construction often remains reliant on sequential, binary merging paradigms. SHC complements these approaches by introducing a fundamentally different organizational principle, namely global structural complexity management. This principle could be integrated with learned representations in future work to jointly optimize for feature learning and coherent structure formation.

While deep clustering methods have achieved impressive results by learning powerful representations, the hierarchical construction process in many of these approaches still often relies on a sequential, binary merging paradigm or post-hoc processing of learned flat clusters. A highly valuable direction for future work is the integration of the SCEntropy principle with deep representation learning, potentially creating architectures that learn features optimized for structural coherence from the outset.

Second, the computational complexity of the current SCEntropy calculation, which scales with the fourth power of the set size $O(N^4)$, presents a challenge for scaling to extremely large datasets (A detailed Analysis of computational complexity is provided in the Supplementary Information (Section 3.2)). This is a recognized trade-off for achieving high-fidelity structural discovery. We envision several paths to mitigate this. Future work will explore efficient approximations, such as sampling-based estimation of the similarity variance or leveraging low-rank approximations of the similarity matrix. Furthermore, in many practical applications, such as the class-level clustering and sentence-level structuring demonstrated here, the number of conceptual elements is manageable. The scalability challenge thus does not diminish the utility of SCEntropy as a powerful tool for understanding and organizing medium-scale conceptual spaces, a critical step in developing structural reasoning.

Beyond these considerations, the current formulation of SCEntropy relies on second-order relational patterns, potentially missing higher-order interactions that could further enrich structural understanding. Developing more sophisticated complexity measures that capture these higher-order relationships would provide a more comprehensive framework for structural analysis.

Future work should explore the integration of SCEntropy with neural-symbolic reasoning frameworks, where structural representations could guide both perceptual processing and logical inference. A promising direction for developing more generally intelligent systems is the application of these principles to other AI domains, particularly reinforcement learning for hierarchical task decomposition and knowledge graph organization. Furthermore, investigating the neural correlates of structural complexity entropy in biological intelligence could provide insights into the fundamental principles of structured information processing across natural and artificial systems.

In conclusion, SCEntropy provides a unified framework for quantifying, discovering, and utilizing hierarchical organization in complex information. By enabling AI systems to autonomously reason about structural relationships, we move closer to machines that not only recognize patterns but understand how those patterns are organized—a crucial step toward artificial intelligence that truly comprehends, rather than merely processes, the complex structure of information in our world.

Methods

Experimental Setup

Image Clustering Experiments: We evaluated our SHC algorithm on the FashionMNIST[33], CIFAR-10[34], and CIFAR-100[34] datasets to demonstrate its efficacy in visual concept discovery. As baseline methods for comparison, we employed standard agglomerative hierarchical clustering with Ward, Average, and Complete linkage criteria[36]. The primary metric for evaluating the quality of the resulting conceptual hierarchies was our proposed Hierarchical Semantic Coherence (HSC) score, defined in Equation (4) of the main text.

Language Generation: A dataset of 100 complex questions spanning diverse domains such as technology, culture, and ethics was constructed. The initial response for each question was generated by calling the Zhipu.AI (GLM-4, https://www.zhipuai.cn/en/) API without any specific prompting. The generated content was then evaluated using BERTScore[43] (reporting Precision, Recall, and F1), ROUGE[44] (1, 2, L), and METEOR[45] to systematically compare outputs generated with and without structural constraints derived from SHC. For the constrained generation tasks, the DeepSeek-R1 model was utilized.

Feature Extraction and SCEntropy Calculation

For feature extraction in the image clustering experiments, we used a ResNet-50 model pre-trained on ImageNet[46], which was subsequently fine-tuned on each specific experimental dataset (FashionMNIST, CIFAR-10, CIFAR-100). After fine-tuning, features for each image were extracted using this model. The feature representation for an entire image class was then defined as the centroid (mean) of the feature vectors of all images within that class. For the natural language generation experiments, each sentence was treated as a distinct element, and its feature vector was obtained using the Sentence-BERT model[47], specifically the all-MiniLM-L6-v2 variant.

The normalized similarity $S_{i,j}$ between two elements (whether image classes or sentences) with feature vectors $V_i$and V_j was calculated based on cosine similarity:

$$S_{i,j} = 0.5 \times \left(1 - cosine\left(V_i, V_j\right)\right). \tag{5}$$

This formulation ensures that $S_{i,j}$ lies within the interval [0, 1], meeting the requirement for SCEntropy calculation. The Structural Complexity Entropy (SCEntropy) for a set was then computed according to Equation (1) in the main text. For computational efficiency, the calculation iterated only over the upper triangular part of the similarity matrix to avoid redundant computations.

Selection of the SCEntropy Threshold

The SCEntropy threshold H_{th} is a key parameter governing the granularity of the hierarchies produced by SHC. For the image clustering experiments on FashionMNIST, CIFAR-10, and CIFAR-100, where clustering was performed on the class-level features (centroids), the thresholds were set as follows: $H_{th} = 0.4$ for FashionMNIST and CIFAR-10, and $H_{th} = 6$ for CIFAR-100. These values were determined by a principled heuristic: we identified a putative "minimal cluster" (a coherent group of classes) and calculated its SCEntropy.

In the second image clustering experiment on CIFAR-100 (Fig. 3), we aimed to validate that a universal principle, derived from the data itself, could effectively govern the construction of the global hierarchy. To this end, we did not use a single fixed threshold. Instead, for each of the 20 human-defined superclasses (e.g., "fish", "flowers"), we calculated its intrinsic SCEntropy, $H_{superclass}$, as the global clustering threshold, which represents the structural complexity of a known, coherent semantic concept.

For the natural language generation experiments, the SCEntropy threshold was empirically set to a fixed value of $H_{th} = 1$.

Text Generation with Structural Constraints

The process for applying SHC-derived structural constraints to guide text generation involved the following steps:

Initial Response and Structure Building: The initial answer to a query was generated by the base LLM (GLM-4). The SHC algorithm was then applied to the sentence embeddings of this response to construct an initial hierarchical content structure tree.

Target Sentence Selection: For each question, we calculated the perplexity (using a GPT-2 model) for each sentence in the initial response. The sentence with the highest perplexity was selected as the target for content expansion.

Constrained Expansion and Iterative Refinement:

a. Initial Expansion: The target sentence was expanded using the LLM (DeepSeek-R1) with the prompt:

"Please ensure the generated expansion sentences maintain similarity to the original sentence and guarantee semantic consistency."

b. Structural Coherence Check: The newly generated sentences were combined with the original text, and the SHC algorithm was run again to build a new content structure. The critical check was whether the newly generated sentences shared a direct common parent node (one hop up in the hierarchy) with the original target sentence in this new structure.

c. Iterative Correction: If the common parent condition was not met, indicating semantic drift, a correction prompt was used to guide the LLM:

"Please avoid repeating the following phrasing, and generate content that is semantically closer to the target sentence: {previous_content[-1]}"

Here, previous_content[-1] refers to the last generated sentence that failed the check. This process (steps a-c) was repeated iteratively until all newly generated sentences shared a direct common parent node with the original target sentence, ensuring the preservation of thematic coherence within the discovered hierarchical structure.

Code Repository Declaration

All code necessary to reproduce the experiments, implement the Structural Complexity Entropy (SCEntropy) metric, and run the Structural Hierarchical Clustering (SHC) algorithm is publicly available at: https://github.com/SCEntropy/SCEntropy_master

References

1. Noor, K. T., Robles-Kelly, A., Zhang, L. Y. *et al.* A consistency-aware deep capsule network for hierarchical multi-label image classification. *Neurocomputing* 604, 128376 (2024).
2. Ruffini, G. & Castaldo, F. Structured Dynamics in the Algorithmic Agent. *Entropy* 27, 90 (2025).
3. Liu, B., Li, X., Zhang, J. *et al.* Advances and challenges in foundation agents: From brain-inspired intelligence to evolutionary, collaborative, and safe systems. Preprint at https://arxiv.org/abs/2504.01990 (2025).
4. Ren, Y., Pu, J., Yang, Z. *et al.* Deep clustering: A comprehensive survey. *IEEE Trans. Neural Netw. Learn. Syst.* 36, 5858–5878 (2024).
5. Müllner, D. Modern hierarchical, agglomerative clustering algorithms. Preprint at https://arxiv.org/abs/1109.2378 (2011).
6. Ackermann, M. R., Blömer, J., Kuntze, D. *et al.* Analysis of agglomerative clustering. *Algorithmica* 69, 184–215 (2014).
7. Tokuda, E. K., Comin, C. H. & Costa, L. F. Revisiting agglomerative clustering. *Physica A* 585, 126433 (2022).
8. Griffiths, A., Robinson, L. A. & Willett, P. Hierarchic agglomerative clustering methods for automatic document classification. *J. Doc.* 40, 175–205 (1984).
9. Murtagh, F. & Legendre, P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J. Classif.* 31, 274–295 (2014).

10. Murtagh, F. & Contreras, P. Algorithms for hierarchical clustering: an overview. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 2, 86–97 (2012).
11. Wang, X. *et al.* HiBERT: detecting the illogical patterns with hierarchical BERT for multi-turn dialogue reasoning. *Neurocomputing* 524, 167–177 (2023).
12. Oyewole, G. J. & Thopil, G. A. Data clustering: application and trends. *Artif. Intell. Rev.* 56, 6439–6475 (2023).
13. Huang, H., Wang, C., Wei, X. & Zhou, Y. Deep image clustering: a survey. *Neurocomputing* 599, 128101 (2024).
14. Mittal, H., Pandey, A. C., Saraswat, M. *et al.* A comprehensive survey of image segmentation: clustering methods, performance parameters, and benchmark datasets. *Multimed. Tools Appl.* 81, 35001–35026 (2022).
15. Kauffmann, J. *et al.* From clustering to cluster explanations via neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* 35, 1926–1940 (2024).
16. Dong, C., Li, Y., Gong, H. *et al.* A survey of natural language generation. *ACM Comput. Surv.* 55, 1–38 (2022).
17. Ji, Z., Lee, N., Frieske, R. *et al.* Survey of hallucination in natural language generation. *ACM Comput. Surv.* 55, 1–38 (2023).
18. Ikotun, A. M., Ezugwu, A. E., Abualigah, L. *et al.* K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Inf. Sci.* 622, 178–210 (2023).
19. Ghazal, T. M. Performances of k-means clustering algorithm with different distance metrics. *Intell. Autom. Soft Comput.* 30, 735–742 (2021).
20. Li, T., Rezaeipanah, A. & El Din, E. S. M. T. An ensemble agglomerative hierarchical clustering algorithm based on clusters clustering technique and the novel similarity measurement. *J. King Saud Univ. Comput. Inf. Sci.* 34, 3828–3842 (2022).
21. Ran, X., Xi, Y., Lu, Y. *et al.* Comprehensive survey on hierarchical clustering algorithms and the recent developments. *Artif. Intell. Rev.* 56, 8219–8264 (2023).
22. Asghar, S., Pei, Q. X., Volpe, G. *et al.* Efficient rare event sampling with unsupervised normalizing flows. *Nat. Mach. Intell.* 6, 1370–1381 (2024).
23. Valerdi, R. & Kohl, R. J. The concept of problem complexity. *Procedia Comput. Sci.* 28, 539–546 (2014).

24. Jumarie, G. Structural entropy, information potential, information balance and evolution in self-organizing systems. *Int. J. Syst. Sci.* 5, 923–942 (1974).
25. Sommaruga, G. (ed.) Formal theories of information: from Shannon to semantic information theory and general concepts of information. (Springer-Verlag, Berlin, 2009).
26. Siai, O., Idoudi, H., Scheuermann, D. *et al.* On measuring structural complexity of a conceptual model. *Proc. Second IASTED Int. Multi-Conf. Autom. Control Inf. Technol.* 193–198 (2005).
27. Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97 (1956).
28. Goldstone, R. L. The role of similarity in categorization: Providing a groundwork. *Cognition* 52, 125–157 (1994).
29. Triantafyllidis, E., Acero, F., Liu, Z. *et al.* Hybrid hierarchical learning for solving complex sequential tasks using the robotic manipulation network roman. *Nat. Mach. Intell.* 5, 991–1005 (2023).
30. Wang, Z., Xu, Y., Wang, D. *et al.* Hierarchical deep reinforcement learning reveals a modular mechanism of cell movement. *Nat. Mach. Intell.* 4, 73–83 (2022).
31. Yuan, K., Sajid, N., Friston, K. *et al.* Hierarchical generative modelling for autonomous robots. *Nat. Mach. Intell.* 5, 1402–1414 (2023).
32. Kim, J. I., Baek, J. W. & Kim, C. B. Hierarchical image classification using transfer learning to improve deep learning model performance for amazon parrots. *Sci. Rep.* 15, 3790 (2025).
33. Smith, R., Johnson, P. & Williams, K. Structural analysis and decomposition of complex systems. *Annu. Rev. Control Robot. Auton. Syst.* 1, 123–145 (2017).
34. Anderson, M. L. Coherent information structure in complex computation. *Cogn. Sci.* 36, 1425–1447 (2012).
35. Kauffmann, J., Dippel, J., Ruff, L. *et al.* Explainable AI reveals Clever Hans effects in unsupervised learning models. *Nat. Mach. Intell.* 7, 1–11 (2025).
36. James, R. G., Emenheiser, J. & Crutchfield, J. P. Reconsidering unique information: Towards a multivariate information decomposition. *2019 IEEE Int. Symp. Inf. Theory (ISIT)* 2554–2558 (2019).
37. Navarro, D. J. Similarity, distance, and categorization: A discussion of Smith's (2006) warning about "colliding parameters". *Psychon. Bull. Rev.* 14, 823–833 (2007).

38. Chen, Y., Wang, Z. & Li, J. The Standard Deviation Score: a novel similarity metric for data analysis. *J. Big Data* 12, 45 (2025).

39. Xiao, H., Rasul, K. & Vollgraf, R. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms. Preprint at https://arxiv.org/abs/1708.07747 (2017).

40. Krizhevsky, A. & Hinton, G. Learning multiple layers of features from tiny images. *University of Toronto Tech. Rep.* (2009).

41. Shumailov, I., Shumaylov, Z., Zhao, Y. *et al.* AI models collapse when trained on recursively generated data. *Nature* 631, 755–759 (2024).

42. Ward, J. H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58, 236–244 (1963).

43. Zhang, T., Kishore, V., Wu, F. *et al.* BERTScore: Evaluating text generation with BERT. Preprint at https://arxiv.org/abs/1904.09675 (2019).

44. Lin, C.-Y. ROUGE: A package for automatic evaluation of summaries. *Proc. ACL 2004 Work. Text Summar. Branches Out* 74–81 (2004).

45. Banerjee, S. & Lavie, A. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. *Proc. ACL Work. Intrinsic Extrinsic Eval. Meas. Mach. Transl. Summar.* 65–72 (2005).

46. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 770–778 (2016).

47. Reimers, N. & Gurevych, I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proc. 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process.* 3982–3992 (2019).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryinformation.docx