American Journal of Information Science and Technology
2026, Vol. 10, No. 1, pp. 35–43
https://doi.org/10.11648/j.ajist.20261001.15

Research Article

The Extrapolation Horizon: Delineating the Boundaries Between Narrow Optimisation and General Intelligence

Partha Majumdar*

Department of Computer Science, Swiss School of Business and Management (SSBM), Geneva, Switzerland

Abstract

This analysis distinguishes between the narrow optimisation of current AI and the extrapolation needed for Artificial General Intelligence (AGI). Today, AI mainly consists of Artificial Narrow Intelligence (ANI), which is good at pattern recognition within its training data. Architectures like the Transformer, with mechanisms such as self-attention and gradient descent, are designed to identify statistical correlations within a specific distribution, such as spam SMS features. But this makes them fragile; they struggle to apply learned rules to slightly different domains, such as email filtering, because their knowledge is limited to the "convex hull" of the training examples. This isn't a scale issue but a structural one-so-called "zero-shot" learning is broad interpolation, not true skill. Conversely, biological cognition emphasises extrapolation-forming abstract, hierarchical schemas from limited experience and applying them to new situations. Human learning, like a child moving from crawling to walking, uses embodied cognition to relate concepts like gravity and momentum to physical reality, enabling effective knowledge transfer. Large Language Models are less data-efficient than humans, who learn language through social and physical grounding. Their failure on benchmarks such as the Abstraction and Reasoning Corpus (ARC-AGI), which test abstract rule induction from a few examples, highlights this gap. Public discussions on Brain-Computer Interfaces are often misleading; these systems are advanced, narrow AI applications for signal processing and decoding motor intentions, not for understanding or reading abstract thoughts. Achieving AGI will require a shift beyond merely scaling current architectures, which face data and energy constraints. The future involves integrating causal reasoning, neuro-symbolic systems, and embodied AI to close the gap between statistical pattern matching and genuine, general understanding.

Keywords

Extrapolation Gap, Narrow Optimisation vs. General Intelligence, Embodied Cognition, Manifold Hypothesis, Causal Reasoning

1. Introduction: The Taxonomy of the Artificial and the Biological

The current epoch of technological advancement is frequently mischaracterised as the dawn of artificial thought. In reality, it is the golden age of Artificial Narrow Intelligence (ANI)-a paradigm defined not by cognitive flexibility, but by hyper-specialised optimisation. As articulated in the foundational critique driving this investigation, we have constructed systems that excel at isolating patterns within specific distributions-such as identifying spam SMS messages-yet fail catastrophically when

*Correspondence: Partha Majumdar (partha.majumdar@hotmail.com)

Received: 12 January 2026; **Accepted:** 21 January 2026; **Published:** 4 February 2026

tasked with applying those learned principles to adjacent domains, such as email filtering, without explicit retraining. This limitation is not merely a quirk of current engineering but a fundamental property of the underlying mathematical architectures.

The prevailing discourse, amplified by industry figureheads, often conflates the fluency of Large Language Models (LLMs) with the reasoning capabilities of the human mind. However, a rigorous analysis reveals a stark divergence in mechanism. Human cognition is rooted in extrapolation-the ability to construct abstract schemas from limited experiences (e.g., a child transitioning from crawling to walking) and apply them to novel environments. In contrast, current AI systems operate through interpolation within high-dimensional manifolds, bounded by the "precise mathematics" of their training objectives.

This report provides an exhaustive analysis of this "Extrapolation Gap." We will dissect the mathematical rigidity of Transformer architectures, contrast them with the fluid dynamics of biological cognitive development, and critically examine claims surrounding Brain-Computer Interfaces (BCIs) such as Neuralink. By synthesising data from computational linguistics, developmental psychology, and neuroscience, we affirm the hypothesis that Artificial General Intelligence (AGI) remains a distant prospect, separated from our current capabilities by decades of necessary innovation in causal reasoning, embodiment, and energetic efficiency.

2. The Computational Substrate: Precise Mathematics of Narrow AI

To understand why an AI cannot spontaneously transfer knowledge from SMS to email, one must look beyond the anthropomorphic interfaces and examine the rigid mathematical operations that constitute "learning" in silicon. The observation that "we have to teach AI the precise mathematics" is technically precise. A neural network does not discover the laws of logic; it minimises a cost function defined by human engineers.

2.1. The Transformer Architecture: Optimisation, Not Comprehension

The dominant architecture of the modern AI landscape is the Transformer. While often described as "reading" text, it is, in functional reality, a sequence processing engine designed to maximise the statistical likelihood of the next token (x_{t+1}) given a context window of previous tokens ($x_1, x_2, ..., x_t$). [27].

2.1.1. The Mechanism of Self-attention

The Transformer's engine is Self-Attention. This mechanism allows the model to weigh the relevance of different words in a sequence to one another. However, this "relevance" is not semantic; it is vector alignment. For every token, the model generates three vectors: a Query (Q), a Key (K), and a Value (V). The attention score is calculated via a specific mathematical operation:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V$$

This formula dictates the model's "focus." The term QK^T represents the dot product between the Query and Key vectors. In a high-dimensional vector space, the dot product measures geometric alignment. If the vector for "Spam" points in the same direction as the vector for "Urgent," their product is high, and the model "attends" to the relationship. [27].

Crucially, this is a closed system. The model does not know what "Spam" *is*; it only knows its vector relationships to other tokens, such as "Winner," "Click," and "Free," within the specific distribution of SMS data. When the domain shifts to email-where "Spam" might be defined by HTML header anomalies or complex phishing narratives-the geometric relationships learned from SMS (short length, specific vocabulary) no longer hold. The "precise mathematics" that optimised the model for SMS creates a rigid manifold that does not encompass the email domain.

2.1.2. The Softmax Bottleneck and the Illusion of Choice

The perception that LLMs "write perfect text" and exhibit creativity is a byproduct of the Softmax Bottleneck. The final layer of an LLM produces a vector of "logits"-raw numerical scores for every word in the model's vocabulary. [14] These logits are converted into probabilities using the softmax function with a temperature parameter (T):

$$P_i = \frac{\exp(zi/T)}{\sum_j \exp(zi/T)}$$

This equation is the gatekeeper of the model's output.

1) Low Temperature ($T \to 0$): The distribution sharpens. The model becomes deterministic, selecting the most probable word from its training data. This results in the "perfect grammar," as the model defaults to the most statistically common (and thus grammatically correct) syntactic structures found in its massive training corpus. [14]
2) High Temperature ($T > 0.7$): The distribution flattens, allowing the model to sample lower-probability words. This stochasticity is often mistaken for "creative extrapolation." In reality, it is randomised interpolation-the model is exploring the "fuzzier" edges of the learned distribution, not stepping outside it to generate novel ideas. [14]

2.2. Gradient Descent: The Rigid Teacher

The process by which the machine "learns" these vector relationships is Backpropagation driven by Gradient Descent. This is the "teaching of precise mathematics."

During training, the model's predictions are compared to the actual data using a Loss Function (typically Cross-Entropy Loss). The error is propagated backwards through the network, and the partial derivative of the error with respect to every weight is calculated. The weights are then adjusted to reduce the error:

$$\theta_{new} = \theta_{old} - \eta \nabla_{\theta} J(\theta)$$

Where η is the learning rate and $\nabla_{\theta} J(\theta)$ is the gradient. This process moulds the network's parameters to fit the training data's manifold perfectly. However, once training stops, the weights are frozen. The model has no mechanism to update its "rules" based on new inputs during inference. It is a static artefact of its training run. To learn a new task (like recognising spam emails), the engineer must intervene, curate a new dataset, and re-run the optimisation process (Fine-Tuning). This is the antithesis of the autonomous extrapolation seen in biological systems.

2.3. Comparative Analysis: Narrow AI Optimisation vs. Biological Learning

The fundamental difference between the "precise mathematics" of AI and human learning can be summarised by the distinction between Curve Fitting and Schema Formation.

Table 1. Comparison of narrow AI and biological intelligence and their implications for extrapolation.

Feature	Narrow AI (e.g., LLMs, Spam Filters)	Biological Intelligence (e.g., Humans)	Implications for Extrapolation
Learning Mechanism	Backpropagation: Global optimisation of weights to minimise error on a specific dataset. [7]	Hebbian Plasticity & Synaptic Consolidation: Local, real-time strengthening of connections based on causal association. [25]	AI requires massive retraining for new tasks; humans adapt in real-time.
Knowledge Representation	High-Dimensional Vectors: Dense, continuous representations based on statistical co-occurrence. [27]	Hierarchical Schemas: Abstract, structured representations of concepts, causality, and relationships. [1]	AI "concepts" are fragile correlations; human schemas are robust and transferable.
Generalisation Type	Interpolation: Predicting values *within* the range of training data. [28]	Extrapolation: Applying rules to entirely new domains *outside* prior experience. [20]	AI fails when data distribution shifts (SMS → Email); humans succeed.
Data Efficiency	Low: Requires trillions of tokens to approximate grammar. [19]	High: Requires minimal exposure (millions of words) to master language. [5]	Humans leverage "priors" and embodiment; AI relies on brute force.
Update Frequency	Static: Weights are frozen after training; no learning during inference. [29]	Continuous: The brain constantly updates its world model with every millisecond of experience. [16]	AI is an archive of the past; humans are adaptive agents of the present.

3. The Geometry of Limitations: Why AI Cannot Extrapolate

To deeply understand the "SMS to Email" failure, we must engage with the geometry of high-dimensional spaces. This section explores why what appears to be learning is often merely sophisticated memory retrieval.

3.1. The Manifold Hypothesis and the Convex Hull

Machine learning theory relies on the Manifold Hypothesis, which posits that real-world data (such as valid English sentences or images of faces) lies on a low-dimensional topological manifold embedded in the high-dimensional input space. "Learning" is the process of mapping this manifold.

When an AI is asked to classify a new SMS message, it is performing interpolation. If the new message falls within the "convex hull" (the geometric envelope) of the training examples, the model can accurately predict the label by averaging the behaviours of nearby training points.

The Failure of Extrapolation:

Extrapolation, in the strict mathematical sense, occurs when a query lies outside the convex hull of the training data-that is, beyond the geometric envelope formed by all previously observed examples. In modern neural networks, this envelope exists in an extremely high-dimensional feature space, often

comprising thousands of dimensions. In such spaces, a counterintuitive but well-established phenomenon emerges: the volume of space outside the convex hull grows exponentially relative to the volume inside it. This asymmetry, known as the Curse of Dimensionality, implies that almost all possible inputs lie outside the region in which the model has empirical support. Consequently, what appears to be a modest generalisation shift from a human perspective can correspond to a vast geometric displacement from the model's standpoint.

Recent theoretical work by LeCun and Balestriero formalises why this geometric reality is fatal to extrapolation. Neural networks, despite their apparent complexity, are fundamentally spline approximators: they construct smooth, continuous functions that interpolate between known data points. Within the region spanned by the training data, these splines behave predictably, producing stable and often highly accurate outputs. Outside that region, however, there are no constraints enforcing meaningful behaviour. The learned function may flatten, oscillate, or diverge arbitrarily, not because the model is malfunctioning, but because the optimisation process never defined how it should behave there. Extrapolation is therefore not merely difficult for neural networks-it is mathematically underdetermined. [28].

This geometric limitation explains why transferring knowledge from "Spam SMS" detection to "Spam Email" detection fails so reliably. Although both are labelled with the same semantic category by humans, they occupy fundamentally different regions of the feature space. Spam SMS messages are characterised by short length, limited vocabulary, and the absence of structural metadata. Spam emails, by contrast, involve longer narratives, HTML formatting, header information, and complex phishing patterns. As a result, spam emails lie far outside the manifold learned from SMS data. The learned decision boundary-optimised with precise mathematics for SMS-does not extend into this new region in any principled way.

Crucially, this failure does not reflect a lack of intelligence in the colloquial sense, nor a deficiency in training scale. The model lacks an abstract or causal concept of "spam" to guide its behaviour in unfamiliar domains. Instead, it possesses a detailed geometric map of SMS spam-a local approximation valid only within a narrowly defined territory. When the domain shifts, the map simply ends. Without an internal theory to bridge the gap between domains, the model has no basis for inferring how the concept should manifest under new surface features. In this sense, the system is not confused; it is blind. Its precise mathematical formulation was never designed to generalise beyond the manifold on which it was optimised.

This distinction between mapping and theorising marks the fundamental divide between narrow artificial intelligence and biological cognition. Humans abstract causal structure from experience, allowing concepts such as deception, intent, or manipulation to transfer seamlessly across media. [22] Neural networks, by contrast, remain bound to the geometry of their training data. Until artificial systems can form representations that transcend local manifolds and encode causal regularities rather than statistical proximity, extrapolation will remain an inherent and unavoidable failure mode rather than a solvable engineering challenge.

3.2. The Illusion of "Zero-Shot" Generalisation

Proponents of the "AI is already AGI" narrative often cite Zero-Shot Learning (ZSL) as evidence of extrapolation. ZSL is the ability of a model to perform a task it was not explicitly trained to do (e.g., GPT-4 translating a language pair it was not fine-tuned on).

However, a closer examination reveals this is often a misnomer.

1) Implicit Pre-training: LLMs are trained on internet-scale datasets (Common Crawl). It is highly probable that the model has seen similar tasks, instructions, or the underlying structure of the "novel" problem during pre-training. It is not generating a new skill; it is retrieving a latent capability. [13]
2) In-Context Learning (ICL): When a user provides a prompt with examples (Few-Shot Prompting), the model uses the attention mechanism to copy the pattern *within the context window*. This is not a weight update (learning); it is a temporary activation state. The model is effectively running a "mini-program" defined by the prompt. This capability, while impressive, is bound by the training distribution. If the prompt requires a logic totally alien to the training data (like the ARC-AGI benchmark), ICL fails. [29]

The "Zero-Shot" capability is, therefore, better understood as Broad Interpolation. The model covers such a vast area of the conceptual space that many tasks fall within its interpolated range. But this is not the infinite adaptability of AGI; it is just a very large, static library. [26]

4. The Biological Standard: Embodiment and Schema Theory

The human child's ability to extrapolate from "crawling to walking" is the gold standard of intelligence. This capability is not magic; it is the result of Embodied Cognition [9] and Schema Theory [18], mechanisms absent in current AI.

4.1. Piagetian Schemas: The Architecture of Transfer

Jean Piaget's theory of cognitive development offers a robust framework for understanding human extrapolation. Piaget defined learning as a dynamic equilibrium between two processes: Assimilation and Accommodation. [25].

1) Assimilation (Interpolation): The child integrates new information into an existing schema. (e.g., Recognising a Great Dane is a "dog" because it fits the "four legs,

barks" schema). Narrow AI does this well.

2) Accommodation (Extrapolation): The child modifies the schema to handle conflicting information. (e.g., Realising a cat is *not* a dog, creating a new "cat" schema, and refining the "animal" super-schema). Narrow AI cannot do this. Its "schemas" (weights) are frozen.

The Hierarchical Advantage:

Humans organise knowledge hierarchically. A "Spam" schema in a human mind is abstract: Spam = Unsolicited + Deceptive + Transactional. Because this schema is abstract, it is independent of media. A human can apply the "Deception" attribute to an SMS, an email, or a knock on the door. AI models operate on surface features (tokens, pixels). They lack the abstract hierarchical layer that permits transfer across distinct feature sets. [1].

4.2. Embodied Cognition: Grounding Knowledge in Physics

The transition from crawling to walking is a prime example of Embodied Extrapolation. [9].

1) The Crawling Phase: The infant learns Sensorimotor Contingencies-the rules governing how sensory inputs change based on motor actions. They learn "optic flow" (how the world appears to move when I move) and "proprioception" (where my limbs are). [20].
2) The Transfer: When learning to walk, the infant does not discard the data from crawling. They extrapolate the *concepts* of balance, gravity, and momentum. The "motor babbling" (random movements) is not truly random; it is constrained by the physics learned during crawling.
3) The AI Deficit: An AI trained on a dataset of crawling videos has learned the *visual pattern* of crawling. It has learned no physics. If you ask it to generate a walking video, it has no underlying model of gravity or mass to guide the transition. It hallucinates because its symbols are ungrounded-they refer to other symbols rather than to physical realities. [12].

4.3. Language Acquisition: The Data Efficiency Paradox

The comparison between child language acquisition and LLM training highlights the inefficiency of disembodied learning.

1) The Child: Achieves fluency with ~10 million words of input.
2) The LLM: Requires ~10 trillion words to achieve comparable fluency. [19].

Why the discrepancy? The child uses Joint Attention and Social Intent. When a parent says, "Look at the dog," the child follows the parent's gaze, triangulates the object, and anchors the word "dog" to the physical entity. The syntax is acquired as a tool to navigate this social-physical reality. [5] The LLM, lacking a body or social goals, must rely on the brute force of the Distributional Hypothesis-inferring meaning solely from the statistical company words keep. It requires orders of magnitude more data to triangulate meaning, because it is solving a much harder problem: learning about the world from text alone, without ever seeing it. [23].

5. The Benchmark of Failure: ARC-AGI and the Limits of Reasoning

The most empirical proof that current AI lacks extrapolation capabilities is found in the Abstraction and Reasoning Corpus (ARC-AGI), developed by François Chollet. This benchmark was explicitly designed to test the definition of intelligence: the ability to efficiently acquire new skills. [3].

5.1. The Nature of the Test

The Abstraction and Reasoning Corpus (ARC-AGI) is deliberately constructed to isolate the core faculty that distinguishes general intelligence from narrow pattern recognition: the ability to infer abstract rules from minimal evidence and apply them to novel situations. Each ARC task consists of a small visual grid composed of coloured squares, presented as a set of input–output pairs. The solver is provided with only two or three such demonstrations and must generate the correct output for a new input that obeys the same underlying transformation. No explicit instructions, labels, or task descriptions are given. The problem must be understood solely through the structure of the examples themselves.

Crucially, the transformations required to solve ARC tasks are not surface-level statistical regularities. They invoke abstract relational concepts such as symmetry, continuation, containment, object persistence, directional motion, or gravity-like behaviour, where objects appear to “fall,” “collide,” or “stop” when encountering boundaries. The solver must therefore move beyond memorising pixel configurations and instead induce a latent rule that governs how objects interact within the grid. This process mirrors few-shot learning in its strongest sense: the task is not to recognise a familiar pattern, but to *construct a hypothesis* about the generative process that produced the examples and then apply that hypothesis to unseen data.

ARC is explicitly designed around the assumption of what François Chollet terms *core knowledge priors*-basic cognitive assumptions about the world that humans possess prior to formal learning. These include expectations such as object permanence (objects continue to exist even when they move or change position), identity preservation (an object remains the same object despite transformation), and the relevance of colour, shape, and spatial adjacency. Such priors are so fundamental that even young children apply them effortlessly. A human solver does not need to be taught that a contiguous block of squares constitutes an “object,” or that symmetry should be preserved unless violated; these assumptions are automatically brought to bear when interpreting the task. [3].

The importance of these priors cannot be overstated. ARC tasks do not reward exhaustive search or brute-force pattern matching; instead, they penalise systems that lack an internal model of objects and relations. Without assumptions about persistence, boundaries, and causal structure, the search space of possible transformations becomes combinatorially intractable. Humans prune this space almost instantly by applying intuitive physical and logical constraints. In contrast, most contemporary AI systems lack these built-in priors and must approximate solutions by statistical similarity to previously seen patterns. When no such pattern exists-as is often the case in ARC-the system fails not due to insufficient compute, but because the benchmark presupposes the abstract representational scaffolding that is absent.

In this sense, ARC-AGI does not test intelligence as performance on a fixed skill, but intelligence as *skill acquisition itself*. Success requires the solver to rapidly form a new internal rule, manipulate symbolic relationships, and generalise that rule beyond the examples provided. This is precisely the mode of reasoning at which humans excel, and narrow AI systems struggle. The benchmark thus serves not as an incremental challenge to existing architectures, but as a diagnostic instrument, revealing the structural limits of systems that rely on interpolation rather than genuine abstraction and extrapolation.

5.2. The Failure of the State-of-the-Art

Despite the massive scaling of models like GPT-5 and o1 (Strawberry), their performance on ARC-AGI reveals a profound deficit.

1) Human Performance: ~85-90% accuracy. Humans easily identify the abstract rule (e.g., "move the blue square until it hits the red wall").
2) AI Performance: SOTA models, even with heavy compute and "Chain of Thought" reasoning, struggle to break 40-50% on the public set and score significantly lower on the private evaluation set.
3) The Why: LLMs attempt to solve ARC by retrieving memorised programs or finding approximate patterns in their training data. They lack a discrete program synthesis engine. They cannot define a new variable or a new logical operator on the fly. When the task requires a logical leap that isn't represented in the training corpus (extrapolation), the model fails.

Recent attempts to "solve" ARC often rely on generating thousands of Python programs and checking them against the examples (Test-Time Compute). While this improves scores, it is a brute-force search rather than an intelligent extrapolation. It is akin to trying every key on a keychain rather than looking at the lock's shape.

6. The Neural Interface Fallacy: Why Neuralink Is Not AGI

Elon Musk's claims that Neuralink represents a path to AGI or "reading thoughts." This is scientifically warranted. The conflation of Signal Processing with Cognitive Understanding is a major source of public confusion.

6.1. The Engineering of Neural Interfaces

To understand what Neuralink does, we must examine the signal processing pipeline, which relies on Narrow AI.

1) Acquisition: The implant records extracellular electrical potentials using micro-electrodes.
2) Spike Sorting: The raw signal is noisy. Narrow AI algorithms (such as Convolutional Neural Networks or PCA-based clustering) are used to detect "spikes" (Action Potentials) and assign them to specific neurons. This is a classification task: *Is this voltage spike noise, Neuron A, or Neuron B?*. [10].
3) Decoding: A decoding algorithm (e.g., a Kalman Filter or a Recurrent Neural Network) maps the *spik*e rate to a desired output, such as the X/Y coordinates of a mouse cursor.

The Narrowness of the Decoder:

This decoder is the definition of Narrow AI. It is trained on a specific dataset: Subject A imagines moving their hand left → Neurons 1, 4, and 9 fire.

1) No Extrapolation: If Subject A decides to control the cursor by imagining *singing a song* instead of moving their hand, the decoder fails. It cannot extrapolate the *intent* "move cursor" from a new neural pattern. It must be explicitly retrained on the "singing" signals. [15].
2) Brittleness: The neural code is non-stationary. As the brain changes (neuroplasticity) or the electrodes shift slightly, the decoder's performance degrades. It requires constant recalibration (retraining), unlike a human who can adapt to a new tennis racket instantly. [32].

6.2. The "Reading Thoughts" Myth

The claim that such devices can "read thoughts" implies access to the semantic content of the mind. This is biologically implausible with current technology.

1) The Location Problem: Neuralink primarily targets the Motor Cortex (movement). Abstract thoughts, memories, and language are distributed across the Prefrontal Cortex, Temporal Lobes, and Hippocampus. Recording the motor cortex tells you *how* the person wants to move, not *why* or *what* they are thinking about. [30].
2) The Translation Problem: There is no universal "Neural Code." The pattern of neurons firing for the concept "Democracy" in Person A's brain is completely different from Person B's. It is shaped by their unique life history and synaptic weights. An AI cannot "learn" a universal dictionary of thought because one does not exist. It can only learn a personalised, narrow dictionary for one user. [31].

6.3. Large Brain Models (LaBraM): Scaling Narrow AI to the Brain

Recent research has introduced Large Brain Models (LaBraM)-Transformers trained on massive datasets of EEG and fMRI data. [11].

1) The Promise: By pre-training on thousands of hours of brain signals, these models learn general representations of neural activity.
2) The Reality: They function exactly like LLMs. They predict the "next token" (next segment of the EEG signal). They are excellent at anomaly detection (i.e., identifying seizures) and state classification (i.e., sleep stages).
3) The Limitation: They do not understand the *content*. A LaBraM can tell you that a subject is "alert" or "processing visual stimuli," but it cannot extrapolate to tell you *what* the subject is seeing unless it has been supervised-trained on that specific image-to-signal pair. It is a statistical signal processor, not a mind reader. [11].

Neuralink and LaBraM are instances of Narrow AI applied to biological data. They process specific data (electrical spikes) to perform one task (cursor control or seizure prediction). They possess no general understanding of the human mind. [24].

7. The Trajectory of AGI: Scaling Walls and Future Horizons

I posit that AGI is "30-40 years away." This estimate is supported by a significant body of research that identifies the structural limitations of the current deep learning paradigm.

7.1. The Limits of Scaling: Chinchilla and Diminishing Returns

The "Scaling Hypothesis"-the idea that simply adding more data and compute will yield AGI-is facing mathematical and physical realities.

1) The Data Wall: The Chinchilla Scaling Laws define the optimal ratio of model size to training data. [6] We are rapidly exhausting the supply of high-quality human text on the internet. Synthetic data (AI training on AI output) risks "Model Collapse," in which the distribution loses variance, and quality degrades. [8].
2) The Energy Wall: The human brain operates on ~20 watts of power. A cluster training a GPT-4 level model consumes gigawatts. This efficiency gap (a factor of millions) suggests that our current algorithms are fundamentally inefficient approximations of intelligence.

7.2. Moravec's Paradox and the Physical Barrier

Moravec's Paradox remains the most stubborn obstacle to AGI. It observes that high-level reasoning (chess, algebra) is computationally cheap, while low-level sensorimotor skills (walking, folding laundry) are computationally expensive. [21].

1) Why it Matters: AGI implies the ability to navigate the physical world. If an AI cannot perform the "simple" task of clearing a dinner table without breaking dishes (a task a 5-year-old can do), it cannot be considered general.
2) The Implication: We have solved the "easy" problems (generating text). We are only now beginning to address the "hard" problems (robotics and causal physics). This requires a shift from disembodied LLMs to Embodied AI, a field that is decades behind language modelling in maturity.

7.3. The Timeline of Paradigm Shifts

The history of AI is cyclic. We are currently at the peak of the "Deep Learning Summer." The transition to AGI will likely require a new paradigm that integrates:

1) Causal Inference: Systems that model cause-and-effect (Judea Pearl's Ladder of Causality) rather than just correlation. [2].
2) Neuro-Symbolic AI: Architectures that combine the learning capability of neural networks with the logical rigour of symbolic systems (to solve problems like ARC-AGI). [3].
3) Neuromorphic Computing: Hardware that mimics the brain's spiking architecture to achieve energy efficiency. [4].

Given that these technologies are largely in the research phase (Technology Readiness Level 1-4), a timeline of 30-40 years for their maturation and integration into a coherent AGI system is a scientifically grounded projection. [17].

8. Conclusion

The scepticism regarding the current state of AI is not only justified but also aligned with the field's deepest technical realities. The distinction between Narrow AI and AGI is not merely one of scale; it is one of fundamental topology.

Narrow AI, exemplified by LLMs and current BCI decoders, relies on the precise mathematics of interpolation. It constructs rigid manifolds within high-dimensional spaces, optimising weights to minimise error on specific distributions. It excels where data is abundant and rules are static, but it is fundamentally brittle. It cannot step outside its training hull to identify spam in a new medium or decode a thought it hasn't explicitly mapped.

In contrast, biological intelligence is defined by extrapolation. Humans leverage hierarchical schemas, embodied grounding, and causal reasoning to bridge the gap between the known and the unknown. The child's transition from crawling to walking is a triumph of transferring physical principles to a new motor context-a feat no current AI can replicate without extensive retraining.

The "reading of thoughts" by machines remains a fantasy of

extrapolation-projecting the success of motor decoding onto the vastly more complex domain of semantic thought without the necessary scientific bridge. Until we move beyond the "Next Token Prediction" dogma and build systems that can model the universe's causal structure rather than just its statistical structure in text, AGI will remain a futuristic aspiration. We are not on the verge of creating a mind; we are perfecting the ultimate mirror.

Abbreviations

AGI	Artificial General Intelligence
AI	Artificial Intelligence
ANI	Artificial Narrow Intelligence
ARC-AGI	Abstraction and Reasoning Corpus – Artificial General Intelligence
BCI	Brain-Computer Interface
EEG	Electroencephalogram
fMRI	Functional Magnetic Resonance Imaging
GPT	Generative Pre-trained Transformers
HTML	Hyper-Text Markup Language
ICL	In-Context Learning
LaBraM	Large Brain Model
LLM	Large Language Model
PCA	Principal Component Analysis
SMS	Short Messaging Service
ZSL	Zero-Shot Learning

Author Contributions

Partha Majumdar is the sole author. The author read and approved the final manuscript.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Ansari, N. A. (2019, June). *Role and Importance of Schemas in Pedagogy and Learning: A Cognitive Approach*. Journal of Communication and Cultural Trends. Retrieved January 12, 2026, from https://doi.org/10.32350/jcct.11.03

[2] Bareinboim, E., Correa, J. D., Ibeling, D., & Icard, T. (2022, March 4). *On Pearl's Hierarchy and the Foundations of Causal Inference*. Probabilistic and Causal Inference. Retrieved January 12, 2026, from https://doi.org/10.1145/3501714.3501743

[3] Chollet, F. (2019, November 25). *On the Measure of Intelligence*. Cornell University. Retrieved January 12, 2026, from https://arxiv.org/abs/1911.01547

[4] Cui, S., Lee, D., & Wen, D. (2024, December 4). *Toward brain-inspired foundation model for EEG signal processing: Our opinion*. Frontiers in Neuroscience. Retrieved January 12, 2026, from https://doi.org/10.3389/fnins.2024.1507654

[5] Cuskley, C., Woods, R., & Flaherty, M. (2024, August 31). *The Limitations of Large Language Models for Understanding Human Language and Cognition*. Open Mind. Retrieved January 12, 2026, from https://doi.org/10.1162/opmi_a_00160

[6] Feng, T., Jin, C., Liu, J., Zhu, K., Tu, H., Cheng, Z., Lin, G., & You, J. (2024, May 16). *How Far Are We From AGI?* ArXiv. Retrieved January 12, 2026, from https://arxiv.org/html/2405.10313v1

[7] Grum, M. (2023, September 14). *Learning Representations by Crystallized Back-Propagating Errors*. ResearchGate. Retrieved January 12, 2026, from https://doi.org/10.1007/978-3-031-42505-9_8

[8] Hoffmann, J., Borgeaud, S., & Mensch, A. (2022, March 29). *Training Compute-Optimal Large Language Models*. Cornell University. Retrieved January 12, 2026, from https://doi.org/10.48550/arXiv.2203.15556

[9] Hoffmann, M. (2025, May 15). *Embodied AI in Machine Learning -- is it Really Embodied?* Cornell University. Retrieved January 12, 2026, from https://arxiv.org/abs/2505.10705

[10] Horton, P. M., Nicol, A. U., Kendrick, K. M., & Feng, J. F. (2007, February 15). *Spike sorting based upon machine learning algorithms (SOMA)*. Journal of Neuroscience Methods. Retrieved January 12, 2026, from https://doi.org/10.1016/j.jneumeth.2006.08.013

[11] Jiang, W. B., Zhao, L. M., & Lu, B. L. (2024, May 29). *Large Brain Model for Learning Generic Representations with Tremendous EEG Data in BCI*. ArXiv. Retrieved January 12, 2026, from https://arxiv.org/html/2405.18765v1

[12] Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2013, December 16). *Crawling and walking infants see the world differently*. Child Development. Retrieved January 12, 2026, from https://doi.org/10.1111/cdev.12206

[13] Luca, C. (2023, August). *Challenges and Limitations of Zero-Shot and Few-Shot Learning in Large Language Models*. ResearchGate. Retrieved January 12, 2026, from https://www.researchgate.net/publication/388920473_Challenges_and_Limitations_of_Zero-Shot_and_Few-Shot_Learning_in_Large_Language_Models

[14] Masarczyk, W., Ostaszewski, M., Cheng, T. S., Trzciński, T., Lucchi, A., & Pascanu, R. (2025, June 2). *Unpacking Softmax: How Temperature Drives Representation Collapse, Compression and Generalization*. ArXiv. Retrieved January 12, 2026, from https://arxiv.org/html/2506.01562v1

[15] Meyer, L. M., Zamani, M., Rokai, J., & Demosthenous, A. (2024, November 14). *Deep learning-based spike sorting: A survey*. Journal of Neural Engineering. Retrieved January 12, 2026, from https://doi.org/10.1088/1741-2552/ad8b6c

[16] Moscovitch, D. A., Moscovitch, M., & Sheldon, S. (2023, February 16). *Neurocognitive Model of Schema-Congruent and -Incongruent Learning in Clinical Disorders: Application to Social Anxiety and Beyond*. Perspectives on Psychological Science. Retrieved January 12, 2026, from https://doi.org/10.1177/17456916221141351

[17] Müller, V. C., & Bostrom, N. (2025, August 9). *Future progress in artificial intelligence: A survey of expert opinion*. Cornell University. Retrieved January 12, 2026, from https://doi.org/10.48550/arXiv.2508.11681

[18] Pankin, J. (2013). *Schema Theory*. Massachusetts Institute of Technology. Retrieved January 12, 2026, from https://web.mit.edu/pankin/www/Schema_Theory_and_Concept_Formation.pdf

[19] Pantcheva, M. (2023, September 19). *How do LLMs and humans differ in the way they learn and use language*. Retrieved January 12, 2026, from https://www.rws.com/blog/large-language-models-humans/

[20] Radovanovic, M., & Sommerville, J. A. (2025, July 16). *Embodied Cognition in Child Development*. Oxford Research Encyclopedias. Retrieved January 12, 2026, from https://doi.org/10.1093/acrefore/9780190236557.013.886

[21] Rotenberg, V. S. (2017, February 20). *Moravec's Paradox: Consideration in the Context of Two Brain Hemisphere Functions*. Activitas Nervosa Superior. Retrieved January 12, 2026, from https://doi.org/10.1007/BF03379600

[22] Safron, A., Hipólito, I., & Clark, A. (2023, November 14). *Editorial: Bio A. I. - from embodied cognition to enactive robotics*. PubMed Central. Retrieved January 12, 2026, from https://pmc.ncbi.nlm.nih.gov/articles/PMC10682788/

[23] Sahlgren, M. (2008). *The distributional hypothesis*. Italian Journal of Linguistics. Retrieved January 12, 2026, from https://www.researchgate.net/publication/228385506_The_distributional_hypothesis

[24] Saif-ur-Rehman, M., Ali, O., Dyck, S., Lienkämper, R., Metzler, M., Parpaley, Y., Wellmer, J., Liu, C., Lee, B., Kellis, S., Andersen, R., Iossifidis, I., Glasmachers, T., & Klaes, C. (2021, February 5). *SpikeDeep-classifier: A deep-learning based fully automatic offline spike sorting algorithm*. Journal of Neural Engineering. Retrieved January 12, 2026, from https://doi.org/10.1088/1741-2552/abc8d4

[25] Ser, J. D., Lobo, J. L., Müller, H., & Holzinger, A. (2025, May 19). *World Models in Artificial Intelligence: Sensing, Learning, and Reasoning Like a Child*. ArXiv. Retrieved January 12, 2026, from https://arxiv.org/html/2503.15168v1

[26] Towns, A. (2024, December 24). *What Is Zero Shot Learning? Benefits and Limitations*. Retrieved January 12, 2026, from https://learn.g2.com/zero-shot-learning

[27] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023, August 2). *Attention Is All You Need*. Cornell University. Retrieved January 12, 2026, from https://arxiv.org/abs/1706.03762

[28] Webb, T. W., Dulberg, Z., Frankland, S. M., Petrov, A. A., O'Reilly, R. C., & Cohen, J. D. (2023, September 6). *Learning Representations that Support Extrapolation*. Cornell University. Retrieved January 12, 2026, from https://doi.org/10.48550/arXiv.2007.05059

[29] Xiao, C., Zhang, P., Han, X., Xiao, G., Lin, Y., Zhang, Z., Liu, Z., & Sun, M. (2024, May 28). *InfLLM: Training-Free Long-Context Extrapolation for LLMs with an Efficient Context Memory*. Cornell University. Retrieved January 12, 2026, from https://arxiv.org/abs/2402.04617

[30] Xie, Y. (2025, May 21). *A multiscale brain emulation-based artificial intelligence framework for dynamic environments*. Scientific Reports. Retrieved January 12, 2026, from https://doi.org/10.1038/s41598-025-01431-2

[31] Zhang, J. (2024, October 1). *Artificial Intelligence vs. Human Intelligence: Which Excels Where and What Will Never Be Matched*. UTHealth Houston McWilliams School of Biomedical Informatics. Retrieved January 12, 2026, from https://sbmi.uth.edu/blog/2024/artificial-intelligence-versus-human-intelligence.htm

[32] Zhang, X., Ma, Z., Zheng, H., Li, T., Chen, K., Wang, X., Liu, C., Xu, L., Wu, X., Lin, D., & Lin, H. (2020, June 15). *The combination of brain-computer interfaces and artificial intelligence: Applications and challenges*. Medical Artificial Intelligent Research. Retrieved January 12, 2026, from https://doi.org/10.21037/atm.2019.11.109