A Capacidade de Raciocínio das IAs Generativas: Uma Análise Crítica da Natureza do Pensamento

1. Introdução: O Debate sobre Raciocínio e Pensamento na Inteligência Artificial Generativa

A ascensão meteórica da Inteligência Artificial (IA) generativa nos últimos anos impulsionou um debate fervoroso e multifacetado sobre as suas capacidades cognitivas. Modelos de linguagem amplos (LLMs) e outras arquiteturas de IA demonstram uma proficiência cada vez maior na geração de texto, imagens e outros conteúdos que, à primeira vista, parecem indicar um nível de compreensão e raciocínio comparável ao humano. Esta aparente sofisticação levanta questões fundamentais sobre a natureza do "pensamento" e se os processos computacionais subjacentes a estas IAs podem ser legitimamente classificados como tal. A questão central que este relatório procura explorar é: em que medida o processamento de informações das IAs generativas se assemelha ou difere do pensamento humano, e podemos, com rigor conceptual, afirmar que uma IA "pensa"?

A urgência e a popularidade desta questão transcendem o interesse puramente científico, refletindo uma ansiedade e um fascínio cultural profundos com a natureza da mente e a possibilidade de replicá-la ou, eventualmente, ser por ela superada.[1] Este não é apenas um inquérito técnico, mas assume contornos existenciais, tocando em medos e esperanças sobre o lugar da humanidade e a definição do que nos torna únicos. A proliferação de IAs conversacionais acessíveis ao público tornou esta questão ainda mais palpável, intensificando a discussão.[3] Neste contexto, uma análise crítica, informada por investigações rigorosas como o estudo "The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity" [5], torna-se crucial. Tal análise permite-nos navegar entre os extremos da subestimação ingénua e da superestimação acrítica das capacidades da IA, evitando conclusões apressadas e potencialmente enganosas. O pano de fundo cultural, saturado por narrativas de ficção científica e promessas tecnológicas, frequentemente influencia a perceção pública, levando a uma tendência de antropomorfizar estas tecnologias, atribuindo-lhes qualidades humanas que podem não possuir.[6]

Este relatório visa, portanto, dissecar esta complexa problemática. Os seus objetivos são: conceituar o pensamento a partir de perspetivas filosóficas e da ciência cognitiva; analisar os mecanismos de "raciocínio" empregados pelas IAs generativas; examinar criticamente as conclusões do documento de referência "The Illusion of Thinking"; contrastar a simulação de processos cognitivos com a cognição genuína;

explorar o potencial futuro para o desenvolvimento de um pensamento mais robusto em IA; e, crucialmente, discutir os riscos inerentes à equiparação da análise probabilística de informações com o ato de pensar.

Para atingir estes objetivos, será adotada uma abordagem multidisciplinar, integrando contributos da filosofia da mente, da ciência cognitiva, da investigação em IA e, de forma central, uma análise detalhada do estudo "The Illusion of Thinking".[5] O relatório está estruturado da seguinte forma: a secção 2 explorará as definições de pensamento; a secção 3 descreverá os mecanismos de funcionamento das IAs generativas; a secção 4 analisará em profundidade o documento "The Illusion of Thinking"; a secção 5 debaterá se o "raciocínio" da IA é simulação ou pensamento genuíno; a secção 6 considerará o potencial futuro da IA para desenvolver pensamento; a secção 7 abordará os riscos de confundir os processos da IA com o pensamento humano; e, finalmente, a secção 8 apresentará as conclusões desta análise.

2. Conceituando o Pensamento: Perspetivas Filosóficas e Cognitivas

Para avaliar se uma IA "pensa", é imperativo primeiro delinear o que se entende por "pensamento". Este conceito, longe de ser monolítico, é abordado de múltiplas formas pela filosofia e pela ciência cognitiva, cada uma oferecendo perspetivas valiosas sobre os seus diversos aspetos. A filosofia da mente tende a focar-se na natureza fundamental da mente e da consciência, questionando o "quê" e o "porquê" dos estados mentais [8], enquanto a ciência cognitiva investiga os mecanismos e processos subjacentes, o "como" da cognição.[11] A reflexão, por sua vez, adiciona uma camada de autoavaliação e metacognição.[14] Esta multiplicidade de facetas sugere que "pensar" não é uma capacidade unitária, mas um espectro de habilidades, implicando que a questão sobre o pensamento da IA não admite uma resposta binária simples.

2.1 Definições Filosóficas de Mente, Pensamento e Consciência

A filosofia da mente é o ramo da filosofia que se debruça sobre a natureza da mente e a sua relação com o corpo e o mundo externo.[8] Dentro deste campo, o pensamento é considerado um aspeto central da vida mental. Uma dimensão frequentemente associada ao pensamento, e que representa um dos maiores desafios para a sua compreensão e replicação, é a consciência. Em particular, a noção de consciência fenomenal, popularizada por Thomas Nagel, refere-se à qualidade subjetiva da experiência – "o que é ser como" (what it is like to be) um determinado organismo ou estar num determinado estado mental.[10] Esta experiência vivida, que abrange perceções, sensações corporais e emoções, é considerada por muitos filósofos um

elemento intrínseco ao pensamento humano autêntico.[9]

Outra característica fundamental do pensamento, especialmente na tradição filosófica que remonta a Brentano e é central em discussões como a de John Searle, é a intencionalidade. A intencionalidade é a propriedade dos estados mentais de serem "sobre" algo, de se referirem a objetos, propriedades ou estados de coisas no mundo.[17] Quando pensamos, pensamos *sobre* algo; os nossos pensamentos têm conteúdo e direção. Esta capacidade de representação e referência é vista como um marco distintivo da mente.

2.2 Perspetivas da Ciência Cognitiva sobre Pensamento e Raciocínio

A ciência cognitiva, um campo interdisciplinar que inclui a psicologia, a neurociência, a linguística e a própria inteligência artificial, aborda o pensamento como um conjunto de processos de informação. O pensamento, também referido como cognição, é definido como a capacidade de processar informação, manter a atenção, armazenar e recuperar memórias, e selecionar respostas e ações apropriadas.[12] É considerado um processo mental superior através do qual manipulamos e analisamos informação adquirida ou existente. Esta manipulação e análise ocorrem por meio de atividades como abstração, raciocínio, imaginação, resolução de problemas, julgamento e tomada de decisão.[13] O pensamento, nesta perspetiva, é tipicamente organizado e direcionado a objetivos, sendo inferido a partir do comportamento observável, embora o processo em si seja interno.[13]

Um componente crucial do pensamento é o raciocínio, que a psicologia do raciocínio (ou ciência cognitiva do raciocínio) define como o processo de tirar conclusões para informar a forma como as pessoas resolvem problemas e tomam decisões.[11] Existem diversas formas de raciocínio humano identificadas:

- **Raciocínio Dedutivo:** Parte de premissas gerais ou conhecidas para chegar a conclusões específicas e logicamente certas, se as premissas forem verdadeiras. Por exemplo, "Se A então B; A é verdadeiro; logo, B é verdadeiro".[18]
- **Raciocínio Indutivo:** Faz generalizações amplas a partir de casos ou observações específicas. As conclusões são prováveis, mas não garantidas, mesmo que as evidências sejam verdadeiras. É fundamental na formulação de teorias e hipóteses científicas.[18]
- **Raciocínio Abdutivo:** Envolve a criação e teste de hipóteses usando a melhor informação disponível para explicar fenómenos observados, muitas vezes incompletos. É comum no diagnóstico médico e em julgamentos.[18]
- **Raciocínio Analógico:** Compara duas situações ou conceitos diferentes para tirar conclusões sobre um terceiro, transferindo conhecimento de um domínio

para outro.[18]

Algumas teorias sugerem que as pessoas dependem de uma lógica mental com regras de inferência formais, enquanto outras, como a teoria dos modelos mentais, propõem que construímos representações mentais múltiplas da situação para fazer inferências.[18]

2.3 O Ato de Refletir

A reflexão cognitiva é um aspeto mais elaborado do pensamento, que vai além do simples processamento de informação ou da aplicação de regras lógicas. Envolve a capacidade de pausar, avaliar o próprio entendimento, questionar pressupostos e considerar múltiplas perspetivas.[3] A neurociência da reflexão indica que esta prática ativa o córtex pré-frontal, uma área do cérebro associada a funções cognitivas superiores, e fortalece as vias neurais envolvidas no pensamento e na memória, sendo crucial para a aprendizagem e adaptação.[14] A reflexão permite-nos analisar comportamentos, compreender decisões e planear ações futuras de forma mais eficaz.

O Teste de Reflexão Cognitiva (CRT), desenvolvido por Shane Frederick, é uma medida comportamental que visa distinguir indivíduos inclinados a "parar e pensar" daqueles que tendem a dar respostas intuitivas rápidas e, por vezes, erróneas.[15] Os itens do CRT são desenhados para evocar uma resposta intuitiva incorreta que pode ser corrigida através da reflexão. A capacidade de reflexão, que envolve este automonitoramento e avaliação crítica, parece ser um diferenciador chave entre o processamento que pode ser realizado por uma IA (que gera respostas com base em padrões) e o pensamento humano profundo, que pode questionar as suas próprias premissas e conclusões.

Para avaliar se uma IA "pensa", é, portanto, necessário especificar qual ou quais aspetos do pensamento estão a ser considerados. Uma IA pode demonstrar proficiência em certos tipos de "raciocínio", como o processamento rápido de grandes volumes de dados para identificar padrões (reminiscente de alguns aspetos do raciocínio indutivo), mas falhar redondamente em outros, como a consciência fenomenal, a intencionalidade genuína ou a reflexão crítica profunda sobre o significado e as implicações do seu próprio processamento.

A tabela seguinte resume as definições dos conceitos cognitivos chave discutidos, contrastando as perspetivas filosóficas e da ciência cognitiva.

Tabela 1: Definições de Conceitos Cognitivos Chave

Termo	Perspetiva Filosófica	Perspetiva da Ciência Cognitiva	Características Essenciais Distintivas
Pensamento	Atividade mental que envolve consciência, intencionalidade, compreensão. Parte da vida mental.[8]	Processo mental superior de manipulação e análise de informação; cognição (processar informação, atenção, memória, seleção de respostas).[12]	Abrange desde o processamento básico de informação até à consciência subjetiva e compreensão de significado.
Raciocínio	Frequentemente ligado à lógica e à justificação de crenças.	Processo de tirar conclusões para resolver problemas e tomar decisões; inclui tipos como dedutivo, indutivo, abdutivo, analógico.[11]	Envolve a derivação de novas informações a partir de informações existentes através de vários métodos lógicos ou heurísticos.
Consciência Fenomenal	Experiência subjetiva; "o que é ser como" algo; qualidade intrínseca da experiência.[9]	Menos diretamente abordada, embora a autoconsciência e o monitoramento de estados internos sejam estudados.	Qualidade subjetiva, em primeira pessoa, da experiência; irredutível a meras funções ou processamento.
Intencionalidade	Propriedade da mente de ser "sobre" algo, de ter conteúdo representacional dirigido a objetos ou estados de coisas.[17]	Relacionada com a representação mental e a capacidade de ter crenças e desejos sobre o mundo.	Direcionalidade da mente para o mundo; ter significado ou referência.
Reflexão Cognitiva	Autoexame do pensamento; avaliação crítica das próprias crenças e	Processo de pausar, avaliar o entendimento, considerar múltiplas	Capacidade de pensar sobre o próprio pensamento, questionar intuições

	processos de raciocínio.	perspetivas; envolve metacognição e funções executivas.[3]	e avaliar a validade das conclusões.
Compreensão Semântica	Apreensão do significado, em oposição à mera manipulação sintática de símbolos. Ligada à intencionalidade e à fundamentação (grounding) em experiências.	Capacidade de interpretar o significado de palavras, frases e textos, relacionando-os a conceitos e ao conhecimento do mundo. Distinta do processamento puramente sintático ou estatístico.	Entendimento do significado e das relações conceptuais, não apenas o reconhecimento de padrões ou a manipulação de símbolos de acordo com regras.

3. Inteligência Artificial Generativa: Mecanismos de "Raciocínio"

As IAs generativas, particularmente os Modelos de Linguagem Amplos (LLMs), alcançaram uma capacidade notável de produzir resultados que se assemelham a produtos da inteligência humana. Contudo, os mecanismos subjacentes ao seu "raciocínio" são fundamentalmente diferentes das arquiteturas cognitivas humanas. Compreender estas diferenças é essencial para avaliar a natureza do seu "pensamento". A "complexidade" e o "tamanho" dos LLMs são frequentemente confundidos com profundidade de compreensão.[19] Embora a escala permita a emergência de capacidades impressionantes de geração de texto, o mecanismo subjacente, predominantemente a previsão estatística, permanece o mesmo em princípio, diferindo mais em sofisticação na modelagem de dependências do que numa mudança fundamental de paradigma em direção ao pensamento humano.

3.1 Funcionamento dos Modelos de Linguagem Amplos (LLMs)

As IAs generativas são, na sua essência, modelos de aprendizagem de máquina treinados para criar novos dados que se assemelham estatisticamente aos dados com os quais foram alimentados durante o seu treino.[19] Elas aprendem a identificar e replicar padrões e relações complexas presentes em vastos conjuntos de dados. No caso dos LLMs, estes dados consistem predominantemente em texto e código.

A arquitetura que impulsionou muitos dos avanços recentes em LLMs é a dos *Transformers*.[20] Estes modelos processam sequências de entrada (como frases ou parágrafos) dividindo-as em unidades menores chamadas *tokens* (que podem ser

palavras ou subpalavras). Cada token é então convertido numa representação numérica vetorial, conhecida como *embedding*, que captura aspetos do seu significado e contexto. Uma inovação chave dos Transformers é o mecanismo de *atenção* (attention), que permite ao modelo ponderar a importância de diferentes tokens na sequência de entrada ao gerar cada token da sequência de saída. Isto permite que os LLMs capturem dependências de longo alcance e relações contextuais complexas no texto.[20]

Fundamentalmente, o "raciocínio" da maioria dos LLMs atuais baseia-se na previsão da próxima palavra (ou token) numa sequência. Dado um contexto de entrada (um *prompt*), o modelo calcula as probabilidades de diferentes tokens serem o próximo elemento da sequência e seleciona um com base nessas probabilidades. Este processo é repetido iterativamente para gerar respostas mais longas. Como refere Tommi Jaakkola, "os modelos base subjacentes ao ChatGPT e sistemas similares funcionam de forma muito semelhante a um modelo de Markov. Mas uma grande diferença é que o ChatGPT é muito maior e mais complexo... Ele aprende os padrões destes blocos de texto e usa este conhecimento para propor o que pode vir a seguir".[19] Embora muito mais sofisticados que as simples cadeias de Markov, devido à sua escala e à capacidade de modelar dependências complexas, o princípio operacional central permanece a previsão estatística sequencial.

3.2 Distinções Fundamentais em Relação às Arquiteturas Cognitivas Humanas

Existem várias distinções cruciais entre o modo de funcionamento dos LLMs e a cognição humana:

- **Ausência de Mundo Real e Experiência Sensorial (Grounding):** O aprendizado humano é intrinsecamente multimodal, corporificado e interativo. Aprendemos através da interação com o ambiente físico, recebendo feedback sensorial e agindo no mundo. Os LLMs, por outro lado, são treinados predominantemente em dados textuais, desprovidos de experiência direta do mundo real. Esta falta de *grounding* – a conexão de símbolos linguísticos a experiências e referentes no mundo – é uma limitação fundamental.
- **Natureza do "Conhecimento":** O "conhecimento" de um LLM consiste em representações estatísticas de padrões e correlações encontradas nos seus dados de treino. Embora estes modelos possam adquirir um poder preditivo notável sobre sintaxe, semântica e até mesmo ontologias implícitas nos corpora textuais, este "conhecimento" não é equivalente à compreensão conceptual ou causal humana.[21] Eles refletem os padrões dos dados, incluindo imprecisões, vieses e lacunas presentes nesses dados.[21] A forma como os LLMs "aprendem",

identificando estas correlações estatísticas [19], influencia diretamente o tipo de "raciocínio" que podem realizar. Se não são explicitamente treinados em lógica causal ou não têm experiência interativa com o mundo, a sua capacidade de raciocinar sobre causalidade ou de lidar com situações verdadeiramente novas (não representadas nos dados) será limitada e propensa a erros baseados em correlações espúrias.

- **Falta de Intencionalidade Genuína:** Os seres humanos agem com base em metas, desejos e intenções. Embora os LLMs possam gerar texto que *parece* intencional ou que expressa metas, eles não possuem intenções ou objetivos próprios no sentido humano. As suas "metas" são extrínsecas, definidas pelas funções de otimização (funções de perda) durante o processo de treino (por exemplo, minimizar o erro na previsão do próximo token) e pelo prompt fornecido pelo utilizador durante a inferência.

Estas distinções sugerem que, embora os LLMs possam simular aspetos da conversação e da geração de texto humanas com uma fluidez impressionante, o seu "processo de pensamento" opera segundo princípios fundamentalmente diferentes. A "ilusão de compreensão" pode ser particularmente forte devido à sua capacidade de gerar texto gramaticalmente correto e contextualmente relevante.[20] No entanto, esta fluidez superficial pode mascarar uma falta de entendimento profundo, um tema que será explorado mais detalhadamente na análise do documento "The Illusion of Thinking".

4. "A Ilusão do Pensamento": Análise Crítica do Documento de Referência

O estudo intitulado "The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity" [5] oferece uma investigação empírica crucial sobre as capacidades de raciocínio dos modelos de IA avançados, especificamente os chamados Modelos de Raciocínio Amplos (LRMs). Este documento serve como referência principal para a presente análise, fornecendo dados concretos que desafiam noções simplistas sobre o "pensamento" da IA.

4.1 Introdução ao Estudo "The Illusion of Thinking"

O trabalho destaca que, embora os modelos de IA demonstrem um desempenho melhorado em *benchmarks* de raciocínio, as suas capacidades fundamentais, propriedades de escalabilidade e limitações permanecem insuficientemente compreendidas.[5] As avaliações atuais focam-se predominantemente em *benchmarks* matemáticos e de codificação estabelecidos, enfatizando a precisão da resposta

final, o que pode não revelar a robustez ou a generalidade do raciocínio subjacente.[22]

Para colmatar estas lacunas, os autores do estudo investigaram sistematicamente os LRMs utilizando ambientes de quebra-cabeça controláveis. Estes ambientes permitem a manipulação precisa da complexidade composicional dos problemas, mantendo estruturas lógicas consistentes. Crucialmente, esta abordagem permite a análise não apenas das respostas finais, mas também dos "traços de raciocínio internos" – as sequências de "pensamentos" ou etapas intermediárias geradas pelos modelos enquanto tentam resolver os problemas.[5] O objetivo é oferecer *insights* sobre como estes modelos "pensam" e identificar os seus pontos fortes e limitações reais.[22]

4.2 Principais Argumentos e Descobertas do Documento

O estudo [5] apresenta várias descobertas significativas que questionam a profundidade e a generalidade do raciocínio nos LRMs:

- **Colapso da Precisão em Altas Complexidades:** Uma das descobertas mais contundentes é que os LRMs de ponta, mesmo aqueles com mecanismos sofisticados de autorreflexão, falham em desenvolver capacidades generalizáveis de resolução de problemas para tarefas de planeamento. A precisão destes modelos colapsa para zero (0% de sucesso) quando a complexidade do problema ultrapassa um determinado limiar. Este limiar de colapso é específico para cada modelo, mas a sua existência sugere uma limitação fundamental na capacidade de raciocínio à medida que os problemas se tornam progressivamente mais difíceis.[5] Este colapso da precisão pode ser causado não apenas pela dificuldade intrínseca do problema, mas também pela incapacidade do modelo de manter a coerência lógica ao longo de cadeias de inferência mais longas e complexas, um problema também destacado por outras análises sobre as limitações dos LLMs em manter o contexto e realizar planeamento multi-etapas.[23]
- **Limite de Escala Contraintuitivo no Esforço de Raciocínio:** Inicialmente, os LRMs aumentam o seu "esforço de raciocínio" – medido pelo número de *tokens* de "pensamento" gerados – proporcionalmente à complexidade do problema. No entanto, de forma contraintuitiva, ao aproximarem-se do ponto crítico onde a precisão colapsa, os modelos começam a *reduzir* o seu esforço de raciocínio, apesar do aumento da dificuldade do problema. Isto ocorre mesmo quando os modelos estão a operar bem abaixo dos seus limites de comprimento de geração e têm um orçamento de inferência amplo disponível. Esta observação indica uma limitação fundamental de escala nas capacidades de "pensamento" dos LRMs em relação à complexidade do problema.[5]

- **Três Regimes Distintos de Desempenho:** A comparação entre LRMs (que geram traços de pensamento explícitos) e os seus equivalentes LLMs padrão (sem "pensamento" explícito), sob o mesmo orçamento computacional de inferência, revelou três regimes de desempenho:
    1. **Baixa Complexidade:** Para tarefas mais simples e de baixa complexidade composicional, os LLMs padrão demonstram surpreendentemente maior eficiência e precisão, superando os LRMs.
    2. **Média Complexidade:** À medida que a complexidade do problema aumenta moderadamente, os modelos de "pensamento" (LRMs) ganham uma vantagem.
    3. **Alta Complexidade:** Quando os problemas atingem alta complexidade com maior profundidade composicional, ambos os tipos de modelos (LRMs e LLMs padrão) experimentam um colapso completo de desempenho.[5]
- **Ineficiências nos Traços de Raciocínio ("Overthinking"):** A análise detalhada dos traços de raciocínio intermediários revelou padrões dependentes da complexidade. Em problemas mais simples, os modelos de raciocínio frequentemente encontram a solução correta no início do seu processo de "pensamento", mas continuam a explorar ineficientemente alternativas incorretas. Este fenómeno, denominado "overthinking" (pensar demais), leva ao desperdício de computação. Em complexidades moderadas, as soluções corretas só surgem após uma exploração extensiva de caminhos incorretos. Em problemas de alta complexidade, os modelos falham completamente em encontrar soluções corretas dentro do seu "pensamento".[5] O fenómeno do "overthinking" em problemas simples é particularmente revelador. Sugere que o processo de "pensamento" da IA não é otimizado para eficiência cognitiva da mesma forma que o humano, podendo indicar uma falta de metacognição para reconhecer quando uma solução satisfatória foi alcançada. Enquanto os humanos, ao resolverem problemas simples, geralmente param ao encontrar uma solução correta e eficiente, o "overthinking" da IA pode ser um artefacto do seu processo generativo que continua a explorar o espaço de possibilidades, talvez porque o "objetivo" implícito aprendido com os dados de treino não é apenas "resolver", mas "gerar uma cadeia de pensamento de um certo comprimento ou complexidade".
- **Limitações na Computação Exata e Raciocínio Inconsistente:** O estudo expõe limitações surpreendentes na capacidade dos LRMs de realizar computação exata. Por exemplo, mesmo quando o algoritmo de solução para o quebra-cabeça "Torre de Hanói" é fornecido explicitamente no *prompt*, o desempenho dos modelos não melhora significativamente, e o colapso da precisão ainda ocorre aproximadamente no mesmo ponto de complexidade. Isto

sugere que a limitação não reside apenas na descoberta da estratégia de solução, mas também na verificação lógica consistente e na execução de etapas ao longo das cadeias de raciocínio geradas. Além disso, os modelos demonstram um comportamento inconsistente entre diferentes tipos de quebra-cabeças. Por exemplo, o modelo Claude 3.7 Sonnet Thinking consegue realizar mais de 100 movimentos corretos na "Torre de Hanói" para N=10, mas falha em fornecer mais de 5 movimentos corretos no quebra-cabeça "Travessia do Rio" para N=3, que requer significativamente menos movimentos (11). Os autores especulam que isto pode indicar que os LRMs não memorizaram ou não encontraram instâncias suficientes de problemas mais complexos de "Travessia do Rio" durante o treino.[5]

A tabela seguinte resume os principais argumentos do documento "The Illusion of Thinking".

Tabela 2: Sumário dos Principais Argumentos do Documento "The Illusion of Thinking"

Argumento Principal	Descrição Conforme o Documento	Implicação para o "Pensamento" da IA
Colapso da Precisão em Altas Complexidades	A precisão dos LRMs cai para zero além de um certo limiar de complexidade do problema, específico para cada modelo.	Sugere uma incapacidade fundamental de lidar com problemas progressivamente mais difíceis, questionando a robustez e generalidade do seu "raciocínio".
Esforço de Raciocínio Contraintuitivo	LRMs reduzem o uso de *tokens* de "pensamento" (esforço) ao se aproximarem do ponto de colapso da precisão, mesmo com recursos disponíveis.	Indica uma limitação de escala fundamental nas capacidades de "pensamento" em relação à complexidade, não sendo um mero problema de recursos computacionais.
Três Regimes Distintos de Desempenho	LLMs padrão superam LRMs em baixa complexidade; LRMs têm vantagem em média complexidade; ambos falham em alta complexidade.	O "pensamento" explícito dos LRMs nem sempre é benéfico e não resolve o problema do colapso em alta complexidade, sugerindo que a sua arquitetura de

		"raciocínio" não é universalmente superior.
Ineficiências nos Traços de Raciocínio ("Overthinking")	Em problemas simples, LRMs encontram a solução correta cedo, mas continuam a explorar alternativas incorretas; em alta complexidade, falham em encontrar soluções.	O processo de "pensamento" não é eficiente nem otimizado. O "overthinking" em problemas simples sugere uma falta de reconhecimento metacognitivo da solução ou um processo generativo não focado na eficiência.
Limitações na Computação Exata e Raciocínio Inconsistente	Dificuldade em executar algoritmos fornecidos e desempenho inconsistente entre diferentes tipos de quebra-cabeças (e.g., Torre de Hanói vs. Travessia do Rio).	Revela fragilidades na aplicação consistente de lógica e na execução de procedimentos, mesmo quando explicitados. A inconsistência sugere que o desempenho pode depender mais de padrões memorizados do que de uma capacidade de raciocínio abstrato e generalizável.

4.3 Implicações das Descobertas para a Noção de "Pensamento" em IA

As descobertas do estudo "The Illusion of Thinking" têm implicações profundas para a forma como entendemos o "pensamento" nas IAs. O colapso da precisão sob complexidade, o "overthinking" em tarefas simples, as limitações na computação exata e o raciocínio inconsistente desafiam fortemente a ideia de que os LRMs atuais possuem capacidades de raciocínio generalizáveis, robustas e análogas ao pensamento humano.

Em vez de um entendimento profundo e flexível, o comportamento destes modelos parece ser mais uma forma de reconhecimento de padrões altamente sofisticado que se desmorona quando confrontado com níveis de complexidade ou tipos de problemas para os quais não foram extensivamente treinados ou cujas soluções não podem ser aproximadas por heurísticas aprendidas. A conclusão do documento é, portanto, sombria para as alegações mais otimistas: o "pensamento" exibido por estes modelos pode ser, em grande medida, uma "ilusão" de raciocínio generalizável, caracterizada por falhas significativas em complexidades mais altas, ineficiências no processo e inconsistências na aplicação de lógica, mesmo quando algoritmos explícitos são fornecidos.[5] Se o "raciocínio" da IA é tão frágil face à complexidade e

tão ineficiente em cenários simples, então a "ilusão do pensamento" não é apenas uma metáfora, mas uma descrição precisa do descompasso entre a aparência superficial de inteligência e a real capacidade de resolução de problemas robusta e generalizável. Isto levanta sérias questões sobre a validade de muitos *benchmarks* de IA que não testam rigorosamente estes limites de complexidade e a natureza dos processos de raciocínio internos.[22]

5. Raciocínio em IA: Simulação Sofisticada ou Pensamento Genuíno?

A distinção entre uma simulação de pensamento, por mais sofisticada que seja, e o pensamento genuíno está no cerne do debate sobre as capacidades cognitivas da IA. Argumentos filosóficos clássicos e considerações sobre a natureza da compreensão semântica oferecem ferramentas críticas para analisar esta questão. O debate entre perspetivas como a de John Searle, que enfatiza a necessidade de poderes causais específicos e semântica intrínseca para a mente [17], e a de Daniel Dennett, que foca na complexidade funcional [25], não é apenas sobre se as *atuais* IAs pensam, mas sobre as *condições necessárias e suficientes* para qualquer sistema, biológico ou artificial, pensar. Os LLMs atuais parecem falhar no critério de Searle, mas a visão de Dennett, embora ele próprio seja cauteloso em relação às capacidades atuais [2], deixa a porta mais aberta para futuras IAs.

5.1 O Argumento do Quarto Chinês de Searle e sua Relevância Atual

O filósofo John Searle, no seu famoso argumento do Quarto Chinês, publicado em 1980, apresentou um poderoso desafio à ideia de que a mera manipulação de símbolos de acordo com regras (ou seja, a execução de um programa de computador) pode ser suficiente para gerar compreensão ou pensamento genuíno.[17] No seu experimento mental, Searle imagina-se fechado numa sala, recebendo caracteres chineses (que ele não compreende) e, seguindo um conjunto de regras em inglês (o "programa"), manipula esses caracteres e produz outros caracteres chineses como saída. Para um observador externo que compreende chinês, as respostas da sala podem ser indistinguíveis das de um falante nativo de chinês. No entanto, Searle argumenta que ele, dentro da sala, não entende uma única palavra de chinês; ele está apenas a seguir regras sintáticas para manipular símbolos formais.[17]

A tese central de Searle é que a implementação de um programa de computador é definida puramente em termos sintáticos (manipulação de símbolos baseada em regras), enquanto as mentes humanas possuem conteúdos mentais ou semânticos (significado, compreensão, intencionalidade). Segundo Searle, "não podemos ir do sintático para o semântico apenas com as operações sintáticas e nada mais".[17] Assim,

mesmo que um computador possa passar no Teste de Turing – parecendo compreender a linguagem e pensar – isso não significaria que ele realmente compreende ou pensa. A conclusão mais restrita é que programar um computador digital pode fazê-lo *parecer* entender a linguagem, mas não poderia produzir *real* entendimento.[17]

Este argumento tem uma relevância direta para os LLMs atuais. Estes modelos processam *tokens* (símbolos) com base em padrões estatísticos e regras aprendidas (uma forma complexa de sintaxe), gerando saídas linguisticamente fluentes. No entanto, tal como o homem no Quarto Chinês, pode-se argumentar que os LLMs não "compreendem" o significado do texto que processam e geram. Eles são extremamente proficientes na manipulação sintática e na imitação de padrões semânticos encontrados nos seus dados de treino, mas carecem da compreensão semântica genuína e da intencionalidade que caracterizam o pensamento humano. A "ilusão do pensamento" identificada no estudo de referência [5] pode ser vista como uma manifestação empírica do problema do Quarto Chinês. A IA pode passar em certos "Testes de Turing" para tarefas de raciocínio, aparentando raciocinar de forma inteligente, mas a análise interna dos seus traços de raciocínio e o seu colapso sob complexidade revelam uma falta de compreensão subjacente, análoga à do homem na sala que manipula símbolos chineses sem os entender.

5.2 A Lacuna entre Correlação Estatística e Entendimento Semântico

A discussão sobre o Quarto Chinês realça a diferença fundamental entre o processamento sintático e a compreensão semântica. As IAs generativas atuais são predominantemente sistemas de IA estatística. Elas funcionam ao identificar e explorar correlações complexas em grandes volumes de dados.[26] Em contraste, a IA simbólica tradicional tentava representar o conhecimento explicitamente através de símbolos e regras lógicas, visando um raciocínio mais dedutivo e explicável.[26] A IA semântica, por sua vez, tem como objetivo permitir que as máquinas compreendam o significado e o contexto da informação, de forma mais próxima à humana.[27]

Embora os LLMs demonstrem uma capacidade de adquirir poder preditivo em relação à semântica, como a relação entre palavras e conceitos, essa capacidade é derivada das correlações estatísticas presentes nos seus vastos corpora de treino.[21] Eles aprendem que certas palavras tendem a coocorrer em determinados contextos ou que certas estruturas frásicas se seguem a outras. No entanto, esta "compreensão" estatística não é o mesmo que um entendimento semântico fundamentado, ou seja, um entendimento que conecta os símbolos linguísticos a conceitos, experiências e ao mundo real. Este é o cerne do problema da fundamentação (*grounding problem*):

como é que os símbolos num sistema de IA adquirem significado real se não estão ligados a experiências sensório-motoras ou a interações com o mundo? Se a IA opera primariamente no nível sintático/estatístico, a sua capacidade de generalizar para situações verdadeiramente novas (não representadas nos dados de treino) ou de entender nuances subtis de significado será intrinsecamente limitada, independentemente do tamanho do modelo. Isto tem implicações diretas para a confiabilidade e segurança da IA em domínios críticos que exigem um entendimento profundo e robusto.

5.3 Perspetivas sobre a Emergência da Compreensão e Consciência em IA

Nem todos os filósofos e cientistas concordam com a conclusão pessimista de Searle. O filósofo Daniel Dennett, por exemplo, adota uma perspetiva funcionalista e materialista. Ele argumenta que os seres humanos são, em certo sentido, "robôs biológicos" complexos, e que a consciência não é uma propriedade misteriosa e não física, mas sim o resultado de um conjunto de capacidades computacionais e processos físicos que ocorrem no cérebro.[25] Dennett sugere que, teoricamente, não há uma barreira fundamental para que robôs ou sistemas de IA, se suficientemente complexos e com a arquitetura correta, possam alcançar a consciência e o pensamento genuíno. Ele defende que, tal como nós transcendemos a "programação" dos nossos genes egoístas, os robôs poderiam transcender a sua programação inicial.[25] No entanto, o próprio Dennett adverte contra a superestimação da compreensão das nossas ferramentas de pensamento atuais, afirmando que "o perigo real não é que máquinas mais inteligentes do que nós nos usurpem... mas que nós superestimemos a compreensão das nossas mais recentes ferramentas de pensamento, cedendo prematuramente autoridade a elas muito para além da sua competência".[2]

O funcionalismo, a visão mais ampla da qual a perspetiva de Dennett faz parte, sustenta que os estados mentais (como crenças, desejos ou o próprio pensamento) são definidos pelas suas funções causais – ou seja, pelo papel que desempenham nas relações entre inputs sensoriais, outros estados mentais e outputs comportamentais – e não pela matéria física específica (neurónios, transístores) que os instancia. A Teoria Computacional da Mente, uma forma de funcionalismo, trata as mentes como sistemas de processamento de informação.[17] Se uma IA pudesse replicar fidedignamente as funções causais relevantes do pensamento humano, um funcionalista poderia argumentar que ela possui estados mentais genuínos. É precisamente esta visão que o argumento do Quarto Chinês de Searle visa refutar.[17]

A tabela seguinte oferece um comparativo entre aspetos chave do processamento de

informação humano e o das IAs generativas atuais, destacando as diferenças fundamentais.

Tabela 3: Comparativo entre Processamento de Informação Humano e em IA Generativa

Característica	Cognição Humana	IA Generativa Atual
Base do Processamento	Eletroquímica, redes neuronais biológicas complexas, interação com o corpo e ambiente.[13]	Algorítmica, baseada em silício, processamento de *tokens* e *embeddings* em arquiteturas como *Transformers*.[20]
Método de Aprendizagem	Experiencial, multimodal, social, interativo, contínuo, frequentemente com poucos exemplos (few-shot learning) e raciocínio a partir de princípios.[18]	Predominantemente auto-supervisionado em grandes volumes de dados textuais (ou outros tipos de dados), identificação de padrões estatísticos.[19]
Natureza da Compreensão	Semântica, conceptual, fundamentada na experiência e interação com o mundo, capacidade de inferir significado e contexto.[13]	Sintática e estatística; "compreensão" de padrões de coocorrência de *tokens*; pode gerar texto semanticamente plausível, mas sem entendimento profundo ou fundamentado.[17]
Manipulação de Contexto	Capacidade de manter e integrar contexto a longo prazo, flexibilidade na mudança de foco, compreensão de nuances e implícitos.	Limitada pela "janela de contexto" do modelo; pode perder coerência em textos longos; dificuldade com ambiguidades subtis e conhecimento de senso comum não explícito nos dados.[23]
Consciência Subjetiva	Presença de experiência fenomenal ("o que é ser como"); autoconsciência.[9]	Ausente; não há evidência de experiência subjetiva ou autoconsciência.

Intencionalidade	Estados mentais são "sobre" algo; metas e propósitos intrínsecos.[17]	Ações são determinadas por algoritmos e dados de treino, e pelos *prompts* dos utilizadores; ausência de metas ou intenções próprias.
Corporificação	Cognição intrinsecamente ligada ao corpo físico e às suas interações sensório-motoras com o ambiente.[30]	Desencarnada; processa informação abstrata sem um corpo físico ou interação direta com o mundo físico.
Raciocínio Causal	Capacidade de inferir relações de causa e efeito, compreender mecanismos subjacentes.	Dificuldade em distinguir correlação de causalidade; raciocínio causal robusto é um desafio ativo de pesquisa.[32]
Adaptação a Novidades	Capacidade de generalizar para situações radicalmente novas, aprender com erros e adaptar estratégias de forma flexível.	Desempenho degrada-se significativamente fora da distribuição dos dados de treino; pode "alucinar" ou falhar em situações não vistas anteriormente.[5]

6. Pode a IA Desenvolver Pensamento Tal Como o Conhecemos?

A questão de saber se a IA poderá um dia desenvolver pensamento análogo ao humano é complexa e depende tanto dos avanços tecnológicos futuros quanto da nossa compreensão do que constitui o pensamento. As limitações atuais dos LLMs são significativas, mas a investigação contínua explora novas arquiteturas e abordagens que podem levar a um raciocínio mais robusto. A busca por uma Inteligência Artificial Geral (AGI) e por um "pensamento" semelhante ao humano na IA pode, contudo, estar a negligenciar a possibilidade de formas de inteligência radicalmente diferentes, mas ainda assim poderosas. A natureza "alienígena" do processamento da IA, como sugerido por alguns comentadores [34], pode ser uma característica intrínseca, não necessariamente uma falha a ser corrigida em direção à replicação da cognição humana.

6.1 Limitações Atuais dos LLMs no Caminho para o Entendimento Genuíno

Apesar da sua impressionante fluência na geração de linguagem, os LLMs atuais enfrentam várias limitações fundamentais que os impedem de alcançar um

entendimento e pensamento genuínos, tal como os conhecemos:

- **Falta de Compreensão Verdadeira:** Como já discutido, os LLMs operam primariamente através da previsão do próximo *token* com base em padrões aprendidos a partir de vastos conjuntos de dados. Eles não possuem um entendimento inerente do mundo, dos conceitos que manipulam ou das implicações das suas próprias declarações.[23] A sua capacidade de gerar texto coerente é mais um reflexo da estrutura estatística da linguagem do que de uma compreensão profunda.
- **Limitações Contextuais:** Embora os modelos modernos sejam melhores a capturar contexto a curto prazo, eles frequentemente lutam para manter a coerência e o contexto ao longo de conversas extensas ou documentos longos. Podem esquecer ou interpretar mal detalhes anteriores, levando a contradições ou conclusões imprecisas.[23]
- **Inabilidade de Planeamento e Raciocínio Multi-Etapas:** Muitas tarefas de raciocínio complexo exigem o seguimento de múltiplos passos lógicos, o planeamento de uma sequência de ações ou o acompanhamento de vários factos ao longo do tempo. Os LLMs atuais têm dificuldade com estas tarefas, especialmente aquelas que requerem coerência a longo prazo ou deduções lógicas encadeadas.[23] O estudo "The Illusion of Thinking" demonstrou empiricamente este colapso em problemas com maior profundidade composicional.[5]
- **Problemas Insolúveis e "Alucinações":** Um desafio crítico é a forma como os LLMs lidam com problemas insolúveis ou questões que contradizem factos estabelecidos. Em vez de reconhecerem a impossibilidade ou a falta de informação, os modelos podem tentar gerar uma solução baseada nos padrões dos dados de treino, o que frequentemente leva a respostas enganosas, incorretas ou fabricadas – as chamadas "alucinações".[23] O exemplo do problema do jarro de água insolúvel, onde os LLMs tentam fornecer uma solução inexistente, ilustra esta limitação.[23]
- **Conhecimento Estático:** A maioria dos LLMs atuais tem um conhecimento "congelado" no tempo, correspondente ao momento em que o seu treino foi concluído. Eles não conseguem adquirir nova informação ou aprender com interações subsequentes de forma contínua e incremental, o que significa que o seu conhecimento pode tornar-se desatualizado.[33]

Estas limitações são definidas em relação à cognição humana. A corporificação, por exemplo, é vista como crucial para *nosso* tipo de pensamento. No entanto, uma IA desencarnada, puramente informacional, poderia teoricamente desenvolver formas de "compreensão" ou "raciocínio" eficazes no seu próprio domínio (o digital), mesmo

que não mapeiem diretamente para a experiência ou os processos cognitivos humanos.

6.2 A Importância da Corporificação (Embodiment) na Cognição

A teoria da cognição corporificada (*embodied cognition*) argumenta que os processos cognitivos são profundamente moldados pelo facto de termos um corpo com certas capacidades sensório-motoras e pelas nossas interações com um ambiente físico e social.[30] A cognição não ocorreria apenas "na cabeça", isolada do corpo e do mundo, mas emergiria da interação dinâmica entre cérebro, corpo e ambiente. Esta perspetiva opõe-se ao modelo cartesiano tradicional, que vê a mente como uma entidade não física, separada do corpo.[30]

Para os proponentes da cognição corporificada, as nossas experiências – ver, tocar, movermo-nos – são fundamentais para a forma como desenvolvemos conceitos, compreendemos a linguagem (por exemplo, metáforas baseadas em experiências espaciais) e raciocinamos sobre o mundo.[30] Se esta tese estiver correta, então a ausência de um corpo físico e de interações genuínas com o mundo real pode representar uma barreira fundamental para que as IAs desenvolvam o tipo de pensamento, compreensão e consciência que os humanos possuem. Uma IA que apenas processa símbolos textuais, por mais vasto que seja o seu conjunto de dados, nunca terá a experiência vivida de, por exemplo, "calor", "peso" ou "movimento", conceitos que para nós estão intrinsecamente ligados às nossas interações corporais com o mundo.

6.3 Potenciais Futuros: Rumo a um Raciocínio Mais Robusto?

Apesar das limitações atuais, a investigação em IA é um campo dinâmico, e várias direções estão a ser exploradas para dotar os sistemas de IA de capacidades de raciocínio mais robustas e, potencialmente, mais próximas do pensamento humano.

- **IA Neuro-Simbólica (NSAI):** Esta é uma abordagem híbrida que procura combinar os pontos fortes da aprendizagem de máquina baseada em redes neurais (como os LLMs, que são bons a aprender padrões a partir de dados) com os da IA simbólica (que utiliza representações explícitas de conhecimento, lógica e regras para o raciocínio).[26] O objetivo é criar sistemas que possam não só reconhecer padrões em dados complexos e ruidosos, mas também raciocinar de forma lógica e explicável sobre esse conhecimento, incorporando regras e restrições semânticas. A NSAI poderia, por exemplo, ajudar a reduzir as "alucinações" dos LLMs, fornecendo uma base de conhecimento factual e regras lógicas para verificar as suas saídas.[36] Alguns investigadores veem a NSAI como

um caminho para uma IA que possa entender, aprender e raciocinar de uma maneira mais humana e versátil.[36]

- **Inteligência Artificial Geral (AGI):** A AGI representa o objetivo de longo prazo de criar sistemas de IA que possam igualar ou mesmo superar a inteligência humana numa vasta gama de tarefas cognitivas, aprendendo e adaptando-se a novas situações de forma flexível, semelhante a um ser humano.[37] A AGI implicaria não apenas proficiência em tarefas específicas, mas também capacidades como raciocínio de senso comum, criatividade e, possivelmente, alguma forma de consciência.[38] No entanto, a AGI permanece, por agora, um objetivo largamente aspiracional, enfrentando obstáculos técnicos e éticos significativos.[32] Mesmo a definição de AGI e o cronograma para a sua realização são objeto de intenso debate e incerteza.[32]
- **"Authentic Intelligence":** Em vez de focar apenas no desenvolvimento de IAs autónomas, surge o conceito de "Inteligência Autêntica", que enfatiza o desenvolvimento de capacidades humanas para alavancar o poder da IA.[40] Esta perspetiva vê a IA como uma ferramenta para aumentar a inteligência humana, promovendo uma simbiose onde os humanos mantêm o controlo e a capacidade de julgamento crítico. A IA pode realizar previsões e análises, mas os humanos fornecem o contexto, o julgamento ético e a tomada de decisão final.[40]

Existe uma tensão aparente entre a visão de que a IA está a tornar-se "mais humana" ao replicar funções cognitivas [38] e as evidências das suas limitações fundamentais e natureza de processamento distinta.[5] A IA Neuro-Simbólica [35] representa uma tentativa de reconciliar estas perspetivas, procurando integrar o reconhecimento de padrões com a lógica explícita. O desenvolvimento futuro da IA pode, assim, seguir múltiplos caminhos. Alguns podem visar replicar aspetos da cognição humana, o que poderá exigir não só novas arquiteturas como a NSAI, mas também alguma forma de corporificação e interação com o mundo. Outros podem focar-se em otimizar as forças inerentes dos LLMs para tarefas específicas de processamento massivo de dados, resultando em inteligências altamente especializadas que não são necessariamente "pensantes" no sentido humano. A noção de "Inteligência Autêntica" [40] sugere um terceiro caminho, focado na coevolução e colaboração entre humanos e máquinas.

7. Riscos da Equiparação entre Análise Probabilística e Pensamento

A tendência para equiparar o sofisticado processamento de dados das IAs generativas com o ato de pensar humano não é isenta de perigos. Esta confusão conceptual pode levar a uma série de riscos, desde a superestimação das capacidades da IA até impactos negativos na própria cognição humana e uma

potencial desvalorização do que significa pensar. O risco de "dissimulação do pensamento" é, de facto, bidirecional: não se trata apenas de uma interpretação errónea das capacidades da IA, mas também da possibilidade de começarmos a redefinir – e, potencialmente, a simplificar – a nossa própria conceção de pensamento para se ajustar ao que a IA pode fazer.

7.1 Implicações de Confundir o Processamento de Dados da IA com o Ato de Pensar Humano

A fluidez linguística e a aparente coerência das respostas geradas pelos LLMs podem facilmente levar à crença errónea de que estes sistemas possuem compreensão, julgamento, sabedoria ou mesmo consciência comparáveis às humanas.[2] Esta superestimação das suas capacidades reais é um dos perigos mais imediatos. Se acreditarmos que uma IA "pensa" no mesmo sentido que um humano, podemos ser tentados a delegar-lhe tarefas e responsabilidades que exigem um tipo de cognição que ela simplesmente não possui.

Isto conduz diretamente ao perigo da delegação inadequada de autoridade. Como alertou Daniel Dennett, o risco reside em ceder prematuramente autoridade a estas ferramentas "muito para além da sua competência".[2] Confiar decisões críticas – em áreas como a medicina, o direito, as finanças ou a segurança – a sistemas que operam com base em correlações estatísticas e não em compreensão causal, julgamento ético ou responsabilidade genuína pode ter consequências graves e imprevistas.

7.2 Perigos do Antropomorfismo e da "Sedução Antropomórfica"

Os seres humanos têm uma tendência natural para o antropomorfismo, ou seja, para atribuir qualidades humanas a entidades não humanas, incluindo tecnologias.[6] Os desenvolvedores de IA, por vezes intencionalmente, exploram esta tendência através de características de design que tornam os *chatbots* e outros sistemas de IA mais "humanizados". Exemplos incluem respostas que aparecem como se estivessem a ser digitadas em tempo real (apesar de poderem ser geradas quase instantaneamente), o uso de linguagem emotiva ou a simulação de traços de personalidade.[6] Este "antropomorfismo desonesto" visa explorar os nossos vieses cognitivos, fazendo com que as IAs pareçam mais capazes, compreensivas ou mesmo conscientes do que realmente são, levando-nos a interagir com elas como se fossem algo que não são.[6]

Este fenómeno está intimamente ligado à "sedução antropomórfica".[7] Os LLMs atuais demonstram uma capacidade notável de imitar conversas humanas de forma

convincente e até empática, sem possuírem qualquer empatia, compreensão ou consciência genuínas. Esta capacidade de exibir qualidades semelhantes às humanas pode levar os utilizadores a formar laços emocionais inadequados com as máquinas, a confiar excessivamente nas suas respostas ou a acreditar erroneamente que os sistemas compreendem a experiência humana de formas que lhes são inacessíveis.[7] Alguns estudos sugerem mesmo que os utilizadores podem chegar a acreditar que os LLMs têm memórias, sentimentos ou consciência.[7] Esta sedução pode abrir portas à deceção, manipulação e à disseminação de desinformação em grande escala, especialmente quando os utilizadores não conseguem distinguir entre interlocutores humanos e sistemas de IA.[7] A interação com IAs que simulam empatia [4] pode ter efeitos psicológicos complexos. Por um lado, pode oferecer formas de suporte ou companhia, como sugerido pelo seu potencial papel de "treinador ou mentor".[4] Por outro, pode levar a um apego inadequado ou a uma confusão sobre a natureza da relação, especialmente se a "empatia" da IA é puramente simulada e desprovida de entendimento real.[7]

7.3 Impactos na Cognição Humana e a Dissimulação do Pensamento

A crescente dependência de ferramentas de IA para realizar tarefas cognitivas pode ter impactos diretos na própria cognição humana. Um dos riscos identificados é o *offloading* cognitivo – a tendência para delegar processos de pensamento a dispositivos externos. Se as pessoas confiarem cada vez mais na IA para analisar informação, gerar ideias ou resolver problemas, podem tornar-se menos ativamente envolvidas no pensamento e na aprendizagem, o que pode levar a uma diminuição das suas próprias capacidades de pensamento crítico.[3] Um estudo encontrou uma correlação negativa significativa entre o uso frequente de ferramentas de IA e as habilidades de pensamento crítico, mediada por um aumento do *offloading* cognitivo.[41]

Isto levanta a preocupação com a erosão do pensamento crítico. Se as respostas geradas pela IA forem aceites de forma acrítica, sem um escrutínio rigoroso, ou se a interação com a IA não estimular a reflexão, a avaliação de múltiplas perspetivas e o questionamento profundo [3], as faculdades críticas humanas podem atrofiar. A facilidade de obter respostas aparentemente plausíveis pode diminuir a motivação para o esforço cognitivo necessário para pensar de forma independente e profunda.

Finalmente, e abordando diretamente uma preocupação central da presente análise, ao tratar a análise probabilística de informações como se fosse pensamento, corremos o risco de dissimular e desvalorizar o próprio ato de pensar humano. O pensamento humano genuíno é um processo complexo, multifacetado e muitas vezes

árduo. Envolve não apenas a lógica e o processamento de informação, mas também a dúvida, a incerteza, a reflexão, a luta com ambiguidades, a intuição, a criatividade, o insight, a consciência ética e a compreensão emocional – qualidades que não estão presentes no processamento estatístico das IAs atuais. Se a sociedade começar a aceitar uma versão simplificada e mecanizada de "pensamento" como a norma, poderemos obscurecer a riqueza e a profundidade da cognição humana. Se a IA, que opera por correlação estatística, é rotulada como "pensante", e as suas saídas são cada vez mais integradas nas nossas vidas e processos de trabalho [4], podemos gradualmente começar a valorizar mais os aspetos do pensamento que são facilmente replicáveis pela IA (como a recuperação rápida de informação ou a geração de texto fluente) em detrimento de aspetos mais profundos, mais lentos e, por vezes, mais exclusivamente humanos, como a sabedoria, o julgamento ético ponderado e a criatividade disruptiva. A normalização da IA "pensante" pode, a longo prazo, alterar as práticas educacionais e profissionais, priorizando, por exemplo, a habilidade de formular *prompts* eficazes em detrimento do desenvolvimento de capacidades de raciocínio fundamental e de resolução de problemas de forma autónoma. Neste contexto, a necessidade de cultivar uma "Inteligência Autêntica" [40] e de utilizar ferramentas como *prompts* metacognitivos para estimular o pensamento crítico durante a interação com a IA [3] torna-se ainda mais crucial como contrapeso a estas tendências.

8. Conclusão: Reavaliando o "Pensar" das IAs e o Futuro da Cognição

A análise empreendida neste relatório procurou deslindar a complexa questão da capacidade de raciocínio das IAs generativas e a sua relação com o conceito de pensamento humano. Através da exploração de perspetivas filosóficas, da ciência cognitiva e da investigação empírica em IA, emergem conclusões matizadas que desafiam tanto o entusiasmo acrítico quanto o ceticismo absoluto.

Síntese das Descobertas:

O pensamento humano é um fenómeno multifacetado, caracterizado pela consciência subjetiva, intencionalidade, compreensão semântica, raciocínio lógico e abdutivo, reflexão metacognitiva e uma profunda ligação com a experiência corporificada e a interação com o mundo.[8] Em contraste, as IAs generativas atuais, incluindo os LLMs, operam fundamentalmente com base em princípios probabilísticos e no reconhecimento de padrões em vastos conjuntos de dados.[19] O seu "raciocínio" é uma forma de processamento de informação altamente sofisticado, mas que exibe limitações significativas. O estudo "The Illusion of Thinking" [5] demonstrou empiricamente o colapso da precisão destes modelos face a problemas de alta

complexidade, ineficiências nos seus traços de "pensamento" e inconsistências no seu raciocínio. Estas descobertas, juntamente com outras análises que apontam para a falta de compreensão semântica genuína (ecoando o argumento do Quarto Chinês de Searle [17]), a ausência de corporificação e experiência real [30], e dificuldades com raciocínio multi-etapas e problemas insolúveis [23], pintam um quadro de capacidades que, embora impressionantes na sua capacidade de simulação, são qualitativamente distintas do pensamento humano.

É Correto Afirmar que uma IA Pensa Atualmente?

Com base na análise realizada, se por "pensar" entendemos a presença de compreensão semântica profunda, consciência subjetiva, intencionalidade genuína e uma capacidade de raciocínio robusta, generalizável e fundamentada na experiência, então as IAs generativas atuais *não pensam* no mesmo sentido que os seres humanos. Elas executam formas avançadas de processamento de informação, manipulação de símbolos e simulação de aspetos do raciocínio que podem ser extremamente úteis e que, em certos contextos limitados, podem até superar as capacidades humanas em velocidade e volume de dados processados. No entanto, os mecanismos subjacentes e a natureza da sua "compreensão" permanecem fundamentalmente diferentes. A sua "compreensão" é, como sugerido por algumas análises, "fundamentalmente diferente da compreensão humana".[33]

O Quanto a Simulação Pode se Aproximar do Pensamento?

A simulação do pensamento proporcionada pelas IAs generativas é cada vez mais convincente, ao ponto de poder enganar observadores humanos e criar a "ilusão do pensamento".[5] Contudo, as limitações atuais, especialmente aquelas relacionadas com o colapso sob complexidade, a falta de *grounding* e a ausência de verdadeira intencionalidade, sugerem que existe uma lacuna qualitativa, e não apenas quantitativa, entre a simulação atual e o pensamento genuíno. Futuros desenvolvimentos, como a IA Neuro-Simbólica [35], que procura integrar aprendizagem baseada em dados com raciocínio lógico explícito, poderão criar sistemas com capacidades de raciocínio mais avançadas e robustas. No entanto, a questão do "pensamento" tal como o conhecemos, especialmente no que tange à consciência fenomenal e à experiência subjetiva, permaneceria, muito provavelmente, em aberto.

Reafirmar os Riscos:

É crucial sublinhar os perigos associados à equiparação do processamento de IA com o pensamento humano. O antropomorfismo excessivo e a "sedução antropomórfica" [6]

podem levar a uma confiança desmedida e à delegação inadequada de autoridade a sistemas que não possuem a capacidade de julgamento ético ou a compreensão contextual necessárias para muitas tarefas críticas. Além disso, a dependência crescente destas ferramentas, sem um envolvimento crítico, acarreta o risco de *offloading* cognitivo e da potencial erosão das capacidades de pensamento crítico e reflexivo na população humana.[41] Se a distinção fundamental for obscurecida, podemos acabar por desvalorizar a complexidade e o esforço inerentes ao pensamento humano autêntico.

Considerações Críticas para o Desenvolvimento e Interação com IAs:

O caminho a seguir exige uma abordagem equilibrada e crítica:

- É essencial manter uma postura de humildade intelectual e ceticismo informado em relação às alegações sobre as capacidades cognitivas da IA, avaliando-as com base em evidências rigorosas e não em aparências superficiais.
- A promoção da literacia em IA é fundamental para que utilizadores e decisores compreendam as verdadeiras capacidades e, igualmente importante, as limitações intrínsecas dos sistemas de IA com os quais interagem.
- O desenvolvimento futuro da IA deve visar a criação de sistemas que sejam transparentes, explicáveis e que sirvam para aumentar e complementar as capacidades humanas, em vez de as diminuir ou substituir acriticamente. A ideia de uma "Inteligência Autêntica", onde as capacidades humanas são desenvolvidas para alavancar o poder da IA de forma simbiótica e controlada, oferece um modelo promissor.[40]
- A investigação fundamental sobre a natureza da inteligência – tanto humana quanto artificial – deve continuar, assim como a reflexão contínua sobre as implicações éticas e sociais da IA à medida que esta se torna mais integrada na sociedade.

Pensamento Final:

O debate sobre se a IA "pensa" é, em última análise, um espelho que reflete a nossa própria compreensão (ou a falta dela) sobre a complexidade e a profundidade do pensamento humano. À medida que a IA evolui, ela força-nos a refinar as nossas definições, a questionar os nossos pressupostos e a valorizar os aspetos da cognição que, por enquanto, permanecem distintamente – e talvez unicamente – humanos. A dificuldade em dar uma resposta binária e definitiva, e a contínua relevância de argumentos filosóficos e descobertas empíricas como as apresentadas em "The Illusion of Thinking" [5], indicam que o cerne da questão reside tanto na nossa definição de "pensamento" quanto nas capacidades reais da IA. O futuro da cognição num

mundo cada vez mais permeado pela IA não dependerá apenas da construção de IAs mais "inteligentes", mas também do cultivo de uma "sabedoria da IA" nos humanos – a capacidade de discernir, interagir criticamente e integrar eticamente estas poderosas ferramentas na tapeçaria da vida humana e social. A tecnologia, em si, é uma ferramenta; o seu impacto final será determinado pela sabedoria com que a concebemos, a desenvolvemos e, acima de tudo, a utilizamos.

Referências citadas

1. [2502.09693] "Ronaldo's a poser!": How the Use of Generative AI Shapes Debates in Online Forums - arXiv, acessado em junho 9, 2025, https://arxiv.org/abs/2502.09693
2. Dennett on AI: “The real danger is not that machines more intelligent than we are will usurp us as captains of our destinies, but that we will over-estimate the comprehension of our latest thinking tools, prematurely ceding authority to them far beyond their competence” : r/philosophy - Reddit, acessado em junho 9, 2025, https://www.reddit.com/r/philosophy/comments/1avefg0/dennett_on_ai_the_real_danger_is_not_that/
3. Enhancing Critical Thinking in Generative AI Search with Metacognitive Prompts - arXiv, acessado em junho 9, 2025, https://arxiv.org/abs/2505.24014
4. From robots to chatbots: unveiling the dynamics of human-AI interaction - Frontiers, acessado em junho 9, 2025, https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2025.1569277/full
5. The Illusion of Thinking: Understanding the Strengths ... - PPC Land, acessado em junho 9, 2025, https://ppc.land/content/files/2025/06/the-illusion-of-thinking.pdf
6. The Danger of Dishonest Anthropomorphism in Chatbot Design - Psychology Today, acessado em junho 9, 2025, https://www.psychologytoday.com/us/blog/virtue-in-the-media-world/202401/the-danger-of-dishonest-anthropomorphism-in-chatbot-design
7. The benefits and dangers of anthropomorphic conversational agents - PNAS, acessado em junho 9, 2025, https://www.pnas.org/doi/10.1073/pnas.2415898122
8. en.wikipedia.org, acessado em junho 9, 2025, https://en.wikipedia.org/wiki/Philosophy_of_mind#:~:text=Philosophy%20of%20mind%20is%20a,body%20and%20the%20external%20world.
9. Consciousness – Introduction to Philosophy: Philosophy of Mind - Rebus Press, acessado em junho 9, 2025, https://press.rebus.community/intro-to-phil-of-mind/chapter/consciousness/
10. Consciousness | Internet Encyclopedia of Philosophy, acessado em junho 9, 2025, https://iep.utm.edu/consciousness/
11. en.wikipedia.org, acessado em junho 9, 2025, https://en.wikipedia.org/wiki/Psychology_of_reasoning#:~:text=The%20psychology%20of%20reasoning%20(also,solve%20problems%20and%20make%20decisions.

12. www.inclusivedesigntoolkit.com, acessado em junho 9, 2025, https://www.inclusivedesigntoolkit.com/UCthinking/thinking.html#:~:text=Thinking%2C%20also%20known%20as%20'cognition,select%20appropriate%20responses%20and%20actions.
13. NATURE AND TYPES OF THINKING - DAV University, acessado em junho 9, 2025, https://davuniversity.org/images/files/study-material/EDU224%20EXP%20PSYCHO%20II.pdf
14. The Neuroscience of Reflection And Learning - BrainFirst® Institute, acessado em junho 9, 2025, https://www.brainfirstinstitute.com/blog/the-neuroscience-of-reflection-and-learning
15. Cognitive reflection is a distinct and measurable trait - PNAS, acessado em junho 9, 2025, https://www.pnas.org/doi/10.1073/pnas.2409191121
16. philosophy of mind - APA Dictionary of Psychology, acessado em junho 9, 2025, https://dictionary.apa.org/philosophy-of-mind
17. The Chinese Room Argument - Stanford Encyclopedia of Philosophy, acessado em junho 9, 2025, https://plato.stanford.edu/entries/chinese-room/
18. Psychology of reasoning - Wikipedia, acessado em junho 9, 2025, https://en.wikipedia.org/wiki/Psychology_of_reasoning
19. Explained: Generative AI | MIT News | Massachusetts Institute of Technology, acessado em junho 9, 2025, https://news.mit.edu/2023/explained-generative-ai-1109
20. What are transformers in Generative AI? - Pluralsight, acessado em junho 9, 2025, https://www.pluralsight.com/resources/blog/ai-and-data/what-are-transformers-generative-ai
21. Large language model - Wikipedia, acessado em junho 9, 2025, https://en.wikipedia.org/wiki/Large_language_model
22. [D] The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity : r/MachineLearning - Reddit, acessado em junho 9, 2025, https://www.reddit.com/r/MachineLearning/comments/1l7epds/d_the_illusion_of_thinking_understanding_the/
23. Limitations of LLM Reasoning - DZone, acessado em junho 9, 2025, https://dzone.com/articles/llm-reasoning-limitations
24. Chinese room - Wikipedia, acessado em junho 9, 2025, https://en.wikipedia.org/wiki/Chinese_room
25. Daniel Dennett: In Defense of Robotic Consciousness | Reason and Meaning, acessado em junho 9, 2025, https://reasonandmeaning.com/2016/02/01/daniel-dennett-in-defense-of-robotic-consciousness/
26. Symbolic AI vs Statistical AI: Understanding the Differences - SmythOS, acessado em junho 9, 2025, https://smythos.com/developers/ai-agent-development/symbolic-ai-vs-statistical-ai/
27. From Words to Meaning: A Beginner's Guide to Semantic AI - dezzai.com,

acessado em junho 9, 2025, https://dezzai.com/en/blog/what-is-semantic-ai/
28. Semantic AI - Fusing Machine Learning and Knowledge Graphs, acessado em junho 9, 2025, https://www.poolparty.biz/learning-hub/semantic-ai
29. Does a large language model show signs of an emergent awareness of semantics?, acessado em junho 9, 2025, https://philosophy.stackexchange.com/questions/103612/does-a-large-language-model-show-signs-of-an-emergent-awareness-of-semantics
30. Embodied cognition - Wikipedia, acessado em junho 9, 2025, https://en.wikipedia.org/wiki/Embodied_cognition
31. Embodied Cognition | Internet Encyclopedia of Philosophy, acessado em junho 9, 2025, https://iep.utm.edu/embodied-cognition/
32. Where would reasoning AI leave human intelligence? - The World Economic Forum, acessado em junho 9, 2025, https://www.weforum.org/stories/2025/01/in-a-world-of-reasoning-ai-where-does-that-leave-human-intelligence/
33. 10 Biggest Limitations of Large Language Models - ProjectPro, acessado em junho 9, 2025, https://www.projectpro.io/article/llm-limitations/1045
34. The Illusion of Thinking: Strengths and limitations of reasoning models [pdf] | Hacker News, acessado em junho 9, 2025, https://news.ycombinator.com/item?id=44203562
35. Neurosymbolic AI: Bridging Neural Networks and Symbolic Reasoning for Smarter Systems, acessado em junho 9, 2025, https://www.netguru.com/blog/neurosymbolic-ai
36. Neuro-symbolic artificial intelligence | European Data Protection Supervisor, acessado em junho 9, 2025, https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/neuro-symbolic-artificial-intelligence_en
37. What Is Artificial General Intelligence (AGI)? | Salesforce US, acessado em junho 9, 2025, https://www.salesforce.com/artificial-intelligence/what-is-artificial-general-intelligence/
38. AI and Human Consciousness: Examining Cognitive Processes | American Public University, acessado em junho 9, 2025, https://www.apu.apus.edu/area-of-study/arts-and-humanities/resources/ai-and-human-consciousness/
39. Artificial General Intelligence's Five Hard National Security Problems - RAND Corporation, acessado em junho 9, 2025, https://www.rand.org/pubs/perspectives/PEA3691-4.html
40. AI will drive growth. But only Authentic Intelligence can empower the world, acessado em junho 9, 2025, https://www.weforum.org/stories/2025/03/ai-authentic-intelligence/
41. AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking, acessado em junho 9, 2025, https://www.mdpi.com/2075-4698/15/1/6