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

A Preprint

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

August 12, 2025

Abstract

We introduce a system of instrumental and internally defined criteria for distinguishing levels of AI.

Let $\mathcal{I}$ denote an intelligence system and let $\mathcal{D}(\mathcal{I})$ denote the ensemble of data generated by $\mathcal{I}$ under standardized observational conditions. The proposed classification is not based on task performance or functional scope. Instead, it is defined through intrinsic structural invariants of $\mathcal{D}(\mathcal{I})$. In particular, we consider (among many) intrinsic dimension, topological and complexity characteristics. The resulting classification is based on internal criteria and differs from the familiar classifications of Weak AI and Strong AI.

We define a hierarchy $\{\mathsf{Level}_k\}_{k\geq 0}$ based on measurable structural properties of $\mathcal{D}(\mathcal{I})$.

Level_0 corresponds to Weak Artificial Intelligence (conventional narrow AI systems, mainstream AI, including specialized weak AI, and others).

Level_1 corresponds to Strong Artificial Intelligence (natural or artificial).

Level_2 corresponds to Super-Strong Artificial Intelligence. Higher levels extend this hierarchy inductively.

We provide a formal operational definition of Strong Artificial Intelligence as the class of systems $\mathcal{I}$ which generated data $\mathcal{D}(\mathcal{I})$ satisfy specified intrinsic structural invariants.

This definition distinguishes Strong AI from both Strong Natural Intelligence and from Super-Strong Artificial Intelligence within the same formal framework.

We present initial experimental data confirming the applicability of these criteria to differentiate intelligence levels.

Keywords AGI · Artificial General Intelligence · Strong AI · Super Strong AI · Complexity · Weak Ai · Hard Problem of Consciousness · Cognition · Whitney's Theorem · Intrinsic Dimension · Data topology

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 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.

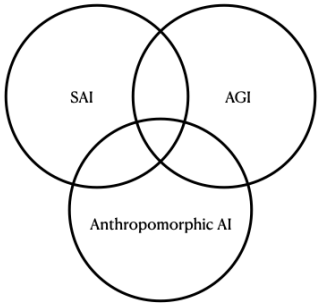

Fig. 1: Schematic representation of AIs.

We argue that this identification is fundamentally flawed. AGI, as commonly defined, is a behavioral 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. That is, neither Strong AI (SAI) nor Artificial General Intelligence (AGI) necessarily implies anthropomorphic intelligence (AAI). At the same time, 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. See 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 of intelligence and advance the hypothesis that intelligence is characterized by the statistical and geometric properties of the data it generates and the complexity and structure of the informational objects 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, as extracted from data, 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.

Animal intelligence is an adaptive mechanism optimized for survival and navigation within a three-dimensional physical environment. Consequently, 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.

Invoking the Whitney embedding theorem [1, 2], which guarantees that an n-dimensional manifold can be embedded into a space of dimension $2n + 1$, we obtain an upper bound of $2 \times 3 + 1 = 7$ for the effective dimensionality of semantic representations required for such tasks. This suggests that the functional semantic spaces instantiated in higher mammalian brains are constrained to dimensions on the order of seven.

Human intelligence constitutes a qualitative extension beyond this regime, as it operates over abstract semantic constructs that are not reducible to just physical navigation. Accordingly, the associated semantic spaces must exceed the animal bound. Empirical evidence supports this claim: both linguistic structures and neural data exhibit intrinsic dimensionalities significantly above this threshold [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15].

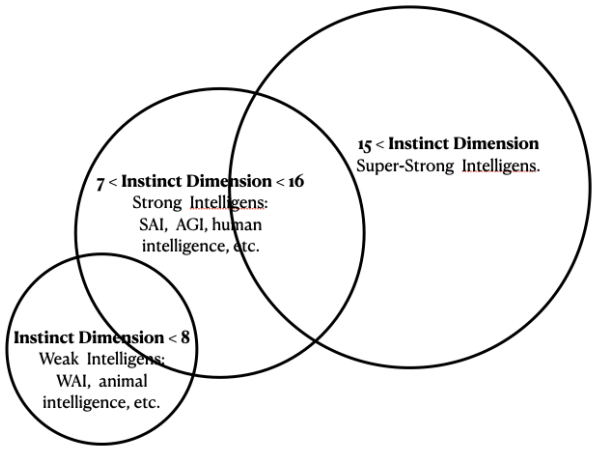

Fig. 2: Hierarchy of increasingly powerful forms of intelligence.

Interpreting the value 7 as both an upper bound for animal intelligence and a lower bound for human-level semantic processing [1], we obtain the next critical threshold $2 \times 7 + 1 = 15$.

Exceeding this boundary naturally corresponds to super-strong intelligence. Continuing this construction yields a hierarchy of increasingly powerful forms of intelligence. See Fig. 2.

Crucially, this hierarchy is non-anthropocentric: it does not rely on human-specific cognitive features, but instead arises from general geometric and dynamical constraints on information processing systems. We emphasize that this dimensional criterion can be used to characterize intelligences in general.

Note that the partially ordered scheme (Fig. 2) implies the existence of transitional regimes, i.e., systems capable of transitioning between distinct states and operating across multiple levels of intelligence. In this sense, intelligence may be viewed as a structured continuum, with transitions between regimes analogous to phase transitions in complex systems.

Such transitions may correspond to qualitative changes in the intrinsic structural properties of the generated data (e.g., intrinsic dimensionality, topology, and dynamical complexity), 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 (See Fig. 3). Different types of intelligence correspond to different sets of homological features and distinct topological classes, enabling a classification framework grounded in topological data analysis of large-scale structures.

Finally, intelligence manifests dynamically through trajectories evolving over these semantic manifolds. The complexity of such trajectories (quantified via measures of predictability, entropy, and chaos) constitutes a third, independent criterion. Together, these three invariants (dimension, topology, dynamics) define a measurable and theoretically grounded taxonomy of intelligence.

Importantly, these quantities are not merely abstract constructs but are empirically accessible. Measurements of intrinsic dimensionality for both human language and neural human brain signals yield values on the order of 9, indicating that human intelligence statistically exceeds the animal regime while remaining below the threshold associated with super-strong intelligence (naturally, this is a statistical observation, and for some individuals this value may be higher). This observation further supports the existence of a hierarchy of intelligence levels beyond the human domain.

[1] Evolutionary constraints impose a requirement for animals to operate on three-dimensional objects embedded within a space of dimension seven. By analogy, human cognition operates on higher-order semantic objects of dimension approximately seven, embedded in spaces of higher dimensionality. This reflects a qualitative extension of the representational capacity beyond that required for physical interaction.

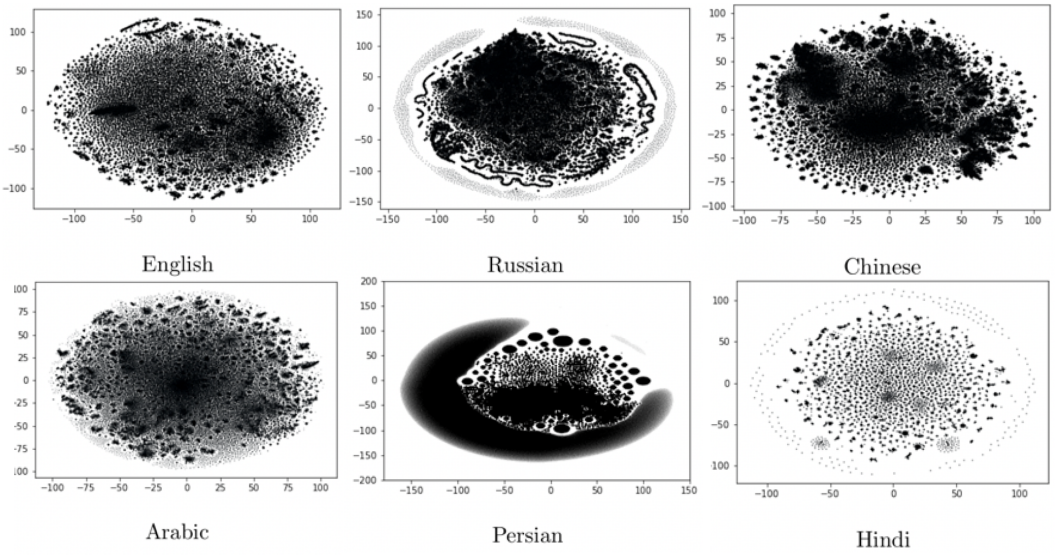

Fig. 3: Projection of languages embedding manifold

This perspective necessitates a strict demarcation between SAI and AGI. The framework introduced in "A Definition of AGI" [24] provides a comprehensive behavioral benchmark, quantifying performance across a set of cognitive tasks and introducing an "AGI score" as an aggregate metric. While this approach is methodologically sound, it remains fundamentally limited: it captures functional competence but not structural complexity.

An AI system may achieve high performance across all such benchmarks without undergoing a qualitative transition in its underlying informational geometry. In this sense, AGI remains within the same regime as advanced weak AI. Crucially, the cognitive capabilities enumerated in AGI frameworks are, in principle, simulable and can be reproduced without altering the intrinsic dimensionality, topology, or dynamical complexity of the underlying system.

We therefore conclude that AGI and SAI correspond to fundamentally different notions: the former is defined operationally through task performance, whereas the latter is defined intrinsically through the geometry and dynamics of generated information.

The remainder of the paper is organized as follows:

- Section 2 reviews related work, including existing definitions of AI, current AI systems, and empirical observations of strong intelligence in language and brain data.
- Section 3 introduces the fundamental characteristics of strong intelligence.
- Section 4 presents the concepts of intrinsic dimension, data topology, and complexity as structural invariants of intelligence.
- Section 5 discusses implications of the proposed framework, including the relationship between language and cognition, and outlines directions for future research.
- Section 6 concludes the paper.

2 Related Work

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 such as [25, 26], as well as by a growing body of work demonstrating emergent reasoning capabilities in large-scale architectures [27, 28]. These developments are often interpreted as evidence of a gradual approach toward human-level intelligence.

However, this interpretation 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.

As argued in the Introduction, the decisive factor is not the sophistication of transformations per se, but the capacity of the semantic space in which these transformations occur. Reasoning, in this framework, is constrained by the geometry and intrinsic dimensionality of the underlying representational manifold.

Empirical evidence supports this distinction. In the work "Spot the Bot" [5], 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. In particular, the authors demonstrate that machine-generated text occupies a lower-dimensional manifold compared to human-generated language, making it statistically distinguishable. This suggests that scaling alone may lead to denser sampling within a constrained semantic space, rather than to a qualitative expansion of that space.

More broadly, the current state of AI can be characterized as a combination of large-scale generative models with increasingly sophisticated training regimes and architectural refinements. While these systems exhibit impressive performance across a wide range of benchmarks, their operation remains largely confined to transformations within fixed representational regimes.

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.

The absence of such structural criteria in existing frameworks leads to a fundamental ambiguity. In particular, widely adopted definitions of AGI, such as the one proposed in [24], rely on aggregating performance across multiple cognitive domains into a unified "AGI score." While this approach provides a useful operational benchmark, it remains inherently behavioral and task-oriented. As a result, it does not capture the underlying structural complexity of intelligence.

Consequently, prevailing approaches do not offer operational criteria that distinguish Weak AI from Strong AI. This lack of differentiation often leads to a conceptual conflation of SAI with AGI, which we argue constitutes a methodological misalignment. An AI system may achieve high performance across a wide spectrum of tasks (thereby qualifying as AGI) without undergoing a qualitative transition in its intrinsic informational geometry.

At the same time, AGI represents an important and practically relevant direction of research, focused on generality and versatility across tasks. However, without a theory grounded in intrinsic dimensionality, topology, and dynamical complexity, AGI cannot serve as a sufficient proxy for Strong AI.

2.1 Available Definitions of AI

The diversity of interpretations of weak and strong artificial intelligence can, in practice, be reduced to two principal definitional strategies.[2]

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

Definition 1 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:

[2]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.

Definition 2 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.

2.2 Available AI Systems

Contemporary artificial intelligence systems, including advanced Large Language Models (LLMs) such as GPT, Gemini, and Claude, represent highly sophisticated instances of Weak AI (or Narrow AI). More recent developments, often referred to as Large Reasoning Models (LRMs), including systems such as DeepSeek-R1 and reasoning-enhanced variants of Claude and Gemini(e.g., o3-mini, DeepSeek-R1, Claude-3.7-Sonnet-Thinking), demonstrate significantly improved performance in tasks requiring multi-step inference and structured problem solving.

Despite these advances, such systems remain confined to transformations within learned representational spaces and therefore do not exhibit the structural properties associated with Strong AI.

The term Strong AI is frequently used interchangeably with Artificial General Intelligence (AGI), which shifts the focus toward generality across tasks. However, 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 specialized but operate at a level of complexity comparable to Strong Natural Intelligence.

The framework proposed in "A Definition of AGI" [24] formalizes AGI as performance across a set of cognitive benchmarks, introducing a quantitative "AGI score" that aggregates breadth and depth of capabilities. This approach is methodologically sound and practically useful, as it enables standardized evaluation of AI systems.

However, such frameworks remain fundamentally behavioral. They assess what a system can do, but not what the system is in terms of its intrinsic informational structure. As a consequence, high performance across cognitive benchmarks does not imply a qualitative transition from Weak AI to Strong AI.

Indeed, the key characteristics commonly attributed to Strong AI (e.g., generalization, problem-solving ability, 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.

This observation highlights the absence of fundamental criteria for identifying Strong Intelligence, whether natural or artificial. The reliance on extensive panels of cognitive tests obscures rather than resolves this issue, as such tests measure functional performance rather than intrinsic complexity.

Given the distinction between AGI and SAI, it is both justified and necessary to adopt a different evaluative framework for Strong AI. In particular, the identification of Strong Intelligence requires criteria that capture qualitative differences in the structure and dynamics of information processing, rather than quantitative differences in task performance.

In this work, we therefore define Strong AI not through an enumeration of cognitive abilities, but through fundamental structural characteristics that distinguish strong from weak intelligence.

2.3 Observation of Strong Intelligence on language data

A promising direction in the search for operational criteria of strong intelligence is the analysis of intrinsic dimensionality of linguistic structures. Language, as a primary product of natural intelligence, provides a natural empirical substrate for probing the geometry of semantic representations.

Recent works [3, 4, 5] 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 can be treated as a metric measure space (X, d, μ).

We employ three complementary notions of dimensionality:

2.3.1 Topological dimension

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

2.3.2 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.

2.3.3 Intrinsic dimension

Intrinsic dimension is defined as a mapping [29, 30]:

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

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

2.3.4 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 (SVD):

$$X \approx U_k \Sigma_k V_k^\top$$

Alternatively, neural embeddings such as CBOW (Word2Vec) are defined through:

$$\max \prod_t P(w_t \mid \text{context}(w_t))$$

The embedding of an n-gram is defined as the concatenation of embeddings of its constituent words.

2.3.5 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:

$$w_{ij} = \|x_i - x_j\|$$

Let T be the minimum spanning tree (MST). Define:

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

According to the Schweinhart estimator [31], for a d-dimensional distribution:

$$\mathbb{E}[L_\gamma(X_N)] \sim N^{\frac{d-\gamma}{d}}$$

which allows estimation of the fractal dimension d.

The Brito estimator [32] provides an estimate of topological dimension based on MST statistics and Bayesian inference over candidate dimensions.

2.3.6 Empirical result

A large-scale V.A. Gromov study across 70 languages from 21 language families demonstrates that:

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

for both word-level and n-gram representations, across embedding methods [4].

Crucially, this value is non-integer (indicating fractal structure), stable across languages, and invariant under representation choices.

2.3.7 Interpretation

The convergence of multiple independent estimators toward a stable, non-integer value of intrinsic dimensionality provides strong evidence that language occupies a fractal semantic space of dimension approximately 9.

If language is treated as an observable projection of internal semantic structures, then 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.

2.4 Observation of Strong Intelligence on 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 [10, 11, 12, 13, 16, 17, 18, 19, 20, 21, 23?]).

Let $\{x_i(t)\}_{i=1}^{M}$ denote multichannel neural recordings. We model (Bernadotte brain signal model, [10, 11, 12]) these signals 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.

2.4.1 MSSA reconstruction and oscillator structure

To extract these components, we employ Multichannel Singular Spectrum Analysis (MSSA) [12, 13]. Let $X(t)$ denote the multichannel signal. Construct the trajectory (Hankel) matrix:

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

Applying singular value decomposition:

$$\mathcal{H} = U \Sigma V^\top$$

yields a decomposition into principal components corresponding to oscillatory modes.

Theorem 1 (Recovery of exp-sinusoidal components via MSSA) [12, 13] 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 window length L and signal separability, MSSA reconstructs each exp-sinusoidal component as a pair (or low-rank block) of singular components in $\mathcal{H}$, and the original signal can be recovered as a 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.

2.4.2 Intrinsic dimension of thought

Applying intrinsic dimension estimators to the reconstructed phase-space trajectories (obtained via MSSA or delay embeddings), we obtain:

$$\dim_{\text{intrinsic}}(\text{thought}) \approx D_{\text{thought}}$$

where D_{thought} is empirically observed to be [12, 13]:

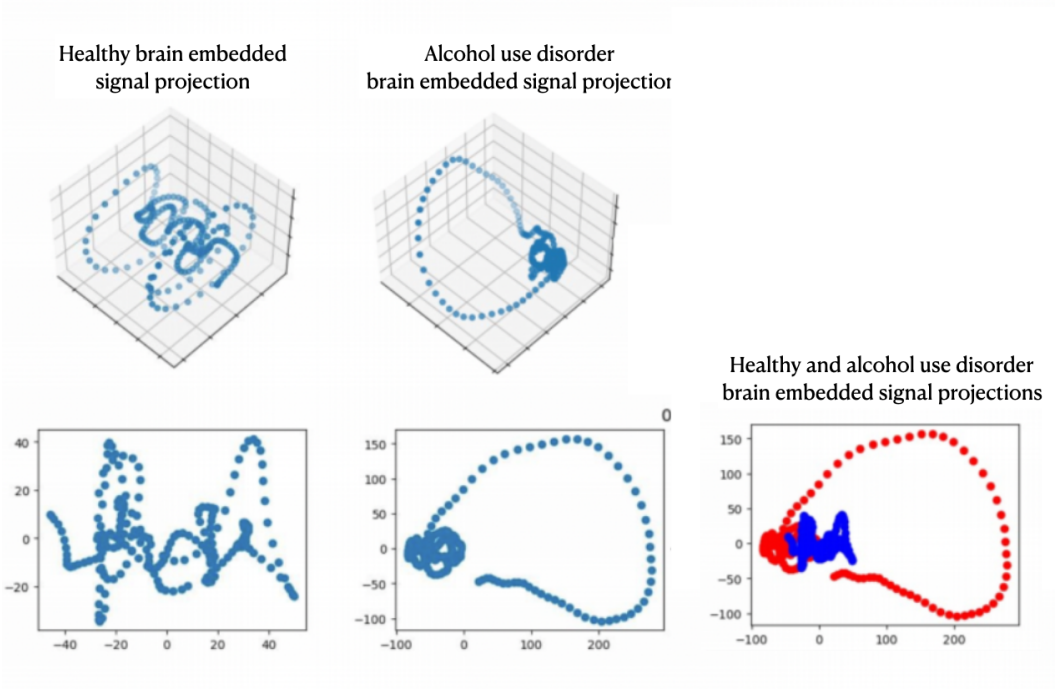

Fig. 4: 3D and 2D projections of healthy and alcohol use disorder brain embedded (obtained via MSSA) projections.

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

In Bernadotte's work, it is shown that a healthy brain produces an electromagnetic signal with intrinsic dimensionality in the range of 8–9, whereas in pathological states the brain exhibits dimensionality below 8 (See Fig. 4) [12, 13] .

Importantly, this value consistently exceeds the intrinsic dimension of language:

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

2.4.3 Interpretation and core result

These findings support a hierarchical relationship between observable linguistic structures and underlying cognitive dynamics:

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

Language thus provides a lower-bound projection of the semantic space, while neural dynamics reveal a higher-dimensional structure corresponding to internal cognitive processes.

This observation motivates the hypothesis of an ultra-narrow bottleneck, whereby high-dimensional cognitive representations undergo compression to an approximately nine-dimensional manifold when projected into the domain of consciousness responsible for speech production. In this sense, linguistic expression may be interpreted as a constrained information bottleneck, limiting the effective dimensionality of observable semantic structures.[3]

The transition from weak to strong intelligence can therefore be characterized as a transition between regimes of intrinsic dimensionality. Systems operating within low-dimensional manifolds may achieve high task performance, yet remain structurally limited. In contrast, strong intelligence corresponds to the emergence of high-dimensional, dynamically structured semantic spaces.

[3]Quisque verbum falsum est (lat.).

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.

3 Fundamental Characteristic of Strong 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.

That is, moving from the concepts of intelligence and consciousness that are difficult to formalize, we go into the area of data-science and, therefore, we can apply such formal Fundamental Characteristics of data as Data Structure (Data Topology), Intrinsic Dimension, and Complexity.

Next, we need to make some essential statements.

1. Definition 3 First, Weak AI, trained on data from SNI, has an ontological secondary nature in relation to SNI, therefore, the average Intrinsic Dimension and Complexity of the Weak AI (the Intrinsic Dimension and Complexity of it's outgoing data) D_{WAI} will be strictly less than the average Intrinsic Dimension and Complexity of SNI (the Intrinsic Dimension and Complexity of it's outgoing data) D_{SNI}:
$$D_{WAI} < D_{SNI}, \tag{1}$$
2. Definition 4 Second, Strong AI, trained on data from SNI, is ontologically equal to SNI, therefore, the average Intrinsic Dimension and Complexity of the Strong AI (it's outgoing data) D_{SAI} will be comparable to average Intrinsic Dimension and Complexity of SNI (it's outgoing data – we will omit this from now on) D_{SNI}:
$$D_{SAI} \approx D_{SNI}, \tag{2}$$

This provides a basic definition, while the topological structure of AI-generated data (see sections below) enables a further classification of intelligence systems, both weak and strong.

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 or self-consciousness), and toward other components of the same system. Differences in dimensionality between these components allow for a more refined classification of intelligence systems according to the types of consciousness they generate.

Thus, math theory and some understanding the ontology of Strong Intelligence allows us to find such an Fundamental Characteristic of Strong Intelligence and build our definition of Strong AI on it.

Of course, we could choose other Fundamental Characteristics to describe Strong Intelligence (whether natural or artificial). However, these three allow us to look at the data from different perspectives. Data Topology tells us what kinds of structures (components, loops, voids) exist. Intrinsic Dimension reflects the dimensionality of the effective diversity of the data and limits how many independent structures the data can have. Data Complexity quantifies the richness of those structures, combining dimension and topology of the object often through entropy, description, or algorithmic complexity.

Moreover, language, visual images, and other products of Strong Intelligence are justifiably classified as complex systems, or more precisely, as dynamic systems, and thus, further we will talk about Intrinsic Dimension of Dynamic Systems.

4 Intrinsic Dimension, Data Topology, and Complexity of Strong Intelligence

Before we dive into methods for estimating the Fundamental Characteristics of data, let's look at how these concepts are related and how they allow us to look at the data holistically. The concept of the Intrinsic Dimension, Data Topology, and Complexity of a given data can be defined in various ways, depending on assumptions of the data nature.

Definition 5 Intrinsic Dimension of a dataset (or a probability distribution) is the minimal number of parameters (variables or coordinates) needed to represent the data.

Definition 6 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 relation.

Definition 7 Complexity of a dataset X is a measure of the richness and diversity of its structural organization, determined jointly by its intrinsic dimension and topological features.

In particular, complexity increases with:

- the intrinsic dimension, which determines the number of independent degrees of freedom;
- the number and diversity of topological features (connected components, loops, voids);
- the variability and distribution of data within the underlying space.

These three concepts: Intrinsic Dimension, Complexity, and Data Topology can be linked into a single mathematical framework by considering data as a set in a metric space, onto which different topological structures are superimposed; while Complexity can be expressed in terms of both intrinsic dimension and topology:

1. Combinatorial complexity C_{comb} in terms of intrinsic dimension

   Let $X \subset \mathbf{R}^n$ be a dataset, with intrinsic dimension $ID(X) = d$, which is the minimal d such that X lies approximately on a manifold $M \subset \mathbf{R}^n$ of dimension d. The number of degrees of freedom is d. If each degree of freedom has m discrete states, then the number of possible configurations (Combinatorial complexity) grows exponentially in d:

   $$C_{comb}(X) \sim m^d, \tag{3}$$

2. When complexity C_{ent} is interpreted as entropy H, and data X are distributed on a d-dimensional manifold M with characteristic scale $\mathbf{R}$, then:

   $$C_{ent}(X) \propto H(X) \approx d \log R, \tag{4}$$

3. Topological Complexity C_{top} may be captured by both Intrinsic Dimension and Data Topology – topological invariants such as Betti numbers β_k describe connected components (β_0), loops (β_1), voids (β_2), etc.

   Intrinsic Dimension $ID(X) = d$ bounds the highest nontrivial homology group (Betti numbers β_i) of a $X \subset \mathbf{R}^n$ with intrinsic dimension $ID(X) = d$:

   $$\forall k \geq d : \beta_k(X) = 0 \tag{5}$$

   Topological Complexity C_{top} may be captured by the number of independent topological structures (components, loops, voids):

   $$C_{top}(X) = \sum_{i=0}^{d} \beta_i(X) \tag{6}$$

4. Algorithmic (Kolmogorov) Complexity C_{alg} binds Intrinsic Dimension $ID(X) = d$ and program length a compressed description. If data $X \subset \mathbf{R}^N$ lie on a low-dimensional manifold ($d << N$), then a compressed description C_{alg} exists with program length:

   $$C_{alg} \approx O(d) \tag{7}$$

   Kolmogorov complexity C_{alg} is maximal, if data X approach uniform noise in $\mathbf{R}^N$ ($d \approx N$).

While Intrinsic Dimension limits how many independent structures the data can have, and Data Topology tells us what kinds of structures exist, the Complexity $C(X)$ quantifies the richness of those structures, combining dimension and topology:

$$C(X) = F\Big(d, \{\beta_k(X)\}_{k=0}^{d}\Big), \tag{8}$$

the Intrinsic Dimension can be provided as a Fractal formulation: $d = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)}$, where N_ε is the number of ε-balls required to cover X.

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 the existence of 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, and therefore the null hypothesis should be that of a homeomorphic relationship between functional or morpho-functional spaces.

The apparent contradiction between smooth structures and empirical irregularity may be resolved through the use of minimal invariant manifolds (MIMs) and mappings between them.

The possibility of absence of any structural relationship appears unlikely.

For certain dimensions, Thurston's geometrization theorem states that any manifold can be decomposed into submanifolds corresponding to a finite set of canonical geometries.

We suggest a possible analogy between these geometric structures and those arising in functional spaces during cognitive processes. One of the submanifolds may be associated, via a homeomorphic mapping, with linguistic structures, while other submanifolds may correspond to cognitive processes not directly related to language (e.g., meditation or prayer).

5.2 Intrinsic Dimension of Super Strong Intelligence

When discussing the dimensionality of Strong Intelligence, we have to start from our empirical data, which shows that both the signals of the human brain and language are eight-dimensional, and from logical conclusions regarding the space with which intelligent beings operate.

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 the Whitney embedding theorem, we obtain that processing such information requires embedding into a space of dimension 7. This value can be interpreted as an upper bound for animal intelligence and a lower bound for human intelligence.

Extending this reasoning, embedding a 7-dimensional semantic space requires a space of dimension $2 \cdot 7 + 1 = 15$. Thus, dimensionality exceeding this threshold may correspond to super-strong intelligence.

These theoretical considerations are consistent with empirical observations obtained from language and brain data.

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 also evolve through interaction with increasingly complex systems.

From this perspective, evolution may be interpreted as an expansion of semantic dimensionality.

Another possible scenario involves the integration of human cognition with AI systems via neurointerfaces, leading to the emergence of new forms of intelligence. In this case, the evolution of such systems would also be characterized by increasing dimensionality of semantic spaces.

5.4 Higher-Dimensional Cognitive States

Currently, humans may only transiently experience higher-dimensional cognitive states. However, continued interaction with high-dimensional AI systems may lead to stabilization of such states through mechanisms of neuroplasticity.

This may result in the formation of stable high-dimensional attractors in cognitive dynamics, corresponding to new modes of thinking.

At the same time, degradation toward lower-dimensional states is also possible, as observed in certain pathological or altered conditions.

6 Conclusions

The distinction between different levels of intelligence should be based on formal statistical and geometric characteristics, rather than on task categories defined by human-centric relevance.

The fundamental measure of complexity of the spaces in which different types of natural and artificial intelligence operate is their intrinsic dimensionality. The application of the Whitney embedding theorem allows constructing a hierarchy of dimensional intervals. The first interval may be associated with animal cognition, the second with human intelligence, and the third with super-strong intelligence. This classification is independent of whether the intelligence is natural or artificial.

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. This value lies close to the lower bound (~ 7) of the human-related interval.

It may be hypothesized that in certain cognitive states (e.g., meditative or contemplative states), the effective dimensionality may increase and approach higher values within this interval. However, such statements require further experimental validation. Conversely, states that do not involve higher cognitive functions may correspond to a transition toward lower-dimensional regimes associated with animal-like cognition.

This observation is consistent with the hypothesis that intrinsic dimensionality reflects a characteristic property of human intelligence and is not dependent on a specific language.

At present, we observe a transition from predominantly biosocial evolution (which has historically characterized the development of Homo sapiens) 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.

Advances in computational capabilities and neurointerface technologies open the possibility of actively influencing this process. In particular, they suggest that expansion of cognitive capabilities may be understood not only metaphorically but also in terms of measurable changes in the structure of semantic spaces.

Another possible scenario of this co-evolutionary process involves increasing integration between human cognition and artificial intelligence systems. In such cases, the evolution of hybrid systems may also be characterized by the properties of the semantic spaces in which they operate.

At the same time, such developments may be associated with significant challenges, including potential social and biological consequences of artificially induced cognitive states. These aspects require careful study and further interdisciplinary investigation.

7 Funding Declaration

The study was implemented in the framework of the Basic Research Program at HSE University (HSE-BR-2025-001).

8 Declaration

The authors declare no competing interests.

Data availability statement: not applicable.

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 intrin- sic dimensionality and topological analysis of language structures, performed the language-related experiments, contributed to the theoretical interpretation of the results, and participating in drafting the manuscript. Both authors reviewed, edited, and approved the final manuscript.

REFERENCES

[1] Hassler Whitney. “Differentiable Manifolds.” Annals of Mathematics. 1936, Vol 31(3), pp. 645–680.

[2] John M. Lee. “Introduction to Smooth Manifolds.” 2012. Springer. doi: 10.1007/978-1-4419-9982-5 ISBN: 978-1-4419-9981-8

[3] 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, 2025. doi: 10.1155/cplx/9659172

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

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

[6] 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

[7] 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

[8] 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 Vol. 15(3), pp. 675-690. doi: 10.20537/2076-7633-2023-15-3-675-690

[9] 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

[10] 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, Turkiye, 2026, pp. 621-626, doi: 10.1109/ICARA69401.2026.11480398.

[11] I. Menshikov, N. Elfimov and 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, Turkiye, 2026, pp. 589-593, doi: 10.1109/ICARA69401.2026.11480292.

[12] Alexandra Bernadotte, Ivan Menshikov. “Brain Oscillator Intrinsic Dimension marks consciousness, recovery mode, and failed downshifting in coma.” Aicumene. Preprint. 2026. DOI: 10.13140/RG.2.2.31610.25288 Available: https://research.aicumene.com/papers/boid-coma/

[13] Alexandra Bernadotte, Ivan Menshikov. “Multichannel Singular Spectrum Analysis.” Aicumene. Preprint. 2026. Available: https://research.aicumene.com/papers/MSSA/

[14] Alexandra Bernadotte. “Oscillatory Model and Topological Structure of Brain Signals.” Aicumene. Preprint. Available: https://research.aicumene.com/papers/OscModel/

[15] Alexandra Bernadotte. “Topology-driven classification of time series.” bioRxiv 2026.04.25.720787. 2026. doi: 10.64898/2026.04.25.720787

[16] Chen Tao, Huang Haiyun, Pan Jiahui, Li Yuanqing. “An EEG-based brain-computer interface for automatic sleep stage classification.” 2018. 1988-1991. 10.1109/ICIEA.2018.8398035.

[17] 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.

[18] Luaute J., Morlet D., Mattout J.. "BCI in patients with disorders of consciousness: clinical perspectives." Ann Phys Rehabil Med. 2015. Vol 58(1), pp 29-34. doi: 10.1016/j.rehab.2014.09.015. Epub 2015 Jan 8. PMID: 25616606.

[19] 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. 10.52866/ijcsm.2023.04.04.006.

[20] Noirhomme Quentin, Kitney Richard, Macq Benoit. "Single-Trial EEG Source Reconstruction for Brain–Computer Interface." IEEE transactions on bio-medical engineering. 2008, Vol 55. 1592-601. 10.1109/TBME.2007.913986.

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

[22] Fruitet J., Clerc M.. "Reconstruction of cortical sources activities for online classification of electroencephalographic signals." Annu Int Conf IEEE Eng Med Biol Soc. 2010. 6317-20. doi: 10.1109/IEMBS.2010.5627713. PMID: 21097168.

[23] V. M. Buchstaber. "Time Series Analysis and Grassmannians." Applied problems of Radon transform, Amer. Math. Soc. Transl. 1994. Vol 2 (162), pages 1–17.

[24] 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:2510.18212v2

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

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

[27] Wei et al.. "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models." NeurIPS, 2022.

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

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

[30] Elizaveta Levina, and Peter J. Bickel. "Maximum Likelihood Estimation of Intrinsic Dimension." Advances in Neural Information Processing Systems (NeurIPS), 2005, pp. 777–784.

[31] Benjamin Schweinhart. "Fractal Dimension and the Persistent Homology of Random Geometric Complexes." arXiv preprint arXiv:1808.02196, 2019.

[32] M. R. Brito, E. L. Chavez, A. J. Quiroz, J. E. Yukich. "Connectivity of the mutual k-nearest-neighbor graph in clustering and outlier detection." Statistics and Probability Letters, 1997. 35(1), pp. 33–42.

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