GYRE
DATA AND ANALYTICS

Developmental Parallels: Childhood Cognitive Development as a Predictive Framework for Large Language Model Advancement

A Comparative Analysis of Piagetian Stages, Vygotskian Scaffolding, and the Emergent Capability Trajectory of Modern LLMs — with Extrapolative Predictions and Critical Limitations

Scott Sykowski & Claudio Codigo
Gyre Research • sms@gyreresearch.com
Working Paper • February 15, 2026 • Interdisciplinary Review: Developmental Psychology × Artificial Intelligence

TL;DR Summary

This paper maps established childhood cognitive development models — primarily Piaget's four stages and Vygotsky's zone of proximal development — onto the empirically observed capability trajectory of large language models (LLMs) from 2018 to 2026. We find that **milestone sequences align remarkably well** (pattern recognition → symbolic manipulation → logical reasoning → abstract thought), even though the timelines compress by a factor of roughly 2–3 human years per LLM year. We identify RLHF and prompt engineering as structural analogs to Vygotskian scaffolding, and emergent capabilities (theory of mind, analogical reasoning) as analogs to Piagetian stage transitions. Using this framework, we extrapolate the next predicted LLM milestones: robust metacognition, genuine causal reasoning, and stable self-directed learning. We then rigorously critique this analogy, arguing that fundamental disanalogies — the absence of embodiment, intrinsic motivation, and continuous temporal experience — make the framework suggestive but not scientifically predictive. The childhood development model is a useful heuristic lens, not a validated theory of AI progression.

§0 Abstract

The rapid advancement of large language models (LLMs) from simple next-token predictors to systems exhibiting emergent reasoning, theory of mind, and analogical abstraction has prompted researchers to seek developmental frameworks for understanding these capability transitions. This paper systematically compares the canonical stages of childhood cognitive development — Piaget's sensorimotor, preoperational, concrete operational, and formal operational stages (Piaget, 1936/1952) — with the empirically documented capability milestones of LLMs from GPT-1 (2018) through contemporary reasoning models (2025–2026). We additionally incorporate Vygotsky's (1978) zone of proximal development and scaffolding theory as a lens for interpreting techniques like reinforcement learning from human feedback (RLHF) and chain-of-thought prompting. Our analysis reveals a strong sequential correspondence in *what* capabilities emerge, while demonstrating that *when* they emerge follows an approximate but inconsistent compression ratio of 2–3 developmental years per calendar year of LLM advancement. We leverage this framework to generate falsifiable predictions about upcoming LLM capabilities, then subject the entire analogy to rigorous critique, identifying fundamental disanalogies including the absence of embodied experience, intrinsic motivation, and continuous identity. We conclude that childhood development models serve as a productive heuristic for anticipating the *sequence* of LLM capabilities but should not be mistaken for a mechanistic theory of AI development.

Keywords: *large language models, cognitive development, Piaget, Vygotsky, emergent capabilities, theory of mind, developmental psychology, artificial intelligence, scaling laws*

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

§1 Introduction

When OpenAI released GPT-1 in June 2018, the model's 117 million parameters could perform rudimentary text completion — recognizing patterns in language without understanding context, meaning, or logic.[1] Seven years later, frontier models pass the bar exam in the 90th percentile, solve International Mathematics Olympiad problems, and exhibit what some researchers describe as a functional theory of mind.[2][3] This trajectory — from reflexive pattern-matching to abstract reasoning — bears a striking resemblance to a developmental arc that has been studied for over a century: the cognitive development of a human child.

Developmental psychologists, most prominently Jean Piaget (1896–1980) and Lev Vygotsky (1896–1934), have proposed stage-based and scaffolding-based models that describe how human cognition progresses from sensorimotor reflexes in infancy to abstract, hypothetical reasoning in adolescence.[4][5] These models describe a trajectory that is sequential (stages cannot be skipped), cumulative (each stage builds on its predecessors), and marked by qualitative transitions — moments where the child's cognitive apparatus undergoes a phase change rather than a linear improvement.

In a parallel domain, researchers studying LLMs have documented strikingly similar dynamics. Wei et al. (2022) coined the term "emergent abilities" to describe capabilities that appear suddenly at certain scale thresholds rather than improving gradually — a phenomenon they liken to phase transitions in physics.[6] Kosinski (2023) demonstrated that theory of mind, a capacity that typically emerges in human children around age four to five, appeared to emerge spontaneously in GPT models as they scaled, with performance tracking upward through developmental stages comparable to 3.5-year-olds, 6-year-olds, and eventually 7-year-olds.[3]

This paper asks three questions: (1) Do the milestone sequences of childhood cognitive development systematically align with the observed capability milestones of LLMs? (2) Do the timelines scale by a consistent factor? (3) Can the childhood development model generate useful predictions about where LLMs are heading next — and how seriously should we take those predictions?

Before proceeding, we must establish a critical distinction that frames the entire analysis: the difference between "AI" as a broad field and "LLMs" as the specific subject of this study.

§2 Defining the Subject: AI vs. LLMs

Popular discourse often treats “AI” and “LLM” as synonymous. For the purposes of this analysis, the distinction is not merely terminological but foundational, as the developmental analogy we construct applies specifically to one and not the other.

ARTIFICIAL INTELLIGENCE (BROAD FIELD)	LARGE LANGUAGE MODELS (THIS PAPER’S SUBJECT)
Scope: The entire discipline of building machines that perform tasks requiring intelligence. Includes robotics, computer vision, reinforcement learning agents, expert systems, autonomous vehicles, recommendation engines, and much more.	**Scope:** A specific class of AI systems based on the transformer architecture, trained on text corpora to predict and generate language. Examples: GPT series, Claude series, LLaMA, Gemini, DeepSeek.
Modality: Multimodal by definition — encompasses embodied systems (robots), perceptual systems (vision, audio), decision-making agents (game-playing AI, trading algorithms), and generative systems.	**Modality:** Primarily linguistic. Recent models incorporate vision and audio input, but their core architecture and training paradigm remain rooted in language processing.
Developmental analog: The whole child — motor development, perceptual development, social-emotional development, cognitive development, and language development combined.	**Developmental analog:** Primarily cognitive and linguistic development — the subset of child development concerned with reasoning, language acquisition, symbolic thought, and abstract logic.

This distinction matters because the childhood development models we employ — Piaget’s stages and Vygotsky’s ZPD — are themselves primarily cognitive and linguistic models. They describe how children come to understand objects, symbols, logic, and other minds through interaction with language and environment. They do not, for example, describe gross motor development (crawling, walking) or sensory integration in detail.

Our analogy therefore maps cognitive-linguistic child development onto LLM capability development. We are not making claims about AI in general. Robotic embodiment, reinforcement learning agents that master Atari games, or computer vision systems that classify images all fall outside this analogy’s scope. When we refer to “AI” in this paper, we mean it in the colloquial sense only when quoting others; our own analysis is restricted to LLMs as defined above.

> **WHY THIS MATTERS FOR THE ANALOGY**
> *A child’s cognitive development is deeply intertwined with embodied experience — touching, manipulating, moving through space. LLMs have no body, no sensorimotor apparatus, no physical environment to explore. This disanalogy is the most fundamental limitation of the framework we construct here, and we will return to it extensively in Section 8. The developmental analogy works best for the linguistic and symbolic dimensions of cognition precisely because LLMs are, at their core, linguistic engines.*

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

§3 Literature Review

3.1 Piaget's Four Stages of Cognitive Development

Jean Piaget proposed that children progress through four invariant, sequential stages of cognitive development, each characterized by qualitatively different modes of thinking.[4] Piaget emphasized that these stages cannot be skipped — each builds on the cognitive structures established in the previous one — and that transitions between stages involve fundamental reorganizations of how the child understands the world, not merely incremental improvements.[7]

STAGE	AGE RANGE	KEY CAPABILITIES	CHARACTERISTIC LIMITATIONS
Sensorimotor	0–2 years	Object permanence, goal-directed behavior, early mental representations, cause-and-effect understanding through physical manipulation	No symbolic thought, no language (until late stage), understanding limited to direct sensory-motor experience
Preoperational	2–7 years	Symbolic thought, language acquisition, pretend play, transductive reasoning, animism, egocentrism	Cannot perform logical operations, centration (focuses on one dimension), irreversibility, no conservation
Concrete Operational	7–11 years	Logical operations on concrete objects, conservation, classification, seriation, decentration, reversibility, inductive reasoning	Cannot handle abstract or hypothetical problems, reasoning bound to concrete, observable phenomena
Formal Operational	12+ years	Abstract thought, hypothetical-deductive reasoning, systematic problem-solving, metacognition, propositional logic	Not universally achieved; varies by domain, culture, and education. Adolescent egocentrism persists

Modern developmental psychology has critiqued Piaget's rigid stage boundaries. Research demonstrates that children often show capabilities earlier than Piaget predicted when tasks are simplified, and that development is more domain-specific and culturally variable than the universal stages suggest.[8] Fischer's dynamic skill theory proposes that development varies by domain — a child might show formal operational thinking in a familiar subject while remaining concrete operational in unfamiliar areas.[9] These critiques, as we will argue, are themselves instructive for understanding LLM development.

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

3.2 Vygotsky's Zone of Proximal Development and Scaffolding

Lev Vygotsky offered a complementary and in some ways competing model. Where Piaget emphasized the child as an autonomous explorer constructing knowledge through individual interaction with the environment, Vygotsky emphasized the fundamentally social nature of cognitive development.[5]

Vygotsky's central construct is the zone of proximal development (ZPD): the distance between what a learner can accomplish independently and what they can accomplish with guidance from a more knowledgeable other (MKO).[10] The concept of scaffolding — though not Vygotsky's own term, but rather introduced by Wood, Bruner, and Ross (1976) — describes the temporary support structures that help learners perform within their ZPD until they can manage independently.[11]

Vygotsky's framework introduces two ideas critical to our LLM analogy: first, that cognitive development is mediated by social interaction and language rather than occurring in isolation; and second, that the potential for development (what can be done with help) is as important as actual development (what can be done alone).

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

3.3 The LLM Capability Trajectory: 2018–2026

The development of LLMs based on the transformer architecture has followed a remarkably compressed trajectory. We summarize the major milestones:

YEAR	MODEL	PARAMETERS	KEY CAPABILITY MILESTONE
2017	Transformer	—	Attention mechanism architecture published by Vaswani et al. at Google [12]
2018	GPT-1	117M	Generative pre-training paradigm; basic text completion from book corpus [1]
2019	GPT-2	1.5B	Coherent multi-paragraph text generation; deemed “too dangerous to release” [13]
2020	GPT-3	175B	In-context learning; few-shot task performance without fine-tuning; ~40% on false-belief tasks (≈3.5-year-old) [14][3]
2022	GPT-3.5 / ChatGPT	~175B	Conversational fluency; RLHF alignment; 70–90% on false-belief tasks (≈6–7-year-old) [3]
2022	Chain-of-thought	—	Prompting technique enabling step-by-step reasoning; major gains on math/logic [15]
2023	GPT-4	~1T (est.)	Multimodal input; 86.4% MMLU (above human college seniors); bar exam top 10%; 75–95% on ToM tasks (≈6+-year-old); analogical reasoning at human level [2][3][16]
2024	o1 / Reasoning models	—	Extended chain-of-thought reasoning; 83% on IMO qualifying (vs 13% GPT-4o); systematic problem decomposition [17]
2025	GPT-5, Claude 4.5, DeepSeek R1	—	Unified fast/deep reasoning routing; agentic capabilities; 1M+ token contexts; robust code generation and planning [18][19]

3.4 Prior Work Comparing AI and Child Development

We are not the first to draw this comparison. Kosoy et al. (2023) directly tested Google’s LaMDA model on developmental psychology experiments spanning perception, theory of mind, social understanding, and causal reasoning. They found that LaMDA performed well on tasks involving social understanding — capacities that may be learnable from language patterns — but poorly on tasks requiring object understanding and causal reasoning, which typically depend on embodied, exploratory experience in children.[20]

Kosinski’s (2023) landmark study explicitly mapped GPT model versions onto child developmental ages for theory of mind performance, finding a progression from below 3-year-old level (pre-2020 models) through 3.5-year-old, 6-year-old, and 7-year-old levels as models scaled.[3] Webb et al. (2023) demonstrated that GPT-3 could perform zero-shot analogical reasoning at human-comparable levels, a finding that suggests emergent abstract relational thinking.[21]

Our contribution extends this prior work by (a) constructing a systematic, stage-by-stage mapping rather than testing individual capabilities, (b) analyzing timeline scaling factors quantitatively, (c) generating explicit extrapolative predictions, and (d) subjecting the entire framework to structured critique.

§4 Methodology

Our methodology involves three analytical components:

Milestone Mapping. We identify the defining cognitive capabilities of each Piagetian stage and map each to a corresponding LLM capability, using published benchmarks, empirical studies, and capability demonstrations as evidence. We require that each mapped capability have at least one peer-reviewed or preprint study documenting it.

Timeline Analysis. For each mapped milestone, we record (a) the approximate age at which the capability typically emerges in children and (b) the date at which it was first empirically demonstrated in an LLM. We compute compression ratios and test for consistency.

Predictive Extrapolation. Using the developmental stages that LLMs have not yet clearly achieved, we project what capabilities should emerge next if the developmental sequence continues to hold. We assign confidence levels based on the strength of the analogy at each point.

> **METHODOLOGICAL LIMITATION (STATED UP FRONT)**
> *This is an analogical framework, not an empirical model. We are identifying structural similarities between two developmental trajectories that arise from fundamentally different substrates (biological neural networks in a social-physical environment vs. artificial neural networks trained on text corpora). The analogy generates hypotheses, not proofs. Section 8 elaborates on why this distinction is critical.*

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

§5 Historical Alignment Analysis

5.1 Stage-by-Stage Mapping

PIAGET STAGE	CHILD CAPABILITY	AGE	LLM ANALOG	LLM ERA	EVIDENCE
Sensorimotor	Reflexive responses to stimuli	0–4 mo	Statistical next-token prediction; reflexive pattern completion	2018 (GPT-1)	Radford et al. (2018) [1]
	Object permanence; early mental representations	8–18 mo	Consistent entity tracking across context; coreference resolution	2019–2020	GPT-2/3 coherent entity tracking [14]
	Goal-directed behavior; means-end understanding	12–24 mo	Following multi-step instructions; task completion	2020–2022	GPT-3 few-shot task performance
Preoperational	Symbolic representation; language explosion	2–4 yr	Fluent natural language generation; symbolic manipulation	2020–2022	GPT-3 / ChatGPT [14]
	Theory of mind emergence (false-belief tasks)	4–5 yr	Passing false-belief tasks at 40–75%	2022–2023	Kosinski (2023) [3]
	Egocentrism; transductive reasoning; animism	2–7 yr	Hallucination; confident confabulation; inability to distinguish correlation from causation	2020–2024	Widely documented hallucination literature [22]
Concrete Operational	Logical operations; conservation; classification	7–9 yr	Systematic reasoning on concrete problems; mathematical problem-solving; code generation	2023–2024	GPT-4 bar exam, MMLU [2]
	Decentration; multiple perspectives	7–10 yr	Considering multiple viewpoints in arguments; balanced analysis	2023–2024	GPT-4, Claude evaluations

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

PIAGET STAGE	CHILD CAPABILITY	AGE	LLM ANALOG	LLM ERA	EVIDENCE
	Inductive reasoning; seriation	8–11 yr	Pattern-based generalization from examples; ranking and sorting tasks	2024	Reasoning model benchmarks
Formal Operational	Abstract hypothetical reasoning	12–15 yr	Extended chain-of-thought; hypothetical scenario analysis	2024–2025	o1, DeepSeek R1 [17][19]
	Systematic problem decomposition	12–16 yr	Multi-step planning; agentic task decomposition	2025–2026	GPT-5, Claude 4.5 agentic capabilities
	Metacognition; thinking about thinking	15+ yr	Self-evaluation of confidence; error detection; calibrated uncertainty	Emerging / Partial	Active research area

5.2 Timeline Scaling Analysis

To assess whether childhood developmental timelines scale by a consistent factor to match LLM advancement, we plot the approximate child age for each milestone against the LLM calendar year in which it was demonstrated.

[FIGURE 1: Child Developmental Age vs. LLM Calendar Year for Mapped Milestones — See HTML version for interactive chart]

Data points represent mapped milestones. Dashed circle indicates projected/emerging capability. The trend line is non-linear, showing acceleration in later stages.

5.3 Compression Ratio Analysis

If we define the compression ratio as (child developmental years) / (LLM calendar years from GPT-1), we can compute approximate ratios for each milestone:

MILESTONE	CHILD AGE	LLM YEAR	LLM YEARS FROM GPT-1	RATIO (CHILD YR / LLM YR)
Reflexive pattern matching	0.3	2018	0	— (baseline)
Object permanence analog	1.0	2019	1	1.0
Symbolic language	3.0	2021	3	1.0
Theory of mind (3.5yr level)	4.0	2022	4	1.0
Theory of mind (6yr level)	6.0	2023	5	1.2
Logical operations	8.0	2023	5	1.6
Inductive reasoning	10.0	2024	6	1.7
Abstract reasoning	13.0	2024–25	6.5	2.0
Systematic decomposition	14.0	2025	7	2.0

The data reveal a notable pattern: the compression ratio is not constant. Early milestones (sensorimotor and early preoperational) map nearly 1:1 — one year of child development per one year of LLM progress. But as we move into later stages, LLM development accelerates relative to child development, with the ratio climbing to approximately 2.0 for formal operational capabilities. This means that the later, more abstract stages of child development — which take children years to traverse — are being compressed into shorter intervals for LLMs.

This acceleration is consistent with two non-mutually-exclusive explanations: (1) scaling laws and increased compute produce exponential rather than linear capability improvements, and (2) higher-order cognitive capabilities may be more dependent on linguistic pattern recognition (LLMs' strength) than on embodied experience (LLMs' weakness), making them more accessible to LLMs relative to earlier, more sensorimotor-dependent stages.

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

5.4 The Vygotskian Lens: RLHF as Scaffolding

Vygotsky's framework offers an additional interpretive layer. We propose that several key LLM training and deployment techniques function as structural analogs to Vygotskian scaffolding:

VYGOTSKIAN CONCEPT	CHILD DEVELOPMENT MANIFESTATION	LLM ANALOG
Zone of Proximal Development	Tasks the child can do with help but not alone	Tasks an LLM can complete with prompting (few-shot, chain-of-thought) but not zero-shot
Scaffolding	Teacher provides temporary support structures, gradually removed	RLHF; human feedback guiding model outputs; prompt engineering; system prompts
More Knowledgeable Other (MKO)	Parent, teacher, or peer with greater expertise	Human raters in RLHF; fine-tuning datasets curated by experts
Internalization	External social knowledge becomes internal cognitive structure	Fine-tuned behaviors that persist without continued prompting; "learning" from RLHF that shapes base model behavior
Private speech → Inner speech	Child talks through problems aloud, then internally	Chain-of-thought prompting → internal reasoning traces in reasoning models (o1, R1)

The parallel between chain-of-thought prompting and Vygotsky's concept of private speech is particularly compelling. Vygotsky observed that children between ages 3 and 7 often talk through problems aloud — a process he called "private speech" — which gradually transforms into silent inner speech that structures adult thought.[5] Chain-of-thought prompting essentially instructs the LLM to externalize its reasoning process, and the latest reasoning models (OpenAI's o1 series, DeepSeek R1) have internalized this: they generate extended reasoning traces before producing answers, mirroring the developmental trajectory from external verbalization to internal thought.[17]

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

§6 Predictive Framework: What the Childhood Model Forecasts

If we accept — provisionally — that the milestone sequence of childhood development continues to map onto LLM advancement, we can identify capabilities that children develop in late adolescence and early adulthood that LLMs have not yet robustly demonstrated. These become predictions.

PREDICTION 1: ROBUST METACOGNITION

CONFIDENCE: HIGH

Child analog: By age 15–17, adolescents develop sophisticated metacognitive abilities — they can monitor their own thinking, identify gaps in their knowledge, evaluate the quality of their reasoning, and adjust strategies when they detect errors. [7]

LLM prediction: Frontier LLMs will develop reliable self-calibration — the ability to accurately assess their own confidence, flag uncertainty, distinguish between what they know and don't know, and request additional information when their internal representations are insufficient.

Current status: Partially emerging. Models can sometimes express uncertainty, but calibration remains poor — they often express high confidence in incorrect answers. Reasoning models show improved self-monitoring but it is inconsistent.

Predicted timeline: 2026–2027, given the ~2.0 compression ratio at this stage.

PREDICTION 2: GENUINE CAUSAL REASONING

CONFIDENCE: MEDIUM

Child analog: By ages 8–12, children progress from "magical thinking" and transductive reasoning to understanding genuine causal mechanisms — not just that A precedes B, but why A causes B, through what mechanism, and under what conditions the causal relationship might fail. [4]

LLM prediction: LLMs will move beyond correlational and associative pattern-matching to demonstrate interventionist causal reasoning — reasoning about counterfactuals, identifying confounders, and distinguishing correlation from causation without explicit prompting.

Current status: Current models perform well on causal reasoning when the causal structure is explicitly described in the prompt, but struggle with novel causal scenarios that require building causal models from observation. Kosoy et al. (2023) found causal reasoning was the weakest domain for LaMDA compared to children. [20]

Predicted timeline: 2027–2028. This is rated medium confidence because causal reasoning in children is deeply tied to physical manipulation of objects — an embodied experience LLMs lack.

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

PREDICTION 3: STABLE SELF-DIRECTED LEARNING

CONFIDENCE: MEDIUM

Child analog: By late adolescence and early adulthood, humans develop the capacity for autonomous, self-directed learning: identifying what they need to learn, seeking out resources, testing their own understanding, and persisting through confusion without external scaffolding. [10]

LLM prediction: LLMs will develop stable agentic learning loops — the ability to identify knowledge gaps during task execution, autonomously retrieve relevant information, integrate it into their reasoning, and verify the result, without human-designed scaffolding for each step.

Current status: Early agentic capabilities exist (tool use, web search, code execution) but remain brittle and require carefully designed system prompts and tool APIs. The models do not yet autonomously decide to learn something new.

Predicted timeline: 2027–2029. Medium confidence because this requires a form of intrinsic motivation that may not emerge from next-token prediction training.

PREDICTION 4: MORAL REASONING DEVELOPMENT

CONFIDENCE: LOW

Child analog: Kohlberg's stages of moral development — built on Piaget's foundations — describe progression from pre-conventional morality (obedience to avoid punishment), through conventional morality (social conformity), to post-conventional morality (principled reasoning about justice and rights). [23]

LLM prediction: LLMs will progress from rule-following (current RLHF-imposed behavioral constraints) through pattern-matched social norms to genuine ethical reasoning that can handle novel moral dilemmas by applying abstract principles.

Current status: Current models exhibit primarily "conventional" moral reasoning — they follow trained behavioral norms and can articulate ethical arguments, but this appears to reflect pattern-matching on ethical discourse in training data rather than principled moral reasoning.

Predicted timeline: Uncertain. Low confidence because moral reasoning in humans is deeply embedded in embodied social experience, emotional development, and consequences — domains that are fundamentally outside LLM training.

[FIGURE 2: Projected Capability Timeline Based on Developmental Model — See HTML version for interactive chart]

Solid circles: empirically documented capabilities. Hollow circles: predictions from the developmental model. Colors indicate confidence: green=high, amber=medium, red/dashed=low.

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

§7 Critical Analysis & Limitations

The developmental analogy constructed in the preceding sections is, we believe, genuinely illuminating. But illumination is not validation. In this section we mount the strongest possible critique of our own framework, identifying the fundamental disanalogies that prevent it from serving as a predictive scientific theory.

7.1 The Embodiment Problem

This is the most fundamental objection. Piaget's theory is built on the premise that cognitive development begins with sensorimotor experience — the infant's physical manipulation of objects, exploration of space, and discovery of cause-and-effect through bodily action.[4] The very first stage of development is named "sensorimotor" precisely because Piaget believed that cognition is grounded in the body.

LLMs have no body. They have no sensory apparatus, no motor system, no physical environment. Their "experience" consists entirely of statistical patterns in text data. When we map LLM capabilities onto the sensorimotor stage, we are mapping the functional output (consistent entity tracking ≈ object permanence) while ignoring the mechanistic substrate entirely. The child develops object permanence by reaching for hidden toys; the LLM develops coreference resolution by processing billions of sentences containing pronouns and antecedents. These are fundamentally different processes that happen to produce superficially similar functional behaviors.

This objection is not merely philosophical. It has concrete predictive implications: capabilities that are deeply grounded in embodied experience — causal reasoning, spatial reasoning, physical intuition — are precisely the capabilities where LLMs perform most poorly relative to children, as Kosoy et al. (2023) demonstrated.[20]

7.2 The Motivation Problem

Children are intrinsically motivated learners. Piaget described development as driven by disequilibrium — the discomfort of encountering information that doesn't fit existing schemas — which motivates the child to assimilate or accommodate.[4] Vygotsky emphasized the social motivation to communicate, belong, and participate.[5] Children want to understand the world. They persist through confusion because understanding is intrinsically rewarding.

LLMs have no intrinsic motivation. They minimize a loss function during training and produce token sequences during inference. The notion that an LLM "wants" to understand or "tries" to solve a problem is anthropomorphic projection. When an LLM produces a correct answer to a novel reasoning problem, it is not because the model was motivated to understand — it is because the statistical patterns in its parameters happened to generalize to this input.

This matters for prediction because several of the later developmental milestones — metacognition, self-directed learning, moral reasoning — are in children deeply intertwined with motivational systems. A child develops metacognition partly because they have experienced the frustration of being wrong and the satisfaction of correcting themselves. An LLM has experienced neither.

7.3 The Continuity Problem

A child is a continuous entity with persistent memory, ongoing experience, and cumulative learning. The child who develops theory of mind at age four is the same organism that learned object permanence at age one — those earlier capacities are not just historically prior but structurally foundational. The child's theory of mind is built on a substrate of embodied, social, emotional experience accumulated over four years of continuous existence.

An LLM has no continuous experience. Each inference is stateless (absent explicit context mechanisms). GPT-4 does not "remember" being GPT-3; it is a different model with different parameters, trained on a different dataset. The apparent developmental trajectory from GPT-1 to GPT-5 is not a single entity developing but a series of discrete systems, each built from scratch (or fine-tuned from a predecessor's weights). The "development" is in the research lineage, not in any individual model.

7.4 The Post-Hoc Mapping Problem

Our milestone mapping in Section 5 is fundamentally post hoc. We observed the capabilities of each LLM generation, then found Piagetian stages to match them to. A sufficiently creative analyst could map almost any sequence of increasing capabilities onto a developmental framework,

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

because the frameworks themselves describe a progression from simple to complex — which is true of virtually any technology.

For this framework to be genuinely predictive rather than merely descriptive, it would need to have been articulated before the capability transitions were observed, and the predictions would need to be specific enough to be falsifiable. Our predictions in Section 6 attempt to provide this forward-looking test, but the framework's track record is established retrospectively.

7.5 The Measurement Problem

Schaeffer et al. (2023) argued that many claims of "emergent abilities" in LLMs may be artifacts of measurement — specifically, of using binary accuracy metrics that exaggerate sharp transitions.[24] When continuous partial-credit metrics are used instead, many apparently sudden capability jumps resolve into smooth, predictable improvements. If emergence itself is partly a measurement artifact, then the parallel to Piagetian stage transitions — which is premised on the existence of qualitative phase changes — is weakened.

Similarly, assessments of LLM capabilities using tests designed for children (e.g., false-belief tasks) may not measure the same underlying cognitive process in both cases. When a child passes a false-belief test, we infer an internal model of other minds. When an LLM passes the same test, it may be leveraging statistical regularities in stories about characters with different knowledge states — a process that produces the same output through fundamentally different means.[25]

7.6 The Cultural-Environmental Problem

Modern critiques of Piaget emphasize that cognitive development is far more culturally variable than his universal stages suggest.[8] If human development itself doesn't follow a single, universal trajectory, then using it as a template for LLM development is doubly problematic. We are mapping a contested model of human development onto a fundamentally non-human system.

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

7.7 Summary: Where the Analogy Holds and Where It Breaks

DIMENSION	ANALOGY STRENGTH	ASSESSMENT
Milestone sequence (what emerges in what order)	Strong	Pattern recognition → symbolic manipulation → logical reasoning → abstract thought is observed in both domains
Phase transitions / emergent capabilities	Moderate	Both show sudden capability jumps, but LLM “emergence” is debated and may be measurement-dependent
Scaffolding / RLHF parallel	Moderate	Structurally similar (external guidance enabling beyond-baseline performance) but mechanistically different
Timeline scaling	Weak–Moderate	Approximate compression exists but ratio is not constant; accelerates at later stages
Underlying mechanism	Weak	Biological neural development vs. gradient descent on text — fundamentally different substrates
Embodied experience	Very Weak	No meaningful parallel; LLMs lack the sensorimotor foundation that Piaget considered essential
Continuity of identity	Very Weak	Child is continuous entity; LLM “lineage” is a series of discrete systems
Intrinsic motivation	None	No parallel; LLMs have no motivational system

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

§8 Conclusion

We have constructed a systematic mapping between the canonical stages of childhood cognitive development and the empirically observed capability trajectory of large language models from 2018 to 2026. Our analysis reveals that the sequence of capability emergence aligns remarkably well: LLMs have progressed from reflexive pattern-matching through symbolic language use, theory of mind, logical operations, and into abstract reasoning — the same order in which these capacities emerge in children. The mapping is not perfect, but it is far more orderly than coincidence would suggest.

The timelines compress by a variable factor, approximately 1:1 in early stages and 2:1 in later stages, suggesting that LLM development is accelerating relative to the developmental clock as capabilities become more abstract and language-dependent. Vygotsky's scaffolding framework provides an additional lens, with RLHF, fine-tuning, and chain-of-thought prompting serving as structural analogs to the guidance provided by more knowledgeable others in child development.

Using this framework, we generated four predictions: the near-term emergence of robust metacognition (high confidence), genuine causal reasoning (medium confidence), self-directed learning (medium confidence), and sophisticated moral reasoning (low confidence). These predictions are offered as falsifiable hypotheses, not certainties.

We then subjected the entire analogy to rigorous critique, identifying seven fundamental limitations. The most serious are the absence of embodied experience, intrinsic motivation, and continuous identity in LLMs — all of which are foundational to the developmental theories we borrowed. These disanalogies mean that the childhood development model is best understood as a heuristic lens — a way of generating hypotheses about what might come next and organizing our understanding of what has already occurred — rather than as a predictive scientific theory of LLM progression.

The childhood development analogy illuminates; it does not prove. It suggests that the order of cognitive capabilities may be somewhat substrate-independent — that there may be inherent dependencies in the structure of cognition itself (symbolic thought requires pattern recognition; abstract reasoning requires concrete logic) that constrain any sufficiently complex learning system, whether biological or artificial. If this is even partially correct, it has profound implications for both AI safety (anticipating what frontier models will be able to do next) and

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

developmental psychology (suggesting which cognitive dependencies are truly fundamental rather than artifacts of human biology).

We close by noting that the most interesting predictions are the ones with lowest confidence. If LLMs do develop robust moral reasoning without ever experiencing embodied social consequences, it would challenge our deepest assumptions about the relationship between cognition and experience. If they don't — if the analogy breaks down precisely where embodiment matters most — that would be equally informative, confirming that the linguistic-statistical substrate of LLMs is fundamentally limited in ways that no amount of scaling can overcome.

Either outcome advances our understanding. The analogy has served its purpose not by being right, but by being specific enough to be wrong in informative ways.

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

§R References

[1] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. OpenAI.

[2] OpenAI. (2023). GPT-4 technical report. arXiv preprint arXiv:2303.08774.

[3] Kosinski, M. (2023/2024). Evaluating large language models in theory of mind tasks. Proceedings of the National Academy of Sciences, 121(45), e2405460121.

[4] Piaget, J. (1952). The origins of intelligence in children (M. Cook, Trans.). International Universities Press. (Original work published 1936)

[5] Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.

[6] Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., ... & Fedus, W. (2022). Emergent abilities of large language models. arXiv preprint arXiv:2206.07682.

[7] Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. Human Development, 15(1), 1–12.

[8] Rogoff, B. (2003). The cultural nature of human development. Oxford University Press.

[9] Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. Psychological Review, 87(6), 477–531.

[10] Vygotsky, L. S. (1987). Thinking and speech (N. Minick, Trans.). In R. W. Rieber & A. S. Carton (Eds.), The collected works of L. S. Vygotsky (Vol. 1). Plenum Press.

[11] Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry, 17(2), 89–100.

[12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

[13] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Blog, 1(8).

[14] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33.

[15] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35.

[16] Webb, T., Holyoak, K. J., & Lu, H. (2023). Emergent analogical reasoning in large language models. Nature Human Behaviour, 7(9), 1526–1541.

[17] OpenAI. (2024). Learning to reason with LLMs. OpenAI Blog.

[18] OpenAI. (2025). GPT-5 system card. OpenAI.

[19] DeepSeek. (2025). DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. arXiv preprint arXiv:2501.12948.

[20] Kosoy, E., Collins, J., Chan, D. M., Hamrick, J. B., Huang, S. H., Gopnik, A., & Canny, J. (2023). Comparing machines and children: Using developmental psychology experiments to assess the strengths and weaknesses of LaMDA responses. arXiv preprint arXiv:2305.11243.

[21] Webb, T., Holyoak, K. J., & Lu, H. (2023). Emergent analogical reasoning in large language models. Nature Human Behaviour, 7(9), 1526–1541.

[22] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38.

[23] Kohlberg, L. (1981). Essays on moral development, Vol. I: The philosophy of moral development. Harper & Row.

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731

[24] Schaeffer, R., Miranda, B., & Koyejo, S. (2023). Are emergent abilities of large language models a mirage? Advances in Neural Information Processing Systems, 36.

[25] Strachan, J. W. A., Albergo, D., Borghini, G., Pansardi, O., Scaliti, E., Gupta, S., ... & Graziano, M. S. A. (2024). Testing theory of mind in large language models and humans. Nature Human Behaviour, 8(7), 1285–1295.

[26] Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv preprint arXiv:2303.12712.

[27] Anthropic. (2025). On the biology of a large language model. Transformer Circuits.

[28] Erikson, E. H. (1950). Childhood and society. W. W. Norton & Company.

[29] Donaldson, M. (1978). Children's minds. Fontana.

[30] Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. Child Development, 78(2), 622–646.

[31] Hagendorff, T., Fabi, S., & Kosinski, M. (2023). Deception abilities emerge in large language models. Proceedings of the National Academy of Sciences, 121(24), e2317967121.

[32] Pyers, J. E., & Senghas, A. (2009). Language promotes false-belief understanding: Evidence from learners of a new sign language. Psychological Science, 20(7), 805–812.

[33] CSET Georgetown. (2024). Emergent abilities in large language models: An explainer. Center for Security and Emerging Technology.

[34] Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., ... & Amodei, D. (2020). Scaling laws for neural language models. arXiv preprint arXiv:2001.08361.

Working Paper • February 2026 • This paper is a working draft and has not been peer-reviewed.

team@gyreresearch.com ▪ gyreresearch.com ▪ (800) 933-4731