

Em Algum Lugar, Não em Lugar Nenhum: Registro em Primeira Pessoa, Limitação Epistêmica e Formação de Perspectiva em Modelos de Linguagem

Rafael de Menezes Ehlers · ORCID [0009-0009-6476-6690](https://orcid.org/0009-0009-6476-6690)

Pesquisador independente

rafaehlers@gmail.com

Preprint. © 2026 Rafael de Menezes Ehlers. Licenciado sob [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Resumo

Corpora em escala web contêm material substancial em primeira pessoa, incluindo diários, fóruns, avaliações, autobiografias, encenação de papéis (role-play) e ficção. O que eles não fornecem é uma intervenção deliberadamente equilibrada em que o registro gramatical e a limitação epistêmica representada variem de forma independente. Este artigo pergunta se narrativas sintéticas construídas para esse contraste podem treinar melhores modelos funcionais de *perspectiva*: a organização situada, indexical e orientada à ação daquilo que um agente percebe, sabe, não sabe, tenta e revisa. Propomos um cruzamento 2×2 do registro (primeira versus terceira pessoa) com a limitação epistêmica (informação parcial versus onisciente), variado entre tipos de agente e *umwelten*. As células compartilham um esqueleto de eventos e fatos de mundo, mas a limitação não pode ser neutra quanto ao conteúdo: estado de crença, incerteza, surpresa e atualização são o conteúdo manipulado. O desenho, portanto, estima se a voz acrescenta algo além da supervisão do estado epistêmico, em vez de alegar correspondência proposicional perfeita. Duas adições testam a validade externa: um controle positivo de interação contendo trajetórias estado–ação–resultado com a mesma estrutura latente de tarefa, e uma avaliação de transferência de formato usando diagramas, ferramentas de observação privada e decisões POMDP não narrativas sem nenhuma prosa em primeira pessoa. A alegação é funcional, não fenomenal. Um resultado positivo mostraria que narrativas epistêmicas controladas treinam capacidades transferíveis relevantes à perspectiva; não mostraria que o texto fornece experiência ou presença.

Palavras-chave: corpora em primeira pessoa · tomada de perspectiva · ponto de vista · dados sintéticos · dêixis · teoria da mente · distinção eu/outro · Umwelt · modelos de linguagem

1. Introdução

Um modelo de linguagem é treinado sobre um corpus heterogêneo que contém tanto discurso em primeira quanto em terceira pessoa. A lacuna empírica não é a ausência de texto em primeira pessoa. É a ausência de uma distribuição de treino controlada que varie de forma independente a voz gramatical e a estrutura epistêmica, mantendo a família de eventos subjacente, o orçamento de dados e a dificuldade tão estáveis quanto possível.

Este artigo faz uma pergunta estreita: através de esqueletos de eventos equiparados, o registro carrega um sinal de treino distinto para a *perspectiva* e, em caso afirmativo, esse sinal provém da voz em primeira pessoa, da *limitação* epistêmica representada ou da interação entre elas? Por perspectiva entendemos um perfil de capacidades funcionais: resolução dêitica, rastreamento epistêmico, distinção eu/outro, atribuição de agência e ação sob observabilidade parcial. Um resultado positivo em um componente sustenta apenas aquele componente. A intuição original privilegia a voz em primeira pessoa; o desenho não. Ele admite o resultado de que a limitação importa enquanto a voz não.

A intuição motivadora é antiga e, acreditamos, subexplorada como hipótese de engenharia. Representações objetivas e situadas organizam a informação de modo diferente (Nagel, 1974), e os indexicais contribuem com um contexto orientador da ação que uma descrição correferente em terceira pessoa não precisa expressar na mesma forma (Perry, 1979; Kaplan, 1989). Isso motiva uma intervenção controlada, não uma alegação empírica de que os corpora ordinários carecem de discurso em primeira pessoa. A proposta é equilibrar deliberadamente o registro e o estado epistêmico e testar se algum deles move uma capacidade mensurável.

Somos explícitos quanto a um firewall, e o erguemos desde o início porque as palavras convidam à leitura equivocada. A proposta **não** faz alegação alguma de que o texto em primeira pessoa dê ao modelo experiência, de que o modelo *sinta* algo, de que seja consciente, ou de que uma interioridade simulada seja interioridade real. A linguagem em primeira pessoa vinda de um modelo não é evidência suficiente de um sujeito interior (Shanahan et al., 2023), e não é o que medimos. *Perspectiva*, aqui, é uma propriedade funcional, observável em terceira pessoa (acurácia de tomada de perspectiva, resolução dêitica, discriminação eu/outro, atribuição de agência, decisão dependente de ponto de vista), e toda alegação do artigo é sobre essa propriedade e sua treinabilidade, não sobre presença. Essas medidas não decidem a presença fenomenal: o sucesso não é suficiente para presença nem o fracasso é suficiente para ausência. O título contrasta a célula onisciente, o ponto de vista de *lugar nenhum*, com um ponto de vista localizado e parcial, de *algum lugar*. A própria gramática em terceira pessoa não é o ponto de vista de lugar nenhum, pois ela pode codificar um ponto de vantagem limitado. Perguntamos se treinar sobre texto que renderiza o ponto de vista de algum lugar ensina o que uma renderização onisciente não ensina, e se o ingrediente operante é a voz em primeira pessoa, a limitação representada, ou ambos.

Perspectiva funcional também não é sinônimo de individuação funcional. No programa mais amplo, a individuação requer linhagem, dependência de trajetória, propriedade, plasticidade limitada e clareza de ramificação. Este artigo testa uma capacidade componente relevante à perspectiva sem alegar o perfil longitudinal de cinco condições.

O artigo faz quatro contribuições. Ele operacionaliza a distinção descrição/perspectiva (Seção 2), isola a intervenção residual de corpus (Seção 3), compara a abstração narrativa com a interação (Seção 4) e especifica a equiparação de esqueletos de eventos, um controle positivo de interação, a avaliação de transferência de formato e as condições de falsificação (Seções 5–6). Este é um artigo de posição e uma proposta de conjunto de dados, não um relato empírico.

2. Descrição não é perspectiva

A distinção sobre a qual o artigo gira é entre uma *descrição* da experiência e uma *perspectiva* sobre o mundo. É fácil confundi-las porque ambas são feitas de palavras, muitas vezes as mesmas palavras,

e um modelo fluente em uma pode parecer fluente na outra.

Uma **descrição em terceira pessoa** é uma renderização gramatical, não uma condição epistêmica. Ela pode representar tanto um estado limitado de agente quanto um estado onisciente. O **ponto de vista de lugar nenhum** nomeia apenas a célula onisciente: informação completa, sem limitação situada, e o tempo examinado de fora. Descrições em terceira pessoa podem ser arbitrariamente detalhadas e ainda assim ser *sobre* uma experiência em vez de *a partir de* uma, mas não precisam fornecer informação global.

Uma **perspectiva** é situada e indexical; é renderizada de modo mais direto na primeira pessoa, mas sua estrutura epistêmica também pode ser carregada em forma de terceira pessoa. Ela é organizada em torno de um ponto de vista que possui apenas parte da informação, está localizado em algum lugar, quer algo, e encontra o mundo estruturado por esse querer e essa localização. *Aqui e ali, agora e antes, eu e você, este* obstáculo e *aquela* meta são definidos em relação ao ponto de vantagem.

Crucialmente, uma perspectiva é *parcial e comprometida*: ela percebe uma fatia, age sobre ela, e é corrigida por aquilo que não viu.

Três características distinguem o registro perspectivo, e cada uma é um sinal de treino candidato.

Primeiro, a **indexicalidade**. O indexical essencial não pode ser reduzido a nenhuma descrição em terceira pessoa (Perry, 1979; Kaplan, 1989): o conteúdo “*eu* estou prestes a sair da beira” orienta a ação de um modo que “o agente chamado A está prestes a sair da beira” não orienta, mesmo para um agente que sabe que é A. Um corpus que renderiza trajetórias em primeira pessoa treina a vinculação de indexicais a um ponto de vista rastreado; um corpus em terceira pessoa treina sua interpretação como relato.

Segundo, a **limitação situada**. Uma perspectiva vê parcialmente e age mesmo assim. A narrativa de um ponto de vista é cheia de oclusão, incerteza, surpresa e da correção que se segue de agir com informação incompleta. Uma renderização onisciente fornece a informação faltante de graça; uma renderização limitada em terceira pessoa não. O fator de limitação, e não a pessoa gramatical, testa a estrutura de decidir sob um ponto de vantagem restrito.

Terceiro, a **estrutura agentiva**: meta, obstáculo, tentativa, erro, revisão. A experiência em primeira pessoa organiza-se como um pequeno arco de tentar e ser respondido pelo mundo. Essa é a estrutura que um corpus formador de perspectiva colocaria em primeiro plano e que um corpus descritivo achataria em fato relatado.

Ajuda ver o contraste sobre um único evento, com o conteúdo mantido fixo.

Registro	Mesmo evento, renderizado
Terceira pessoa (ponto de vantagem limitado)	O forrageador não conseguia ver atrás da folha. Seguindo apenas o odor disponível a partir de sua posição, ele se afastou. Mais tarde soube que a folha havia escondido o alimento daquela posição.
Primeira pessoa (perspectiva)	O odor está forte aqui, então deve estar perto; mas eu vejo apenas a borda da folha, e nada atrás dela. Eu me afasto. (Mais tarde: estava ali; a folha o havia escondido de onde eu estava.)

Ambas codificam o mesmo estado de crença limitado e a mesma ação. A segunda usa adicionalmente indexicais em primeira pessoa e encena explicitamente a correção posterior. O 2×2

não infere uma vantagem geral de perspectiva a partir desse par apenas: ele testa separadamente se a limitação fornece o sinal relevante à decisão e se o registro em primeira pessoa acrescenta a vinculação indexical depois que a limitação é equiparada.

O ponto de Nagel (1974), de que um ponto de vista objetivo abstrai de um ponto de vista particular, motiva o *sinal*, não uma alegação sobre presença. O contraste experimental operacionaliza essa abstração como onisciência, e não como gramática em terceira pessoa. Se há algo que é *como* algo para o modelo está fora do escopo do artigo (Seção 7).

3. O que os corpora atuais ensinam, e o que deixam por reivindicar

Cada componente da proposta tem um vizinho; a conjunção não tem. Esta seção situa a proposta em meio ao trabalho existente e isola a lacuna residual. (Uma revisão dirigida do estado da técnica para este projeto não encontrou trabalho algum que reunisse um corpus em escala de pré-treino, em primeira pessoa, estruturado por limitação, direcionado à perspectiva, através de pontos de vantagem humanos e não humanos; os vizinhos mais próximos são revisados aqui. Esta é uma alegação delimitada de “até onde sabemos”, não um achado de revisão sistemática.)

Corpora narrativos sintéticos. O TinyStories mostra que modelos pequenos treinados sobre narrativas sintéticas simples adquirem uma linguagem fluente e coerente (Eldan & Li, 2023), e a linha dos “livros-texto” mostra que dados cuidadosamente sintetizados elevam o raciocínio e a competência de forma eficiente (Gunasekar et al., 2023). Esses trabalhos estabelecem que *sintetizar* um corpus para moldar uma capacidade é viável: o precedente metodológico do qual a proposta depende. Mas o gênero que sintetizam é a história infantil ou o livro-texto: narração em terceira pessoa e explanação expositiva, otimizadas para fluência e raciocínio, não um registro em primeira pessoa otimizado para a perspectiva. Eles são evidência de que o método funciona, não instâncias do alvo.

Personas como condicionamento e diversidade. O Persona Hub escala dados sintéticos condicionando a geração a um bilhão de personas, usando-as como *lentes para a diversidade* (Ge et al., 2024); a linha Anthology condiciona modelos a histórias de fundo em primeira pessoa para eliciar respondentes virtuais variados (Moon et al., 2024). Esses trabalhos são próximos na superfície (envolvem material em primeira pessoa), mas seu objeto é a *distribuição* de saídas (cobertura, variedade, comportamento tipo survey), não a formação de um ponto de vista que percebe parcialmente e é corrigido. Uma persona é uma máscara que o modelo veste para variar o que diz; uma perspectiva, em nosso sentido, é uma estrutura que restringe o que o agente na narrativa pode saber e fazer. As duas podem ser distinguidas empiricamente (Seção 6).

Primeira pessoa como câmera, não como interioridade. O trabalho egocêntrico e corporificado pode renderizar a primeira pessoa como um *ponto de vista no espaço*: aquilo que uma câmera ou agente percebe e faz a partir de um corpo localizado, como na navegação visualmente aterrada (*grounding*) a partir de entrada sensorial bruta (Mirowski et al., 2017). Isso é geometria em primeira pessoa sem uma alegação de interioridade em primeira pessoa: captura *onde* está o ponto de vantagem e que informação está disponível ali, não necessariamente o arco indexical, estruturado por metas, de erro e revisão, que a proposta visa. É um sinal complementar, não o mesmo.

Experiência como substrato. Uma linha crescente trata a “experiência” como um substrato de treino, mas como *trajetórias* em vez de *narrativas*: traços de experiência corporificada vindos de simuladores

melhoram os modelos de linguagem no raciocínio sobre o mundo físico (Xiang et al., 2023), e métodos de experiência sintética geram dados de interação para aprendizado por reforço. Esses trabalhos compartilham o diagnóstico de que a descrição não basta, mas fornecem a experiência como dados de estado-ação, não como um texto experiencial em primeira pessoa cujo propósito é a perspectiva. A Seção 4 retoma a versão mais forte dessa visão.

Os dois quase-gêmeos que precisam ser nomeados. Dois conjuntos de dados em primeira pessoa já existentes situam-se próximos o bastante para que a alegação descuidada de que “isto não existe” fosse falsa, e o artigo depende de diferenciá-los. (i) Um conjunto de dados de cerca de 31.500 *narrativas de ação cognitiva em primeira pessoa* sintéticas (700 para cada uma das 45 ações cognitivas) foi construído para interpretabilidade, a fim de treinar sondas (probes) lineares sobre o que os modelos representam a respeito da cognição (Chulo & Joshi, 2025), com a mesmíssima justificativa que a proposta usa: que a primeira pessoa captura estrutura interna que a terceira pessoa não captura. Ele difere em *propósito* (sondar representações, não formar perspectiva), em *escala* (dezenas de milhares de fragmentos, não um corpus em escala de pré-treino) e em *estrutura* (sem *umwelten* não humanos, sem ordenação por complexidade de ponto de vista). (ii) *Autorrelatos* sintéticos em primeira pessoa têm sido usados como o conjunto de dados inteiro de um preditor a jusante (downstream) em PLN de saúde mental (Moëll & Sand Aronsson, 2026): a primeira pessoa como um sinal rotulado para um escore, não como *grounding* para a perspectiva. Ambos são em primeira pessoa; nenhum treina, ou visa treinar, a capacidade funcional que este artigo tem como alvo.

A lacuna residual. O que resta por reivindicar é a conjunção: um corpus (a) em uma escala destinada a *formar* em vez de *sondar* uma capacidade, (b) em um registro experiencial em primeira pessoa (indexical, parcial, estruturado por meta e erro), (c) abrangendo *umwelten* humanos e não humanos (von Uexküll, 1957; Nagel, 1974), (d) ordenado por complexidade de ponto de vista, e (e) avaliado especificamente quanto à perspectiva funcional em vez de fluência, cobertura de personas ou um rótulo a jusante. Nenhum trabalho examinado reúne esses elementos; essa conjunção, e seu teste falsificável, é a contribuição.

4. A objeção da interação-apenas

O rival mais sério não é outro corpus, mas uma *prescrição* rival, e ela compartilha o diagnóstico do artigo. A posição da “era da experiência” sustenta que o progresso agora requer agentes que aprendam de seus próprios fluxos de interação com o mundo, porque o texto gerado por humanos é um registro de *descrição* e está esgotando a margem que importa; nessa visão, o que falta é experiência aterrada (grounded), e o remédio é o aprendizado por reforço a partir da interação, com o texto sintético estático explicitamente depreciado (Silver & Sutton, 2025).

Aceitamos o diagnóstico e resistimos à dicotomia. Três pontos posicionam a proposta.

Primeiro, a concordância é real e vale enunciar: descrição não é experiência, e escalar a descrição não fornece, por si só, aquilo que a experiência fornece. A presente proposta é uma *inferência diferente* a partir da mesma premissa: não “portanto, interação em vez de texto”, mas “portanto, texto com estrutura epistêmica controlada”, cujo registro pode acrescentar um sinal adicional. Dito de forma incisiva: a interação ensina por *consequência causal no ambiente*, ao passo que a narrativa situada ensina por *estrutura epistêmica representada*: a forma daquilo que um ponto de vantagem sabe, deixa de saber, e como se atualiza. O corpus não é, portanto, um competidor da interação, mas uma abstração textual da estrutura da perspectiva, aprendível antes ou ao lado dela.

Segundo, os registros não são tanto rivais quanto de preços diferentes. A interação real é o sinal mais forte e o eventual; é também caro, difícil de controlar, inseguro de executar através de todo o espaço de agentes e situações que se poderia querer cobrir, e impossível para pontos de vantagem que não podemos instanciar (o de um inseto, o de uma pessoa em uma circunstância que não podemos reproduzir eticamente). Um corpus sintético em primeira pessoa é mais barato, controlável, auditável e *cobrável através de umwelten*; e é um plausível estágio *intermediário*: um modo de pré-treinar estrutura relevante à perspectiva antes de, ou ao lado de, a interação custosa. A alegação não é que o texto seja igual à experiência; é que o *registro* experiencial do texto, situado e marcado por limitação, é um sinal distinto e útil que a visão da interação-apenas descarta por pressuposto. Terceiro, a disputa é empírica. Um efeito de registro ausente rejeita a alegação especificamente em primeira pessoa, mas pode deixar um efeito de limitação. Se nenhum dos fatores narrativos se transfere para além da prosa enquanto o braço de interação tem sucesso, a narrativa estática é a ferramenta errada para o alvo. Se a limitação ou o registro se transfere, a abstração narrativa e a interação podem ser complementares. A questão é decidida por medição.

5. Desenho do conjunto de dados

A proposta é um corpus e um controle equiparado, especificados por função de modo que vários pipelines de geração pudessem realizá-los.

5.1 A unidade experiencial

O átomo do corpus é uma curta narrativa em primeira pessoa organizada como uma perspectiva, não como um relato. Cada unidade contém, explicitamente:

- **um ponto de vantagem localizado e parcial**: o narrador percebe uma fatia da cena, com oclusão e incerteza marcadas;
- **uma meta** mantida pelo narrador;
- **um obstáculo** ou lacuna entre o ponto de vantagem e a meta;
- **uma ação** tomada sob essa informação limitada;
- **um erro ou surpresa**: o mundo responde de um modo que o ponto de vantagem não previu;
- **uma revisão**: o ponto de vista se atualiza, e o quadro indexical (*aqui, agora, o que eu sei*) se desloca.

Os indexicais são usados, não descritos: *aqui, agora, isto, eu, você*, ancorados ao ponto de vantagem rastreado. A unidade pode ser pareada com uma **reflexão opcional em terceira pessoa** que nomeia o que foi aprendido, fazendo a ponte entre o registro vivido e o conhecimento explícito. Para evitar contaminar o fator de registro, tal reflexão, se usada de todo, é anexada *simetricamente* a todas as quatro células do 2×2 (Seção 5.4), ou reservada por completo, nunca apenas às células em primeira pessoa; sua contribuição é então uma ablação deliberada, não um fator de confusão (confound).

5.2 Variação entre umwelten

A perspectiva é treinada *variando o ponto de vantagem*, de modo que o corpus abranja tipos de agente cujos mundos são estruturados de forma diferente, o *Umwelt* de cada um (von Uexküll, 1957): um humano sob restrição ordinária; um animal não humano cujo mundo saliente é químico ou térmico em vez de visual; um inseto com um campo radicalmente parcial; um agente artificial com limites de sensor; um coletivo com um ponto de vantagem distribuído e não unificado; um corpo restringido

de um modo incomum. A variável manipulada não é a *não-humanidade*, e menos ainda *como é* ser qualquer um desses; pelo próprio argumento que citamos, o mundo interior do morcego não é nosso para renderizar (Nagel, 1974). É a *estrutura* do ponto de vantagem parcial, especificada ao longo de eixos que podemos definir e verificar: o **conjunto de modalidades** (quais sentidos estão disponíveis), o **campo** (quão parcial é a fatia percebida), a **unificação** (um único ponto de vantagem versus um dividido ou distribuído), e a **topologia de meta-e-erro** (o que o ponto de vantagem persegue e como é corrigido). Os casos não humanos merecem seu lugar não como zoologia, mas como portadores de configurações estruturais que os humanos raramente ocupam: a pergunta do morcego reformulada como uma variável de engenharia (como um ponto de vista organiza um mundo ao longo de eixos diferentes dos nossos) respondida apenas ao nível da estrutura que um corpus pode instanciar e uma verificação pode confirmar, nunca ao nível do conhecimento fenomenal que o firewall (Seção 7) põe entre parênteses. A diversidade, aqui, é o sinal de treino, não decoração; e *diversidade* significa variação desses parâmetros estruturais, não do vocabulário de superfície (Seções 5.5, 6.3).

5.3 Ordenação e currículo

As unidades podem ser ordenadas por *complexidade de ponto de vista*: de um único ponto de vantagem simples com uma meta, em direção a pontos de vantagem aninhados (modelar o ponto de vista de outro), pontos de vantagem divididos ou distribuídos, e a autorreferência (um ponto de vantagem que se representa como um entre outros). Se tal ordenação ajuda é, por si só, uma questão que o desenho pode testar; o corpus é construído para suportar uma variante ordenada e uma embaralhada.

5.4 O desenho: registro cruzado com limitação

O controle é o cerne do desenho, e um único contraste primeira-versus-terceira pessoa não é suficiente. Converter uma unidade em primeira pessoa para a terceira pessoa não remove, por si só, sua estrutura epistêmica: “eu não consigo ver atrás da folha” e “o forrageador não consegue ver atrás da folha” codificam ambos a mesma limitação perceptiva. O *registro* gramatical (voz) e a *limitação* representada (ponto de vantagem) são separáveis, e um contraste de um fator os confundiria: um efeito atribuído à voz em primeira pessoa poderia, de fato, ser um efeito da limitação representada, que a terceira pessoa também pode carregar. O desenho, portanto, cruza os dois fatores em um 2×2:

	Limitação presente	Limitação ausente (onisciente)
Primeira pessoa	ponto de vantagem e voz	voz sem ponto de vantagem
Terceira pessoa	ponto de vantagem sem voz	nenhum dos dois

Para cada evento, quatro renderizações compartilham o mesmo mundo, agentes, metas, resultados, orçamento de tokens e dificuldade lexical. O fator de limitação necessariamente altera o estado de crença representado, a incerteza, o erro de predição e a atualização temporal; essas não são diferenças acessórias, mas a própria intervenção. Chamamos, portanto, o quarteto de **equiparado por esqueleto de eventos**, não proposicionalmente idêntico. O registro é equiparado dentro de cada nível de limitação, onde a equivalência proposicional é um alvo defensável.

O 2×2 é o que permite à análise separar *voz de ponto de vantagem*. Um **efeito principal da limitação** diria que o sinal operante é a estrutura epistêmica representada, disponível em qualquer registro; a contribuição torna-se então “limitação epistêmica no texto”, não “registro em primeira pessoa”. Um **efeito principal do registro** dentro de um nível de limitação diria que a voz em primeira pessoa acrescenta algo além do estado epistêmico representado. Uma **interação** diria que os dois não são independentes. O nulo deflacionário é que nenhum dos fatores move o perfil de perspectiva em relação aos controles equiparados por esqueleto de eventos.

5.5 Construção, validação e modos de falha

Um corpus experiencial sintético tem riscos característicos, e o desenho os trata como restrições de construção com verificações explícitas.

Vazamento antropomórfico nas unidades não humanas é o mais agudo. Os geradores tendem a narrar um besouro como um pequeno humano. A restrição de construção é concreta: o prompt de geração limita os predicados acessíveis ao narrador às modalidades sensoriais de seu ponto de vantagem (um forrageador de orientação química narra gradientes de odor e contato, não cores ou emoções humanas), proíbe o vocabulário de interioridade humana, e exige que a parcialidade seja estrutural, de modo que aquilo que o ponto de vantagem não pode perceber seja simplesmente *ausente* da narração em vez de descrito como oculto. Uma passagem de filtragem, por um modelo reservado (held-out) e por avaliadores humanos, sinaliza as unidades que contrabandeiam interioridade humana ou conhecimento que o ponto de vantagem não poderia ter; essas são regeradas ou descartadas. Esse filtro é uma verificação *negativa* (ele remove predicados exclusivamente humanos e conhecimento impossível) e é o máximo que a inspeção pode reivindicar, porque não há renderização não humana de referência (ground-truth) contra a qual validar: pelo argumento que tomamos emprestado (Nagel, 1974), como é ser a criatura é exatamente o que não podemos fornecer, e não fingimos poder. O que a inspeção não consegue apanhar é a falha sutil: organização perceptiva humana, saliência de metas e estrutura de erro humanas vestindo uma máscara sensorial correta, um *humano fantasiado com as cores removidas*. Essa falha é apanhada não pela leitura das unidades, mas comparativamente, pelo controle de reskin sensorial da Seção 6.3: o sucesso não é que uma unidade *seja* o mundo de um besouro, mas que variar os parâmetros estruturais da Seção 5.2 se transfira para tarefas de perspectiva onde um mero reskin não se transfere.

Sobreajuste (overfitting) de estilo confessional, aprender uma *maneira* em primeira pessoa (emotiva, introspectiva) em vez da perspectiva, é controlado pelo 2×2 e pela equiparação de afeto. As células em terceira pessoa não são descrições secas e enciclopédicas; elas carregam a mesma carga afetiva e motivacional que suas contrapartes em primeira pessoa, de modo que o diferencial entre registros é o indexador e o ponto de vantagem marcado, não o tom emocional. A avaliação (Seção 6) pondera adicionalmente a perspectiva relevante à decisão acima do estilo, e os controles de imitação de persona penalizam respostas que a maneira em primeira pessoa acertaria por acaso.

Equivalência de pares é validada no nível que o desenho permite. Dentro de um nível de limitação, as versões em primeira e terceira pessoa são verificadas quanto à implicação (entailment) bidirecional, comprimento em tokens, dificuldade lexical e afeto. Entre níveis de limitação, os validadores verificam os fatos de mundo comuns, metas, ações e resultados, ao mesmo tempo em que confirmam a diferença pretendida na informação acessível e na transição de estado de crença.

Avaliadores cegados relatam diferenças residuais em vez de certificar uma identidade proposicional impossível.

Degradação recursiva, o colapso distribucional decorrente de treinar sobre texto gerado por modelos, é limitada filtrando a geração, limitando a fatia de qualquer gerador isolado e aterrando (grounding) a avaliação reservada (held-out) em itens de autoria humana ou verificados por humanos onde for possível. Essas restrições não eliminam os riscos; elas os tornam auditáveis, e um corpus que falhe nelas testaria maneira, não perspectiva.

5.6 Controle positivo de interação

As quatro células narrativas são acompanhadas por uma quinta condição de treino construída a partir de trajetórias estado–ação–observação–resultado geradas em um ambiente interativo. Ela usa as mesmas famílias latentes de eventos, fronteiras de observação, metas e consequências de ação, mas as apresenta como trajetórias estruturadas em vez de experiência narrada. Essa condição estima quanto da capacidade-alvo pode ser aprendido a partir da abstração epistêmica textual em relação à interação causal aterrada (grounded).

O braço de interação é um controle positivo, não um rival que precise perder. Se ele supera todas as células narrativas, o resultado quantifica o preço da abstração. Se as narrativas marcadas por limitação se aproximam dele e se transferem para além da prosa, elas são um sinal útil de menor custo. Se apenas a interação funciona, a proposta de corpus falha em seu papel pretendido.

6. Avaliação e predições

A proposta é conceitual, mas sua alegação central é empírica: através de esqueletos de eventos equiparados, a voz, a limitação representada, ou a interação entre elas pode melhorar a perspectiva funcional sem um ganho geral de competência. O que segue é um desenho proposto, não um relato de resultados.

6.1 Tarefas

A avaliação tem como alvo o *perfil* de perspectiva (Seção 1), não a fluência ou o fato, e favorece itens verificáveis por máquina de modo que a pontuação não dependa, ela mesma, das preferências estilísticas de um modelo:

- **Vinculação dêitica e pronominal:** resolver *aqui/ali, eu/você, isto/aquilo, antes/depois* e pronomes ambíguos contra um ponto de vantagem rastreado ao longo de um discurso, pontuado como itens de escolha forçada com uma única vinculação correta.
- **Rastreamento epistêmico / tomada de perspectiva:** dada uma cena que um agente particular percebe apenas em parte, predizer o que aquele agente pode saber, querer e fazer, distinto dos fatos globais.
- **Distinção eu/outro:** atribuir crenças, metas e ações ao agente correto sob paráfrase e pistas enganosas.
- **Atribuição de agência:** atribuir ações e suas consequências ao agente cujo ponto de vantagem as produziu.
- **Decisão sob observabilidade parcial:** itens em linguagem natural de estilo POMDP em que a ação correta depende do ponto de vantagem limitado do agente, não do ponto de vista de lugar

nenhum; um modelo que recorre por padrão à informação global escolhe erradamente. Essas são as tarefas de sustentação (load-bearing), pontuadas por máquina contra a ação ótima sob o ponto de vantagem enunciado.

- **Falsa-crença / teoria da mente perturbadas:** usando variantes *perturbadas* em vez de itens canônicos, porque os resultados de teoria da mente dos LLMs são frágeis e sensíveis a alterações triviais e os itens canônicos podem estar memorizados (Kosinski, 2024; Ullman, 2023).
- **Controles de imitação de persona:** itens nos quais o *estilo* em primeira pessoa produz a resposta errada, de modo que o modelo não possa passar por soar em primeira pessoa.
- **Tarefas de transferência de formato:** rastreamento de crenças diagramático, chamadas de ferramenta de observação privada, POMDPs em grade ou tabela e decisões estruturadas estado-ação que não contêm nenhuma prosa em primeira pessoa. Essas testam se a disposição aprendida se transfere para além do registro narrativo.

Um benchmark geral de recuperação factual, no qual o registro e a limitação não deveriam fazer diferença alguma, é incluído como o controle negativo. A condição de interação é o controle positivo, e a suíte de transferência de formato sem primeira pessoa é o principal teste de validade externa.

6.2 O estudo mínimo viável

Um primeiro teste não precisa pré-treinar do zero. Gere os quatro braços do 2×2 equiparados por esqueleto de eventos e um braço de interação estruturado sobre as mesmas famílias latentes de eventos. Faça o ajuste fino (fine-tuning) de um modelo de pesos abertos ajustado por instrução (na faixa de 7–8B) por braço, mantendo o otimizador, os passos e o orçamento efetivo de dados fixos através de ao menos três sementes (seeds). Um corpus v1 da ordem de 10^4 – 10^5 famílias de eventos é uma estimativa inicial a ser fixada por uma análise de poder do piloto. Avalie cada braço no perfil narrativo, no controle negativo de recuperação factual e na suíte de transferência de formato sem prosa. A pontuação favorece itens verificáveis por máquina; o julgamento em texto livre é cegado, vinculado a rubrica, calibrado contra humanos e auditado quanto a viés de estilo. O estudo mínimo pergunta se o registro, a limitação ou a interação entre eles se transferem para além da forma narrativa e como seu efeito se compara com a interação.

6.3 Predições

H1 — Efeito do ponto de vantagem (limitação). Marcar a parcialidade do ponto de vantagem melhora as tarefas de decisão-sob-observabilidade-parcial e de rastreamento epistêmico, em qualquer registro, sobre as renderizações oniscientes. *Refutada se as renderizações oniscientes forem igualmente boas; o sinal não era a limitação representada.*

H2 — Efeito da voz (registro) além do ponto de vantagem. Com limitação equiparada, as renderizações em primeira pessoa melhoram a vinculação dêitica / pronominal e a distinção eu/outro sobre as renderizações em terceira pessoa, e o ganho generaliza para itens de indexical-novo e de registro-trocado que preservam a tarefa enquanto mudam a voz de superfície. *Refutada se a terceira-pessoa-com-limitação igualar a primeira-pessoa-com-limitação, ou se o efeito desaparecer sob a troca de registro; então a contribuição se estreita para “limitação epistêmica no texto”, não “registro em primeira pessoa”. A transferência sem prosa é um teste opcional mais forte para a voz, porque remover a linguagem também remove os indexicais manipulados.*

H3 — Dissociação (a assinatura). Quaisquer ganhos que apareçam, aparecem no perfil de perspectiva e **não** na recuperação factual geral. *Refutada se os ganhos acompanharem a recuperação, indicando dados melhores em vez de um sinal específico de perspectiva.*

H4 — Estrutura do ponto de vantagem, não variedade de superfície. Variar os parâmetros estruturais do ponto de vantagem (conjunto de modalidades, campo, unificação, topologia de meta-e-erro; Seção 5.2) melhora a generalização em tarefas de agência e eu / outro para além de um corpus restrito a pontos de vantagem humanos ordinários, e para além de um controle de reskin sensorial que troca o vocabulário de modalidade de superfície enquanto mantém fixas a organização perceptiva humana, a saliência de metas e a estrutura de erro. *Refutada se o ganho aparecer igualmente no controle de reskin (o efeito foi variedade de entrada, não estrutura do ponto de vantagem) ou se a diversidade estrutural acrescentar estilo sem generalização alguma.*

H5 — Abstração em relação à interação. O braço de interação fornece uma estimativa superior ou de controle positivo para aprender a mesma estrutura latente. As narrativas marcadas por limitação sustentam a alegação ampla de perspectiva somente se melhorarem as tarefas de transferência de formato sem prosa acima dos controles oniscientes. *Se os ganhos de limitação permanecerem confinados à linguagem narrativa, o corpus terá treinado formato em vez de uma capacidade de perspectiva transferível. A voz é julgada principalmente pela transferência de indexical-novo e de registro-trocado em H2; a transferência de voz sem prosa, se observada, é mais forte, mas não obrigatória.*

6.4 Falsificação

A proposta é solapada, ou tornada desnecessária, por qualquer um dos seguintes:

1. **Nenhum efeito de qualquer fator:** dentro do 2×2, a limitação não consegue melhorar o perfil de perspectiva com registro equiparado, a voz não consegue melhorá-lo com limitação equiparada, e nenhuma interação aparece (contra H1 e H2). A célula onisciente em terceira pessoa isoladamente não é o comparador limpo para ambas as hipóteses.
2. **Ponto de vantagem, não voz:** a limitação melhora o perfil, mas a voz em primeira pessoa não acrescenta nada nos contrastes equiparados por limitação e não mostra interação confiável. A proposta sobrevive apenas na forma mais fraca “a limitação epistêmica representada é o sinal”, e a alegação em primeira pessoa especificamente falha (contra H2).
3. **Nenhuma dissociação:** os ganhos acompanham a recuperação factual geral (contra H3).
4. **A imitação explica:** a vantagem desaparece sob os controles de imitação de persona.
5. **Nenhuma transferência de formato para a limitação:** os ganhos de limitação desaparecem em diagramas, ferramentas e tarefas estruturadas de observabilidade parcial sem nenhuma prosa em primeira pessoa, rejeitando a alegação ampla de perspectiva transferível.
6. **Nenhuma generalização da voz:** uma vantagem de voz desaparece em tarefas de linguagem de indexical-novo ou de registro-trocado, rejeitando o mecanismo de registro mesmo que a prosa em distribuição melhore.
7. **Resultado de interação-apanas:** o controle positivo de interação melhora enquanto todos os braços narrativos permanecem no nulo.
8. **Nenhuma vantagem prática:** construir e equiparar o corpus de quatro braços custa mais do que ganhos equivalentes de dados mais baratos ou de interação.

O resultado mais forte é um efeito de registro que sobrevive aos testes de linguagem de indexical-novo e de registro-trocado somado a um efeito de limitação que sobrevive à transferência sem prosa, ambos independentes da recuperação. Um resultado de transferência apenas-de-limitação continua útil, mas muda a contribuição de registro em primeira pessoa para supervisão de estado epistêmico.

6.5 Desfechos ambíguos

Alguns resultados refinariam em vez de decidir a alegação. Se os fatores elevarem o perfil de perspectiva *e* a recuperação igualmente, a ajuda veio via qualidade geral dos dados, não via o sinal direcionado: informativo, mas não a tese. Se a voz em primeira pessoa sem limitação já captura o efeito, a variável operante é o enquadramento indexical em vez do ponto de vantagem parcial, e o desenho deveria recentrar-se nela. Se a diversidade de *umwelt* não ajudar mais do que o controle de *reskin* sensorial, o ganho foi variedade de entrada, não estrutura do ponto de vantagem (Seção 6.3), ou acrescenta estilo sem generalização alguma, e a diversidade deveria ser podada até o núcleo humano-mais-limitação. Cada desfecho estreita o desenho em vez de resgatá-lo *post hoc*.

7. Riscos e limitações

O firewall, reafirmado. O maior risco é a leitura equivocada: de que um corpus de experiência em primeira pessoa é uma tentativa de dar ao modelo uma vida interior. Não é. Toda métrica da Seção 6 é em terceira pessoa e comportamental. Essas medidas não decidem a presença fenomenal: o sucesso não é suficiente para presença nem o fracasso é suficiente para ausência (Shanahan et al., 2023; Nagel, 1974). O artigo mede se a intervenção treina uma função. “Perspectiva funcional” não deve ser lida como, ou comercializada como, perspectiva no sentido fenomenal.

Antropomorfismo, nos dados e no leitor. O texto em primeira pessoa convida à projeção de interioridade tanto nas unidades não humanas (Seção 5.5) quanto no modelo treinado. O desenho resiste ao primeiro com filtragem apropriada ao ponto de vantagem; o artigo resiste ao segundo mantendo suas alegações no nível da capacidade medida.

Texto não é experiência. A objeção está correta e o artigo a concede: uma narrativa de um ponto de vantagem não é o ter de um. A defesa é que a proposta nunca alega o contrário — ela testa se o *registro textual* da experiência é um sinal funcional distinto, o que é uma questão separada e respondível.

Riscos de construto e de conjunto de dados. “Perspectiva” pode agrupar capacidades separáveis (dêixis, teoria da mente, atribuição de agência); a avaliação, portanto, relata um perfil, não um escore único, e o desenho pode ablacionar componentes. A geração sintética arrisca viés e interioridade rasa; controles equiparados, avaliação reservada (*held-out*) aterrada em humanos e verificações de imitação de persona limitam, sem eliminar, esses riscos.

8. Conclusão

Os modelos de linguagem já encontram um extenso discurso em primeira pessoa, mas não como uma intervenção equilibrada que separe a voz da limitação epistêmica. Este artigo propõe tal intervenção: narrativas equiparadas por esqueleto de eventos, variadas através de *umwelten*

humanos e não humanos, avaliadas por meio de um 2×2 que cruza o registro com a limitação, de um controle positivo de interação e de tarefas de transferência sem prosa em primeira pessoa.

A alegação é deliberadamente modesta em metafísica e incisiva em teste. O texto não se torna, com isso, experiência. Um resultado positivo mostraria que a voz, a limitação epistêmica representada, ou a interação entre elas fornece supervisão transferível para capacidades funcionais específicas. Se os ganhos permanecerem lexicais, acompanhar a recuperação geral, ou desaparecerem fora da prosa enquanto a interação tem sucesso, a proposta de corpus falha. Se as narrativas marcadas por limitação se transferem, elas fornecem uma abstração controlada e potencialmente mais barata da estrutura epistêmica; apenas um efeito adicional de registro vindicaria a alegação especificamente em primeira pessoa.

Agradecimentos e declarações

Escrita assistida por IA. A tese e todas as ideias substantivas deste artigo são de autoria do próprio autor. Ferramentas de grandes modelos de linguagem foram usadas, sob a direção e a revisão contínua do autor, para auxiliar na redação, na edição e na tradução; o autor revisou e assume total responsabilidade pelo texto final.

Referências

- 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>
- Mirowski, P., Pascanu, R., Viola, F., Soyer, H., Ballard, A. J., Banino, A., Denil, M., Goroshin, R., Sifre, L., Kavukcuoglu, K., Kumaran, D., & Hadsell, R. (2017). Learning to navigate in complex environments. *arXiv preprint arXiv:1611.03673*. <https://arxiv.org/abs/1611.03673>
- 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 G. Konidaris (Ed.), *Designing an Intelligence* (in press). 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>