Quando as Máquinas "Pensam" Antropomorfização no Discurso sobre IA

Riscos Conceptuais e Proposta de Léxico Não Antropomórfico para PLN em Português

Anabela Barreiro
Linguateca Lisboa
anabela.barreiro.papers@gmail.com

Abstract

A antropomorfização de sistemas de Inteligência Artificial tornou-se particularmente relevante em contextos de Processamento de Linguagem Natural em português, onde expressões como "*o modelo compreende*" ou "*o sistema alucina*" podem gerar equívocos conceptuais, contribuindo para uma percepção errada das capacidades dos modelos. Este artigo propõe um enquadramento terminológico para descrever sistemas de Processamento de Linguagem Natural em português sem recurso a metáforas antropomórficas, apresentando um conjunto de reformulações linguísticas destinadas a melhorar a precisão conceptual e a literacia em Inteligência Artificial.

Palavras-chave: Processamento de Linguagem Natural, Modelos de Linguagem de Grande Escala, Expressões Antropomórficas, Inteligência Artificial, Compreensão Linguística, Terminologia Funcional, Educação e Literacia em Inteligência Artificial, Comunicação Científica

1 Introdução

O rápido desenvolvimento de modelos de linguagem de grande escala (LLMs) tem ampliado significativamente o papel da Inteligência Artificial (IA) na investigação, no ensino e no discurso público. Sistemas de Processamento de Linguagem Natural (PLN) são cada vez mais descritos em termos acessíveis e metafóricos, tanto em contextos académicos como na comunicação mediática. Expressões antropomórficas associadas a predicados verbais como *compreender*, *raciocinar* ou *decidir* (e.g., *o modelo compreende*, *o sistema decide* ou *a IA alucina*) tornaram-se comuns. Embora possam funcionar como simplificações pedagógicas, estas formulações introduzem frequentemente ambiguidade conceptual e podem levar a interpretações incorretas das capacidades reais dos sistemas, alimentando uma percepção inflacionada da sua inteligência ou compreensão, sugerindo implicitamente que estes sistemas possuem formas de compreensão ou raciocínio análogas às humanas.

A preocupação com a antropomorfização da IA não é recente. Desde a década de 1970, Weizenbaum (1976) alertava para os riscos de atribuir qualidades humanas a programas computacionais. Estudos filosóficos subsequentes reforçam a distinção entre simulação de comportamento inteligente e compreensão genuína (Searle, 1980; Haugeland, 1985; Fodor, 1983; Putnam, 1975), evidenciando que sistemas estatísticos podem gerar respostas linguísticas coerentes sem modelar verdadeiramente a semântica nem instanciar experiências conscientes (Bender et al., 2021; Mitchell, 2019; Marcus and Davis, 2019; Goldberg, 2019). Neste contexto, Lerchner (2026) enfatiza que simular comportamentos cognitivos não equivale a instanciar processos conscientes, reforçando a necessidade de precisão terminológica.

Apesar da relevância destes debates, a investigação sobre o uso de linguagem antropomórfica em contextos educativos e científicos em português permanece limitada. Para preencher esta lacuna, realizámos um levantamento exploratório de expressões antropomórficas em artigos recentes de PLN, documentação técnica e materiais pedagógicos. As ocorrências foram identificadas manualmente e categorizadas segundo o tipo de atribuição cognitiva implícita, por exemplo, compreensão, raciocínio ou tomada de decisão. Com base nesta análise, propomos um quadro terminológico que reformula descrições antropomórficas em descrições funcionais e computacionais.

Este trabalho procura responder à seguinte questão de investigação: *de que forma o uso de linguagem antropomórfica no discurso sobre IA e modelos de linguagem influencia a interpretação das capacidades desses sistemas, e como pode um léxico funcional alternativo contribuir para maior precisão conceptual na descrição de processos de*

Proceedings of the 17th International Conference on Computational Processing of Portuguese (PROPOR 2026)
Vol. 1, pages 1027–1031

PLN em português?

O trabalho apresenta três contribuições principais. Primeiro, discute criticamente o uso de linguagem antropomórfica no discurso sobre IA e modelos de linguagem, articulando perspectivas da filosofia da mente, da ética da IA e da linguística cognitiva. Segundo, apresenta uma observação empírica exploratória do uso de expressões antropomórficas em textos de IA e PLN em português, identificando padrões recorrentes de atribuição de propriedades cognitivas a sistemas computacionais. Terceiro, propõe um léxico funcional não antropomórfico para descrever processos típicos de PLN em português, acompanhado de exemplos ilustrativos destinados a promover maior rigor conceptual, literacia crítica em IA e precisão terminológica na comunicação científica e pedagógica.

O artigo está organizado da seguinte forma. A Secção 2 sumariza trabalhos importantes na área da antropomorfização e metáforas no discurso sobre IA. A Secção 3 descreve a abordagem metodológica adoptada neste estudo. A Secção 4 apresenta o uso de linguagem antropomórfica em textos de PLN em português e discute os impactos pedagógicos e epistemológicos desse fenómeno. A Secção 5 apresenta a proposta de um léxico não antropomórfico acompanhado de exemplos ilustrativos. Finalmente, a Secção 6 sintetiza as principais conclusões e aponta direcções para trabalho futuro.

2 Literatura Relevante

A tendência para descrever sistemas computacionais em termos antropomórficos tem sido discutida desde as primeiras décadas da investigação em Inteligência Artificial. Weizenbaum (1976) alertou para os riscos epistemológicos e sociais de atribuir qualidades humanas a programas computacionais. No domínio da filosofia da mente, o argumento da *Chinese Room* de Searle (1980) reforçou a distinção entre manipulação simbólica e compreensão genuína, mostrando que um sistema pode produzir respostas linguísticas apropriadas sem possuir entendimento semântico.

Debates contemporâneos sobre IA e modelos de linguagem mantêm esta distinção entre comportamento observável e processos cognitivos reais. Bender et al. (2021) argumentam que LLMs reproduzem padrões estatísticos presentes nos dados de treino, não constituindo sistemas capazes de compreender linguagem no sentido humano. De forma semelhante, Mitchell (2019) e Marcus and Davis (2019) sublinham que muitos sistemas descritos como "inteligentes" operam essencialmente através de correlações estatísticas. Discussões recentes sobre consciência artificial reforçam igualmente a diferença entre simulação comportamental e instância de processos cognitivos (Lerchner, 2026).

Do ponto de vista da linguística cognitiva, as metáforas moldam a interpretação conceptual dos fenómenos descritos (Lakoff and Johnson, 1980). Assim, quando sistemas de IA são sistematicamente descritos em termos humanos, torna-se mais provável que lhes sejam atribuídas propriedades como compreensão ou intencionalidade (Birhane, 2021). Apesar da existência desta literatura internacional, o impacto da antropomorfização no discurso técnico e pedagógico sobre PLN em português permanece ainda pouco estudado.

3 Metodologia

Embora o presente trabalho inclua uma observação exploratória de exemplos linguísticos, o seu objetivo principal é conceptual e terminológico. Assim, em vez de realizar uma avaliação empírica quantitativa, o artigo procura analisar criticamente o papel da linguagem na descrição de sistemas de IA e propor um enquadramento terminológico alternativo para o contexto do português. Este tipo de contribuição é complementar a estudos empíricos sobre modelos de linguagem, focando-se sobretudo na precisão conceptual e na comunicação científica.

3.1 Levantamento e Análise das Expressões Antropomórficas

Realizámos um levantamento exploratório de expressões antropomórficas na literatura recente de PLN e em materiais educativos relacionados com tecnologias de linguagem para português. Foram examinados três tipos de fontes frequentemente utilizados por investigadores, estudantes e profissionais destas áreas: textos recentes e amplamente acessíveis em contextos académicos e educativos tais como artigos científicos e pré-publicações sobre PLN, documentação técnica associada a bibliotecas ou modelos de linguagem, e textos de divulgação científica sobre IA publicados em português e encontrados na internet. O objetivo desta análise foi identificar padrões recorrentes de atribuição de propriedades cognitivas a sistemas computacionais.

As ocorrências identificadas foram classificadas segundo o tipo de atribuição cognitiva implícita, in-

cluindo categorias como compreensão, raciocínio e tomada de decisão. Esta classificação permitiu observar de que modo tais expressões são utilizadas para descrever processos essencialmente computacionais.

3.2 Critérios de identificação

A análise consistiu na identificação manual de expressões linguísticas que atribuem propriedades cognitivas ou agentivas a sistemas computacionais. Foram considerados exemplos de linguagem antropomórfica enunciados que sugerem processos tipicamente associados à cognição humana, incluindo expressões relacionadas com **compreensão** (*o modelo compreende*, *o sistema entende*), **aprendizagem** ou **conhecimento** (*o modelo aprende*, *o sistema sabe*), **decisão** ou **raciocínio** (*o modelo decide*, *a IA raciocina*), bem como **experiência** ou **percepção** (*a IA alucina*, *o modelo lembra-se*). Estas expressões foram registadas como exemplos ilustrativos de antropomorfização sempre que eram utilizadas para descrever diretamente o funcionamento ou o comportamento de sistemas de PLN.

Mesmo nesta análise exploratória, verificou-se uma presença recorrente destas formulações em diferentes tipos de textos, incluindo documentação técnica e materiais pedagógicos. Em muitos casos, expressões antropomórficas aparecem como simplificações explicativas destinadas a tornar os sistemas mais intuitivos para os leitores. Contudo, tais formulações também podem contribuir para interpretações equivocadas sobre a natureza estatística dos modelos de linguagem, reforçando a tendência para lhes atribuir capacidades cognitivas humanas.

3.3 Limitações

Esta observação preliminar apresenta várias limitações. Em particular, a seleção das fontes não segue ainda um protocolo sistemático de amostragem, e a identificação das expressões foi realizada manualmente sem recurso a ferramentas de anotação ou análise automática de corpus.

Estudos futuros poderão desenvolver esta análise através da construção de corpora especializados de textos de IA e PLN em português, permitindo medir quantitativamente a frequência, distribuição e evolução temporal de expressões antropomórficas em diferentes géneros discursivos.

4 Impacto da antropomorfização

A linguagem antropomórfica pode influenciar significativamente a forma como sistemas de IA são interpretados por estudantes, investigadores e utilizadores. Ao atribuir capacidades como compreensão, raciocínio ou decisão a sistemas estatísticos, cria-se frequentemente uma discrepância entre o funcionamento real dos modelos e a percepção das suas capacidades.

Este fenómeno tem três consequências principais. Primeiro, pode produzir uma **compreensão distorcida da IA**, na qual processos probabilísticos são interpretados como formas de cognição genuína (Talbot, 2019). Segundo, pode gerar **confiança excessiva nas saídas dos sistemas**, fenómeno frequentemente descrito como viés de automatização, levando utilizadores a aceitar respostas sem verificação crítica (Brennen and Kreiss, 2020; Zhou, 2022). Terceiro, a antropomorfização pode criar **ambiguidade na atribuição de responsabilidade**, sugerindo implicitamente que decisões problemáticas são tomadas pela máquina e não pelos sistemas e processos humanos que a configuram (Coeckelbergh, 2020; Amershi et al., 2019).

Deste modo, a escolha terminológica na descrição de sistemas de PLN não é neutra. O uso de descrições funcionais pode contribuir para alinhar o discurso científico com as capacidades reais dos modelos e promover uma literacia crítica em IA.

A Figura 1 ilustra expressões antropomórficas que podem induzir a atribuição de capacidades cognitivas a sistemas estatísticos, conduzindo a interpretações equivocadas das suas capacidades. O esquema ilustra também como o uso de um léxico funcional não antropomórfico pode reduzir essa discrepância conceptual.

5 Proposta de léxico não antropomórfico para PLN em português

5.1 Processamento linguístico

Em tarefas frequentemente descritas como compreensão de texto, recomenda-se evitar formulações como *o modelo compreende a pergunta*. Uma descrição mais precisa consiste em afirmar que o sistema *identifica padrões linguísticos na entrada e gera uma resposta consistente com padrões observados nos dados de treino*. Por exemplo, perante a pergunta *Qual é a capital de Portugal?*, o modelo consegue gerar a resposta *Lisboa*, refletindo regularidades estatísticas presentes nos da-

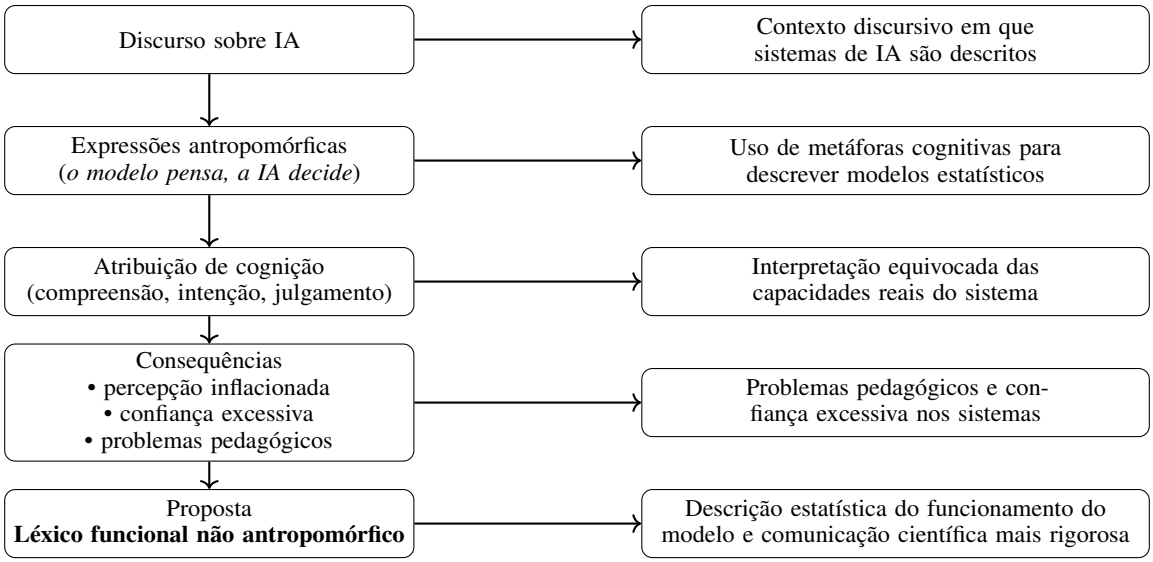

Figure 1: Esquema conceptual do efeito da linguagem antropomórfica no discurso sobre IA. O fluxo vertical representa o encadeamento entre descrições antropomórficas, atribuição de cognição e consequências interpretativas. As caixas laterais explicitam os mecanismos associados a cada etapa.

dos e não compreensão semântica no sentido humano.

5.2 Inferência e seleção de saídas

Expressões como *o modelo decide* ou *a IA raciocina* podem sugerir processos deliberativos inexistentes. Em termos funcionais, os sistemas de PLN realizam inferência estatística, selecionando sequências linguísticas com maior probabilidade segundo os parâmetros do modelo. Assim, numa pergunta lógica simples, como *Se João é mais alto que Maria e Maria é mais alta que Ana, quem é a mais baixa?*, a resposta correta resulta da aplicação de padrões aprendidos e não de raciocínio consciente.

5.3 Erros e saídas não fundamentadas

A popularização do termo "alucinação" ilustra bem o problema da antropomorfização. Em vez de afirmar que *a IA alucina*, é mais adequado descrever que *o modelo gera uma saída não fundamentada nos dados disponíveis*. Tais erros resultam da natureza probabilística do processo de geração e de limitações nos dados de treino, e não de criatividade ou imaginação do sistema.

Para sistematizar a proposta de transformações terminológicas do nosso trabalho, a Tabela 1 apresenta exemplos representativos de reformulações de expressões antropomórficas frequentemente utilizadas no discurso sobre IA e PLN, acompanhadas de alternativas descritivas que procuram refletir com maior precisão os processos computacionais envolvidos.

6 Conclusão

Este artigo argumentou que a antropomorfização no discurso sobre IA e PLN constitui uma fonte recorrente de distorção conceptual. Ao atribuir propriedades cognitivas como compreensão, raciocínio ou julgamento moral a modelos estatísticos, o discurso técnico pode criar interpretações equivocadas das capacidades reais dos sistemas.

Para mitigar esse problema, propusemos um léxico funcional não antropomórfico para descrever processos de PLN em português. A utilização de descrições baseadas em processos computacionais, como modelação estatística, inferência probabilística ou classificação, permite alinhar a linguagem científica com o funcionamento real dos modelos e contribuir para maior literacia crítica em IA.

Trabalhos futuros poderão aprofundar esta proposta através da construção de corpora de textos de IA em português, permitindo analisar quantitativamente padrões de antropomorfização e a evolução terminológica na área.

References

Saleema Amershi and 1 others. 2019. Guidelines for Human–AI Interaction. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Expressão antropomórfica	Descrição funcional	Processo
Compreensão e interpretação de texto		
o modelo compreende	o modelo identifica padrões no texto	modelação estatística
o sistema entende a pergunta	o sistema processa padrões linguísticos	processamento linguístico
Aprendizagem e ajuste de parâmetros		
o sistema aprende	o modelo ajusta parâmetros durante o treino	optimização paramétrica
Inferência e decisão		
o modelo decide	o sistema seleciona saída com maior probabilidade	inferência probabilística
a IA raciocina	o sistema aplica inferência estatística	representação estatística
o modelo pensa	o sistema executa operações algorítmicas	processamento computacional
Memória e conhecimento		
o modelo sabe	o modelo codifica regularidades dos dados de treino	representação estatística
o modelo lembra	o modelo retém padrões aprendidos	codificação paramétrica
Erros ou saídas não fundamentadas		
a IA alucina	o modelo gera saída não fundamentada nos dados	erro de geração
o modelo julga	o sistema classifica segundo padrões aprendidos	classificação

Table 1: Exemplos de reformulação de expressões antropomórficas em descrições funcionais no contexto de PLN.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, pages 610–623. ACM.

Abeba Birhane. 2021. The Dangers of Non-Neutral AI Metaphors. *Patterns*, 2(11):1–3.

S. Brennen and Daniel Kreiss. 2020. Automation Bias and Algorithmic Authority in Education. *Information, Communication Society*, 23(14):2105–2121.

Mark Coeckelbergh. 2020. Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics*, 26:2051–2068.

Jerry A. Fodor. 1983. *The Modularity of Mind: An Essay on Faculty Psychology*. MIT Press, Cambridge, MA.

Yoav Goldberg. 2019. Assessing BERT's Syntactic Abilities. *arXiv preprint arXiv:1901.05287*.

John Haugeland. 1985. *Artificial Intelligence: The Very Idea*. MIT Press, Cambridge, MA.

George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. University of Chicago Press.

Alexander Lerchner. 2026. The Abstraction Fallacy: Why AI Can Simulate but Not Instantiate Consciousness. *Philosophy and Theory of Artificial Intelligence*. Preprint available at PhilPapers.

Gary Marcus and Ernest Davis. 2019. Rebooting AI Research. *Communications of the ACM*, 62(10):36–38.

Melanie Mitchell. 2019. *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus and Giroux, New York.

Hilary Putnam. 1975. The Meaning of "Meaning". *Minnesota Studies in the Philosophy of Science*, 7:131–193.

John R. Searle. 1980. Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3(3):417–457.

Christine Talbot. 2019. The Psychology of Anthropomorphism: How Do We Think About Non-Humans? *Annual Review of Psychology*, 70:139–161.

Joseph Weizenbaum. 1976. *Computer Power and Human Reason: From Judgment to Calculation*. W. H. Freeman, San Francisco.

Jing Zhou. 2022. Overtrust in AI: Psychological Mechanisms and Educational Risks. *Computers & Education*, 184:104495.