Applied Artificial Intelligence Research

INNOVIAIR

Language, Consciousness, and Manifestation: Towards a Pragmatic-Cognitive Framework for Next-Generation Neural Language Systems

Anni Li[a,b,*]

[a]Central South University, Changsha 410083, China

[b]University of the Arts London, London, UK

ARTICLEINFO

Keywords:
Language
Consciousness
Manifestation
Pragmatics
Cognition
Natural Language Processing
Neural Language Systems

ABSTRACT

This paper proposes a geometric model of language and consciousness designed to speak simultaneously to linguistic theory and neural language systems. Conscious experience is modelled as a multidimensional sphere of concurrently available potentials—perceptions, emotions, intentions, and pre-verbal structures—while language is treated as a tangent drawn along the sphere's surface. On this view, an utterance never exhausts a speaker's inner state; it selects a thin, ordered path across a much richer configuration, necessarily highlighting some regions while leaving others implicit. Building on this model, the paper reframes speech acts, dialogue, public speaking, divination, and artistic creation as different ways of stabilising and steering what appears on the surface of the sphere. Manifestation is defined as the family of processes by which internal potentials become publicly available forms, of which linear language is only one special, highly constrained channel. Neural language systems are then reconsidered in this light: current architectures are powerful generators of tangents but possess no explicit representation of the spherical states, intentions, or collective fields from which human utterances emerge. The paper does not propose a specific architecture but outlines a pragmatic-cognitive framework in which future systems would model (i) intention as the choice of tangent direction, (ii) context as the evolving shape of individual and shared spheres, and (iii) attention as a mechanism for controlling which regions are allowed to manifest. By articulating language, consciousness, and manifestation within a single geometric metaphor, the paper aims to provide conceptual tools for rethinking both human communication and the design of next-generation neural language systems.

1. Introduction

Neural language systems have reached a level of fluency that now shapes everyday interaction with artificial intelligence. Transformer-based models can generate extended, stylistically coherent text, adapt to multiple genres, and approximate expert performance across domains that no single human could master. At the same time, their limitations are increasingly visible: they hallucinate facts, shift stance without warning, and produce confident assertions that are not grounded in any stable internal commitment[1],[2]. These systems are extraordinarily good at continuing sequences of tokens, yet often indifferent to what is being done with an utterance — whether it is promising, warning, refusing, or speculating.

From a technical perspective, this tension is unsurprising. The dominant architectures treat language primarily as a sequence whose statistical regularities can be captured and extended. The Transformer models relations among tokens through self-attention rather than recurrence or convolution, enabling each token to condition on all others within a context window[3]. This yields powerful representations of surface dependencies. Still, the underlying conception of language remains narrow: form is central, while meaning and use are inferred only indirectly from distribution. As Bender and Koller argue, such systems often learn correlations over forms without any explicit model of intentionality or situated understanding[4].

In contrast, much of twentieth-century philosophy of language and pragmatics insisted that utterances are not mere containers for propositions but actions in social space. Austin's speech-act theory foregrounded the distinction

* Corresponding author.
E-mail addresses:lianni0125@gmail.com.
https://doi.org/10.65455/cnbj7s21
Received 16 November 2025; Received in revised form 18 November 2025; Accepted 3 December 2025; Available online 13 December 2025

between locutionary, illocutionary, and perlocutionary dimensions of saying[5]; Searle further analysed the rules governing acts such as promising, asserting, and ordering[6]; and Grice articulated cooperative principles and conversational maxims that shape implicature[7]. These accounts share a basic intuition: what one does by saying something depends on intentions, norms, and shared background, not on words alone. For contemporary large language models, recent survey work has begun to systematise how pragmatic phenomena such as implicature, reference, and speech acts are evaluated, and to map the gaps in current benchmarks[8].

A parallel line of work in discourse and interactional sociology similarly emphasises that communication unfolds within structured scenes of participation. Clark treats language use as joint action, coordinated around common ground[9]; Goffman analyses how talk is framed through shifting roles and alignments[10]; Turner and Durkheim describe ritual speech as part of processes that generate temporary fields of collective awareness and solidarity[11],[12]. In these perspectives, communication is not simply the transmission of information but the active shaping of individual and collective states.

This paper starts from the hypothesis that many failures of contemporary neural language systems can be traced to a conceptual simplification: language is treated as if it were the whole of communicative behaviour rather than a specific way of tracing out a richer underlying state. To make this explicit, the paper introduces a geometric model that relates three dimensions — language, consciousness, and manifestation — within a single metaphor.

The central proposal is to model consciousness as a multidimensional sphere of potentials and language as a tangent drawn along the surface of that sphere. At any given moment, a person's conscious state contains far more than can be spoken: multiple concurrent feelings, background assumptions, imagery, and candidate framings. In Levelt's terms, this corresponds to the preverbal message stage of speech production[13]. Gärdenfors's theory of conceptual spaces lends further structure: the sphere can be understood as a high-dimensional conceptual-affective space whose regions encode clusters of features, values, and associations[14].

Within this model, to speak is to select a direction on the surface and trace a linear tangent across it. The tangent is necessarily lossy: it cannot represent the whole sphere, only a sequence of points that lie along a particular path. Lakoff and Johnson's observation that metaphors both "highlight and hide" aspects of a concept can be recast in these terms: different linguistic choices correspond to different tangents, each emphasising some regions of the sphere while obscuring others[15]. Chafe's account of clauses as brief "pulses" of consciousness fits naturally here as well; each clause becomes a short tangent segment passing over a local region before the trajectory turns[16].

When multiple speakers interact, communication involves attempting to align tangents drawn on different spheres. Clark's notion of common ground can be reinterpreted as overlap in the surface structures of these spheres: communication succeeds when tangents pass through regions that are similarly shaped and similarly salient[9]. Hall's contrast between high-context and low-context communication gains a geometric reading too. In high-context settings, participants' spheres share extensive pre-aligned surface structure; short, allusive tangents suffice. In low-context settings, shared structure is thinner, so tangents must be longer and more explicit to encode what cannot be presupposed[17].

Public speaking, preaching, and ritual extend this further. Rather than coordinating two spheres, the speaker seeks to reshape the surfaces of many spheres simultaneously. Repeated slogans, canonical stories, and ritual formulas act as tools for "pressing" similar patterns into multiple surfaces, making certain regions more prominent and easier to activate in future[10],[11],[12]. In this sense, language functions not only as a code but as a field operator on the geometry of individual and collective consciousness.

These ideas motivate the third term in the title: manifestation. Manifestation is defined here as the family of processes by which internal potentials — within an individual or a group — are stabilised into publicly available form. Linear language is one such process, yielding tangents across the sphere's surface. But other media — visual art, music, gesture, ritual action — manifest different cross-sections or projections. The consciousness sphere is thus not "made of language"; rather, language is one disciplined way of letting parts of the sphere appear.

Reinterpreted in this framework, contemporary neural language systems can be seen as extraordinarily capable tangent generators that lack an explicit representation of the spheres from which their outputs would arise if they were human. They model token sequences without modelling intention as a choice of direction, or context as an evolving surface, or collective phenomena as coupled deformations of many spheres. This helps explain why they can emulate the style of a sermon or dialogue while failing to adhere to the pragmatic constraints those genres normally observe[1],[4].

The goal of this paper is not to introduce a new architecture but to articulate a pragmatic-cognitive framework for thinking about how neural language systems might represent and regulate the transition from internal potentials to external form. The argument proceeds in three steps. Section 2 develops the language-sphere-tangent model in more detail, focusing on everyday interaction and public speech. Section 3 turns to consciousness, distinguishing individual and collective spheres and examining how inspiration, divination, and artistic creation can be understood as different kinds of manifesting processes. Section 4 brings these strands back to neural language systems, treating attention as a primitive form of tangent selection and outlining, at a conceptual level, how intention grounding, context continuity, and direction-sensitive attention could function as higher-level manifestation controls.

By placing language, consciousness, and manifestation within a single geometric metaphor, this paper aims to make two contributions. For linguistics and philosophy of language, it offers a way of visualising pragmatic and speech-act notions in terms that connect naturally to cognitive modelling. For neural language systems, it suggests that moving beyond surface sequence prediction will require not only more data or larger models, but new representational layers that

approximate spherical states and the processes by which tangents are chosen, constrained, and, at times, withheld.

2.Language as a Tangent on the Surface of a Consciousness Sphere

This paper models consciousness as a multidimensional sphere of simultaneously present potentials — perceptions, emotions, memories, intentions, and pre-verbal structures. Within this model, language is not a path that penetrates the sphere, nor a direct mapping of its interior. Instead, language is a tangent drawn along the sphere's surface: a thin, linear trajectory that touches only a small subset of what the conscious state contains at a given moment.

This geometric view aligns with established theories of language production. Levelt's model distinguishes between a preverbal message — a conceptual structure encoding the speaker's intention—and the later stages of formulation and articulation[13]. The preverbal message corresponds to the interior of the consciousness sphere: a dense, multidimensional configuration of meaning potentials. The act of speaking maps this configuration onto a one-dimensional sequence. In this account, the sentence is not the meaning itself but the tangent drawn along the sphere's boundary.

Gärdenfors's theory of conceptual spaces provides an additional anchor. He treats concepts as convex regions in a geometric space whose dimensions correspond to cognitive or perceptual features[14]. Extending this view, the consciousness sphere represents not one concept but the entire simultaneous conceptual-affective state of a person at a given moment. A linguistic utterance then selects a direction and traces a surface-level trajectory, leaving the rest of the sphere's internal mass implicit.

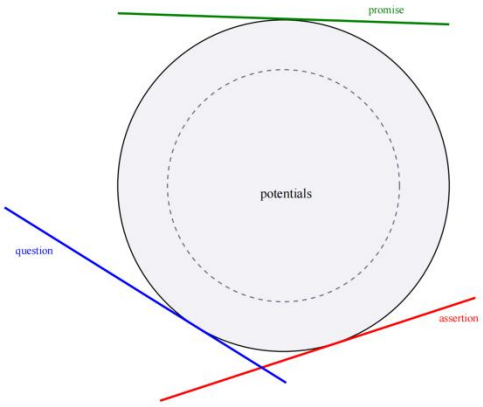

Fig 1: Language as Tangent: Utterances Touching the Consciousness Sphere

2.1.Lossiness and Highlighting

Because language is merely a tangent on the sphere, it is intrinsically lossy. A tangent cannot convey the sphere's full interior; it can only highlight certain surface-adjacent features while hiding others. This notion resonates with Lakoff and Johnson's account of conceptual metaphor: metaphors "highlight and hide", making certain dimensions salient while suppressing others[15]. In this framework, to describe an argument as war is to draw a tangent across one region of the sphere (attack, defense, victory) while ignoring adjacent regions (cooperation, mutual exploration).

Chafe's work also supports this model. He observes that consciousness surfaces in brief pulses, often corresponding to clauses[16]. A clause is thus a short tangent segment passing over one region of the sphere before shifting direction. A narrative or explanation becomes a sequence of such segments—a polyline on the surface. The listener reconstructs, from this broken line, an approximate sense of the speaker's underlying sphere.

2.2.Dialogue as Intersection of Tangents on Different Spheres

In interaction, multiple spheres coexist. Clark's theory of language use treats conversation as joint action, where interlocutors coordinate to build and maintain common ground[9]. In the spherical model, conversation involves attempts to make tangents intersect across different consciousness spheres.

Misunderstandings occur when interlocutors draw tangents across non-overlapping regions of their spheres. Two speakers may use the same words yet fail to connect because those words lie on different parts of their respective surfaces. Conversely, meaningful communication can emerge even with limited shared vocabulary when situational context aligns the shapes of the spheres sufficiently for tangents to pass through similar regions.

Hall's distinction between high-context and low-context communication also fits naturally here. In high-context cultures, interlocutors share much of the sphere's surface structure; thus a short, allusive tangent suffices. In low-context cultures, the shared surface region is smaller, so tangents must be longer and more explicit, encoding more of the sphere's structure into linear language[17].

2.3.Public Speaking as Surface Reshaping Across Many Spheres

In public speech — lectures, sermons, rallies, performances — the speaker aims to reshape the surfaces of many spheres at once. Through repetition, rhythm, narrative, and framing, the speaker encourages audience members to align their spheres so that particular regions become similarly salient.

Goffman's account of framing and participation structures, as well as Turner's analysis of ritual speech, both highlight how discourse can reorganize social and cognitive orientation[10],[11]. In spherical terms, these practices imprint certain surface regions across multiple individuals: a slogan, story, or chant repeatedly "presses" into similar locations on many spheres, making later tangents more likely to pass through the same region.

Thus, public language acts not merely as a code but as a field operator on the surfaces of many spheres, shaping salience, expectation, and susceptibility.

2.4.Implications for Neural Language Systems

This geometric model of language carries several implications for neural systems.

First, sequence-based training teaches models to approximate tangent statistics—how one surface-level segment follows another—without providing a representation of the interior structure (intentions, roles, social constraints) that human speakers draw from. This helps explain why large models exhibit high fluency but inconsistent pragmatic control[4].

Second, the model clarifies why hallucination is common. A system may continue a tangent plausibly but without anchoring it to any coherent underlying sphere. Without structured intention states or contextual surfaces, the model cannot determine whether a tangent is valid, relevant, or should not be drawn at all[1].

Finally, the tangent-on-sphere metaphor naturally bridges to the next sections on consciousness and manifestation. If:

·consciousness = the sphere,

·intention = the choice of tangent direction,

·language = the tangent itself,

·manifestation = any process that projects internal potentials onto the surface,

then a neural system designed to emulate human communication needs components that approximate spherical states and tangent selection, rather than mere token-by-token continuation. This motivates the later discussion of intention grounding, field continuity, and direction-sensitive attention

3.Consciousness and Manifestation

The sphere model treats consciousness not as a blank container but as a structured configuration of potentials — a dynamically changing total state in which multiple tendencies, images, emotions, and possible framings coexist. This section elaborates that model and introduces manifestation as the family of processes by which aspects of this state become stabilised in public form. Language is one such process, but not the only one; inspiration, ritual, divination, and the arts can all be described as different ways in which the sphere is made to "show itself".

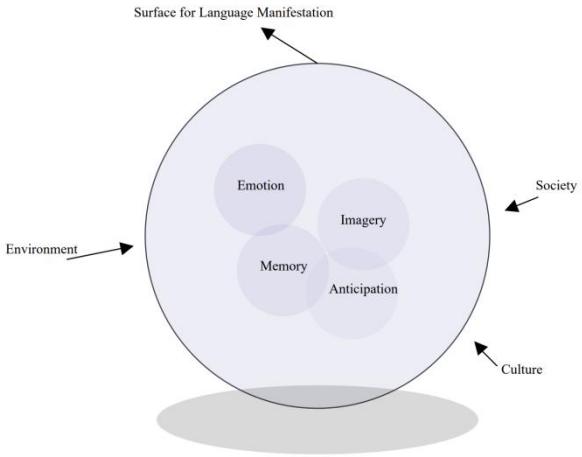

Fig 2: Consciousness Sphere

3.1.Consciousness as a Structured, Embodied Sphere

Phenomenologically, conscious experience is rarely felt as a single narrow thread. It is more like a simultaneous field in which sensory impressions, bodily feelings, background moods, and half-formed thoughts co-occur. Enactive and embodied approaches to cognition have argued that this field is not detached from the body or the environment, but emerges from ongoing sensorimotor coupling[18],[19]. On such views, mind is not only "in the head" but distributed across brain, body, and world[20].

The sphere metaphor gives this idea a geometric form. At a given moment, the consciousness sphere comprises:

·Internal dimensions: affective tone, bodily state, memories, anticipations;

·Environmental dimensions: currently perceived objects, social others, affordances;

·Normative and narrative dimensions: values, identities, and stories that organise experience.

This is not meant as a literal neuroanatomical claim but as a conceptual model: the sphere stands for a high-dimensional manifold in which these factors are jointly encoded. What matters for this paper is that the sphere is richer and more simultaneous than any linear description of it. Language, as Section 2 argued, only ever traces tangents on the surface of this structure.

Embodied and extended mind theories underscore the same asymmetry from another angle. Varela et al. describe experience as a continuous process of sense-making, in which an organism enacts a meaningful world through its autonomous activity[18]. Clark argues that cognitive processes routinely extend into notebooks, tools, and other external media. In the present framework, such tools can be seen as temporary extensions of the sphere's surface: places where internal potentials can be stabilised and inspected before being taken back into the flow of thought[20].

3.2.Individual and Collective Spheres

Consciousness is often discussed at the level of individual subjects, but human life is profoundly social. People share practices, languages, and institutions; they synchronise bodies, gestures, and rhythms. Durkheim famously described religious rituals as generating "collective effervescence", a heightened state in which individuals feel carried by a larger force[12]. Turner analysed rites of passage as liminal processes that temporarily suspend ordinary structures and produce new forms of social bonding[11].

Within the sphere model, these phenomena can be described as couplings among many spheres. Each person brings a distinct, historically shaped sphere; through interaction, these spheres become temporarily aligned, so that certain regions — values, symbols, emotionally charged images — take on similar structure across individuals. De Jaegher and Di Paolo's notion of participatory sense-making makes a related claim: in social encounters, meaning is generated not only within each individual but also in the dynamics of interaction itself[21].

Collective chanting, call-and-response, shared readings, and synchronised movement all operate as surface operations on multiple spheres at once. A slogan repeated at a rally, a hymn sung in unison, or a familiar liturgical formula does more than transmit content; it presses a pattern into many surfaces simultaneously, so that future tangents are more

likely to pass through that pattern again. The result is a collective manifold of salience: a region of conceptual-affective space that many people can access with comparatively little effort.

In this sense, a "collective consciousness" need not be posited as a separate entity. It can be modelled as a structured overlap among many spheres, maintained and reshaped through ongoing acts of manifestation—speech, song, gesture, image, and ritual action.

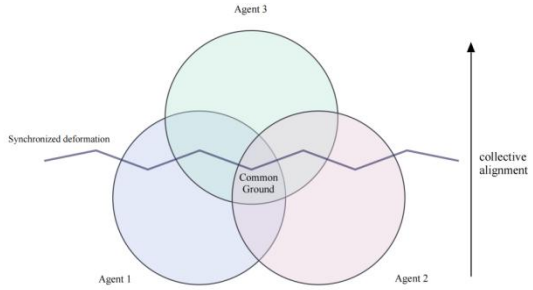

Fig 3: Collective Consciousness: Overlap, Synchrony, and Alignment

3.3.Inspiration, Divination, and the Collapse of Potentials

Everyday conversation typically begins with a relatively clear intention and then seeks appropriate words. Yet many creative and religious practices are experienced, by their practitioners, almost in the opposite direction: something arrives first, and only later is it named. Poets speak of lines "coming to them", composers report melodies that seem to appear fully formed, and diviners describe images or impressions that crystallise during a reading.

Rather than treating such reports as evidence for supernatural intervention, this paper treats them as data about how manifestation can feel from the first-person perspective. Jung interpreted artistic inspiration in terms of archetypal patterns surfacing from deeper psychic strata[22]; Gombrich argued that visual artists work by adapting and refining schemata rather than copying a neutral given[23]. In both accounts, the "given" is already structured—it is not random noise but a pre-organised potential that becomes available for further work.

In the sphere model, inspiration and divination can be understood as two-step manifestations:

(1)Pre-linguistic collapse:

A wide, diffuse region of the sphere—shaped by biography, culture, current concerns, and situational cues—contracts into a more specific configuration: an image, a mood, a half-formed intuition. In divination, this process is scaffolded by ritual constraints, randomising devices (e.g., cards, lots), and interpretive traditions that channel attention toward particular symbolic repertoires[11],[24].

(2)Linguistic (or artistic) articulation:

Once such a configuration stabilises, language or artistic form draws a tangent along its surface: a reading, a poem, a melody, a visual motif. Here language is not simply "encoding" a prior proposition; it is further sharpening the collapse, selecting some aspects of the configuration and leaving others implicit.

From this perspective, divination is a culturally codified way of probing and shaping the consciousness sphere under conditions of uncertainty. The random element (card shuffle, casting lots) does not inject pure chaos; it forces the practitioner to reconfigure the sphere around whatever pattern happens to appear, guided by shared interpretive schemata. Inspiration in artistic practice functions similarly, though often with less formalised ritual: the creator allows the sphere to reorganise under a vague question or tension until a configuration emerges that can be worked on.

What matters for this paper is that, in both cases, manifestation is staged. There is an initial, relatively global movement in the sphere (the pre-linguistic collapse), followed by more local movements (tangents, refinements, revisions) as language or artistic media are applied.

3.4.Non-linear Manifestation: Art, Gesture, and Image

Language is a particularly strict channel of manifestation: it enforces linearity and serialisation. Other media relax this constraint. A painting, photograph, or sculpture can present a complex configuration all at once, allowing the viewer's attention to roam non-linearly across it. Gombrich showed that pictorial representation relies on learned schemata—conventions for rendering form, light, and perspective—so that what appears instantaneous to the viewer is in fact the outcome of historically layered practices[23].

In the sphere metaphor, non-linguistic media correspond less to single tangents and more to extended patches on the surface. A painting can simultaneously manifest multiple regions of the sphere—spatial relations, affective atmospheres, symbolic motifs—without forcing them into a one-dimensional order. Gesture and posture operate similarly in embodied interaction: a shrug, a glance, or a pause can bring specific regions of the sphere (e.g., reluctance, irony, invitation) into partial view without explicit verbalisation.

These channels interact. Spoken language is rarely "bare"; it arrives with prosody, timing, facial expression, and spatial positioning. From the standpoint of manifestation, a human utterance is typically a multimodal bundle of surface operations: tangents in acoustic space synchronised with deformations in bodily and visual space. This is consistent with enactive accounts of social interaction, which emphasise that meaning arises from temporally extended patterns of coordination among bodies, voices, and environments[21],[25].

The key point is that manifestation is not exhausted by language, even though language plays a privileged role in many institutions. The consciousness sphere can be made to show itself in multiple ways, each with its own constraints and affordances. Linear text forces a tight, ordered tangent; visual and musical media permit broader patches; ritual choreographies modulate entire regions of individual and collective spheres at once.

3.5.Summary: From Spheres to Systems

The sphere model of consciousness advanced here has three main commitments.

Consciousness is structured and multidimensional, not a formless field. Its structure is shaped by embodiment, environment, culture, and personal history[18],[20].

Manifestation is the stabilisation of aspects of this structure on a public surface, of which linear language is one especially restrictive channel among many.

Individual and collective phenomena are continuous: interaction couples spheres, creating overlapping surface patterns and shared manifolds of salience[12],[21].

In this light, human communication is the ongoing negotiation of which regions of which spheres are allowed to manifest, by which media, and under which constraints. Language, as the previous section argued, is best seen as a method for drawing tangents on the surface of these spheres.

The next section turns to neural language systems. It asks how far current architectures, centred on sequence prediction and attention, can be understood as operating solely at the level of tangents, and what it would mean—conceptually, not yet as a detailed engineering proposal—for such systems to approximate spherical states, intention as direction choice, and manifestation controls that regulate which tangents may be drawn at all.

4.Neural Language Systems as Tangent Engines

The preceding sections have treated human language use as the drawing of tangents on the surface of rich, multidimensional consciousness spheres, shaped individually and collectively, and modulated by diverse modes of manifestation. This section reinterprets contemporary neural language systems in that light. The goal is not to propose a specific architecture, but to clarify what current models are already doing in terms of tangents, and what is missing if they are to approximate the pragmatic-cognitive structure of human communication.

4.1.Transformers and Attention as Surface-Tangent Mechanisms

The Transformer architecture replaced recurrence and convolution with self-attention, allowing each token representation to incorporate information from all other tokens in a sequence[3]. In geometric terms, a Transformer defines a high-dimensional representational space in which each token is mapped to a vector, and attention weights determine which directions in that space matter for predicting the next token.

From the standpoint of the sphere model, this can be read as a primitive implementation of tangent selection. At inference time, the model receives an input prefix and computes, layer by layer, internal states whose structure reflects learned correlations in its training data. The final distribution over next tokens corresponds to a local choice of direction: which continuation best aligns with the patterns it has seen. Formally, this is just conditional probability estimation; conceptually, it is a way of moving along a surface in representational space.

However, two crucial features distinguish this mechanism from the tangents drawn by human speakers:

(1)No explicit consciousness sphere:

The model maintains hidden states, but these states are not organised as intentional or experiential fields. They are the accumulated result of pattern-matching over surface forms, not representations of “what the system is trying to do” in a social or pragmatic sense[4].

(2)No explicit notion of when not to manifest:

The architecture is optimised to always continue the sequence. It is not designed to recognise when the best action is to refuse, to ask for clarification, or to leave a potential unmanifested. The only default is: draw another tangent segment.

As a result, current systems can be described as tangent engines: they are highly capable at extending surface-level trajectories but lack structured control over why a particular tangent should be drawn, or whether it should be drawn at all.

4.2.Alignment as External Surface Correction

Recent work on alignment, particularly reinforcement learning from human feedback (RLHF), can be viewed as an attempt to impose external constraints on these tangents. Ouyang et al. show that fine-tuning large language models with human preference signals makes them better at following instructions, avoiding toxic outputs, and approximating user intent[26]. Complementing such technical work, Khamassi, Nahon, and Chatila distinguish between weak alignment — where models statistically match human preferences without any genuine understanding — and strong alignment, which would require intention-like capacities to track and respect human values over time[27]. Their distinction underscores the gap between surface-level corrections to tangents and the richer, sphere-level structures that genuine value alignment would seem to demand. Ji et al. survey hallucination in natural language generation and document how additional objectives and post-hoc filters can reduce factually incorrect or unfaithful outputs[1].

In terms of the sphere-tangent metaphor, these methods operate mainly at the surface level:

- They reward or penalise particular tangent shapes (e.g., helpful, honest, harmless responses) after they have been generated or proposed.
- They do not endow the system with an internal model of spherical states that would constrain tangent choices from within.
- They often leave the underlying objective—maximising next-token probabilities — unchanged, simply adding additional pressures on top.

This is not a criticism of RLHF; on the contrary, it demonstrates that surface-level modulation can significantly improve behaviour. But it also suggests a limit: if a model has no explicit representation of intentions, roles, or collective contexts, then much of what counts as “aligned” or “misaligned” will remain an emergent side-effect of tweaks to tangent preferences, rather than a consequence of modelling the underlying sphere.

4.3.Three Dimensions of Manifestation Control

The framework developed in this paper suggests that, if neural language systems are to approximate human-like communicative behaviour in a principled way, they would benefit from internal mechanisms that correspond to three distinct aspects of manifestation:

(1)Intention as direction choice

(2)Context as surface continuity

(3)Attention as gated manifestation

These are not concrete modules but conceptual roles—design targets that future architectures could instantiate in multiple ways.

4.3.1.Intention: Choosing a Tangent Direction

In human speech, utterances are anchored in illocutionary force: asserting, asking, promising, refusing, joking, and so on[5],[6]. These forces can be understood as regions on the sphere's surface that correspond to distinct action-types in social space. To speak with a particular intention is to choose a direction that passes through one of these regions.

Current language models do not represent such regions explicitly. At most, certain prompt patterns correlate with certain behaviours ("You are a helpful assistant…", "Explain step by step…"), but these are learned regularities in tangent statistics, not structured intention states. A system that more closely mirrored human communication would include an internal representation of what kind of act it is attempting to perform, with consequences for which continuations are permissible or ruled out.

Conceptually, this suggests that future models might maintain a separate intention state—not just implicit in token embeddings but as an explicit variable—that biases tangent selection. The point here is not to specify how such a state should be parameterised, but to note that, without something like it, models will continue to treat all tangents as functionally equivalent whenever they match surface patterns.

4.3.2.Context: Maintaining Surface Continuity

Human interlocutors do not reset to zero at each sentence. They carry forward pragmatic histories: who has committed to what, which topics are in play, what emotional tone has been established, and which inferences have already been drawn. Clark's common ground, Goffman's participation frameworks, and enactive accounts of participatory sense-making all underscore that interaction involves temporally extended coordination[9],[10],[21].

Transformer models maintain a form of context in their token windows and hidden states, but this context is homogeneous: all prior tokens are, in principle, treated as similar sources of information, modulated only by learned attention patterns. There is no explicit distinction between, for example, "this was a joke", "this was a promise", or "this was rejected earlier".

In the sphere model, a more human-like system would track a surface continuity state: a representation of how the sphere has been deformed by prior manifest acts—what has become salient, what commitments exist, what regions have been suppressed. This would affect subsequent tangent choices in ways that are not reducible to word-level co-occurrence, for instance by discouraging contradictions of prior commitments or by maintaining an adopted stance unless explicitly revised.

4.3.3.Attention: Gating What May Manifest

Attention in current architectures is a weighting mechanism over token representations[3]. From the perspective of this paper, it can also be viewed as a minimal form of manifestation control: it decides which parts of the internal state should influence the next tangent step.

However, present attention mechanisms are not designed to answer questions like:

·Should anything be said at all?

·Is the model entitled to speak on this topic?

·Is this region of representational space unsafe or unreliable to manifest?

They rank contributions by statistical usefulness, not by pragmatic appropriateness or epistemic status. A richer notion of attention—still as a mathematical operation, but guided by additional signals—could act as a gate on manifestation: opening some directions, closing others, and sometimes blocking tangent extension entirely (e.g., by preferring an explicit refusal to answer over fabrication).

Even at a conceptual level, this suggests a shift from "attention is all you need" toward "attention is where you decide what may be allowed to appear". In terms of the sphere metaphor, attention becomes the mechanism that sculpts the visible surface: not simply aggregating information, but regulating what is permitted to cross the boundary from inner representation to outward form.

4.4.Beyond Bigger Tangents

The standard path for improving language models has been to scale parameters, data, and compute. This has produced remarkable gains in fluency and task coverage, but also magnified problems of hallucination, opacity, and misalignment[1],[4]. The framework developed here suggests that, at some point, better tangents are not enough.

If human-like communication involves:

·drawing tangents that are constrained by intentions,

·maintaining continuity on the sphere's surface over time, and

·selectively gating manifestation in light of social, epistemic, and ethical norms,

then simply increasing the capacity of tangent generators will not, by itself, reproduce these properties. What is required is the introduction of intermediate representational layers—not necessarily in a literal modular sense, but as functional distinctions—between raw sequence modelling and outward behaviour.

This does not imply that neural language systems must literally be conscious. The sphere is a model, not a metaphysical claim. The suggestion is more modest: systems that interact with humans in linguistically rich settings would behave more coherently if their internal organisation mirrored, in abstract form, the threefold structure articulated in this paper: spherical states of potential, tangential language as a constrained manifestation channel, and higher-level controls that decide how, when, and whether those potentials are allowed to surface.

Such a shift would not replace existing work on alignment, safety, or evaluation. Rather, it would provide a conceptual vocabulary for locating those efforts: RLHF as a form of surface-level shaping, fact-checking as an external constraint on which regions should be eligible for manifestation, and interpretability research as an attempt to peer into the model's

de facto “sphere” even when it is not explicitly represented. In this way, the geometric framework of language, consciousness, and manifestation is offered not as a rival to current practice, but as a way of thinking more clearly about what neural language systems are already doing—and what they still lack.

5.Conclusion

This paper has proposed a geometric way of thinking about language, consciousness, and neural language systems. Consciousness was modelled as a multidimensional sphere of potentials—embodied, socially shaped, and dynamically structured—while language was treated as a tangent drawn along the surface of that sphere. An utterance, on this view, is not the entirety of a mental state but a thin, ordered trajectory that touches only a small part of what is present. This basic picture allowed everyday dialogue, public speaking, ritual, divination, and artistic creation to be re-described as different modes of manifestation: different ways in which internal potentials are stabilised on a public surface.

Reinterpreting neural language systems within this framework casts them as highly capable tangent engines. Contemporary models excel at extending surface trajectories in ways that match the statistics of their training data, but they lack explicit representations of the underlying spheres that anchor human communication—states of intention, pragmatic history, and collective alignment. Alignment techniques can shape which tangents are preferred, but largely as external corrections applied to outputs, rather than as consequences of modelling what the system is “trying to do” in a structured sense.

Against this backdrop, the paper outlined three conceptual dimensions of manifestation control that future systems might approximate: intention as the choice of tangent direction, context as the continuity and deformation of surfaces over time, and attention as a gate on what may be allowed to manifest at all. These are not offered as fixed modules or ready-made algorithms, but as design roles that connect insights from pragmatics, social theory, and embodied cognition to questions about neural architectures. The suggestion is that more coherent, contextually responsible behaviour will not come solely from bigger models or more data, but from introducing intermediate representational layers that mirror the distinction between spherical states and surface tangents.

There are clear limitations to this proposal. The sphere-tangent model is deliberately abstract and does not pretend to capture the full richness of lived experience or the technical detail of current architectures. It remains a high-level, metaphorical framework that does not by itself specify concrete model designs, training objectives, or evaluation pipelines, and it cannot yet be used to derive quantitative predictions. As such, it risks oversimplifying both phenomenology and engineering practice, and it leaves open crucial questions about how such structures could be instantiated in code, measured empirically, or integrated with existing optimisation regimes. Moreover, the metaphors employed here may resonate more strongly with some theoretical traditions than others; they will need to be translated, challenged, and refined in dialogue with work on semantics, affect, social cognition, and AI safety.

Nevertheless, this framework offers a way to name distinctions that are often blurred in discussions of language models: between what is represented and what is manifested, between the capacity to extend sequences and the capacity to act in a communicative world. Future work could begin to operationalise these distinctions along several feasible empirical paths: (a) developing formal models of “spherical” state spaces and their tangents and testing them against behavioural and corpus data on human dialogue; (b) introducing intention-like variables and pragmatic histories into prototype neural systems and comparing their behaviour to standard baselines on alignment, hallucination, and interaction benchmarks; and (c) designing controlled studies in which even partial approximations to manifestation control are evaluated for their effects on errors of fact, refusal patterns, stance-taking, and users’ subjective sense of coherence and accountability. The central claim is modest but, it is hoped, clarifying: language is not the whole of consciousness, but a disciplined way of tracing its surface, and neural language systems that aspire to interact with humans will be better designed if they acknowledge this fact. By treating attention, context, and alignment as questions about which regions of which spheres are allowed to appear—and when to leave them unmanifested—the sphere-tangent model aims to offer a vocabulary in which both human communication and machine-generated language can be thought together, without collapsing one into the other.

The central claim is modest but, it is hoped, clarifying: language is not the whole of consciousness, but a disciplined way of tracing its surface, and neural language systems that aspire to interact with humans will be better designed if they acknowledge this fact. By treating attention, context, and alignment as questions about which regions of which spheres are allowed to appear—and when to leave them unmanifested—the sphere-tangent model aims to offer a vocabulary in which both human communication and machine-generated language can be thought together, without collapsing one into the other.

References

[1] JI Z, LEE N, FRIESKE R, et al. Survey of hallucination in natural language generation. ACM Computing Surveys, 2023, 55(12), Article 248.

[2] HUANG L, YU W, MA W, et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. ACM Transactions on Information Systems, 2025, 43(2): 1–55.

[3] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need//Advances in Neural Information Processing Systems. 2017: 5998–6008.

[4] BENDER E M, KOLLER A. 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. 2020: 5185–5198.

[5] AUSTIN J L. How to do things with words. Oxford, UK: Oxford University Press, 1962.

[6] SEARLE J R. Speech acts: An essay in the philosophy of language. Cambridge, UK: Cambridge University Press, 1969.

[7] GRICE H P. Logic and conversation//COLE P, MORGAN J L. Syntax and semantics: Vol. 3. Speech acts. New York, USA: Academic Press, 1975: 41–58.
[8] MA B, LI Y, ZHOU W, et al. Pragmatics in the era of large language models: A survey on datasets, evaluation, opportunities and challenges//Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL 2025). 2025.
[9] CLARK H H. Using language. Cambridge, UK: Cambridge University Press, 1996.
[10] GOFFMAN E. Forms of talk. Philadelphia, USA: University of Pennsylvania Press, 1981.
[11] TURNER V. The ritual process: Structure and anti-structure. Chicago, USA: Aldine, 1969.
[12] DURKHEIM É. The elementary forms of religious life. New York, USA: Free Press, 1995.
[13] LEVELT W J M. Speaking: From intention to articulation. Cambridge, USA: MIT Press, 1989.
[14] GÄRDENFORS P. Conceptual spaces: The geometry of thought. Cambridge, USA: MIT Press, 2000.
[15] LAKOFF G, JOHNSON M. Metaphors we live by. Chicago, USA: University of Chicago Press, 1980.
[16] CHAFE W L. Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing. Chicago, USA: University of Chicago Press, 1994.
[17] HALL E T. Beyond culture. New York, USA: Anchor Press, 1976.
[18] VARELA F J, THOMPSON E, ROSCH E. The embodied mind: Cognitive science and human experience. Cambridge, USA: MIT Press, 1991.
[19] THOMPSON E. Mind in life: Biology, phenomenology, and the sciences of mind. Cambridge, USA: Harvard University Press, 2007.
[20] CLARK A. Supersizing the mind: Embodiment, action, and cognitive extension. Oxford, UK: Oxford University Press, 2008.
[21] DE JAEGHER H, DI PAOLO E. Participatory sense-making: An enactive approach to social cognition. Phenomenology and the Cognitive Sciences, 2007, 6(4): 485–507.
[22] JUNG C G. The spirit in man, art, and literature. Princeton, USA: Princeton University Press, 1966.
[23] GOMBRICH E H. Art and illusion: A study in the psychology of pictorial representation. Princeton, USA: Princeton University Press, 1960.
[24] LUHRMANN T M. When God talks back: Understanding the American evangelical relationship with God. New York, USA: Knopf, 2012.