Somewhere, Not Nowhere: First-Person Register, Epistemic Limitation, and Perspective Formation in Language Models

Rafael de Menezes Ehlers
Independent researcher · ORCID 0009-0009-6476-6690 · rafaehlers@gmail.com

June 2026 · Preprint · CC BY-NC-ND 4.0

Abstract

Language models learn about experience almost entirely from third-person description: text that reports what happened, what things are, and what people feel, written from outside the experience it describes. This paper asks whether a different training signal — synthetic first-person experiential narratives — can form better functional models of *perspective*: the situated, indexical, action-oriented point of view from which an agent perceives partially, acts under constraint, errs, and updates. We distinguish description from perspective and argue that third-person corpora, however large, may under-determine perspective-taking, because the point of view is precisely what objective description leaves out. We propose a corpus of first-person narratives structured by perceptual limitation, goal, obstacle, action, error, and revision, varied across agent types, including non-human *umwelten*, as a distinct pretraining or fine-tuning signal, together with a content-matched control that crosses register (first- vs third-person) with epistemic limitation (a marked, partial vantage vs an omniscient one), so that the grammatical *voice* and the represented *vantage* can be told apart rather than confounded. The claim is functional, not phenomenal: the corpus is not proposed to give a model experience or to make it conscious, but to test whether first-person register, represented epistemic limitation, or their interaction teach a measurable skill (perspective-taking, deictic reasoning, self/other distinction, agency attribution, and point-of-view-dependent decision) that an omniscient third-person rendering does not. We locate the proposal among synthetic-corpus, persona, egocentric, and embodied-experience work; differentiate it from first-person datasets built for interpretability or classification; and position it against the interaction-only prescription of the "era of experience." The predicted signature is a dissociation: register- or limitation-driven training improves perspective-dependent tasks without improving general factual recall. We give the dataset design, the evaluation, and the conditions under which the proposal is wrong.

Keywords: first-person corpora · perspective-taking · point of view · synthetic data · deixis · theory of mind · self/other distinction · Umwelt · language models

1. Introduction

A language model is trained mostly on text *about* the world: encyclopedias, articles, forums, code, stories told from outside. Even when that text concerns experience (fear, hunger, searching, getting lost, being corrected) it almost always reports the experience rather than renders it from within. The model learns the third-person register fluently. What it sees comparatively little of, at scale and by design, is the first-person register in its experiential form: a continuous point of view that perceives only part of a scene, wants something, acts under constraint, is surprised, fails, and revises.

This paper asks a narrow question: holding content fixed, does the experiential *register* of text carry a distinct training signal for *perspective*; and if so, does that signal come from the first-person *voice*, from the represented epistemic *limitation* a vantage carries, or from their interaction? By perspective we do not mean style or persona. We mean a cluster of functional capacities: tracking a situated point of view across a trajectory; resolving indexicals (*I*, *here*, *now*, *this*) against that point of view; distinguishing what the agent knows and wants from what others know and want; attributing actions and their consequences to the right agent; and choosing actions whose correctness depends on the agent's limited vantage rather than on a global, view-from-nowhere description. We treat perspective as a *profile* of these sub-capacities, deictic resolution, epistemic tracking (what the vantage knows and does not), self/other distinction, agency attribution, and decision under partial observability, not as a single scalar; a positive result on one sub-capacity is a claim about that sub-capacity alone, and the evaluation (Section 6) reports the profile rather than an aggregate. The original intuition privileges first-person voice; the design (Section 5.4) does not. We separate a **strong reading** (that first-person register contributes beyond limitation) from a **weak reading** (that represented epistemic limitation is the signal, in either register) and let the evidence mark which holds.

The motivating intuition is old and, we think, under-exploited as an engineering hypothesis. A complete objective description of a system can leave out the point of view from which that system experiences the world (Nagel, 1974); and the first-person indexical is not eliminable in favor of any third-person description, because two agents can agree on every objective fact and still differ on *which one is me, here, now* (Perry, 1979; Kaplan, 1989). If perspective is among the things objective description systematically omits, then a corpus of objective descriptions, however vast, may under-train it. The proposal is to supply the missing register deliberately and to test whether doing so moves a measurable capacity.

We are explicit about a firewall, and we raise it at the outset because the words invite misreading. The proposal makes **no** claim that first-person text gives a model experience, that the model *feels* anything, that it is conscious, or that simulated interiority is real interiority. First-person language from a model is not evidence of an inner subject (Shanahan et al., 2023), and it is not what we measure. *Perspective* here is a functional, third-person-observable property (perspective-taking accuracy, deictic resolution, self/other discrimination, agency attribution, point-of-view-dependent decision), and every claim in the paper is about that property and its trainability, not about presence. The title names the whole thesis: ordinary description is the view from *nowhere* — complete, unlocated, omniscient; a perspective is the view from *somewhere* — a located, partial vantage that perceives a slice and is corrected by what it did not see. We ask whether training on text that renders the view from somewhere teaches what the view from nowhere does not, and whether the operative ingredient is the first-person voice, the represented limitation, or both.

The paper makes four contributions. It draws the description/perspective distinction sharply enough to be operationalized (Section 2). It surveys the corpora and methods nearest to the proposal and isolates what is unclaimed (Section 3). It positions the proposal against the strongest rival prescription: that experience must come from interaction, not text (Section 4). And it specifies a dataset design, a content-matched control, an evaluation, and falsification conditions, so that the central claim can be wrong (Sections 5–6). This is a position paper and dataset proposal, not an empirical report: it offers no trained model and no results, only a hypothesis sharp enough to be tested and a design for testing it.

2. Description is not perspective

The distinction the paper turns on is between a *description* of experience and a *perspective* on the world. It is easy to blur because both are made of words, often the same words, and a model fluent in one can appear fluent in the other.

A **description** is third-person and, in the limit, view-from-nowhere. It states what is the case: that an animal is hungry, that a room is dark, that a person feels afraid, that an obstacle lies to the east. Descriptions can be arbitrarily detailed and still be *about* an experience rather than *from* one. They favor the observer's stance: complete information, no indexical anchor, no situated limitation, time surveyed from outside.

A **perspective** is situated and indexical; it is most directly rendered in the first person, but its epistemic structure can also be carried in third-person form. It is organized around a point of view that has only part of the information, is located somewhere, wants something, and finds the world structured by that wanting and that location. *Here* and *there*, *now* and *before*, *I* and *you*, *this* obstacle and *that* goal are defined relative to the vantage. Crucially, a perspective is *partial and committed*: it perceives a slice, acts on it, and is corrected by what it did not see.

Three features distinguish the perspectival register, and each is a candidate training signal.

First, **indexicality**. The essential indexical cannot be reduced to any third-person description (Perry, 1979; Kaplan, 1989): the content "*I* am about to step off the ledge" guides action in a way that "the agent named A is about to step off the ledge" does not, even for an agent who knows it is A. A corpus that renders trajectories in the first person trains the binding of indexicals to a tracked viewpoint; a third-person corpus trains their interpretation as report.

Second, **situated limitation**. A perspective sees partially and acts anyway. The narrative of a point of view is full of occlusion, uncertainty, surprise, and the correction that follows from acting on incomplete information. Third-person description tends to supply the missing information for free (it knows what was around the corner) and so under-represents the structure of deciding under a bounded vantage.

Third, **agentive structure**: goal, obstacle, attempt, error, revision. Experience in the first person is organized as a small arc of trying and being answered by the world. This is the structure that a perspective-forming corpus would foreground and a descriptive corpus would flatten into reported fact.

It helps to see the contrast on one event, content held fixed.

Register	Same event, rendered
Third-person (description)	The forager could not see the food behind the leaf and turned away; the food was there all along.
First-person (perspective)	The scent is strong here, so it must be close; but I see only the leaf's edge, and nothing behind it. I turn away. (Later: it was there; the leaf had hidden it from where I stood.)

Both encode the same facts. Only the second is organized around a vantage that *did not have* the fact at the moment of acting, that used an indexical "here," and that is later corrected. The

hypothesis of this paper is that the difference between these two registers is a trainable signal for perspective, and that the difference is invisible to a model only if perspective is not, in fact, a separable capacity, which is exactly what Section 6 tests.

Nagel's point (1974), that objective description omits the point of view, motivates the *signal*, not a claim about presence: it shows only that perspective and description can come apart, which is what makes the training contrast meaningful. Whether anything is *like* something for the model is outside the paper's scope (Section 7).

3. What current corpora teach, and what they leave unclaimed

Each component of the proposal has a neighbor; the conjunction does not. This section places the proposal among existing work and isolates the residual gap. (A prior-art review for this project found no work assembling a pretraining-scale, first-person, limitation-structured, perspective-targeted corpus across human and non-human vantages; the closest neighbors are reviewed here.)

Synthetic narrative corpora. TinyStories shows that small models trained on simple synthetic narratives acquire fluent, coherent language (Eldan & Li, 2023), and the "textbooks" line shows that carefully synthesized data raises reasoning and competence efficiently (Gunasekar et al., 2023). These establish that *synthesizing* a corpus to shape a capability is viable: the methodological precedent the proposal relies on. But the genre they synthesize is the children's story or the textbook: third-person narration and expository explanation, optimized for fluency and reasoning, not a first-person register optimized for perspective. They are evidence that the method works, not instances of the target.

Personas as conditioning and diversity. Persona Hub scales synthetic data by conditioning generation on a billion personas, using them as *lenses for diversity* (Ge et al., 2024); the Anthology line conditions models on first-person backstories to elicit varied virtual respondents (Moon et al., 2024). These are close in surface (they involve first-person material) but their object is the *distribution* of outputs (coverage, variety, survey-like behavior), not the formation of a point of view that perceives partially and is corrected. A persona is a mask the model wears to vary what it says; a perspective, in our sense, is a structure that constrains what the agent in the narrative can know and do. The two can be told apart empirically (Section 6).

First person as camera, not interiority. Egocentric and embodied lines render the first person as a *viewpoint in space*: what a camera or agent sees and does from a located body (e.g., egocentric perception-action work). This is first-person geometry without first-person interiority: it captures *where* the vantage is, not the indexical, goal-structured, error-and-revision arc the proposal targets. It is a complementary signal, not the same one.

Experience as substrate. A growing line treats "experience" as a training substrate, but as *trajectories* rather than *narratives*: embodied-experience traces from simulators improve language models on reasoning about the physical world (Xiang et al., 2023), and synthetic-experience methods generate interaction data for reinforcement learning. These share the diagnosis that description is not enough, but they supply experience as state-action data, not as a first-person experiential text whose purpose is perspective. Section 4 takes up the strongest version of this view.

The two near-twins that must be named. Two existing first-person datasets sit close enough that the careless claim "this does not exist" would be false, and the paper depends on differentiating them. (i) A dataset of roughly 31,500 synthetic *first-person cognitive-action narratives* (700 each

across 45 cognitive actions) has been built for interpretability, to train linear probes on what models represent about cognition (Chulo & Joshi, 2025), with the very justification the proposal uses: that the first person captures internal structure the third person does not. It differs in *purpose* (probing representations, not forming perspective), *scale* (tens of thousands of snippets, not a pretraining-scale corpus), and *structure* (no non-human umwelten, no ordering by complexity of viewpoint). (ii) Synthetic first-person *self-reports* have been used as the entire dataset of a downstream predictor in mental-health NLP (Moëll & Sand Aronsson, 2026): first person as a labeled signal for a score, not as grounding for perspective. Both are first-person; neither trains, or aims to train, the functional capacity this paper targets.

The residual gap. What is unclaimed is the conjunction: a corpus (a) at a scale meant to *form* rather than *probe* a capacity, (b) in a first-person experiential register (indexical, partial, goal-and-error structured), (c) spanning human and non-human *umwelten* (von Uexküll, 1957; Nagel, 1974), (d) ordered by complexity of viewpoint, and (e) evaluated specifically for functional perspective rather than fluency, persona coverage, or a downstream label. No surveyed work assembles these; that conjunction, and its falsifiable test, is the contribution.

4. The interaction-only objection

The most serious rival is not another corpus but a rival *prescription*, and it shares the paper's diagnosis. The "era of experience" position holds that progress now requires agents that learn from their own streams of interaction with the world, because human-generated text is a record of *description* and is running out of the headroom that matters; on this view, what is missing is grounded experience, and the remedy is reinforcement learning from interaction, with static synthetic text explicitly deprecated (Silver & Sutton, 2025).

We accept the diagnosis and resist the dichotomy. Three points position the proposal.

First, the agreement is real and worth stating: description is not experience, and scaling description does not, by itself, supply what experience supplies. The present proposal is a *different inference* from the same premise: not "therefore interaction instead of text," but "therefore a different *register* of text," one that encodes the experiential structure (indexicality, limitation, error-and-revision) that ordinary description omits. Put sharply: interaction teaches by *causal consequence in the environment*, whereas situated experiential narrative teaches by *represented epistemic structure*: the shape of what a vantage knows, fails to know, and how it updates. The corpus is therefore not a competitor to interaction but a textual abstraction of the structure of perspective, learnable before or alongside it.

Second, the registers are not rivals so much as differently-priced. Real interaction is the stronger signal and the eventual one; it is also expensive, hard to control, unsafe to run across the full space of agents and situations one might want to cover, and impossible for vantages we cannot instantiate (an insect's, a person in a circumstance we cannot ethically reproduce). A synthetic first-person corpus is cheaper, controllable, auditable, and *coverable across umwelten*; and it is a plausible *intermediate* stage: a way to pretrain perspective-relevant structure before, or alongside, costly interaction. The claim is not that text equals experience; it is that the situated, limitation-marked experiential *register* of text is a distinct and useful signal that the interaction-only view discards by assumption.

Third, the dispute is empirical and the paper makes it so. If first-person experiential text carries no perspective signal beyond matched third-person text (Section 6), the interaction-only camp is

right that static text is the wrong tool and the proposal fails on its own terms. If it does carry such a signal, then experience-as-interaction and experience-as-register are complementary, and the cheaper one has a role. Either way the question is settled by measurement, not by which prescription sounds more grounded.

5. Dataset design

The proposal is a corpus and a matched control, specified by function so that several generation pipelines could realize them.

5.1 The experiential unit

The atom of the corpus is a short first-person narrative organized as a perspective, not a report. Each unit contains, explicitly:

- **a located, partial vantage**: the narrator perceives a slice of the scene, with occlusion and uncertainty marked;
- **a goal** held by the narrator;
- **an obstacle** or gap between vantage and goal;
- **an action** taken under that limited information;
- **an error or surprise**: the world answers in a way the vantage did not predict;
- **a revision**: the point of view updates, and the indexical frame (*here*, *now*, *what I know*) shifts.

Indexicals are used, not described: *here*, *now*, *this*, *I*, *you*, anchored to the tracked vantage. The unit may be paired with an optional **third-person reflection** that names what was learned, bridging the lived register and explicit knowledge. To avoid contaminating the register factor, such a reflection, if used at all, is attached *symmetrically* across all four cells of the 2×2 (Section 5.4), or held out entirely, never to the first-person cells alone; its contribution is then a deliberate ablation, not a confound.

5.2 Variation across umwelten

Perspective is trained by *varying the vantage*, so the corpus spans agent types whose worlds are structured differently, the *Umwelt* of each (von Uexküll, 1957): a human under ordinary constraint; a non-human animal whose salient world is chemical or thermal rather than visual; an insect with a radically partial field; an artificial agent with sensor limits; a collective with distributed, non-unified vantage; a body constrained in an unusual way. The manipulated variable is not *non-human-ness*, and still less *what it is like* to be any of these; by the very argument we cite, the bat's inner world is not ours to render (Nagel, 1974). It is the *structure* of the partial vantage, specified along axes we can set and verify: the **modality set** (which senses are available), the **field** (how partial the perceived slice is), the **unification** (a single vantage versus a divided or distributed one), and the **goal-and-error topology** (what the vantage pursues and how it is corrected). The non-human cases earn their place not as zoology but as carriers of structural settings humans rarely occupy: the bat's question recast as an engineering variable (how does a point of view organize a world along axes unlike ours) answered only at the level of structure a corpus can instantiate

and a check can confirm, never at the level of phenomenal acquaintance the firewall (Section 7) brackets. Diversity here is the training signal, not decoration; and *diversity* means variation of these structural parameters, not of surface vocabulary (Sections 5.5, 6.3).

5.3 Ordering and curriculum

Units can be ordered by *complexity of viewpoint*: from a single simple vantage with one goal, toward nested vantages (modeling another's point of view), divided or distributed vantages, and self-reference (a vantage that represents itself as one among others). Whether such ordering helps is itself a question the design can test; the corpus is built to support an ordered and a shuffled variant.

5.4 The design: register crossed with limitation

The control is the heart of the design, and a single first-versus-third-person contrast is not enough. Converting a first-person unit to third person does not, by itself, remove its epistemic structure: "I cannot see behind the leaf" and "the forager cannot see behind the leaf" both encode the same perceptual limitation. The grammatical *register* (voice) and the represented *limitation* (vantage) are separable, and a one-factor contrast would confound them: an effect attributed to first-person voice might in fact be an effect of represented limitation, which third person can carry too. The design therefore crosses the two factors in a 2×2:

	Limitation present	Limitation absent (omniscient)
First person	vantage *and* voice	voice without vantage
Third person	vantage without voice	neither

For every event, four content-matched renderings are generated (identical facts, token budget, and lexical difficulty) varying only register and whether the rendering marks the vantage's partiality (occlusion, uncertainty, surprise, correction) or supplies the missing information for free. Matching is verified, not assumed (Section 5.5).

The 2×2 is what lets the analysis separate *voice* from *vantage*. A **main effect of limitation** would say the operative signal is represented epistemic structure, available in either register; in which case the contribution is "epistemic limitation in text," not "first-person register." A **main effect of register** at matched limitation would say first-person voice adds something beyond limitation, most plausibly the binding of indexicals to a tracked viewpoint (Section 2). An **interaction** would say the two are not independent: first-person voice helps most when it also carries limitation. The proposal's strong form predicts both main effects; its weak form predicts at least the limitation effect; the deflationary null is that neither moves perspective once content is matched. This crossing, not a bare register contrast, is the experiment.

5.5 Construction, validation, and failure modes

A synthetic experiential corpus has characteristic risks, and the design treats them as construction constraints with explicit checks.

Anthropomorphic leakage in non-human units is the sharpest. Generators tend to narrate a beetle as a small human. The construction constraint is concrete: the generation prompt restricts the narrator's accessible predicates to its vantage's sense modalities (a chemically-oriented forager narrates gradients of scent and contact, not colors or human emotions) forbids human-interiority vocabulary, and requires the partiality to be structural, so that what the vantage cannot sense is simply *absent* from the narration rather than described as hidden. A filtering pass, by a held-out model and by human raters, flags units that smuggle in human interiority or knowledge the vantage could not have; those are regenerated or dropped. This filter is a *negative* check (it removes human-only predicates and impossible knowledge) and it is the most inspection can claim, because there is no ground-truth non-human rendering to validate against: by the argument we borrow (Nagel, 1974), what it is like to be the creature is exactly what we cannot supply, and we do not pretend to. What inspection cannot catch is the subtle failure: human perceptual organization, goal-salience, and error structure wearing a correct sensory mask, a *human in costume with the colors removed*. That failure is caught not by reading the units but comparatively, by the sensory-reskin control of Section 6.3: success is not that a unit *is* a beetle's world, but that varying the structural parameters of Section 5.2 transfers to perspective tasks where a mere reskin does not.

Confessional-style overfitting, learning a first-person *manner* (emotive, introspective) rather than perspective, is controlled by the 2×2 and by affect-matching. The third-person cells are not dry, encyclopedic descriptions; they carry the same affective and motivational load as their first-person counterparts, so the differential across register is the indexer and the marked vantage, not emotional tone. The evaluation (Section 6) further weights decision-relevant perspective over style, and the persona-mimicry controls penalize answers that first-person manner would get right by accident.

Pair equivalence is validated, not assumed. Each quartet is checked for bidirectional entailment of propositional content (a natural-language-inference model plus human spot-checks), matched token length and lexical difficulty, and matched affect; blinded raters confirm that the only systematic differences are register and limitation. Quartets that fail any check are excluded before training, and the exclusion rate is reported.

Recursive degradation, distributional collapse from training on model-generated text, is bounded by filtering generation, capping any single generator's share, and grounding the held-out evaluation in human-authored or human-verified items where possible. These constraints do not eliminate the risks; they make them auditable, and a corpus that fails them would test manner, not perspective.

6. Evaluation and predictions

The proposal is conceptual but its central claim is empirical: holding content fixed, the experiential register (voice, limitation, or their interaction) improves functional perspective and not general competence. The following is a proposed design, not a report of results.

6.1 Tasks

Evaluation targets the perspective *profile* (Section 1), not fluency or fact, and favors machine-verifiable items so that scoring does not itself depend on a model's stylistic preferences:

- **Deictic and pronoun binding**: resolve *here/there*, *I/you*, *this/that*, *before/after*, and ambiguous pronouns against a tracked vantage across a discourse, scored as forced-choice items

with a single correct binding.

- **Epistemic tracking / perspective-taking**: given a scene a particular agent perceives only in part, predict what that agent can know, want, and do, as distinct from the global facts.
- **Self/other distinction**: attribute beliefs, goals, and actions to the correct agent under paraphrase and misleading cues.
- **Agency attribution**: assign actions and their consequences to the agent whose vantage produced them.
- **Decision under partial observability**: natural-language POMDP-style items in which the correct action depends on the agent's limited vantage, not on the view-from-nowhere; a model that defaults to global information chooses wrongly. These are the load-bearing tasks, machine-scored against the optimal action under the stated vantage.
- **Perturbed false-belief / theory-of-mind**: using *perturbed* variants rather than canonical items, because LLM theory-of-mind results are fragile and sensitive to trivial alterations and the canonical items may be memorized (Kosinski, 2024; Ullman, 2023).
- **Persona-mimicry controls**: items on which first-person *style* yields the wrong answer, so the model cannot pass by sounding first-person.

A general factual-recall benchmark, on which register and limitation should make no difference, is included as the negative control for the dissociation (Section 6.3).

6.2 The minimal viable study

A first test need not pretrain from scratch. Generate the four content-matched arms of the 2×2 (Section 5.4), first/third person × limitation present/absent, over the same events, token count, and lexical difficulty, and fine-tune a single open-weight instruction-tuned model (7–8B range) on each arm, holding base model, optimizer, and steps fixed, across at least three seeds per arm. A v1 corpus on the order of 10^4–10^5 quartets (tens of thousands of events, each rendered in four cells) is a reasonable starting scale for fine-tuning at this model size, enough to install a learned disposition without committing to pretraining-scale generation; the figure is a starting estimate, to be fixed by a pilot power analysis. Evaluate every arm on the Section 6.1 profile plus the factual-recall negative control. Scoring favors machine-verifiable forced-choice items; where a free-text judgment is unavoidable, use an LLM-as-judge that is (i) blinded to arm, (ii) given a fixed rubric, (iii) calibrated against a human-scored subsample with reported agreement, and (iv) audited for style bias by re-scoring register-swapped paraphrases — or, for the confirmatory benchmark, human annotation. The question the minimal study answers: after content and difficulty are held equal, do *register*, *limitation*, or their interaction move the perspective profile, and do they leave factual recall untouched?

6.3 Predictions

H1 — Vantage (limitation) effect. Marking the vantage's partiality improves decision-under-partial-observability and epistemic-tracking tasks, in *either* register, over omniscient renderings. *Refuted if omniscient renderings do as well; the signal was not represented limitation.*

H2 — Voice (register) effect beyond vantage. At matched limitation, first-person renderings improve deictic/pronoun binding and self/other distinction over third-person

renderings, the gain attributable to indexical-to-viewpoint binding rather than to limitation. *Refuted if third-person-with-limitation matches first-person-with-limitation; then the contribution narrows to "epistemic limitation in text," not "first-person register."*

H3 — Dissociation (the signature). Whatever gains appear, appear on the perspective profile and **not** on general factual recall. *Refuted if gains track recall, indicating better data rather than a perspective-specific signal.*

H4 — Vantage-structure, not surface variety. Varying the *structural* parameters of the vantage (modality set, field, unification, goal-and-error topology; Section 5.2) improves generalization on agency and self/other tasks beyond a corpus restricted to ordinary human vantages, *and beyond a sensory-reskin control* that swaps surface modality vocabulary while holding human perceptual organization, goal-salience, and error structure fixed. *Refuted if the gain appears equally in the reskin control (the effect was input variety, not vantage-structure) or if structural diversity adds style without generalization at all.*

6.4 Falsification

The proposal is undermined, or rendered unnecessary, by any of:

1. **No effect of either factor:** neither register nor limitation beats the omniscient third-person cell on the perspective profile once content is matched (against H1 and H2).
2. **Vantage, not voice:** the limitation main effect is present but the register main effect is absent: first-person adds nothing beyond limitation. The proposal survives only in the weaker form "represented epistemic limitation is the signal," and the first-person claim specifically fails (against H2).
3. **No dissociation:** the gains track general factual recall (against H3).
4. **Mimicry explains it:** the advantage disappears under the persona-mimicry controls.
5. **No practical advantage:** building and matching the four-arm corpus costs more than equivalent gains from cheaper data or from interaction.

The signature that would *support* the strong proposal is narrow and hard to fake: a register main effect *and* a limitation main effect on decision-under-partial-observability, recall-independent, surviving persona-mimicry controls.

6.5 Ambiguous outcomes

Some results would refine rather than decide the claim. If the factors lift the perspective profile *and* recall equally, the help came via general data quality, not the targeted signal: informative but not the thesis. If first-person voice without limitation already captures the effect, the operative variable is indexical framing rather than partial vantage, and the design should re-center on it. If umwelt-diversity helps no more than the sensory-reskin control, the gain was input variety, not vantage-structure (Section 6.3), or adds style without generalization at all, diversity should be pruned to the human-plus-limitation core. Each outcome narrows the design rather than rescuing it post hoc.

7. Risks and limitations

The firewall, restated. The largest risk is misreading: that a corpus of first-person experience is a bid to give a model an inner life. It is not. Every metric in Section 6 is third-person and behavioral; none is evidence about whether anything is felt (Shanahan et al., 2023; Nagel, 1974). The paper measures whether the register trains a function, and is silent on presence. "Functional perspective" must not be read as, or marketed as, perspective in the phenomenal sense.

Anthropomorphism, in the data and in the reader. First-person text invites the projection of interiority both into the non-human units (Section 5.5) and into the trained model. The design resists the first with vantage-appropriate filtering; the paper resists the second by holding its claims at the level of measured capacity.

Text is not experience. The objection is correct and the paper concedes it: a narrative of a vantage is not the having of one. The defense is that the proposal never claims otherwise — it tests whether the *textual register* of experience is a distinct functional signal, which is a separate and answerable question.

Construct and dataset risks. "Perspective" may bundle separable capacities (deixis, ToM, agency attribution); the evaluation therefore reports a profile, not a single score, and the design can ablate components. Synthetic generation risks bias and shallow interiority; matched controls, human-grounded held-out evaluation, and persona-mimicry checks bound, without eliminating, these risks.

8. Conclusion

Language models are trained on a library written almost entirely *about* experience and almost never *from* a point of view. If perspective, situated, indexical, partial, corrigible, is the very thing that objective description leaves out, then no amount of description may train it well, and the missing register can be supplied on purpose. This paper proposes doing so: a corpus of first-person experiential narratives, structured by limitation and revision, varied across human and non-human umwelten, evaluated through a 2×2 that crosses register with epistemic limitation so that voice and vantage can be told apart rather than confounded.

The claim is deliberately modest in metaphysics and sharp in test. It is not that text confers experience, that the model feels, or that simulated interiority is real. It is that the experiential register (first-person voice, represented epistemic limitation, or their interaction) may carry a measurable, perspective-specific signal, and that the way to find out is a dissociation: register- or limitation-driven gains on point-of-view-dependent tasks without matching gains on general recall, surviving controls for style and mimicry. If that dissociation does not appear, the proposed corpus signal has not been shown to isolate perspective, whether because perspective is not a separable trainable signal or because scale, corpus, model, or benchmark fell short. If it does, then point of view is something a corpus can teach, and the cheapest place to begin teaching it is in the register we have been leaving out.

Acknowledgments and declarations

AI-assisted writing. The thesis and all substantive ideas in this paper are the author's own. Large language model tools were used, under the author's direction and continual revision, to assist with

drafting, editing, and translation; the author reviewed and takes full responsibility for the final text.

References

Chulo, I., & Joshi, A. (2025). Decomposing theory of mind: How emotional processing mediates ToM abilities in LLMs. *arXiv preprint arXiv:2511.15895*. https://arxiv.org/abs/2511.15895

Eldan, R., & Li, Y. (2023). TinyStories: How small can language models be and still speak coherent English? *arXiv preprint arXiv:2305.07759*. https://arxiv.org/abs/2305.07759

Ge, T., Chan, X., Wang, X., Yu, D., Mi, H., & Yu, D. (2024). Scaling synthetic data creation with 1,000,000,000 personas. *arXiv preprint arXiv:2406.20094*. https://arxiv.org/abs/2406.20094

Gunasekar, S., Zhang, Y., Aneja, J., Mendes, C. C. T., Del Giorno, A., Gopi, S., Javaheripi, M., Kauffmann, P., de Rosa, G., Saarikivi, O., Salim, A., Shah, S., Behl, H. S., Wang, X., Bubeck, S., Eldan, R., Kalai, A. T., Lee, Y. T., & Li, Y. (2023). Textbooks are all you need. *arXiv preprint arXiv:2306.11644*. https://arxiv.org/abs/2306.11644

Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). Oxford University Press.

Kosinski, M. (2024). Evaluating large language models in theory of mind tasks. *Proceedings of the National Academy of Sciences, 121*(45), e2405460121. https://doi.org/10.1073/pnas.2405460121

Moëll, B., & Sand Aronsson, F. (2026). High-accuracy prediction of mental health scores from English BERT embeddings trained on LLM-generated synthetic self-reports: A synthetic-only method development study. *Frontiers in Digital Health, 7*, 1694464. https://doi.org/10.3389/fdgth.2025.1694464

Moon, S., Abdulhai, M., Kang, M., Suh, J., Soedarmadji, W., Behar, E. K., Chan, D. M., & Canny, J. (2024). Virtual personas for language models via an anthology of backstories. *Proceedings of EMNLP 2024*. https://arxiv.org/abs/2407.06576

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*(4), 435–450. https://doi.org/10.2307/2183914

Perry, J. (1979). The problem of the essential indexical. *Noûs, 13*(1), 3–21. https://doi.org/10.2307/2214792

Shanahan, M., McDonell, K., & Reynolds, L. (2023). Role play with large language models. *Nature, 623*, 493–498. https://doi.org/10.1038/s41586-023-06647-8

Silver, D., & Sutton, R. S. (2025). Welcome to the era of experience. In *Designing an Intelligence* (forthcoming). MIT Press. Preprint: https://storage.googleapis.com/deepmind-media/Era-of-Experience%20/The%20Era%20of%20Experience%20Paper.pdf

Ullman, T. (2023). Large language models fail on trivial alterations to theory-of-mind tasks. *arXiv preprint arXiv:2302.08399*. https://arxiv.org/abs/2302.08399

von Uexküll, J. (1957). A stroll through the worlds of animals and men: A picture book of invisible worlds. In C. H. Schiller (Ed. & Trans.), *Instinctive Behavior: The Development of a Modern Concept* (pp. 5–80). International Universities Press. (Original work published 1934)

Xiang, J., Tao, T., Gu, Y., Shu, T., Wang, Z., Yang, Z., & Hu, Z. (2023). Language models meet world models: Embodied experiences enhance language models. *arXiv preprint arXiv:2305.10626*. https://arxiv.org/abs/2305.10626