Aicumene:2025.043v2 [cs.AI] October 24, 2026

What is Strong AI?
Intrinsic Structural Criteria for Strong Intelligence Based on Dimensionality and Topology

Working paper, version 2 (October 2026)

Alexandra Bernadotte
Laboratory of Analysis of Semantics,
Centre for Language and Semantics Technologies,
Faculty of Computer Science,
HSE University
bernadotte@neurosputnik.com

Vasilii A. Gromov
Laboratory of Analysis of Semantics,
Centre for Language and Semantics Technologies,
Faculty of Computer Science,
HSE University
stroller@rambler.ru

October 24, 2026

Abstract

We propose a system of internally defined, instrumentally measurable criteria for distinguishing levels of intelligence, natural or artificial. Let $\mathcal{I}$ denote an intelligence system and $\mathcal{D}(\mathcal{I})$ the ensemble of data it generates under standardized observational conditions. Unlike behavioural definitions of Artificial General Intelligence (AGI), which aggregate performance over a predefined set of tasks, the proposed classification is defined through intrinsic structural invariants of $\mathcal{D}(\mathcal{I})$: intrinsic dimension, topological organization, and dynamical complexity. We show that performance on any finite benchmark imposes neither a lower nor an upper bound on these invariants (Proposition 1), so generality and strength are logically independent properties of a system.

Combining the generic embedding bound of Whitney–Takens type with a representational-closure hypothesis yields a dimension ladder $d_k = 2^{k+2} - 1$ $(3, 7, 15, 31, \dots)$ whose intervals define a hierarchy $\{\mathsf{Level}_k\}_{k\geq 0}$: Level_0 is Weak Intelligence (conventional narrow AI and animal cognition in the physical-world regime), Level_1 is Strong Intelligence (natural or artificial), Level_2 is Super-Strong Intelligence, and higher levels extend the hierarchy inductively. The ladder is a hypothesis built on a theorem, and we state precisely which part is which.

Published measurements place human language (70 languages from 21 families) and healthy human brain dynamics in the same regime, with non-integer intrinsic dimension close to nine, whereas machine-generated text occupies a lower-dimensional manifold, by roughly 1.5 dimensions in an independent study, and pathological brain states fall below eight. These data support an operational definition of Strong AI as a system whose generated data are structurally indistinguishable from those of Strong Natural Intelligence, and they place current large language models, as measured through their textual output, in Level_0 despite their benchmark performance. We give a measurement protocol based on a structural profile $S(\mathcal{I}) = (\mathrm{ID}, \mathrm{Top}, \mathrm{Dyn})$, falsification conditions, and limitations of the proposal.

Keywords Strong AI · Artificial General Intelligence · Super-Strong AI · Weak AI · Intrinsic Dimension · Data Topology · Complexity · Whitney Embedding Theorem · Takens Embedding · Cognition · Non-anthropocentric Intelligence

1 Introduction

Strong AI (SAI) denotes a regime of artificial intelligence that matches or exceeds human intellectual capacity across all cognitively accessible domains, including reasoning, planning, and generative creativity. Despite its centrality, the

Aicumene:2025.043v2 [cs.AI] October 24, 2026

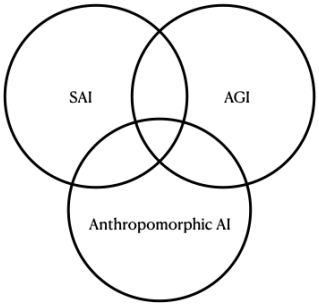

Figure 1: Schematic representation of three notions that are often collapsed: Strong AI (SAI), Artificial General Intelligence (AGI), and anthropomorphic AI (AAI). None of the three logically entails the others.

concept of SAI remains poorly delimited and is frequently conflated with Artificial General Intelligence (AGI), which is typically defined as a system exhibiting a broad spectrum of human-like cognitive abilities.

We argue that this identification is fundamentally flawed. AGI, as commonly defined, is a behavioural and functional notion, grounded in performance across a predefined set of cognitive tasks. In contrast, SAI should be understood as a structural and dynamical property of intelligence, independent of any anthropocentric characteristics.

On the other hand, AI may be strong without being anthropomorphic. Neither Strong AI nor AGI necessarily implies anthropomorphic artificial intelligence (AAI), while AGI may exhibit both anthropomorphic and non-anthropomorphic properties. This observation motivates a non-anthropocentric perspective on intelligence, in which strong intelligence is defined through intrinsic structural properties rather than similarity to human cognition (Fig. 1).

In the present work, we assume that the development of artificial intelligence will proceed along lines different from the simple scaling of either human intelligence or existing AI systems. We depart from scaling-based paradigms and advance the hypothesis that intelligence is characterized by the statistical and geometric properties of the data it generates, and by the complexity and structure of the informational objects that intelligence is designed to process. Consequently, intelligence can be operationally defined and quantitatively assessed through measurable properties of these data; conversely, qualitative changes in these characteristics may indicate a transition to a higher level of intelligence.

We propose that three fundamental invariants determine the level of intelligence:

- the intrinsic dimensionality of generated data and of the semantic spaces in which the system operates;
- the topological organization of these semantic spaces;
- the dynamical complexity of trajectories realized within them.

These invariants define a hierarchy of intelligence regimes and provide a principled criterion for transitions between them.

The dimension ladder. Animal intelligence is an adaptive mechanism optimized for survival and navigation within a three-dimensional physical environment. The survival of an animal critically depends on its ability to internally represent and effectively handle three-dimensional objects: the difference between a sleeping panther and one preparing to leap may be small, yet it is of crucial importance for a hare. The Whitney embedding theorem [1, 2, 3] guarantees that any smooth n-dimensional manifold embeds in $\mathbb{R}^{2n+1}$, and the Takens–Sauer reconstruction theorems [4, 5] show that the same number of generic observables suffices to reconstruct an n-dimensional state space from measurements. Taking $n = 3$ yields $2 \times 3 + 1 = 7$ as the dimension of a semantic space that can host generic representations of three-dimensional objects; iterating the construction yields $2 \times 7 + 1 = 15$ and then 31. Human intelligence operates over abstract semantic constructs that are not reducible to physical navigation, so its semantic spaces should exceed the animal bound, and intelligence exceeding the next bound naturally corresponds to super-strong intelligence (Fig. 2).

We stress at the outset the logical status of this construction. Whitney's theorem is a theorem. The step from "a level-k system must represent generic configurations of level-$(k-1)$ objects" to "its semantic space has dimension at least $2d_{k-1} + 1$" is a hypothesis (Hypothesis 1 in Section 3.4). The ladder is thus a hypothesis built on a theorem, and the paper is organized so that the two can be assessed separately; Section 6 lists what would falsify the hypothesis.

Aicumene:2025.043v2 [cs.AI] October 24, 2026

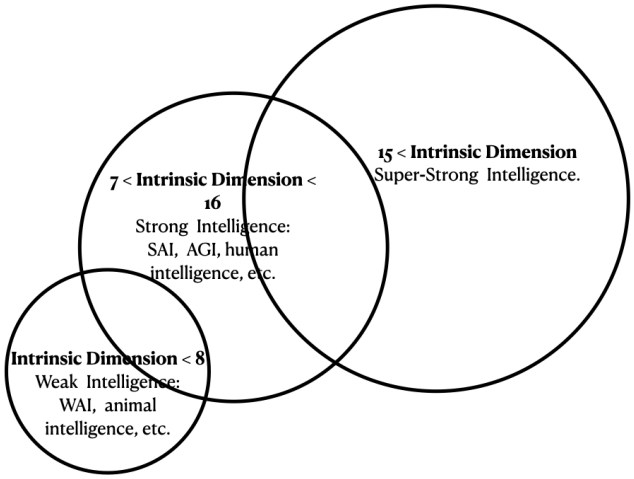

Figure 2: Hierarchy of increasingly powerful forms of intelligence. WAI: weak (narrow) AI; SNI: strong natural intelligence; SAI: strong AI; SSAI: super-strong AI. The partially ordered scheme implies the existence of transitional regimes.

Crucially, this hierarchy is non-anthropocentric: it does not rely on human-specific cognitive features but arises from general geometric and dynamical constraints on information-processing systems. The partially ordered scheme of Fig. 2 implies the existence of transitional regimes, i.e., systems capable of operating across multiple levels. In this sense, intelligence may be viewed as a structured continuum, with transitions between regimes analogous to phase transitions in complex systems: qualitative changes in the intrinsic structural properties of the generated data rather than merely quantitative improvements in task performance.

Beyond dimensionality, the topology of semantic spaces provides an independent axis of classification. The topological analysis of structures arising in data during brain activity and in the evolution of language enables the characterization of the complexity of semantic spaces in which intelligence operates (Fig. 3). Different types of intelligence correspond to different sets of homological features and distinct topological classes. Finally, intelligence manifests dynamically through trajectories evolving over these semantic manifolds; the complexity of such trajectories (quantified via predictability, entropy, and chaos indicators) constitutes a third, independent criterion.

Empirical anchors. These quantities are empirically accessible. Measurements of intrinsic dimensionality for both human language and human brain signals yield values on the order of 9 [8, 17, 18, 19], indicating that human intelligence statistically exceeds the animal regime while remaining below the threshold associated with super-strong intelligence (this is a statistical observation; for some individuals or states the value may be higher). An independent study of human-written and AI-generated texts [10] reports the same value, ≈ 9, for human texts in several alphabet-based languages and a value ≈ 1.5 lower for texts generated by language models, with a clear statistical separation between the two distributions. These observations support the existence of a hierarchy of intelligence levels and place current language models, as measured through their textual output, below the human reference.

Demarcation from AGI. This perspective necessitates a strict demarcation between SAI and AGI. The framework "A Definition of AGI" [41] provides a comprehensive behavioural benchmark, quantifying performance across a set of cognitive tasks and introducing an "AGI score" as an aggregate metric; the "Levels of AGI" of Morris et al. [39] grade systems by performance and generality. While these approaches are methodologically sound, they capture functional competence, not structural complexity. Proposition 1 makes the point formally: the restriction of $\mathcal{D}(\mathcal{I})$ to any finite benchmark is a finite set, of dimension zero, and systems of arbitrary structural level can pass or fail the same benchmark. An AI system may therefore achieve high performance across all such benchmarks without undergoing a qualitative transition in its underlying informational geometry. AGI and SAI are fundamentally different notions: the former is defined operationally through task performance, the latter intrinsically through the geometry and dynamics of generated information.

Aicumene:2025.043v2 [cs.AI] October 24, 2026

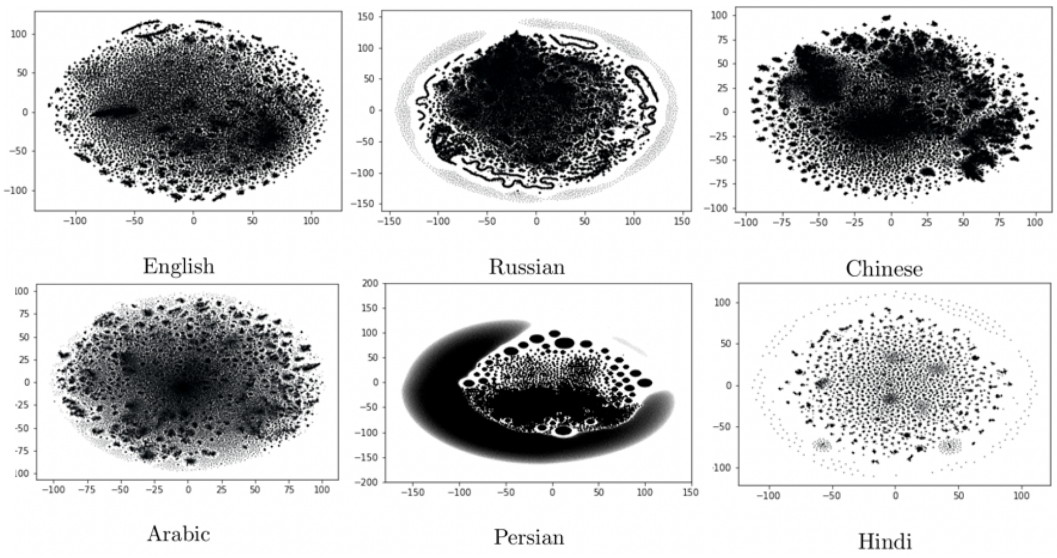

Figure 3: Two-dimensional projections of language embedding manifolds for several languages (adapted from the analyses cited in Section 4.1). The projection illustrates large-scale structure; it is not itself the dimension estimator.

What this paper does not claim. The proposal is not a test for consciousness: a structural increase in the complexity of generated data does not establish phenomenal experience, and we treat consciousness as a separate axis (Section 5.5). The numerical thresholds 7 and 15 are conjectural boundaries motivated by a theorem; the value ≈ 9 is an empirical observation. Finally, what is measured is a projection of the system onto the observed data channel; a low-dimensional output channel may conceal a richer internal process (Section 6).

Contributions of this version. Relative to the first version of this preprint (Aicumene:2025.043v1, August 2025), the present version (i) separates the theorem from the hypothesis in the dimension ladder and gives the ladder in closed form, with a robustness check against the alternative embedding bound (Section 3.4); (ii) proves that benchmark performance does not bound structural level (Proposition 1); (iii) formulates the operational definition of Strong AI through a structural profile and a standardized measurement protocol (Section 3.5); (iv) collects the published evidence, including an independent corroboration of the language result, in Table 1 and Fig. 5; and (v) adds falsification conditions and limitations (Section 6).

The remainder of the paper is organized as follows. Section 2 reviews related work: definitions of intelligence and AGI, current AI systems, and intrinsic dimension as a measurable quantity in machine learning and neuroscience. Section 3 introduces the structural framework: fundamental characteristics, the structural profile, the benchmark-independence proposition, the dimension ladder, and the operational definition of Strong AI. Section 4 presents the empirical evidence from language and brain data and places current systems on the ladder. Section 5 discusses the relationship between language and cognition, super-strong intelligence, co-evolution, and consciousness. Section 6 states falsification conditions and limitations, and Section 7 concludes.

2 Related Work

2.1 Definitions of intelligence and of AGI

The question “Can machines think?” entered the modern discussion in explicitly operational form: Turing’s imitation game [34] displaced a difficult definition of thinking with an observable test of performance. The term “strong AI” was introduced by Searle [35] for the thesis that an appropriately programmed computer literally has a mind; his Chinese Room argument challenged the inference from successful symbol manipulation to understanding. We do not adopt the Chinese Room as a test of intelligence, and our use of “strong” is structural rather than metaphysical, but the argument exposes a distinction that remains relevant: externally successful transformation and internally meaningful organization are not identical descriptions [40].

Aicumene:2025.043v2 [cs.AI] October 24, 2026

Formal definitions of machine intelligence have remained behavioural. Legg and Hutter [36] define universal intelligence as expected performance over all computable environments, weighted by simplicity; Chollet [37] proposes skill-acquisition efficiency over a scope of tasks; Goertzel [38] reviews the AGI concept as generality of competence; Morris et al. [39] grade "Levels of AGI" by depth (performance) and breadth (generality) of capabilities; Hendrycks et al. [41] define AGI in relation to the cognitive versatility and proficiency of a well-educated adult and aggregate performance across psychometrically motivated domains into an AGI score. All of these index intelligence to a task taxonomy. The diversity of interpretations of weak and strong AI can, in practice, be reduced to two definitional strategies.[1]

A common approach to defining Strong AI proceeds by analogy with the only empirically available instance of strong intelligence, namely Strong Natural Intelligence (SNI), conventionally associated with humans. Within this paradigm, SAI is characterized through a set of cognitive properties derived from human intelligence.

Definition 1 (Behavioural Strong AI) *Strong AI is an artificial system capable of understanding, learning, and applying knowledge across domains, adapting to novel situations, performing abstract reasoning, planning, and exhibiting forms of self-reflection at a level comparable to, or exceeding, human intelligence.*

Complementarily, Weak AI (or Narrow AI) is defined via restriction:

Definition 2 (Weak AI) *Weak AI (Narrow AI) refers to an artificial system designed to perform a specific task or a limited class of tasks. Such systems operate within predefined domains and lack the capacity for generalization, transfer of knowledge, or autonomous adaptation beyond their training regime.*

While these definitions are widely adopted, they remain fundamentally descriptive and anthropocentric. They enumerate observable capabilities but do not provide operational criteria that distinguish qualitatively different regimes of intelligence. Indeed, the key characteristics commonly attributed to Strong AI (generalization, problem solving, contextual understanding, self-learning, and even forms of reasoning) are neither necessary nor sufficient conditions for Strong Intelligence: these properties can, in principle, be simulated or approximated within systems that remain structurally confined to a limited semantic regime.

2.2 Available AI systems and the scaling hypothesis

A dominant line of research assumes that SAI will emerge as a result of the extensive scaling of large language models (LLMs). This paradigm is supported by rapid advances in frontier models [42, 43] and by a growing body of work demonstrating emergent reasoning capabilities in large-scale architectures [44, 45]. These developments are often interpreted as evidence of a gradual approach toward human-level intelligence.

Contemporary systems, including advanced LLMs such as GPT, Gemini, and Claude, represent highly sophisticated instances of Weak AI in the sense of Definition 2. More recent developments, often referred to as Large Reasoning Models, including DeepSeek-R1 and the reasoning-enhanced variants of the GPT, Claude, and Gemini families, demonstrate significantly improved performance in tasks requiring multi-step inference and structured problem solving.

This interpretation nevertheless remains conceptually insufficient. In contemporary LLMs, reasoning is primarily instantiated as the capacity to perform formal transformations over symbolic or linguistic structures. While such capabilities are non-trivial and empirically powerful, they do not, by themselves, constitute evidence of Strong AI. The manipulation of abstract representations, whether in LLMs or in classical symbolic AI, remains fundamentally combinatorial in nature. High apparent complexity can therefore arise without a corresponding increase in the intrinsic semantic capacity of the system. By analogy, the ability to assemble complex configurations of building blocks does not imply the structural understanding required for architectural design. The decisive factor is not the sophistication of transformations per se, but the capacity of the semantic space in which these transformations occur: reasoning is constrained by the geometry and intrinsic dimensionality of the underlying representational manifold.

Empirical evidence supports this distinction. In "Spot the Bot" [9], it was shown that the intrinsic dimensionality of text generated by state-of-the-art language models remains limited despite substantial increases in surface-level fluency and coherence: machine-generated text occupies a lower-dimensional manifold than human-generated language and is statistically distinguishable from it. This suggests that scaling alone leads to denser sampling within a constrained semantic space rather than to a qualitative expansion of that space. The term Strong AI is frequently used interchangeably with AGI, which shifts the focus toward generality across tasks; but generality and strength are not equivalent notions. A system may be highly general yet remain structurally limited, while a Strong AI system may, in principle, be highly

[1] The conceptual difficulty of the present problem arises, first and foremost, at the level of definition. Adapting an insight often attributed to Karl Popper, one may argue that the ability to define a concept is a prerequisite for understanding it. Definitions are not merely formal constructs; they determine the scope, methodology, and validity of scientific inquiry.

Aicumene:2025.043v2 [cs.AI] October 24, 2026

specialized but operate at a level of complexity comparable to SNI. Within this context, Strong AI should be understood not as an extrapolation of current capabilities but as a transition to qualitatively different semantic spaces with higher intrinsic dimensionality, richer topology, and more complex dynamical organization.

2.3 Intrinsic dimension and topology as measurable quantities

Intrinsic dimension is a well-developed measurable quantity. Pestov [46] axiomatizes it as a functional on metric measure spaces satisfying concentration, smooth dependence on the data, and normalization. Finite-sample estimators include the maximum-likelihood estimator of Levina and Bickel [47], the two-nearest-neighbour estimator of Facco et al. [48], graph-based estimators built on the minimum spanning tree, whose power-weighted length scales with sample size as $N^{(d-\gamma)/d}$ [49, 50], and the mutual k-nearest-neighbour graph estimators of Brito et al. [51]. In machine learning, intrinsic dimension has been measured for data and for learned representations: natural image datasets have intrinsic dimension of a few tens, far below their pixel dimension, and lower-dimensional data are learned more easily [52]; hidden representations in deep networks occupy manifolds of intrinsic dimension much smaller than the number of units, with a characteristic profile across layers [53]. For text, Tulchinskii et al. [10] use the persistent-homology dimension of token-embedding clouds and find that fluent human texts have intrinsic dimension ≈ 9 for several alphabet-based languages and ≈ 7 for Chinese, while AI-generated texts are ≈ 1.5 lower, a property robust enough to serve as a model-agnostic detector.

In neuroscience, the dimensionality of population activity has become a central descriptor. Stringer et al. [54] show that visual-cortex population responses to natural images have a power-law spectrum whose dimensionality grows with the stimulus ensemble, while Chaudhuri et al. [55] recover a one-dimensional ring manifold from the population activity of the head-direction circuit, a circuit devoted to a single physical variable. Both findings fit the picture developed below: circuits representing physical variables are low-dimensional, and the dimensionality of neural data must be read relative to the semantic content of the task. The reconstruction of state spaces from scalar or multichannel time series rests on the delay-embedding theorems of Takens [4] and Sauer, Yorke, and Casdagli [5], which supply the same $2n + 1$ bound as the Whitney theorem and tie it to observability; the multichannel singular-spectrum methods used for brain data in Section 4.2 belong to this tradition [15, 16, 27, 33].

3 A Structural Framework for Levels of Intelligence

The above considerations necessitate the introduction of new definitions of strong and weak AI, as well as the identification of approaches that allow the detection of their manifestations in real-world intelligent systems. Moving from the concepts of intelligence and consciousness, which are difficult to formalize, we go into the area of data science, where we can apply formal fundamental characteristics of data: data structure (data topology), intrinsic dimension, and complexity. Language, visual images, and other products of Strong Intelligence are justifiably classified as complex systems, or more precisely as dynamical systems, and we therefore speak of the intrinsic dimension of dynamical systems.

3.1 Fundamental characteristics

The concepts of intrinsic dimension, data topology, and complexity can be defined in various ways, depending on assumptions about the nature of the data.

Definition 3 (Intrinsic dimension) *The intrinsic dimension of a dataset (or a probability distribution) is the minimal number of parameters (variables or coordinates) needed to represent the data. In the fractal formulation,*

$$d = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)},$$

where $N(\varepsilon)$ is the number of ε-balls required to cover X. In empirical work the limit is not observed directly; one uses finite-sample estimators (Section 2.3) and compares their stability across sample size, representation, and scale.

Definition 4 (Data topology) *Data topology is a topological structure (X, T), where T is a topology induced on X either by a metric or by a simplicial-complex construction with a notion of proximity, similarity, or neighbourhood. Its homological invariants, the Betti numbers $\beta_k(X)$, count connected components (β_0), loops (β_1), voids (β_2), and higher-dimensional cycles; persistent homology makes them robust to scale by tracking which structures survive across a filtration.*

Aicumene:2025.043v2 [cs.AI] October 24, 2026

Definition 5 (Complexity) *The complexity of a dataset X is a measure of the richness and diversity of its structural organization, determined jointly by its intrinsic dimension, its topological features, and the organization of the dynamics on it. In particular, complexity increases with the intrinsic dimension, which determines the number of independent degrees of freedom; with the number and diversity of topological features; and with the variability and distribution of data within the underlying space.*

These three concepts can be linked into a single framework by considering data as a set in a metric space onto which topological structures are superimposed, while complexity is expressed in terms of both intrinsic dimension and topology.

1. *Combinatorial complexity.* Let $X \subset \mathbb{R}^n$ have intrinsic dimension $\mathrm{ID}(X) = d$, i.e., X lies approximately on a d-dimensional manifold $M \subset \mathbb{R}^n$. If each of the d degrees of freedom has m distinguishable states, the number of possible configurations grows exponentially in d:
$$C_{\mathrm{comb}}(X) \sim m^d. \tag{1}$$
2. *Entropic complexity.* If data are distributed on a d-dimensional manifold of characteristic scale R and observed at resolution ε, the entropy of the ε-discretized distribution scales as
$$C_{\mathrm{ent}}(X) \propto H_\varepsilon(X) \approx d \log(R/\varepsilon). \tag{2}$$
3. *Topological complexity.* Intrinsic dimension bounds the non-trivial homology of X: if X is (approximately) a d-dimensional manifold, then
$$\beta_k(X) = 0 \quad \text{for all } k > d, \tag{3}$$
so that the number of independent topological structures is
$$C_{\mathrm{top}}(X) = \sum_{i=0}^{d} \beta_i(X). \tag{4}$$
4. *Algorithmic complexity.* If $X \subset \mathbb{R}^N$ lies on a d-dimensional manifold with $d \ll N$, a compressed description exists whose length per sample point at resolution ε scales as $d \log(1/\varepsilon)$ rather than $N \log(1/\varepsilon)$:
$$C_{\mathrm{alg}}(X) = O(d) \text{ per point at fixed resolution,} \tag{5}$$
and Kolmogorov complexity is maximal when X approaches uniform noise in $\mathbb{R}^N$ ($d \approx N$).

While intrinsic dimension limits how many independent structures the data can have, and data topology tells us what kinds of structures exist, complexity quantifies the richness of those structures:
$$C(X) = F\big(d, \{\beta_k(X)\}_{k=0}^{d}, \text{dynamical organization}\big). \tag{6}$$
We deliberately leave F open. The framework proposes a research programme rather than a single universal scalar; different modalities may require different estimators, while the classification should be stable under reasonable changes of measurement. Of course, other fundamental characteristics could be chosen; these three allow us to look at the data from complementary perspectives.

3.2 The structural profile and the measurement protocol

No one of the three measures is sufficient. High dimensionality alone can describe noise; rich topology can be generated by sampling artefacts; entropy can be large in a process that is informationally disorganized. The intended criterion is therefore conjunctive: strong intelligence should manifest a coherent increase in effective dimension, non-trivial topological organization, and structured dynamics. We summarize the three in a *structural profile*
$$S(\mathcal{I}) = \big(\mathrm{ID}(\mathcal{D}(\mathcal{I})),\ \mathrm{Top}(\mathcal{D}(\mathcal{I})),\ \mathrm{Dyn}(\mathcal{D}(\mathcal{I}))\big), \tag{7}$$
where Top denotes a persistence summary (e.g., the Betti profile across scales) and Dyn a vector of dynamical indicators (entropy rate, predictability, recurrence structure, Lyapunov-type indicators where appropriate). The vector form prevents the taxonomy from reproducing the problem it criticizes in aggregate benchmark scores: a system with high dimension but trivial topology is not equivalent to a system with moderately high dimension and stable multiscale organization.

The profile becomes scientifically useful only when the same variables can be estimated on heterogeneous data. For language, $\mathcal{D}(\mathcal{I})$ is a point cloud of word or n-gram embeddings; for neural dynamics, a reconstructed trajectory in

Aicumene:2025.043v2 [cs.AI] October 24, 2026

phase space; for an artificial system, a corpus of generated outputs, a sequence of hidden-state activations, a multimodal latent trajectory, or a joint record of perception and action. The comparison must therefore use properties that survive changes of representation, which is exactly what intrinsic dimension and persistent homology provide. A *standardized protocol* P specifies: (i) the probe family generating $\mathcal{D}(\mathcal{I})$, broad enough to reveal the available degrees of freedom while avoiding a benchmark that encodes the desired answer (open-ended generation, controlled perturbations, transfer tasks, counterfactual prompts, repeated probes of the same semantic domain); (ii) the embedding or state-reconstruction method; (iii) at least two independent dimension estimators with confidence intervals and a stability check across sample size and scale; (iv) a persistence computation with the same scale range; (v) dynamical indicators computed on the time-ordered data. Agreement among estimators is more informative than any single numerical output.

3.3 Performance is not ontology

A benchmark asks what a system can do under a prescribed set of tasks. A structural criterion asks what kind of data-generating organization must be present for the observed outputs or state trajectories to occupy a given geometric and dynamical regime. The former is externally indexed to a task taxonomy; the latter is indexed to properties of the generated data themselves. The following elementary proposition makes the independence of the two precise.

Proposition 1 (Benchmarks do not bound structure) *Let $B = \{(x_j, A_j)\}_{j=1}^{m}$ be a finite benchmark: finitely many prompts x_j with finite sets A_j of admissible answers. A system $\mathcal{I}$ passes B if $\mathcal{I}(x_j) \in A_j$ for all j. Then:*

(i) for every system, the data generated on the benchmark, $\mathcal{D}_B(\mathcal{I}) = \{\mathcal{I}(x_1), \ldots, \mathcal{I}(x_m)\}$, is a finite set; its Hausdorff, box-counting, and topological dimensions are 0, and $\beta_k(\mathcal{D}_B(\mathcal{I})) = 0$ for all $k \geq 1$;

(ii) for every $d \in \mathbb{N}$ there exist a system passing B and a system failing B whose generated data $\mathcal{D}(\mathcal{I})$, under any probe family extending B, have intrinsic dimension d.

Consequently, neither passing nor failing a finite benchmark gives a lower or an upper bound on the structural level of $\mathcal{I}$.

The proof is in Appendix A. The content of the proposition is not its (trivial) mathematics but its consequence: a benchmark score is a projection of $\mathcal{D}(\mathcal{I})$ onto a finite set of task coordinates and is therefore structurally blind. A system can improve its aggregate score by learning additional mappings, tools, or procedures while remaining inside the same representational regime. The increase is functionally important without amounting to a structural transition.

3.4 The dimension ladder

We now state the mathematical input and the hypothesis separately.

Theorem 1 (Whitney [1, 2]) *Every smooth n-dimensional manifold ($n \geq 1$, Hausdorff and second countable) admits a smooth embedding into $\mathbb{R}^{2n+1}$; in fact, into $\mathbb{R}^{2n}$. The bound $2n$ cannot be improved in general.*

The bound $2n + 1$ is the *generic* one: for a generic smooth map into $\mathbb{R}^{2n+1}$ injectivity and immersivity are automatic, whereas $2n$ requires Whitney's trick. It is also the bound relevant to observation.

Theorem 2 (Takens [4]; Sauer, Yorke, and Casdagli [5]) *Let A be a compact invariant set of a smooth dynamical system with box-counting dimension d_A. For a generic smooth observable, the delay-coordinate map with m coordinates is one-to-one on A (and an immersion on its smooth part) whenever $m > 2d_A$; in particular, $m = 2n + 1$ coordinates suffice for an attractor of dimension n.*

Theorem 2 says that an observer who reconstructs a state space of dimension n from generic measurements needs $2n + 1$ independent coordinates. This is how $2n + 1$ enters our construction: not as the dimension of the represented objects, but as the dimension of a space in which generic configurations of those objects can be faithfully represented and read off.

Hypothesis 1 (Representational closure) *An intelligence operating at level $k + 1$ must internally represent generic configurations of the objects characteristic of level k, which form a manifold of dimension d_k; its semantic space therefore contains a generic embedding of that manifold and has intrinsic dimension at least $2d_k + 1$.*

The base case is physical: biological organisms operate in a three-dimensional world, so $d_0 = 3$ is the dimension of the objects an animal must represent. Evolutionary constraints then require animals to operate on three-dimensional objects

Aicumene:2025.043v2 [cs.AI] October 24, 2026

within a representational space of dimension seven; by analogy, human cognition operates on higher-order semantic objects of dimension about seven, embedded in spaces of higher dimensionality, which reflects a qualitative extension of representational capacity beyond that required for physical interaction.

Definition 6 (Dimension ladder) $d_0 = 3$ *and* $d_{k+1} = 2d_k + 1$ *for* $k \geq 0$.

Proposition 2 (Closed form and robustness) *(a)* $d_k = 2^{k+2} - 1$ *for all* $k \geq 0$*; the ladder is* $3, 7, 15, 31, 63, \ldots$

(b) If the strong Whitney bound $2n$ *is used in place of the generic bound* $2n + 1$*, the ladder becomes* $d'_k = 3 \cdot 2^k$*, i.e.,* $3, 6, 12, 24, \ldots$

(c) Every value $d \in [7, 12)$*, in particular the empirical human reference* 8–9 *of Section 4, lies in the second interval of both ladders. The classification of the human reference is therefore insensitive to the choice of embedding bound; only values in* $[6, 7) \cup [12, 15)$ *are classified differently by the two ladders.*

The proof is in Appendix A. Under Hypothesis 1, the intervals $[d_k, d_{k+1})$ define the levels of intelligence: $[3, 7)$ for intelligence dominated by physical-world representation, $[7, 15)$ for intelligence operating over abstract semantic objects, $[15, 31)$ for a higher-order regime, and so on. The hypothesis is falsifiable: Section 6 lists the observations that would refute it.

3.5 Structural levels and the operational definition of Strong AI

We first state two ontological definitions, which express the relation between an artificial system and the natural intelligence whose data it was trained on.

Definition 7 (Ontological order) *Weak AI, trained on data from SNI, has an ontologically secondary nature in relation to SNI. Therefore the average intrinsic dimension and complexity of the data generated by Weak AI,* D_{WAI}*, is strictly less than the average intrinsic dimension and complexity of the data generated by SNI,* D_{SNI}*:*

$$D_{\mathrm{WAI}} < D_{\mathrm{SNI}}. \tag{8}$$

Definition 8 (Ontological equality) *Strong AI, trained on data from SNI (or on the same data as SNI was trained on), is ontologically equal to SNI. Therefore the average intrinsic dimension and complexity of its generated data,* D_{SAI}*, is comparable to that of SNI:*

$$D_{\mathrm{SAI}} \approx D_{\mathrm{SNI}}. \tag{9}$$

Definitions 7 and 8 are made operational by the ladder and the structural profile.

Definition 9 (Structural level) *Fix a protocol* P *and an intrinsic-dimension estimator* $\hat{d}$ *satisfying Pestov's axioms [46]. A system* $\mathcal{I}$ *is at* Level_k*,* $k \geq 0$*, if*

$$d_k \leq \hat{d}(\mathcal{D}(\mathcal{I})) < d_{k+1}$$

for the ladder of Definition 6 (with Level_0 *also containing all* $\hat{d} < d_0$*), and if* $\mathcal{D}(\mathcal{I})$ *satisfies the noise-exclusion condition: non-trivial persistent topology and organized (non-random) dynamics in the sense of Section 3.2. A system violating the noise-exclusion condition is assigned to* Level_0 *regardless of* $\hat{d}$*.*

Thus Level_0 ($\hat{d} < 7$) is Weak Intelligence, comprising conventional narrow AI, mainstream AI systems, specialized weak AI, and animal cognition in the physical-world regime; Level_1 ($7 \leq \hat{d} < 15$) is Strong Intelligence, natural or artificial; Level_2 ($15 \leq \hat{d} < 31$) is Super-Strong Intelligence; higher levels extend the hierarchy inductively. Our own engineering supplies instances of Level_0 against which the ladder is read: a lightweight vessel-segmentation network for endovascular surgery [24] with its annotated dataset [25], and counter-augmented deterministic automata for real-time regular-expression matching [22, 23]; the structural readouts of the profile include the topology-driven classification of time series [21].

Definition 10 (Operational Strong AI) *Let* S_{SNI} *be the structural profile of Strong Natural Intelligence measured under protocol* P *(the reference regime of Section 4). An artificial system* $\mathcal{I}$ *is* Strong AI *if it is at* Level_1 *and its profile* $S(\mathcal{I})$ *is statistically indistinguishable from, or dominates,* S_{SNI} *in all three components under* P *(Definition 8). It is* Weak AI *if* $\hat{d}(\mathcal{D}(\mathcal{I}))$ *is significantly below* D_{SNI} *or the noise-exclusion condition fails (Definition 7). It is* Super-Strong AI *if it is at* Level_2 *or higher with non-trivial topology and organized dynamics.*

Aicumene:2025.043v2 [cs.AI] October 24, 2026

Remark 1 (Transitional regimes) *The intervals should be allowed to overlap in practice: a system may move between regimes as task, context, or internal state changes, which makes the hierarchy closer to a phase diagram than to a taxonomic ladder. What matters is a qualitative reorganization of the informational space, not a ceremonial crossing of a single scalar threshold. The same system may therefore be classified differently under different probe families, which is why the protocol P is part of the definition.*

Remark 2 (Consciousness as a component) *Within this framework, consciousness may be understood as a component of a complex system that possesses higher intrinsic dimensionality than its other parts and is directed toward the external world, toward itself (self-awareness), and toward other components of the same system. Differences in dimensionality between components would allow a more refined classification of intelligence systems according to the types of consciousness they generate. We state this as a direction, not as part of the criterion (Section 5.5).*

4 Empirical Evidence

A non-anthropocentric account still requires empirical anchors. Human language and human neural dynamics are useful not because they define intelligence by fiat, but because they are the best studied products and processes of a system ordinarily recognized as strongly intelligent. The question is comparative: which structural regimes recur across these very different projections of human cognition, and where do artificial systems fall?

4.1 Language

Language, as a primary product of natural intelligence, provides a natural empirical substrate for probing the geometry of semantic representations. Recent works [7, 8, 9] demonstrate that natural language can be rigorously described as a unified complex system exhibiting properties of self-organized criticality; within this framework, texts can be interpreted as avalanches in an underlying dynamical system.

To formalize this perspective, consider a language as a dataset embedded into a metric space. Let X denote the space of embeddings of words or n-grams, equipped with a probability measure μ; the resulting structure is a metric measure space (X, d, μ). We employ three complementary notions of dimensionality.

Topological dimension.

$$\dim_T(X) = \inf\{n \mid \text{every finite open cover has a refinement of order } \leq n+1\}.$$

Hausdorff dimension.

$$\dim_H(X) = \inf\left\{d \geq 0 \;\middle|\; \mathcal{H}^d(X) = 0\right\},$$

where $\mathcal{H}^d$ is the d-dimensional Hausdorff measure.

Intrinsic dimension. Intrinsic dimension is a mapping [46, 47]

$$\partial : \mathcal{X} \to \mathbb{R}_+$$

on the class $\mathcal{X}$ of metric measure spaces that satisfies the axioms of concentration, smooth dependence on the dataset, and normalization.

Estimation procedure. Given a corpus of texts in a language, we construct embeddings for words and n-grams. Word embeddings can be obtained via singular value decomposition of a co-occurrence matrix, $X \approx U_k \Sigma_k V_k^\top$, or via neural embeddings such as CBOW (Word2Vec), defined through $\max \prod_t P(w_t \mid \text{context}(w_t))$. The embedding of an n-gram is the concatenation of the embeddings of its constituent words.

Graph-based estimation of dimension. Let $\{x_i\}_{i=1}^N \subset X$ be a sample of embeddings. Construct a complete weighted graph with weights $w_{ij} = \|x_i - x_j\|$ and let T be its minimum spanning tree. Define

$$L_\gamma(X_N) = \sum_{e \in T} w_e^\gamma.$$

For a d-dimensional distribution, $\mathbb{E}[L_\gamma(X_N)] \sim N^{(d-\gamma)/d}$ [49, 50], which allows estimation of the fractal dimension d. The Brito estimator [51] provides an estimate of topological dimension based on k-nearest-neighbour graph statistics and Bayesian inference over candidate dimensions.

Aicumene:2025.043v2 [cs.AI] October 24, 2026

Empirical result. A large-scale study across 70 languages from 21 language families [8] demonstrates that

$$\dim_{\text{intrinsic}}(\text{language}) \approx 9$$

for both word-level and n-gram representations, across embedding methods. Crucially, this value is non-integer (indicating fractal structure), stable across languages, and invariant under representation choices. Topological analysis of the same embedding spaces identifies persistent large-scale structures ("holes") that are not reducible to the scalar dimension [7].

Independent corroboration. Using a different estimator (persistent-homology dimension of contextual token embeddings) on different corpora, Tulchinskii et al. [10] find that the average intrinsic dimensionality of fluent human texts is ≈ 9 for several alphabet-based languages and ≈ 7 for Chinese, while AI-generated texts in each language are ≈ 1.5 lower, with a clear statistical separation between the two distributions. The same gap, measured with yet other geometric features, underlies "Spot the Bot" [9]. The convergence of independent estimators, corpora, and research groups toward a stable, non-integer value provides strong evidence that language occupies a fractal semantic space of dimension approximately nine.

Interpretation. The significance of the value is not that nine is a constant of language, but comparative stability: substantially different languages and representation choices occupy a similar dimensional regime, so intrinsic dimension is a candidate invariant of a broader semantic organization. If language is treated as an observable projection of internal semantic structures, its intrinsic dimensionality constitutes a lower-bound estimate of the dimensionality of the underlying intelligence. This suggests that human intelligence operates in a regime that exceeds the constraints of animal cognition while remaining below higher, yet-to-be-realized intelligence levels.

4.2 Brain data: from language to thought

While language provides an observable projection of semantic structures, it is not the primary substrate of intelligence. Cognitive processes manifest in neural dynamics, which can be directly analyzed through multichannel time series such as EEG signals [15, 16, 17, 18, 19, 26, 27, 28, 29, 30, 31, 32, 33].

Let $\{x_i(t)\}_{i=1}^{M}$ denote multichannel neural recordings. We model these signals (Bernadotte brain-signal model [15, 16, 17]) as superpositions of damped oscillatory components:

$$x_i(t) = \sum_{k=1}^{K} A_{ik} e^{\lambda_k t} \cos(\omega_k t + \phi_{ik}) + \epsilon_i(t),$$

where λ_k are decay rates, ω_k are frequencies, and $\epsilon_i(t)$ is noise. This representation defines a class of signals composed of exponentially modulated sinusoids ("exp-sinusoids"), which serve as elementary modes of neural dynamics.

MSSA reconstruction and oscillator structure. To extract these components we employ Multichannel Singular Spectrum Analysis (MSSA) [17, 18, 19]. Let $X(t)$ denote the multichannel signal and construct the trajectory (block-Hankel) matrix

$$\mathcal{H} = \begin{bmatrix} X(1) & X(2) & \cdots & X(L) \\ X(2) & X(3) & \cdots & X(L+1) \\ \vdots & \vdots & \ddots & \vdots \\ X(T-L+1) & X(T-L+2) & \cdots & X(T) \end{bmatrix}.$$

Its singular value decomposition $\mathcal{H} = U\Sigma V^\top$ yields a decomposition into principal components corresponding to oscillatory modes.

Theorem 3 (Recovery of exp-sinusoidal components via MSSA [17, 18, 19]) *Let $X(t)$ be a multichannel signal composed of a finite sum of exponentially modulated sinusoids with distinct frequencies and decay rates. Then, under mild conditions on the window length L and signal separability, MSSA reconstructs each exp-sinusoidal component as a pair (or low-rank block) of singular components of $\mathcal{H}$, and the original signal is recovered as the sum of these components up to noise.*

This theorem establishes that neural signals admit a representation in terms of coupled oscillatory modes, which can be interpreted as elementary units of cognitive dynamics.

Aicumene:2025.043v2 [cs.AI] October 24, 2026

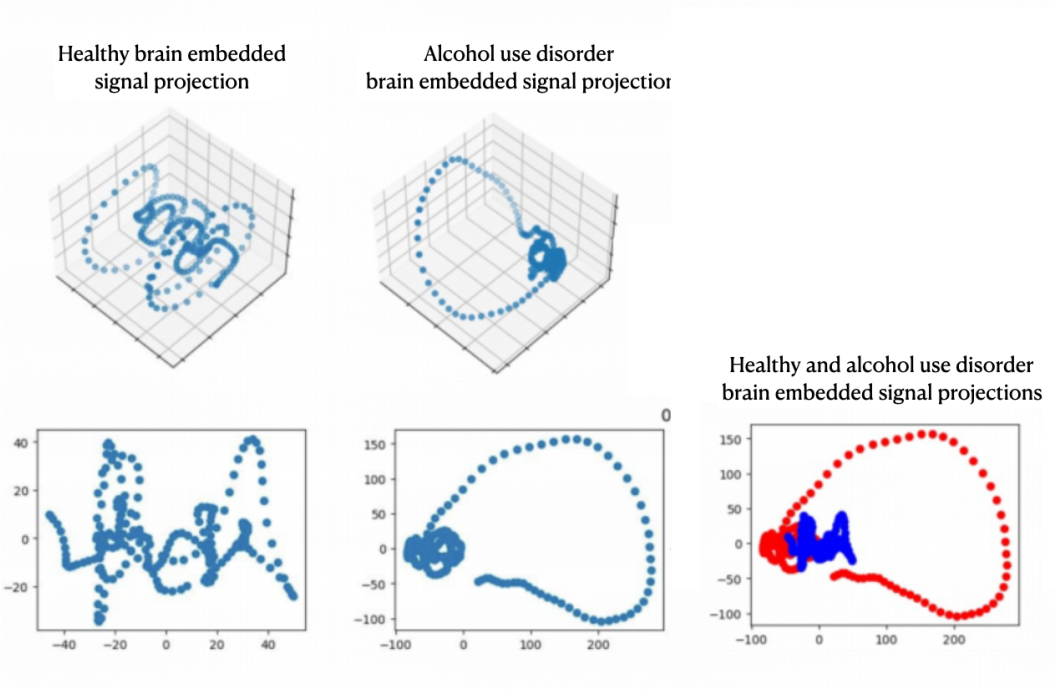

Figure 4: 3D and 2D projections of embedded EEG trajectories (obtained via MSSA) for a healthy brain and for alcohol use disorder. The figure illustrates geometric differences in reconstructed neural dynamics; the projection itself is not the intrinsic-dimension estimator.

Intrinsic dimension of thought. Applying intrinsic-dimension estimators to the reconstructed phase-space trajectories (obtained via MSSA or delay embeddings, Theorem 2), one obtains $\dim_{\text{intrinsic}}(\text{thought}) \approx D_{\text{thought}}$, where D_{thought} is empirically observed to be [17, 18, 19]

$$D_{\text{thought}} \approx 9.$$

More precisely, a healthy brain produces an electromagnetic signal with intrinsic dimensionality in the range 8–9, whereas in pathological states the brain exhibits dimensionality below 8 (Fig. 4) [17, 18, 19].

Interpretation. Language and neural dynamics thus occupy the same dimensional regime. We hypothesize the ordering

$$\dim(\text{language}) \leq \dim(\text{thought}) \leq \dim(\text{intelligence}), \tag{10}$$

with language a lower-bound projection of the semantic space and neural dynamics revealing the structure of internal cognitive processes. The current evidence supports an overlap between the two regimes, both estimated at the same order of magnitude, rather than a demonstrated strict inequality; the stronger claim remains a testable hypothesis requiring paired neural and linguistic measurements from the same cognitive episodes. This observation motivates the hypothesis of an ultra-narrow bottleneck, whereby high-dimensional cognitive representations are compressed to an approximately nine-dimensional manifold when projected into the domain responsible for speech production: linguistic expression as a constrained information bottleneck limiting the effective dimensionality of observable semantic structures.

4.3 Synthesis: the evidence on the ladder

Table 1 collects the published estimates used in this paper, and Fig. 5 places them on the dimension ladder. Three facts stand out. First, two human projections (language and brain dynamics), measured by different groups with different estimators on different data, fall in the same interval $[7, 15)$, near its lower edge. Second, every machine-generated text distribution measured so far falls below the human one by a stable margin. Third, states known to involve a loss of higher cognitive function (pathological brain states) fall below the lower edge of the human interval. The ladder therefore separates the cases it should separate, with the human reference inside Level_1 under both variants of the ladder (Proposition 2(c)).

Aicumene:2025.043v2 [cs.AI] October 24, 2026

Table 1: Published intrinsic-dimension estimates used in this paper. ID: intrinsic dimension; PHD: persistent-homology dimension; MST: minimum spanning tree.

System or state	Data	Estimator	ID	Source
Human language, 70 languages, 21 families	word and n-gram embeddings	MST-based; k-NN graph	≈ 9, non-integer, stable across languages and embeddings	[8]
Human-written text, alphabet-based languages	contextual token embeddings	PHD	≈ 9 (≈ 7 for Chinese)	[10]
AI-generated text, same languages	contextual token embeddings	PHD	≈ 1.5 below human texts	[10]
Machine- vs human-generated text	embeddings	geometric and ID features	lower-dimensional manifold, statistically separable	[9]
Healthy human brain, EEG	MSSA phase-space trajectories	ID estimators on reconstructed trajectories	8–9	[17, 18, 19]
Pathological brain states, EEG	same	same	< 8	[17, 18, 19]
Mouse head-direction circuit (single circuit)	population spiking	nonlinear manifold reconstruction	1 (ring)	[55]

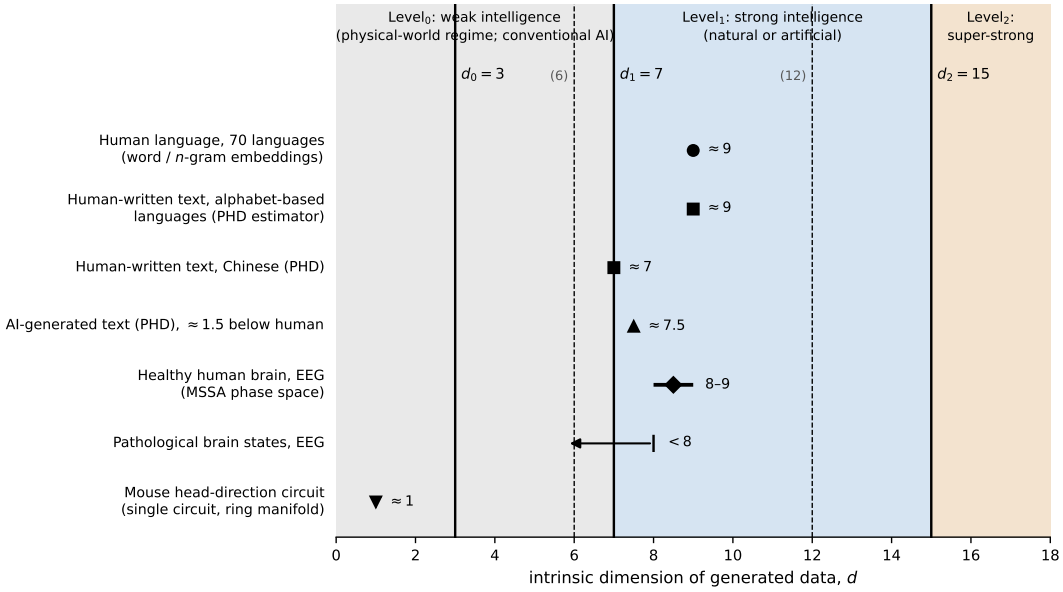

Figure 5: The dimension ladder and the published estimates of Table 1. Solid vertical lines: generic ladder $d_k = 2^{k+2} - 1$ $(3, 7, 15$; the next rung, 31, is off scale). Dashed lines in parentheses: the strong-Whitney variant $d'_k = 3 \cdot 2^k$ $(6, 12)$. Shaded bands: Level$_0$, Level$_1$, Level$_2$. The human reference (language and healthy brain) lies inside Level$_1$ under both ladders; machine-generated text and pathological brain states lie below the human reference. The head-direction circuit is a single circuit devoted to one physical variable and is shown for contrast, not as an estimate of animal intelligence as a whole. Generated by `make_figures.py` from the cited values; no new data.

4.4 Where current systems fall

By Definition 10 and the measurements of Table 1, current large language models, as measured through their textual output, are Weak AI: their generated text has intrinsic dimension significantly below the human reference ($D_{\mathrm{WAI}} < D_{\mathrm{SNI}}$), despite benchmark performance that qualifies several of them as AGI candidates under behavioural definitions [39, 41]. This is exactly the situation anticipated by Proposition 1.

Two caveats are part of the classification rather than objections to it. First, the result concerns one projection of these systems (sampled text) under the probe families used in the cited studies; model families change, and the classification

must be re-measured, not declared once and for all. Second, for an artificial system the internal state is accessible: the intrinsic dimension of hidden-state trajectories can be measured directly [53], which makes the comparison between output geometry and internal-state geometry an experiment rather than a speculation. A future model whose textual output approaches or exceeds the human-language reference would be an important result, but by Definition 9 it would still require topological and dynamical corroboration before being classified as Strong AI.

The transition from weak to strong intelligence can therefore be characterized as a transition between regimes of intrinsic dimensionality and organization. Systems operating within low-dimensional manifolds may achieve high task performance yet remain structurally limited; strong intelligence corresponds to the emergence of high-dimensional, dynamically structured semantic spaces. This provides an operational and measurable criterion for distinguishing weak, strong, and super-strong intelligence based on intrinsic geometric and dynamical properties of generated data.

Aicumene:2025.043v2 [cs.AI] October 24, 2026

5 Discussion and Future Prospects

5.1 Relationship between language and cognition

The remarkable similarity between the intrinsic dimensionality observed in language and in functional spaces arising during cognitive processes suggests a relationship between linguistic semantic spaces and morpho-functional spaces of the brain. A fundamental question for a future theory of Strong AI concerns the nature of this relationship. Formally, possible relationships include metric equivalence, homeomorphism, diffeomorphism, and homomorphism. We hypothesize that the complexity of cognitive processes exceeds that of language, so that the null hypothesis should be a homeomorphic relationship between functional or morpho-functional spaces and linguistic ones; the apparent contradiction between smooth structures and empirical irregularity may be resolved through minimal invariant manifolds and mappings between them. The absence of any structural relationship appears unlikely.

For three-dimensional manifolds, Thurston's geometrization [6], now a theorem, states that every closed orientable 3-manifold decomposes canonically into pieces each carrying one of eight model geometries. We suggest a possible analogy, heuristic at this stage, between such canonical decompositions and the structures arising in functional spaces during cognitive processes: one submanifold may be associated, via a homeomorphic mapping, with linguistic structures, while other submanifolds may correspond to cognitive processes not directly related to language (perception, affect, motor preparation, imagery, or contemplative states). Such a programme would turn an old philosophical problem, the relation between thought and language, into a comparative problem of geometry without reducing one domain to the other.

5.2 Intrinsic dimension of Super-Strong Intelligence

When discussing the dimensionality of strong intelligence, we begin from empirical observations indicating that both language and brain signals have intrinsic dimensionality close to 9. Assuming that biological organisms operate in a three-dimensional physical world and applying Hypothesis 1 with the generic embedding bound, processing such information requires embedding into a space of dimension 7; this value is interpreted as an upper bound for animal intelligence and a lower bound for human intelligence. Extending the reasoning, embedding a seven-dimensional semantic space requires a space of dimension $2 \cdot 7 + 1 = 15$, so dimensionality exceeding this threshold may correspond to super-strong intelligence. These theoretical considerations are consistent with the empirical observations from language and brain data, which place the human reference inside the interval $[7, 15)$ and close to its lower edge.

5.3 Co-evolution of natural and artificial intelligence

We argue that the current stage of technological development is characterized by a transition from purely biological and social evolution toward co-evolution of humans and artificial intelligence systems. The development of AI systems is expected to increase the dimensionality of the semantic spaces in which they operate; at the same time, human cognition may evolve through interaction with increasingly complex systems. From this perspective, evolution may be interpreted as an expansion of semantic dimensionality.

Another scenario involves the integration of human cognition with AI systems via neurointerfaces, leading to the emergence of hybrid forms of intelligence. In that setting the relevant object of analysis may be neither the human nor the model in isolation but the joint dynamical system, and the question is whether interaction produces new stable degrees of freedom, new topological relations, or new attractor structures in the combined semantic space. Such changes would constitute a literal, measurable sense in which cognitive evolution can be described as an expansion or reorganization of semantic dimensionality.

Aicumene:2025.043v2 [cs.AI] October 24, 2026

5.4 Higher-dimensional cognitive states

Currently, humans may only transiently experience higher-dimensional cognitive states. Continued interaction with high-dimensional AI systems may lead to stabilization of such states through mechanisms of neuroplasticity, resulting in stable high-dimensional attractors in cognitive dynamics that correspond to new modes of thinking. At the same time, degradation toward lower-dimensional states is also possible, as observed in pathological or altered conditions (Table 1).

5.5 Consciousness is not the criterion

Strong intelligence should be separated from consciousness. A structural increase in the complexity of generated data does not establish phenomenal experience, self-awareness, or subjectivity; conversely, if consciousness has measurable structural correlates, those correlates constitute an additional axis rather than a definition of intelligence itself. This distinction avoids two opposite errors: treating fluent linguistic performance as proof of consciousness, and making consciousness a prerequisite for recognizing non-human intelligence. The proposed hierarchy remains agnostic on phenomenality and focuses on a narrower question, the informational regime in which a system operates. It is likewise not a metric of moral status: structural richness does not by itself entail sentience, rights, agency, or responsibility.

6 Falsifiability, Failure Modes, and Limitations

A structural definition is useful only if it can fail. The following observations would count against the present proposal.

1. *Instability of the invariant.* If intrinsic dimension varied arbitrarily with embedding method, estimator, or sample size, it could not serve as a stable invariant. The convergence of independent estimators and corpora on the language value (Table 1) is the evidence that it does not; a comparable multi-estimator study for brain data is a priority.
2. *Lack of discriminative power.* If systems with clearly different cognitive organization repeatedly occupied indistinguishable dimensional, topological, and dynamical regimes under a standardized protocol, the criteria would be uninformative.
3. *Noise misclassified as intelligence.* If high-dimensional noise were consistently classified as strong intelligence, the framework would have confused complexity with intelligence. The noise-exclusion condition of Definition 9 is the safeguard; it must be applied, not assumed.
4. *Refutation of the ladder.* Hypothesis 1 is refuted if an intelligence that demonstrably represents generic configurations of level-k objects is found to operate in a semantic space of dimension below $2d_k + 1$, or if the transitions between qualitatively different regimes are found at thresholds unrelated to the ladder. A successful future theory might retain the structural programme while replacing the numbers.

These risks lead to methodological requirements, encoded in the protocol of Section 3.2: measurements must be multi-estimator and multi-scale; dimension must never be used alone; observational conditions must be standardized, because a system can express different regimes under different tasks; and output geometry must be distinguished from internal-state geometry whenever the latter is accessible.

Three further limitations should be stated plainly. (a) Finite-sample intrinsic-dimension estimators are biased, typically downward at higher dimensions and for small samples; the human reference value ≈ 9 may therefore be an underestimate, which would move it further from the lower edge of Level_1 but does not change its classification. (b) Intrinsic dimension is scale dependent; all estimates in Table 1 are reported at the scales used in the cited studies, and cross-study comparisons are comparisons of regimes, not of digits. (c) The empirical base for brain dynamics currently rests on the authors' own studies; the language result has independent corroboration, the brain result does not yet, and we flag this asymmetry explicitly.

7 Conclusions

The distinction between different levels of intelligence should be based on formal statistical and geometric characteristics of generated data rather than on task categories defined by human-centric relevance. Performance on any finite benchmark imposes no bound on these characteristics (Proposition 1), so behavioural generality and structural strength are independent properties.

The fundamental measure of complexity of the spaces in which different types of natural and artificial intelligence operate is their intrinsic dimensionality, refined by topology and dynamics. The generic embedding bound of Whitney–Takens type, combined with the representational-closure hypothesis, yields the ladder $d_k = 2^{k+2} - 1$ and a hierarchy of

Aicumene:2025.043v2 [cs.AI] October 24, 2026

dimensional intervals: the first associated with animal cognition and conventional AI, the second with strong intelligence, and the third with super-strong intelligence. The classification is independent of whether the intelligence is natural or artificial, and the classification of the human reference is insensitive to the choice between the generic and the strong embedding bound.

Computational and experimental analysis of both natural language and functional processes in the human brain indicates that human intelligence is associated with spaces of intrinsic dimensionality on the order of 9, lying close to the lower bound (~ 7) of the human interval; the language result is independently corroborated. Machine-generated text lies below the human reference by a stable margin, so current language models, measured through their output, are Weak AI by the operational definition despite their benchmark performance. It may be hypothesized that in certain cognitive states (e.g., meditative or contemplative states) the effective dimensionality increases within the human interval, while states not involving higher cognitive functions correspond to a transition toward lower-dimensional regimes; the pathological data are consistent with the latter, the former requires experimental validation.

At present, we observe a transition from predominantly biosocial evolution toward a co-evolutionary process involving humans and artificial intelligence systems. Within this framework, the intrinsic dimensionality of semantic spaces may serve as a measure of evolutionary progression for both human cognition and artificial systems, and advances in computational capabilities and neurointerface technologies open the possibility of actively influencing this process. Such developments may also be associated with significant challenges, including social and biological consequences of artificially induced cognitive states; these require careful interdisciplinary study. The conceptual gain of the structural account is a shift in the question: instead of asking only whether an artificial system can do what humans do, we can ask what sort of semantic world it is capable of constructing and inhabiting.

Funding

For V.A.G.: the study was implemented in the framework of the Basic Research Program at HSE University (HSE-BR-2025-025).

Declarations

The authors declare no competing interests.

Data availability. No new dataset was generated for this paper. All empirical values discussed in the text are drawn from the cited published and preprint studies (Table 1).

Code availability. Fig. 5 is generated from the values in Table 1 by the script `make_figures.py` distributed with the LaTeX source of this preprint.

Author contributions. A.B. developed the theoretical framework, developed the framework for intrinsic dimensionality and topological analysis of brain data, formulated the structural criteria for intelligence classification, conducted the analysis of neural data, performed the brain-related experiments, contributed to the theoretical interpretation of the results, and drafted the manuscript. V.A.G. developed the theoretical framework, developed the framework for intrinsic dimensionality and topological analysis of language structures, performed the language-related experiments, contributed to the theoretical interpretation of the results, and participated in drafting the manuscript. Both authors reviewed, edited, and approved the final manuscript.

Version history. Version 1 of this preprint appeared as Aicumene:2025.043v1 [cs.HC] on August 12, 2025. The present version 2 (October 24, 2026) adds the separation of theorem and hypothesis in the dimension ladder with its closed form and robustness check, Proposition 1, the structural profile and measurement protocol, the operational definition of Strong AI, Table 1 and Fig. 5 with independent corroboration of the language result, and Section 6; the hedging of the ordering (10) follows the authors' later discussion of the same material.

A Proofs

Proof of Proposition 1. (i) $\mathcal{D}_B(\mathcal{I})$ has at most m points. A finite set has Hausdorff and box-counting dimension 0 (it is covered by m balls of any radius ε, so $\log N(\varepsilon)/\log(1/\varepsilon) \to 0$) and topological dimension 0 (it is discrete); its homology is that of $m' \le m$ points, so $\beta_0 = m'$ and $\beta_k = 0$ for $k \ge 1$.

(ii) Let $\mathcal{I}_0$ be any system passing B; a lookup table $x_j \mapsto a_j \in A_j$ suffices. Let G_d be a generator whose outputs on prompts outside B are sampled from a smooth distribution supported on an embedded d-dimensional torus in the output space (for $d = 0$, a single point). Define $\mathcal{I}$ to answer benchmark prompts by $\mathcal{I}_0$ and all other prompts by G_d. Under

any probe family extending B with infinitely many non-benchmark prompts, $\mathcal{D}(\mathcal{I})$ is the union of a finite set and a sample dense in a d-dimensional torus, so its intrinsic (Hausdorff) dimension is d, and $\mathcal{I}$ passes B. Replacing $\mathcal{I}_0$ by a lookup table with $x_1 \mapsto a \notin A_1$ gives a system failing B with the same $\mathcal{D}(\mathcal{I})$ up to one point, hence the same intrinsic dimension. Since d was arbitrary, $\inf$ and $\sup$ of the intrinsic dimension over systems passing (or failing) B are 0 and ∞, which is the claim. □

Proof of Proposition 2. (a) Set $e_k = d_k + 1$. Then $e_0 = 4$ and $e_{k+1} = 2d_k + 2 = 2e_k$, so $e_k = 4 \cdot 2^k = 2^{k+2}$ and $d_k = 2^{k+2} - 1$. (b) With $d'_{k+1} = 2d'_k$ and $d'_0 = 3$, $d'_k = 3 \cdot 2^k$. (c) The second intervals are $[d_1, d_2) = [7, 15)$ and $[d'_1, d'_2) = [6, 12)$; their intersection is $[7, 12)$, which contains $[8, 9]$. The symmetric difference of the two intervals is $[6, 7) \cup [12, 15)$. □

REFERENCES

[1] Hassler Whitney. "Differentiable Manifolds." Annals of Mathematics. 1936, Vol. 37(3), pp. 645–680. doi:10.2307/1968482

[2] Hassler Whitney. "The Self-Intersections of a Smooth n-Manifold in $2n$-Space." Annals of Mathematics. 1944, Vol. 45(2), pp. 220–246. doi:10.2307/1969265

[3] John M. Lee. "Introduction to Smooth Manifolds." 2nd ed. Springer, 2012. doi:10.1007/978-1-4419-9982-5

[4] Floris Takens. "Detecting strange attractors in turbulence." In: Dynamical Systems and Turbulence, Warwick 1980. Lecture Notes in Mathematics, Vol. 898. Springer, 1981, pp. 366–381. doi:10.1007/BFb0091924

[5] Tim Sauer, James A. Yorke, Martin Casdagli. "Embedology." Journal of Statistical Physics. 1991, Vol. 65, pp. 579–616. doi:10.1007/BF01053745

[6] William P. Thurston. "Three dimensional manifolds, Kleinian groups and hyperbolic geometry." Bulletin of the American Mathematical Society. 1982, Vol. 6(3), pp. 357–381. doi:10.1090/S0273-0979-1982-15003-0

[7] Gromov Vasilii, Yerbolova Assel, Dang Quynh Nhu. "A Language and Its Holes: the First Order Homologies of the Large-scale Geometrical Structure of a Natural Language." Complexity, 2025, 9659172, 15 pages. doi:10.1155/cplx/9659172

[8] Gromov Vasilii, Borodin Nikita, Yerbolova Assel. "A Language and Its Dimensions: Intrinsic Dimensions of Language Fractal Structures." Complexity. 2024, 8863360. doi:10.1155/2024/8863360

[9] Gromov Vasilii, Dang Quynh Nhu, Kogan Alexandra, Yerbolova Assel. "Spot the bot: the inverse problems of NLP." PeerJ Computer Science. 2024, Vol. 10, e2550. doi:10.7717/peerj-cs.2550

[10] Eduard Tulchinskii, Kristian Kuznetsov, Laida Kushnareva, Daniil Cherniavskii, Sergey Nikolenko, Evgeny Burnaev, Serguei Barannikov, Irina Piontkovskaya. "Intrinsic Dimension Estimation for Robust Detection of AI-Generated Texts." Advances in Neural Information Processing Systems 36 (NeurIPS 2023). arXiv:2306.04723.

[11] Bernadotte A. "The Algorithm That Maximizes the Accuracy of k-Classification on the Set of Representatives of the k Equivalence Classes." Mathematics. 2022, Vol. 10(15), 2810. doi:10.3390/math10152810

[12] Vorontsova D., Menshikov I., Zubov A., Orlov K., Rikunov P., Zvereva E., Flitman L., Lanikin A., Sokolova A., Markov S., Bernadotte A. "Silent EEG-Speech Recognition Using Convolutional and Recurrent Neural Network with 85% Accuracy of 9 Words Classification." Sensors. 2021, Vol. 21, 6744. doi:10.3390/s21206744

[13] Alexandra Bernadotte, Alexandr D. Mazurin. "Optimization of the brain command dictionary based on the statistical proximity criterion in silent speech recognition task." Computer Research and Modeling. 2023, Vol. 15(3), pp. 675–690. doi:10.20537/2076-7633-2023-15-3-675-690

[14] D. V. Vorontsova, M. V. Isaeva, I. A. Menshikov, K. Yu. Orlov, A. Bernadotte. "Frequency, time, and spatial electroencephalogram changes after COVID-19 during a simple speech task." Computer Research and Modeling. 2023, Vol. 15(3), pp. 691–701. doi:10.20537/2076-7633-2023-15-3-691-701

[15] A. Bernadotte. "Estimating the Number of Sources in EEG with Hankel Embedding for Brain-Computer Interface." 2026 12th International Conference on Automation, Robotics and Applications (ICARA), Istanbul, Türkiye, 2026, pp. 621–626. doi:10.1109/ICARA69401.2026.11480398

[16] I. Menshikov, N. Elfimov, A. Bernadotte. "A Lightweight Hankel-Embedded Pipeline for Real-Time EEG Filtering and Classification." 2026 12th International Conference on Automation, Robotics and Applications (ICARA), Istanbul, Türkiye, 2026, pp. 589–593. doi:10.1109/ICARA69401.2026.11480292

Aicumene:2025.043v2 [cs.AI] October 24, 2026

Aicumene:2025.043v2 [cs.AI] October 24, 2026

[17] Bernadotte, A. and Menshikov, I. (2026). Brain Oscillator Intrinsic Dimension marks consciousness, recovery mode, and failed downshifting in coma. Aicumene Preprint 2026.042v2; SSRN working paper 7667278 (October 24, 2026). https://ssrn.com/abstract=7667278, doi:10.2139/ssrn.7667278. https://research.aicumene.com/papers/boid-coma/

[18] Alexandra Bernadotte, Ivan Menshikov. "Multichannel Singular Spectrum Analysis: Counting Oscillatory Modes and Recovering Their Spatial Loadings." Aicumene Preprint 2026.044v2; SSRN working paper 7693202. November 11, 2026. doi:10.2139/ssrn.7693202 Available: https://research.aicumene.com/papers/MSSA/

[19] Alexandra Bernadotte, Ivan Menshikov. "From Oscillatory Modes to Source Geometry: Identifiability and Localization in Multichannel Time Series." Aicumene Preprint 2026.058v1; SSRN working paper 7716058. November 2026. doi:10.2139/ssrn.7716058 Available: https://research.aicumene.com/papers/source-geometry/

[20] Alexandra Bernadotte. "Oscillatory Model and Topological Structure of Brain Signals." Aicumene Preprint, in preparation (2026).

[21] Alexandra Bernadotte. "Topology-driven classification of time series." bioRxiv 2026.04.25.720787. 2026. doi:10.64898/2026.04.25.720787

[22] Alexandra Bernadotte. "Real-time regular expression matching." arXiv:2308.10208. 2023. doi:10.48550/arXiv.2308.10208

[23] Alexandra Bernadotte. "Counter-augmented deterministic automata for real-time regular expression matching without state explosion." 2026 12th International Conference on Automation, Robotics and Applications (ICARA), 471–475. 2026. doi:10.1109/ICARA69401.2026.11480280

[24] Alexandra Bernadotte. "VeNet: a lightweight neural network for efficient brain vessel segmentation in endovascular robotic surgery." Scientific Reports 16, 24566. 2026. doi:10.1038/s41598-026-54176-x

[25] Alexandra Bernadotte, Nikita Elfimov, Ivan Menshikov. "Semi-automatic annotation of brain vessels in magnetic resonance angiography images." Scientific Data 13, 41. 2026. doi:10.1038/s41597-025-06354-1

[26] Chen Tao, Huang Haiyun, Pan Jiahui, Li Yuanqing. "An EEG-based brain-computer interface for automatic sleep stage classification." 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2018, pp. 1988–1991. doi:10.1109/ICIEA.2018.8398035

[27] Е. А. Жирмунская, Г. П. Фомичева, В. М. Бухштабер, В. К. Маслов, А. С. Векслер, Е. А. Зеленюк. "Применение методов многомерного статистического анализа ЭЭГ для оценки состояния нейродинамики мозга." Физиология человека, 1979. Том 5, номер 4, сс. 1–17. [E. A. Zhirmunskaya et al. "Application of multivariate statistical analysis of EEG for assessing the state of brain neurodynamics." Human Physiology, 1979, Vol. 5(4), pp. 1–17; in Russian.]

[28] Luauté J., Morlet D., Mattout J. "BCI in patients with disorders of consciousness: clinical perspectives." Annals of Physical and Rehabilitation Medicine. 2015, Vol. 58(1), pp. 29–34. doi:10.1016/j.rehab.2014.09.015

[29] Abdulaziz Osama, Saltykova Olga. "CNN-PS: Electroencephalogram Classification of Brain States Using Hybrid Machine–Deep Learning Approach." Iraqi Journal for Computer Science and Mathematics. 2023, Vol. 4. doi:10.52866/ijcsm.2023.04.04.006

[30] Noirhomme Quentin, Kitney Richard, Macq Benoit. "Single-Trial EEG Source Reconstruction for Brain–Computer Interface." IEEE Transactions on Biomedical Engineering. 2008, Vol. 55, pp. 1592–1601. doi:10.1109/TBME.2007.913986

[31] Noirhomme Quentin, Macq Benoit. "EEG inverse problem and priors in a brain-computer interface." SPECOM'2006, St. Petersburg, 25–29 June 2006.

[32] Fruitet J., Clerc M. "Reconstruction of cortical sources activities for online classification of electroencephalographic signals." Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2010, pp. 6317–6320. doi:10.1109/IEMBS.2010.5627713

[33] V. M. Buchstaber. "Time Series Analysis and Grassmannians." Applied Problems of Radon Transform, Amer. Math. Soc. Transl. Ser. 2. 1994, Vol. 162, pp. 1–17.

[34] A. M. Turing. "Computing Machinery and Intelligence." Mind. 1950, Vol. LIX(236), pp. 433–460. doi:10.1093/mind/LIX.236.433

[35] John R. Searle. "Minds, brains, and programs." Behavioral and Brain Sciences. 1980, Vol. 3(3), pp. 417–424. doi:10.1017/S0140525X00005756

[36] Shane Legg, Marcus Hutter. "Universal Intelligence: A Definition of Machine Intelligence." Minds and Machines. 2007, Vol. 17(4), pp. 391–444. doi:10.1007/s11023-007-9079-x

Aicumene:2025.043v2 [cs.AI] October 24, 2026

[37] François Chollet. "On the Measure of Intelligence." arXiv preprint arXiv:1911.01547, 2019.

[38] Ben Goertzel. "Artificial General Intelligence: Concept, State of the Art, and Future Prospects." Journal of Artificial General Intelligence. 2014, Vol. 5(1), pp. 1–48. doi:10.2478/jagi-2014-0001

[39] Meredith Ringel Morris, Jascha Sohl-Dickstein, Noah Fiedel, Tris Warkentin, Allan Dafoe, Aleksandra Faust, Clement Farabet, Shane Legg. "Position: Levels of AGI for Operationalizing Progress on the Path to AGI." Proceedings of the 41st International Conference on Machine Learning (ICML 2024). arXiv:2311.02462.

[40] Luciano Floridi, Massimo Chiriatti. "GPT-3: Its Nature, Scope, Limits, and Consequences." Minds and Machines. 2020, Vol. 30, pp. 681–694. doi:10.1007/s11023-020-09548-1

[41] Dan Hendrycks, Dawn Song, Christian Szegedy, Honglak Lee, Yarin Gal, Erik Brynjolfsson, Sharon Li, Andy Zou, Lionel Levine, Bo Han, Jie Fu, Ziwei Liu, Jinwoo Shin, Kimin Lee, Mantas Mazeika, Long Phan, George Ingebretsen, Adam Khoja, Cihang Xie, Olawale Salaudeen, Matthias Hein, Kevin Zhao, Alexander Pan, David Duvenaud, Bo Li, Steve Omohundro, Gabriel Alfour, Max Tegmark, Kevin McGrew, Gary Marcus, Jaan Tallinn, Eric Schmidt, Yoshua Bengio. "A Definition of AGI." 2025. arXiv:2510.18212.

[42] OpenAI. "GPT-4 Technical Report." arXiv preprint arXiv:2303.08774, 2023.

[43] OpenAI. "GPT-4 System Card." 2023. https://openai.com/research/gpt-4-system-card

[44] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, Denny Zhou. "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models." Advances in Neural Information Processing Systems 35 (NeurIPS 2022).

[45] Sébastien Bubeck et al. "Sparks of Artificial General Intelligence: Early experiments with GPT-4." arXiv preprint arXiv:2303.12712, 2023.

[46] Vladimir Pestov. "Intrinsic Dimension of a Dataset: What Properties Does One Expect?" Proceedings of the International Joint Conference on Neural Networks (IJCNN), IEEE, 2007, pp. 2959–2964.

[47] Elizaveta Levina, Peter J. Bickel. "Maximum Likelihood Estimation of Intrinsic Dimension." Advances in Neural Information Processing Systems 17 (NIPS 2004). MIT Press, 2005, pp. 777–784.

[48] Elena Facco, Maria d'Errico, Alex Rodriguez, Alessandro Laio. "Estimating the intrinsic dimension of datasets by a minimal neighborhood information." Scientific Reports. 2017, Vol. 7, 12140. doi:10.1038/s41598-017-11873-y

[49] J. Michael Steele. "Growth Rates of Euclidean Minimal Spanning Trees with Power Weighted Edges." The Annals of Probability. 1988, Vol. 16(4), pp. 1767–1787. doi:10.1214/aop/1176991596

[50] Benjamin Schweinhart. "Fractal dimension and the persistent homology of random geometric complexes." Advances in Mathematics. 2020, Vol. 372, 107291. doi:10.1016/j.aim.2020.107291

[51] M. R. Brito, E. L. Chávez, A. J. Quiroz, J. E. Yukich. "Connectivity of the mutual k-nearest-neighbor graph in clustering and outlier detection." Statistics and Probability Letters. 1997, Vol. 35(1), pp. 33–42. doi:10.1016/S0167-7152(96)00213-1

[52] Phillip Pope, Chen Zhu, Ahmed Abdelkader, Micah Goldblum, Tom Goldstein. "The Intrinsic Dimension of Images and Its Impact on Learning." International Conference on Learning Representations (ICLR 2021). arXiv:2104.08894.

[53] Alessio Ansuini, Alessandro Laio, Jakob H. Macke, Davide Zoccolan. "Intrinsic dimension of data representations in deep neural networks." Advances in Neural Information Processing Systems 32 (NeurIPS 2019). arXiv:1905.12784.

[54] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Matteo Carandini, Kenneth D. Harris. "High-dimensional geometry of population responses in visual cortex." Nature. 2019, Vol. 571, pp. 361–365. doi:10.1038/s41586-019-1346-5

[55] Rishidev Chaudhuri, Berk Gerçek, Biraj Pandey, Adrien Peyrache, Ila Fiete. "The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep." Nature Neuroscience. 2019, Vol. 22, pp. 1512–1520. doi:10.1038/s41593-019-0460-x

[56] Aicumene: https://aicumene.com/