Critical Review of Social Sciences and Humanities
Volume 5, Issue 2, 2025

Artificial Intelligence and Human Language Cognition: An Integrative Comparative Framework

Afshan Ishfaq[1]
Nida Sultan[2]

ABSTRACT

This article presents an integrative conceptual review comparing contemporary artificial intelligence language systems with human language cognition. A targeted, non-systematic search of foundational and recent literature published through July 2026 covered linguistics, psycholinguistics, cognitive science, philosophy of mind, natural language processing, multimodal learning, interpretability, and embodied robotics. Sources were selected based on their direct relevance to five recurring comparative dimensions: representation, acquisition and adaptation, processing, grounding and interaction, and intentionality, consciousness, and responsibility. The synthesis distinguishes functional linguistic competence from representational grounding, phenomenal consciousness, and moral responsibility, thereby avoiding the assumption that fluent performance either proves or refutes human-like understanding. The resulting framework treats AI-human differences as continua rather than fixed binaries and differentiates text-only language models from multimodal, tool-using, interactive, and embodied systems. Current evidence shows substantial convergence in formal language performance, prediction, abstraction, and some internally represented, world-relevant structures. However, it also reveals persistent differences in developmental history, sensorimotor coupling, online social learning, reliability profiles, and the evidence available for subjective experience or autonomous responsibility. These differences support a cautious model of cognitive complementarity: task allocation should depend on capability, verifiability, stakes, contextual sensitivity, and accountability rather than on broad claims that either humans or AI are categorically superior. The framework is interpretive and revisable. It would require revision if future systems demonstrate robust causal grounding, durable autonomous goals, cross-context learning from embodied interaction, or empirically credible indicators of consciousness.

Key Words: artificial intelligence; human language cognition; large language models; grounding; multimodality; intentionality; conceptual review

[1] Head Academics, Institute of Law, Lahore
[2] Lecturer in English, Namal University, Mianwali

Critical Review of Social Sciences and Humanities Volume 5, Issue 2, 2025

1. INTRODUCTION

Large language models (LLMs) have transformed the practical and theoretical landscape of language research. Transformer-based systems can generate extended discourse, translate, summarize, answer questions, follow instructions, and interact through language across many domains (Vaswani et al., 2017; Brown et al., 2020; OpenAI, 2023). Their performance has renewed a long-standing question: what does successful linguistic behavior imply about the cognition that produces it? The question cannot be answered by fluency alone. Similar outputs can arise from different learning histories, internal representations, processing constraints, forms of environmental coupling, and degrees of agency.

The debate is often distorted by treating understanding as a single, all-or-nothing property. Some critiques infer that statistical training objectives make meaningful understanding impossible (Bender & Koller, 2020), while capability-centered accounts emphasize generalization, in-context learning, internal structure, and increasingly multimodal or action-oriented behavior (Driess et al., 2023; Feng & Steinhardt, 2024; Gurnee & Tegmark, 2024). Neither position justifies a simple inference from mechanism to ontology. Statistical learning is also central to human language processing, and human cognition is predictive, fallible, and constrained. Conversely, success on benchmarks does not, by itself, establish lived experience, autonomous intention, or moral responsibility.

This article therefore compares systems along explicit dimensions rather than asking whether AI is intelligent in the abstract. It is an integrative conceptual review rather than an empirical test of consciousness or a systematic review. Its purpose is to organize evidence from cognitive linguistics, psycholinguistics, computational linguistics, cognitive science, and recent AI research within a framework that separates what can be behaviorally tested from what remains representational, philosophical, or normative.

1.1 Conceptual distinctions

Five conceptual distinctions structure the analysis. First, functional linguistic competence refers to observable success in producing or interpreting language for a specified task. Second, representational understanding concerns whether internal states systematically encode entities, relations, situations, and counterfactual structure in ways that support transfer rather than surface imitation. Third, grounding concerns the degree to which linguistic representations are causally connected to perception, action, social interaction, and consequences in an environment. Fourth, intentional agency concerns persistent goal-directed organization: whether a system maintains, revises, and acts on goals across time, rather than merely exhibiting goal-like behavior within a prompted episode. Fifth, phenomenal consciousness concerns subjective experience, whereas moral status and responsibility concern who can be held accountable and whose welfare has ethical weight. These properties may correlate, but they are not interchangeable.

The term intelligence is used here in a limited comparative sense: the capacity to learn, represent, adapt, and perform across language-mediated tasks. Similarly, empathy is divided into functional empathic performance, such as detecting emotion and generating a supportive response, and felt or phenomenal empathy, which would entail an experienced concern for another person. This distinction is necessary because people may experience AI-generated language as supportive even though the system's subjective status remains unsettled (Yin et al., 2024).

Critical Review of Social Sciences and Humanities Volume 5, Issue 2, 2025

1.2 Aim and research questions

The study aims to construct a transparent, non-exhaustive framework for comparing current AI language systems with human language cognition while fairly representing competing theoretical positions. It addresses three questions:

1. Across which dimensions do AI language systems and human language cognition converge, partially overlap, or diverge?
2. What kinds of evidence justify claims about functional competence, representation, grounding, agency, or consciousness, and what remains unresolved?
3. How can these distinctions guide task-level decisions about AI-led, human-led, and combined linguistic work without confusing capability with reliability or responsibility?

2. METHOD OF CONCEPTUAL SYNTHESIS

2.1 Search scope and source selection

A targeted integrative literature search was conducted for this revised manuscript and covered publications available through July 2026. Sources were identified through ACL Anthology, arXiv, OpenReview, major conference proceedings, PubMed, the ACM Digital Library, and publisher platforms for journals in cognitive science, psychology, linguistics, neuroscience, and artificial intelligence. Backward and forward citation tracing was used to connect foundational theories with recent technical and empirical work. Search combinations paired large language model with understanding, meaning, grounding, world model, interpretability, agency, consciousness, multimodal, embodied, social interaction, empathy, and human cognition; and paired human language cognition with embodiment, prediction, working memory, social acquisition, bias, memory distortion, and good-enough processing.

The temporal scope extended from foundational work published between 1950 and 2016 to more recent work published from 2017 through July 2026, with particular emphasis on post-2022 research because model classes have changed rapidly. Sources were included when they met at least one of four criteria: they offered a foundational account of meaning or cognition; reported primary empirical or technical evidence directly relevant to one of the framework dimensions; provided a major philosophical argument that remains influential in the debate; or presented a contemporary system class that weakens a simple text-only AI-human contrast. Tangential application studies and unsupported commentary were excluded. Author-related publications cited in the original manuscript were not used to support broad claims when stronger, more direct evidence was available.

2.2 Analytical procedure and evidence appraisal

Each selected source was assessed according to four elements: the construct being defined, the type of evidence offered, the class of system or human population examined, and the boundary conditions of the claim. Evidence was then grouped into recurring contrasts and areas of overlap. Five dimensions were retained because they appeared across all four contributing disciplines and could be linked to observable or potentially testable indicators: representation; acquisition and adaptation; processing; grounding and interaction; and intentionality, consciousness, and responsibility.

The synthesis does not assign a single numerical quality score across heterogeneous disciplines. Instead, it distinguishes four evidential categories. Engineering evidence shows what a system can do under specified conditions. Behavioral evidence compares outputs or performance. Interpretability evidence concerns internal computational structure and causal mechanisms. Philosophical analysis clarifies conceptual entailments but does not by itself settle

Critical Review of Social Sciences and Humanities Volume 5, Issue 2, 2025

empirical questions. Normative conclusions about responsibility or acceptable use require additional value judgments and domain evidence. Disagreements were retained when the literature supported competing interpretations, as in the debate over emergent abilities and the interpretation of internal world-relevant representations.

2.3 Scope and non-exhaustive character

This is a purposive conceptual synthesis rather than a systematic review. It neither claims exhaustive database coverage nor estimates publication bias or provides a meta-analysis. The framework is designed for contemporary text-only, multimodal, tool-using, interactive, and embodied language systems; it should not be applied to all forms of artificial intelligence as if they constituted a single homogeneous class. Its conclusions are conditional on the systems and evidence reviewed.

Table 1.

Source-to-dimension map for the conceptual synthesis

Dimension	Representative source clusters	Primary evidential question
Representation	Distributional semantics and transformers (Harris, 1954; Mikolov et al., 2013; Vaswani et al., 2017); contextual binding and spatial-temporal representations (Feng & Steinhardt, 2024; Gurnee & Tegmark, 2024); conceptual metaphor and perceptual symbols (Lakoff & Johnson, 1980; Barsalou, 1999).	Do internal states encode relations and support transfer, and how strongly are those representations connected to non-linguistic structure?
Acquisition and adaptation	Human usage-based and joint-attention accounts (Tomasello, 2003; Tomasello et al., 2005); pretraining and instruction tuning (Brown et al., 2020; Ouyang et al., 2022); agent memory and verbal feedback (Park et al., 2023; Shinn et al., 2023).	How is competence acquired, updated, and generalized across data, social interaction, and feedback?
Processing	Human incremental processing and working memory (Just & Carpenter, 1992; Tanenhaus et al., 1995); autoregressive transformers (Vaswani et al., 2017); language-thought dissociation (Mahowald et al., 2024; Fedorenko et al., 2024).	What operations and constraints produce language behavior in real time, and which similarities are functional rather than mechanistic?
Grounding and interaction	Symbol grounding and embodied cognition (Harnad, 1990; Glenberg & Kaschak, 2002); multimodal and robotic systems (Driess et al., 2023; Brohan et al., 2023; Mon-Williams et al., 2025).	How directly are symbols connected to perception, action, consequences, and reciprocal social environments?
Intentionality, consciousness, and responsibility	Communicative intention (Grice, 1957); Chinese Room and ELIZA-effect critiques (Searle, 1980; Weizenbaum, 1976); indicator-based AI-consciousness analysis (Butlin et al., 2023); supportive-response evidence (Yin et al., 2024).	What evidence supports goal-directed function, subjective experience, or accountability, and where must these claims remain separate?

3. THEORETICAL AND EMPIRICAL FOUNDATIONS

3.1 Human language cognition: embodied, predictive, social, and fallible

Cognitive linguistics locates meaning in patterns of categorization, construal, image schemas, and conceptual metaphor rather than in syntax alone (Lakoff & Johnson, 1980; Lakoff, 1987; Langacker, 1987). Embodied accounts propose that conceptual knowledge is partly grounded in perceptual and motor systems (Barsalou, 1999), and behavioral work demonstrates interactions between language comprehension and action-compatible responses (Glenberg & Kaschak, 2002).

These traditions establish the relevance of bodily and environmental history to meaning, but they do not imply that every human judgment is accurate or that embodiment is a sufficient condition for understanding.

Psycholinguistic processing is incremental and predictive. Listeners combine lexical, syntactic, visual, and pragmatic information as an utterance unfolds (Tanenhaus et al., 1995), while working-memory limitations constrain comprehension and production (Just & Carpenter, 1992; Levelt, 1989). Human comprehension may also be shallow or merely "good enough" rather than fully specified (Ferreira et al., 2002). Memory is reconstructive, introspective explanations can be unreliable, and judgment is vulnerable to heuristics and bias (Nisbett & Wilson, 1977; Tversky & Kahneman, 1974; Loftus, 2005). A symmetrical comparison must therefore contrast AI failure modes with actual human limitations rather than with an idealized human agent.

Language acquisition is socially and developmentally embedded. Children learn through joint attention, contingent feedback, gesture, shared activity, and attempts to infer communicative intention (Tomasello, 2003; Tomasello et al., 2005). Humans, however, also learn distributional regularities. The relevant contrast is not statistical AI versus non-statistical humans; it is a difference in the sources, scale, embodiment, developmental constraints, and social organization of learning.

Recent cognitive and neuroscientific research further cautions against equating language with thought. Mahowald et al. (2024) distinguish formal linguistic competence from functional competence, while Fedorenko et al. (2024) argue that human language is primarily a communication system and is partly dissociable from non-linguistic reasoning. This distinction matters because an LLM may model linguistic structure exceptionally well without instantiating the full range of human cognition. At the same time, language models can become useful models of aspects of human behavior: Binz et al. (2025), for example, showed that a language-model-based foundation model could predict behavior across many psychological tasks. Such predictive similarity, however, does not establish mechanistic identity.

3.2 Contemporary AI systems are not a single class

The conventional contrast between text-only LLMs and embodied humans is now incomplete. Text-only LLMs learn from token sequences through prediction objectives and post-training. Multimodal foundation models incorporate images, audio, or video; tool-using agents retrieve information and execute actions; memory-augmented systems retain records across episodes; and embodied systems connect language models to robotic perception and control. GPT-4 was reported as multimodal at the input level (OpenAI, 2023). PaLM-E interleaves textual, visual, and robot-state inputs (Driess et al., 2023), while RT-2 maps vision and language to robot actions represented as tokens (Brohan et al., 2023). More recent embodied systems combine LLM planning with retrieval, force, and visual feedback in uncertain environments (Mon-Williams et al., 2025).

Interactive and agentic architectures also challenge the claim that AI learning is necessarily static or exclusively offline. ReAct couples language-model reasoning with actions that gather information from an environment (Yao et al., 2023). Reflexion stores verbal feedback to improve later attempts without changing model weights (Shinn et al., 2023). Generative-agent architectures combine memory, reflection, and planning across simulated time (Park et al., 2023). Although these systems do not reproduce human development, they occupy intermediate positions between a fixed next-token predictor and an autonomous embodied learner.

Critical Review of Social Sciences and Humanities Volume 5, Issue 2, 2025

3.3 Competing interpretations of representation and emergence

The grounding critique remains consequential. Bender and Koller (2020) argue that form alone does not determine communicative meaning, and Harnad's (1990) symbol-grounding problem asks how symbols acquire non-symbolic content. Searle's (1980) Chinese Room similarly challenges the inference from rule-governed symbol manipulation to understanding. These arguments impose conceptual constraints on interpretation, but they do not directly measure every internal representation in contemporary models.

Empirical interpretability research has identified structures that cannot be adequately described as random surface recombination. Language models can encode context-specific entity-binding information (Feng & Steinhardt, 2024), and representations of space and time can emerge in ways that support linear decoding and intervention (Gurnee & Tegmark, 2024). These findings weaken the strongest version of the claim that internal states contain only local word associations. They do not establish that the encoded structure is human-like, causally grounded in lived experience, or accompanied by awareness.

Claims about abrupt emergent abilities also require qualification. Wei et al. (2022) reported abilities that appeared at scale, but Schaeffer et al. (2023) showed that discontinuous metrics can make smooth improvements appear abrupt. Du et al. (2024) later argued that threshold-like behavior can still be observed when performance is analyzed against pretraining loss. The appropriate conclusion is neither that emergence is universally real nor that it is entirely illusory, but that claims depend on task design, measurement, model family, and the explanatory variable used.

3.4 Functional competence does not settle consciousness or responsibility

Behavioral success may justify limited functional attributions. It may be useful to say that a system understands an instruction when it reliably acts in accordance with it across controlled variations. This operational usage should not be conflated with a claim of phenomenal consciousness. Butlin et al. (2023) propose evaluating AI consciousness through computational indicator properties derived from scientific theories. Their analysis did not identify any current system as conscious, while also rejecting the claim that consciousness is technically impossible in artificial systems. The evidential position is therefore one of non-establishment rather than proof of absence.

Responsibility constitutes a separate issue. A system may produce highly capable outputs without being a suitable bearer of legal or moral accountability. Responsibility depends on governance, control, foreseeability, institutional roles, and the capacity to answer for consequences. These factors are not direct functions of benchmark performance or linguistic fluency.

4. FIVE-DIMENSIONAL COMPARATIVE FRAMEWORK

The framework treats each dimension as a continuum. A system can move along one dimension without moving equally along the others. A multimodal model may gain perceptual links while lacking persistent goals; an agent may display functional planning without credible evidence of phenomenal consciousness; a human may possess lived experience while making an unreliable factual judgment. The dimensions must therefore remain analytically distinct.

4.1 Representation

At one end of the continuum are representations inferred primarily from textual co-occurrence. At intermediate points are internal states that encode entities, relations, space, time, and task structure and that can be causally probed. A system moves further toward grounding when

Critical Review of Social Sciences and Humanities Volume 5, Issue 2, 2025

those representations are systematically coupled to perception and action. Human conceptual representations are shaped by bodily, social, and cultural experience, but they are also abstract, symbolic, and distribution-sensitive. The central question is not whether a vector is meaningful by definition, but rather which regularities it tracks, how causally important those regularities are, and whether they support robust transfer beyond the training distribution.

4.2 Acquisition and adaptation

Text-only pretraining uses massive, largely retrospective datasets and gradient-based optimization. Instruction tuning and human feedback add socially produced preferences but remain components of model training. Memory systems, tool use, and verbal feedback enable limited within-episode or cross-episode adaptation. Human learning uses far less linguistic data but occurs within continuous perception, action, social correction, maturation, and self-generated goals. Empirical comparisons should therefore examine sample efficiency, online updating, transfer after interaction, resistance to catastrophic interference, and whether learning changes future behavior without external retraining.

4.3 Processing

Both humans and LLMs use context and prediction, but their architectures and constraints differ. Human comprehension is incremental, resource-limited, and integrated with perception, emotion, memory, and action. Transformer inference uses learned parameters, attention operations, a context window, and autoregressive output generation. Tool calls and external memory can extend this process. Comparisons should therefore focus on error signatures, sensitivity to context, latency, resource constraints, self-monitoring, and generalization rather than assuming that similar answers imply similar mechanisms.

4.4 Grounding and interaction

Grounding is graded. Text-only systems inherit indirect traces of human experience through language. Multimodal models add cross-modal statistical links. Interactive agents obtain information through actions, and embodied robots receive sensorimotor feedback with real consequences for task success. Human grounding remains broader because it develops through a lifetime of self-maintaining biological activity, social dependence, affect, and action. Nevertheless, the existence of multimodal and embodied systems means that no defensible framework can classify all AI as wholly ungrounded. The empirical question concerns the breadth, causal efficacy, persistence, and self-correcting capacity of this coupling.

4.5 Intentionality, consciousness, and responsibility

Functional intention can be operationalized as goal-directed behavior: maintaining a task objective, selecting actions, and revising a plan. Contemporary agents can implement such functions to varying degrees. Stronger intentionality would require durable, internally maintained goals that organize behavior across contexts and persist independently of a transient prompt. Phenomenal consciousness is a distinct claim for which no accepted decisive test currently exists in AI. Moral responsibility is distinct again and generally remains with people and institutions that design, deploy, authorize, or rely on systems. Evidence in one category should not be used as a shortcut to another.

Critical Review of Social Sciences and Humanities Volume 5, Issue 2, 2025

Table 2.
Operational comparative framework: continua, indicators, and revision conditions

Dimension	Current AI spectrum	Human cognition	Indicators and conditions that would strengthen or revise the distinction
Representation	From text-derived distributional states to causally identifiable entity, spatial, temporal, visual, and action-related representations.	Concepts shaped by linguistic, perceptual, motor, affective, social, and cultural history; substantial individual and cross-linguistic variation.	Causal intervention on internal states; transfer to novel compositions; reliable counterfactual prediction; cross-modal consistency. The distinction weakens if models develop broad, causally effective world representations that generalize beyond familiar distributions.
Acquisition and adaptation	Offline pretraining and post-training; in-context adaptation; external memory; verbal feedback; limited online learning in agents and robots.	Developmental, socially scaffolded, embodied learning with maturation, self-generated exploration, and continuous feedback.	Sample efficiency; durable learning after interaction; autonomous exploration; generalization across environments; retention without retraining. The distinction weakens with robust lifelong learning that integrates social and sensorimotor consequences.
Processing	Transformer computation, finite context, autoregressive generation, optional tools and memory; errors include hallucination, prompt sensitivity, and brittle generalization.	Incremental, working-memory-limited, predictive processing; errors include bias, shallow parsing, confabulation, fatigue, and memory distortion.	Comparative error profiles, process tracing, causal interpretability, stability across paraphrase, metacognitive calibration, and repair after feedback. Similarity is strengthened by shared causal signatures, not output matching alone.
Grounding and interaction	Indirect linguistic traces in text-only models; cross-modal association in multimodal models; active information gathering in agents; physical feedback in embodied systems.	Continuous bodily regulation, perception-action loops, affect, social reciprocity, and consequences across a lifespan.	Action-dependent concept learning; transfer across bodies and settings; correction through physical consequences; reciprocal social coordination. The distinction weakens as coupling becomes broader, persistent, and self-correcting.
Intentionality, consciousness, and responsibility	Prompt-conditioned goals and functional planning can be engineered; durable autonomous goals vary by architecture. No broadly accepted empirical evidence currently	Persistent needs, goals, subjective experience, autobiographical continuity, and participation in social and legal responsibility,	Goal persistence without prompts; self-model continuity; theory-derived consciousness indicators; convergent independent measurements; capacity to explain and answer for actions. The framework must

Critical Review of Social Sciences and Humanities Volume 5, Issue 2, 2025

Dimension	Current AI spectrum	Human cognition	Indicators and conditions that would strengthen or revise the distinction
	establishes phenomenal consciousness. Accountability is assigned through human institutions.	though agency can be impaired or distributed.	change if credible evidence supports artificial subjectivity or new responsibility-bearing institutions.

5. DISCUSSION

5.1 RQ1: Convergence, overlap, and divergence

The strongest convergence occurs in functional language behavior. Contemporary systems can model formal linguistic patterns, maintain discourse context, transform texts, and solve many language-mediated tasks. Interpretability studies also show that at least some internal states track structured information. Multimodal and embodied systems create additional overlap by connecting language with perception and action.

The strongest divergence lies not in the use of statistics itself, but in developmental organization and system-environment coupling. Human language emerges within biological regulation, social dependence, shared attention, emotion, and lifelong action. AI systems vary widely, but most current systems are assembled from pretraining, post-training, prompting, tools, memory, and external controllers. These components can implement functions analogous to learning, planning, or self-correction without reproducing the human developmental pathway. The framework therefore supports differentiated comparison rather than either full equivalence or absolute discontinuity.

5.2 RQ2: What can be inferred from fluent performance?

Fluent, contextually appropriate output supports attributing functional competence when performance is robust, appropriately tested, and reliable under variation. It may also provide evidence of structured internal representation when paired with causal interpretability. It does not independently establish phenomenal consciousness, felt emotion, or moral status. Conversely, an absence of human-like embodiment does not prove that every possible artificial system is incapable of representation or consciousness. The evidential burden should match the claim.

Current evidence concerning comprehension remains mixed. Dentella et al. (2024) found that tested LLMs could be insensitive to underlying meaning on particular tasks, while Hu et al. (2024) reported close alignment with human grammatical judgments when probability-based minimal-pair methods were used. These contrasting findings demonstrate why system version, prompt design, task validity, and comparison baseline must be specified. A single benchmark cannot settle a general question about understanding.

5.3 A symmetrical comparison of reliability

Human judgment should not be assumed to be uniformly grounded, empathic, or accurate. People misremember, rationalize, rely on heuristics, misread syntax, and vary with fatigue, expertise, ideology, and social pressure. AI systems may outperform individual humans on bounded tasks involving speed, consistency, retrieval, or structured transformation. The relevant

Critical Review of Social Sciences and Humanities Volume 5, Issue 2, 2025

basis of comparison is therefore task-specific reliability under documented conditions, not the moral elevation of human cognition or the anthropomorphic elevation of machine output.

AI failure modes also differ from human failure modes. Models can fabricate plausible details, inherit corpus biases, change answers under superficial prompt variation, and present weakly grounded claims with high verbal confidence. Human oversight is effective only when the reviewer has the relevant expertise, sufficient time, access to evidence, and a defined responsibility to verify. Simply placing a person in the loop does not guarantee safety.

5.4 RQ3: Task-level cognitive complementarity

Cognitive complementarity is best understood as a task-routing heuristic rather than a fixed division between machine and human faculties. Five task properties are especially relevant: boundedness, verifiability, reversibility, contextual or relational sensitivity, and accountability. AI-led work is most defensible when its outputs can be checked against explicit criteria and errors are reversible. Combined work is appropriate when generative breadth is useful but evidence, interpretation, or consequences require human review. Human-led work remains necessary when legitimacy, duty of care, contested values, or accountable judgment are central, even if AI assists with preparation.

Table 3

Task-routing heuristic for AI, human, and combined linguistic work

Task property	AI-led or AI-first	Human-AI combined	Human-led with optional AI support
Boundedness	Clearly specified transformation, classification, extraction, or drafting with stable criteria.	Open-ended generation within a defined objective and review process.	Ill-defined problems whose goals or values must be negotiated.
Verifiability	Output can be automatically or quickly checked against authoritative data, rules, or tests.	Some claims are checkable, but interpretation or synthesis requires expertise.	No reliable external check exists and judgment depends on context, values, or lived consequences.
Reversibility and stakes	Errors are low-stakes and easily corrected.	Moderate stakes; staged approval and audit trails are available.	High-stakes or irreversible consequences, especially where rights, welfare, or legal duties are involved.
Relational sensitivity	Minimal need for durable interpersonal trust or shared history.	AI supports communication while a responsible human maintains the relationship.	The act of accountable human presence is itself part of the service or decision.
Accountability	A person or institution has already defined criteria and remains answerable for deployment.	Roles for generation, verification, approval, and escalation are explicitly assigned.	A human professional or institution must exercise and document final judgment.

Critical Review of Social Sciences and Humanities Volume 5, Issue 2, 2025

5.5 Implications for education and emotionally sensitive communication

In education, AI can generate examples, explanations, practice items, formative comments, and alternative phrasings. These are functional advantages, not evidence that such systems can replace learner cognition or teacher responsibility. The central design question is whether AI use preserves opportunities for learners to set goals, retrieve knowledge, formulate ideas, monitor errors, and justify revisions. Empirical studies should compare retention and transfer under different patterns of assistance rather than assuming either inevitable enhancement or inevitable cognitive decline.

In emotionally sensitive communication, AI can sometimes detect emotional cues and generate responses that recipients perceive as supportive. Yin et al. (2024) found that AI-generated responses could make participants feel heard, although the effect weakened when responses were labeled as AI-generated. This finding supports claims about functional empathic performance and user perception; it does not establish felt empathy, therapeutic competence, or independent responsibility. Its use in high-stakes health, safeguarding, or crisis contexts therefore requires domain-specific evidence, clear escalation paths, and accountable human governance. The present conceptual framework does not justify a clinical protocol or a general triage rule.

5.6 Capability, reliability, consciousness, and moral status

Four questions must be considered separately. Can the system perform the task? How reliably does it perform across relevant conditions? Who is responsible for its use and consequences? Does the system possess subjective experience or moral status? A positive answer to the first does not determine the other three. This separation reduces both over-attribution and under-attribution because it permits serious recognition of machine capability without treating competence as proof of consciousness, and it permits philosophical openness without assigning unsupported moral or professional authority.

6. LIMITATIONS AND FUTURE RESEARCH

This framework has several limitations. First, the review is targeted and integrative rather than systematic; source selection may reflect disciplinary biases and unequal access to literature across languages. Second, the evidence base is heterogeneous. Engineering demonstrations, behavioral benchmarks, neural or mechanistic interpretability, philosophical arguments, and normative analysis answer different questions and cannot be reduced to a single evidential score. Third, AI systems change rapidly, and claims about text-only models should not be generalized to multimodal, agentic, or embodied architectures. Fourth, the human side of the comparison is also heterogeneous across languages, cultures, ages, abilities, and social conditions. Fifth, the framework provides no decisive test of consciousness, nor does it claim that current absence of accepted evidence establishes permanent impossibility.

Future research should compare system classes using preregistered tasks designed to discriminate among surface performance, structured representation, and causal grounding. Interpretability studies should test whether identified representations control behavior under intervention and generalize across tasks. Embodied-learning studies should examine whether concepts acquired through action transfer to new settings and remain stable over time. Human-AI studies should compare error distributions, calibration, repair, and accountability, rather than

focusing on accuracy alone. Longitudinal educational work should measure retained competence and transfer after AI-assisted learning. Finally, consciousness research should refine theory-derived indicators and specify what convergent evidence would be sufficient to update judgments about artificial subjectivity.

7. CONCLUSION

This integrative review develops a five-dimensional framework for comparing contemporary AI language systems with human language cognition. The reviewed evidence supports substantial functional convergence and growing intermediate cases: models can encode structured information, adapt through context and external memory, use tools, process multimodal input, and participate in embodied control. These developments render a simple binary between statistical machines and grounded humans increasingly inadequate.

However, the same evidence does not establish full equivalence. Human language cognition remains distinctive in its biological development, pervasive social and sensorimotor history, affective regulation, autobiographical continuity, and established participation in moral and institutional responsibility. Whether these differences permanently constitute differences in kind remains a contestable theoretical interpretation, not a settled empirical fact. For current systems, a narrower conclusion is more defensible: functional competence, representational structure, grounding, agency, consciousness, and responsibility must be evaluated separately.

The framework supports cognitive complementarity, but only as a revisable, task-level principle. AI should be used where its capability is demonstrable, its outputs are verifiable, and accountability is clear; human judgment is indispensable where goals, values, relationships, or consequences require responsible interpretation. Future evidence of robust lifelong embodied learning, autonomous goal formation, causal world modeling, or credible consciousness indicators would require the framework to be revised. Its purpose is not to close the debate, but to make its claims more precise and empirically accountable.

REFERENCES

Barsalou, L. W. (1999). Perceptual symbol systems. Behavioral and Brain Sciences, 22(4), 577-660.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610-623. https://doi.org/10.1145/3442188.3445922

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 5185-5198. https://doi.org/10.18653/v1/2020.acl-main.463

Binz, M., Akata, E., Bethge, M., et al. (2025). A foundation model to predict and capture human cognition. Nature, 644, 1002-1009. https://doi.org/10.1038/s41586-025-09215-4

Brohan, A., Brown, N., Carbajal, J., et al. (2023). RT-2: Vision-language-action models transfer web knowledge to robotic control. arXiv. https://doi.org/10.48550/arXiv.2307.15818

Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.

Butlin, P., Long, R., Elmoznino, E., et al. (2023). Consciousness in artificial intelligence: Insights from the science of consciousness. arXiv. https://doi.org/10.48550/arXiv.2308.08708

Dentella, V., Gunther, F., Murphy, E., Marcus, G., & Leivada, E. (2024). Testing AI on language comprehension tasks reveals insensitivity to underlying meaning. Scientific Reports, 14, 28083. https://doi.org/10.1038/s41598-024-79531-8

Driess, D., Xia, F., Sajjadi, M. S. M., et al. (2023). PaLM-E: An embodied multimodal language model. Proceedings of the 40th International Conference on Machine Learning, 202, 8469-8488.

Du, Z., Zeng, A., Dong, Y., & Tang, J. (2024). Understanding emergent abilities of language models from the loss perspective. Advances in Neural Information Processing Systems, 37, 53138-53167. https://doi.org/10.52202/079017-1683

Fedorenko, E., Piantadosi, S. T., & Gibson, E. A. F. (2024). Language is primarily a tool for communication rather than thought. Nature, 630, 575-586. https://doi.org/10.1038/s41586-024-07522-w

Feng, J., & Steinhardt, J. (2024). How do language models bind entities in context? International Conference on Learning Representations.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. Current Directions in Psychological Science, 11(1), 11-15. https://doi.org/10.1111/1467-8721.00158

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. In Studies in Linguistic Analysis (pp. 1-32). Philological Society.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. Psychonomic Bulletin & Review, 9(3), 558-565.

Grice, H. P. (1957). Meaning. The Philosophical Review, 66(3), 377-388.

Gurnee, W., & Tegmark, M. (2024). Language models represent space and time. International Conference on Learning Representations.

Harnad, S. (1990). The symbol grounding problem. Physica D: Nonlinear Phenomena, 42(1-3), 335-346.

Harris, Z. S. (1954). Distributional structure. Word, 10(2-3), 146-162.

Hu, J., Mahowald, K., Lupyan, G., Ivanova, A., & Levy, R. (2024). Language models align with human judgments on key grammatical constructions. Proceedings of the National Academy of Sciences, 121(36), e2400917121. https://doi.org/10.1073/pnas.2400917121

Johnson, M. (1987). The body in the mind: The bodily basis of meaning, imagination, and reason. University of Chicago Press.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. Psychological Review, 99(1), 122-149.

Lakoff, G. (1987). Women, fire, and dangerous things: What categories reveal about the mind. University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). Metaphors we live by. University of Chicago Press.

Langacker, R. W. (1987). Foundations of cognitive grammar: Volume 1, Theoretical prerequisites. Stanford University Press.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. MIT Press.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. Learning & Memory, 12(4), 361-366. https://doi.org/10.1101/lm.94705

Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2024). Dissociating language and thought in large language models. Trends in Cognitive Sciences, 28(6), 517-540. https://doi.org/10.1016/j.tics.2024.01.011

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv. https://doi.org/10.48550/arXiv.1301.3781

Mon-Williams, R., Li, G., Long, R., Du, W., & Lucas, C. G. (2025). Embodied large language models enable robots to complete complex tasks in unpredictable environments. Nature Machine Intelligence, 7, 592-601. https://doi.org/10.1038/s42256-025-01005-x

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84(3), 231-259. https://doi.org/10.1037/0033-295X.84.3.231

OpenAI. (2023). GPT-4 technical report. arXiv. https://doi.org/10.48550/arXiv.2303.08774

Ouyang, L., Wu, J., Jiang, X., et al. (2022). Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 35, 27730-27744.

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, Article 2, 1-22. https://doi.org/10.1145/3586183.3606763

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

Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3(3), 417-424.

Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K., & Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. Advances in Neural Information Processing Systems, 36.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. Science, 268(5217), 1632-1634.

Tomasello, M. (2003). Constructing a language: A usage-based theory of language acquisition. Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. Behavioral and Brain Sciences, 28(5), 675-691.

Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433-460.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

Wei, J., Tay, Y., Bommasani, R., et al. (2022). Emergent abilities of large language models. Transactions on Machine Learning Research.

Weizenbaum, J. (1976). Computer power and human reason: From judgment to calculation. W. H. Freeman.

Yao, S., Zhao, J., Yu, D., et al. (2023). ReAct: Synergizing reasoning and acting in language models. International Conference on Learning Representations.

Critical Review of Social Sciences and Humanities Volume 5, Issue 2, 2025
Yin, Y., Jia, N., & Wakslak, C. J. (2024). AI can help people feel heard, but an AI label diminishes this impact. Proceedings of the National Academy of Sciences, 121(14), e2319112121. https://doi.org/10.1073/pnas.2319112121