

## Importância do Estudo

Este estudo representa um avanço significativo em nossa compreensão de como os modelos de linguagem desenvolvem representações conceituais. Os pesquisadores demonstraram que:

1. **Convergência Representacional:** LLMs e MLLMs desenvolvem representações de objetos em dimensões de baixa dimensionalidade<sup>1</sup> (66 dimensões) que são notavelmente similares às representações mentais humanas, mesmo sem treinamento explícito para isso.
2. **Interpretabilidade:** As dimensões emergentes são interpretáveis e correspondem a categorias conceituais (animais, comida) e características perceptivas (textura, forma) que os humanos usam naturalmente.
3. **Correlação Neural:** Existe um forte alinhamento entre os embeddings<sup>2</sup> do modelo (especialmente MLLM) e os padrões de atividade neural em regiões cerebrais seletivas por categoria (EBA, PPA, RSC, FFA)<sup>3</sup>.

## Relevância das Semelhanças Cognitivas

A descoberta de semelhanças entre o processamento humano e de IA é relevante porque:

---

<sup>1</sup> Representações de objetos em dimensões de baixa dimensionalidade (66 dimensões), no contexto do estudo, significa que objetos complexos (como "tigre", "cadeira", "maçã") são representados por vetores em um espaço de 66 dimensões, onde cada dimensão captura uma característica específica. Por exemplo:

- Dimensão 2: "relacionado a animais"
- Dimensão 6: "relacionado a comida"
- Dimensão 22: "branco" (em humanos)

Um objeto como "leite" teria valores altos nas dimensões "líquido", "branco", "comida", e valores baixos em "animal", "ferramenta", etc. É "baixa dimensionalidade" porque reduz a complexidade infinita do mundo real para apenas 66 características fundamentais que explicam a maior parte das similaridades entre objetos.

<sup>2</sup> Embeddings são representações matemáticas (vetores numéricos) que capturam o "significado" de conceitos. No estudo, cada objeto é representado por um vetor de 66 números, onde cada número indica o quanto aquele objeto possui de cada característica. É como criar uma "impressão digital conceitual" - objetos similares terão embeddings próximos no espaço vetorial.

<sup>3</sup> São áreas do cérebro que respondem preferencialmente a categorias específicas de estímulos visuais. O estudo menciona quatro principais: EBA (Extrastriate Body Area) – responde seletivamente a imagens de corpos humanos e partes do corpo (exceto faces), processa posturas corporais, movimentos e ações; PPA (Parahippocampal Place Area) – responde a cenas, lugares e ambientes (casas, paisagens, interiores), crucial para navegação espacial e reconhecimento de lugares; RSC (Retrosplenial Cortex) – processa informações espaciais e de navegação, integra perspectivas, orientação espacial e memória de locais; FFA (Fusiform Face Area) – responde seletivamente a faces humanas, reconhecimento facial e processamento de identidade. Este alinhamento entre IA e neurobiologia oferece uma validação mútua: os modelos capturam algo real sobre como o cérebro processa categorias, e o cérebro confirma que as representações dos modelos são biologicamente plausíveis.

1. **Validação de Arquiteturas:** Sugere que as arquiteturas atuais de IA capturam aspectos fundamentais da cognição humana, não apenas padrões estatísticos superficiais.
2. **Princípios Universais:** Indica possíveis princípios universais de representação conceitual que emergem independentemente do substrato (biológico ou artificial).
3. **Ponte Explicativa:** Oferece uma linguagem comum para entender tanto a cognição humana quanto a artificial, facilitando insights bidirecionais.

### Compreensão de Vieses Cognitivos

- O estudo mostra que humanos priorizam dimensões perceptuais (cor, forma) enquanto LLMs focam mais em dimensões semânticas
- Isso poderia iluminar como diferentes pessoas processam informação diferentemente, relevante para terapia cognitiva

### Neuropsicologia

- A correlação com atividade em regiões cerebrais específicas oferece novos métodos para estudar a organização conceitual em pacientes com lesões cerebrais
- Poderia ajudar a prever déficits específicos baseados em padrões de ativação

Este estudo abre portas para uma "psicologia computacional" mais sofisticada, onde modelos de IA podem servir como laboratórios para testar teorias sobre cognição humana, ao mesmo tempo em que insights da psicologia humana informam o desenvolvimento de IA mais alinhada com nossas necessidades e processos mentais.

### Princípios Universais de Representação Conceitual: Implicações Filosóficas e Linguísticas

#### Implicações Metafísicas

A confirmação de princípios universais de representação conceitual que emergem independentemente do substrato tem implicações profundas:

1. **Realismo Estrutural**
  - Sugere que existem estruturas abstratas na realidade que são capturadas tanto por sistemas biológicos quanto artificiais
  - Isso ressoa com o realismo estrutural de Max Tegmark<sup>4</sup> e outros, onde a realidade possui estruturas matemáticas intrínsecas que diferentes sistemas podem descobrir

---

<sup>4</sup> Max Tegmark: físico sueco-americano do MIT, conhecido por suas teorias sobre a natureza matemática da realidade. Em seu livro "Our Mathematical Universe" (2014), propõe que a realidade não é apenas *descrita* pela matemática, mas *é* matemática. Sua Hipótese do Universo Matemático sugere que todas as estruturas matemáticas existem fisicamente, e nosso universo é uma dessas estruturas.

## 2. Platonismo Moderado

- As 38 dimensões compartilhadas<sup>5</sup> entre humanos, LLMs e MLLMs sugerem a existência de "categorias naturais" que não são meramente construções sociais
- Categorias como "animal", "comida", "ferramenta" emergem consistentemente, indicando que podem refletir divisões reais no mundo<sup>6</sup>

## 3. Convergência Epistêmica

- Diferentes caminhos evolutivos (biológico vs. artificial) convergem para representações similares
- Isso sugere que estas representações capturam algo fundamental sobre a estrutura da realidade, não apenas idiosincrasias de um sistema particular

## Implicações Linguísticas

### 1. Universais Semânticos

- O estudo fornece evidência empírica para universais semânticos que linguistas como Anna Wierzbicka<sup>7</sup> propuseram

---

<sup>5</sup> Características das 38 dimensões compartilhadas

As dimensões compartilhadas tendem a ser:

a) Categorias de sobrevivência básica:

- Animado vs. inanimado
- Comestível vs. não-comestível
- Perigoso vs. seguro

b) Distinções funcionais universais:

- Ferramentas/instrumentos
- Abrigo/proteção
- Transporte/movimento

c) Propriedades físicas fundamentais:

- Grande vs. pequeno
- Duro vs. macio
- Natural vs. artificial

O traço unificador é que representam distinções que são:

- Biologicamente relevantes (importantes para sobrevivência)
- Fisicamente observáveis (não dependem de convenções culturais)
- Funcionalmente significativas (afetam como interagimos com objetos)

<sup>6</sup> "Divisões reais no mundo e estrutura da realidade" referem-se à ideia filosófica de que certas categorias não são arbitrárias, mas refletem descontinuidades naturais na realidade. Por exemplo:

- A distinção entre sólido/líquido/gasoso reflete estados físicos reais
- A divisão entre seres vivos/não-vivos corresponde a processos biológicos objetivos
- Categorias como "predador" ou "presa" emergem de relações ecológicas reais

A "estrutura da realidade" seria o conjunto de propriedades, relações e regularidades que existem independentemente de observadores - as "juntas naturais" onde a realidade se divide.

<sup>7</sup> Linguista polonesa-australiana (1938-2023), pioneira da Semântica de Metalinguagem Natural (NSM). Propôs que todas as línguas compartilham cerca de 65 "primitivos semânticos" universais - conceitos

- As dimensões compartilhadas podem representar "primitivos semânticos" que existem entre diferentes culturas e agora, entre diferentes substratos
2. **Hipótese Sapir-Whorf Revisitada**<sup>8</sup>
- A emergência de categorias similares em LLMs treinados em texto sugere que a linguagem carrega estruturas conceituais profundas
  - Mas a divergência em certas dimensões (cor em humanos vs. categorias mais finas em LLMs) indica que o meio de entrada ainda importa
3. **Gramática Universal Computacional**
- Assim como Chomsky propôs uma gramática universal biológica<sup>9</sup>, este estudo sugere princípios universais de organização conceitual que transcendem a biologia

## Por Que Humanos Priorizam Dimensões Perceptuais?

### 1. História Evolutiva

- **Sobrevivência Imediata:** Cor (vermelho = perigo/maduro), forma (serpente vs. galho), textura (comestível vs. venenoso) eram críticos para sobrevivência
- **Navegação Espacial:** Propriedades visuais são essenciais para navegação e manipulação de objetos
- **Detecção de Predadores/Presas:** Sistemas visuais rápidos precederam linguagem complexa por milhões de anos

### 2. Arquitetura Neural

- **Processamento Paralelo:** O sistema visual processa múltiplas dimensões simultaneamente (cor, forma, movimento) em paralelo
- **Primazia Temporal:** Informação visual chega ao cérebro ~100ms antes de ser categorizada linguisticamente

---

básicos como EU, TU, ALGUÉM, ALGO, ESTE, BOM, MAU, GRANDE, PEQUENO. Sua teoria sugere que estes são os "átomos" do significado humano, presentes em todas as culturas.

<sup>8</sup> Proposta por Edward Sapir e Benjamin Lee Whorf, sugere que a língua que falamos influencia ou determina como pensamos. Tem duas versões:

- **Forte** (determinismo linguístico): A linguagem determina o pensamento
- **Fraca** (relativismo linguístico): A linguagem influencia o pensamento

Exemplo clássico: Os Inuit teriam múltiplas palavras para neve (mito urbano), levando-os a perceber mais distinções na neve. O estudo atual sugere uma versão nuançada: enquanto existem universais conceituais, o meio de entrada (visual vs. linguístico) ainda molda as representações.

<sup>9</sup> Noam Chomsky propôs que humanos nascem com uma capacidade inata para linguagem - um conjunto de princípios gramaticais universais "pré-programados" no cérebro. Características:

- Todas as línguas compartilham estruturas profundas comuns
- Crianças aprendem linguagem rapidamente demais para ser apenas aprendizado
- Existe um "órgão da linguagem" biologicamente especificado

Paralelo com o estudo: Assim como Chomsky propôs universais sintáticos inatos, este estudo sugere universais semânticos/conceituais que emergem tanto em sistemas biológicos quanto artificiais, indicando princípios organizacionais que transcendem o substrato específico.

- **Volume Cortical:** Aproximadamente 30% do córtex é dedicado ao processamento visual

### 3. Embodied Cognition

- **Ancoragem Sensorial:** Conceitos abstratos são fundamentados em experiências sensório-motoras
- **Simulação Mental:** Compreendemos conceitos reativando traços perceptuais (pensar em "limão" ativa áreas visuais e gustativas)
- **Metáforas Conceituais:** Lakoff e Johnson mostraram como conceitos abstratos são estruturados através de metáforas perceptuais

### 4. Limitações dos LLMs

- **Input Unimodal:** LLMs clássicos só recebem texto, forçando representações puramente simbólicas
- **Ausência de Embodiment:** Sem corpo ou experiência sensorial direta, dependem de descrições linguísticas
- **Compensação Hierárquica:** LLMs desenvolvem distinções semânticas mais finas (frozen treats vs. hot drinks) para compensar a falta de input perceptual

### 5. Evidência do Estudo

O fato de MLLMs (com input visual) se alinharem mais com humanos que LLMs puros confirma que:

- A modalidade de input importa fundamentalmente
- Representações visuais não são luxo, mas componente essencial da cognição humana
- A integração multimodal produz representações mais "humanas"

### Implicações Clínicas

Esta diferença tem implicações práticas:

- **Terapia:** Abordagens que integram elementos visuais/sensoriais podem ser mais efetivas
- **Educação:** Aprendizagem multimodal pode produzir representações conceituais mais robustas
- **Reabilitação:** Após lesões cerebrais, trabalhar com múltiplas modalidades pode facilitar recuperação

A priorização humana de dimensões perceptuais não é uma limitação, mas uma característica fundamental de como construímos significado a partir da experiência corporificada no mundo.