O que a Inteligência Artificial nos ensina sobre a inteligência

Carlos Barth[9]

A falta que a inteligência nos faz

Em sua primeira onda, a Inteligência Artificial (IA) era uma empreitada dupla. Tratava-se de um projeto tanto científico quanto de engenharia. Os pesquisadores desenvolviam sistemas para simular capacidades mentais e os algoritmos resultantes funcionavam também como hipóteses empíricas para explicar o funcionamento dessas capacidades na mente humana. Não por acaso, a empreitada era frequentemente descrita como um estudo da mente por meio de engenharia reversa (Haugeland, 1997). Essa conexão direta entre sistemas computacionais e cognição biológica só era plausível pela ampla adesão à teoria computacional da mente, no interior do

[9] Pesquisador de pós-doutorado na FAJE - Faculdade Jesuíta de Filosofia e Teologia (MG). Doutor em Filosofia pela UFMG (Filosofia da inteligência artificial, filosofia da Mente e da filosofia das ciências cognitivas).

movimento que ficou conhecido como *cognitivismo*. Capacidades mentais poderiam ser analisadas funcionalmente, e essas funções poderiam ser então modeladas computacionalmente (Haugeland, 1998). Tais modelos serviriam tanto para compreender como a capacidade em questão pode se manifestar em substratos de carbono, quanto para realizá-la em substratos de silício, afinal, para o cognitivismo, ambos podem ser descritos como sistemas computacionais (Newell, 1976).

Nas últimas décadas, o cognitivismo em geral, e a teoria computacional da mente em particular, sofreram sucessivas críticas.[10] Isso não levou ao seu desaparecimento, mas desarticulou a conexão entre a dimensão científica e a dimensão engenheira da IA.[11] Encontrar arquiteturas computacionais que permitam simular feitos da cognição humana não é mais equivalente a formular hipóteses sobre como o aparato cognitivo humano é capaz de realizá-los. Como resultado, a dimensão científica perdeu força, e na IA contemporânea predomina o interesse em gerar aplicações úteis em vez de explicações.

Nesse cenário, a IA não precisa mais se submeter a limitações biológicas ou psicológicas. Ela pode, por exemplo, apelar a arquiteturas computacionais, poder de processamento e volumes de dados incompatíveis com a constituição humana. [12] Isso levou a um

[10] Houve, por exemplo, um reavivamento do interesse pela psicologia ecológica de Gibson (1979), e da ideia geral de que a cognição humana é melhor explicada em termos de sistemas dinâmicos (Van Gelder, 1995). Isso para não falar, claro, da emergência de variedades do programa enativista (Rolla, 2021; Varela; Rosch; Thompson, 1991). Ainda que significativamente distintos, estes programas partilham da aversão à ideia de que a mente é um sistema computacional.

[11] Para um exemplo de abordagem contemporânea da teoria computacional da mente, ver Piccinini, 2021.

[12] Mesmo nos primeiros anos da IA, já havia quem concentrasse seus estudos na inteligência "em si", sem preocupações com a plausibilidade psicológica dos

distanciamento progressivo da inteligência humana como o fenômeno a ser modelado, e a inúmeras tentativas de definir o que poderia ser colocado em seu lugar.

Essa indefinição faz da IA um projeto de engenharia peculiar. Quando um desenvolvedor de *software* se põe a desenhar um sistema para tratar de processos administrativos, contábeis ou jurídicos, ele pode contar com uma especificação clara e independente dos requisitos.[13] Essa especificação pode ser utilizada para avaliar se o sistema desenvolvido de fato alcança os objetivos almejados. Porém, no caso da inteligência, nada remotamente parecido com um conjunto de requisitos claros está disponível. Em que pode consistir, afinal, essa noção de inteligência que norteia o desenvolvimento dos sistemas da IA? Ao deixarmos de lado a inteligência humana como norte, permitimos que toda especificação flerte com os interesses pragmáticos e locais dos desenvolvedores.[14]

Considere, por exemplo, a grande quantidade de diferentes concepções sobre o que constituiria uma inteligência artificial geral (AGI, do inglês *Artificial General Intelligence*). Chollet (2019) argumenta que inteligência deve ser medida em termos de eficiência na aquisição de novas habilidades. Já a OpenAI, empresa por trás do ChatGPT, entende a AGI como um sistema autônomo que se sai melhor do que os seres humanos nas tarefas em que é aplicada (2026). Não faltam exemplos pitorescos, mas ilustrativos do cenário atual, como o de Satya Nadella, CEO da Microsoft, para quem a chegada da

algoritmos desenvolvidos. Contudo, à época isso resultava em distinções menos drásticas que as atuais, tais como a maior disponibilidade de tempo e memória.

[13] Ainda que, na prática, essa especificação nem sempre seja suficientemente clara, ela é ao menos possível.

[14] Hutter; Legg (2007) mapearam setenta concepções diferentes.

AGI será percebida a partir da aceleração do crescimento econômico (Bishop, 2025). Nesse cenário, sem surpresa alguma, há quem afirme que a AGI já foi alcançada (Norvig; Arcas, 2023) e quem a considere fora do alcance de qualquer tecnologia conhecida (Smith, 2019). Essa pluralidade de alvos torna difícil identificar avanços concretos na IA, pois, não raro, o que é descrito como um avanço significativo constitui na verdade uma revisão dos requisitos e objetivos.

Embora contemporaneamente a promessa de se alcançar uma AGI mobilize boa parte dos investimentos em IA, é comum que se tente desviar dessa polêmica reduzindo o escopo das aplicações. Em vez de mirar num único critério amplo, busca-se modelar capacidades associadas a domínios particulares. No interior desses domínios, é possível mensurar avanços por meio de *benchmarks*, conjuntos de testes que tentam capturar o desempenho humano típico em determinadas tarefas. Há, por exemplo, *benchmarks* dedicados a problemas matemáticos (Cobbe et al., 2021), reconhecimento de imagens (Deng et al., 2009) e mesmo senso comum (Talmor et al., 2019).

Não obstante, o problema central permanece. Em que medida é possível afirmar que uma melhor performance significa avanço rumo à emulação adequada da capacidade que inspira aquele conjunto de testes? Um desempenho superior num conjunto de testes matemáticos não necessariamente implica um avanço no projeto mais amplo de emular capacidades matemáticas. Afinal, não faltam exemplos de sistemas que se utilizam de atalhos ou heurísticas capazes de mimetizar os resultados adequados apenas em certos contextos. Um LLM (*Large Language Model*) como o ChatGPT, por exemplo, até consegue jogar xadrez, mas conforme a partida avança, aumentam as chances de ele realizar movimentos ilegais, deixando

claro que suas decisões não são guiadas sequer pelas regras do jogo. Além disso, embora possamos dizer que os testes sejam inspirados em manifestações de capacidades humanas, eles não necessariamente capturam aquilo que as constitui.

Permanece também em aberto a questão sobre qual a relação entre inteligência e tais capacidades cognitivas específicas. É a inteligência um mero agregado delas? Se sim, quais? Seria estranho tomar como essenciais as capacidades mensuradas por *benchmarks* matemáticos, por exemplo. Mesmo pessoas que nunca estudaram matemática são inteligentes, faltando-lhes apenas estudo nesse domínio. Ou talvez, em vez de um agregado, a inteligência seja um certo modo de navegar todos os domínios considerados, isto é, talvez exista um algoritmo de inteligência aplicável a qualquer domínio, tal como sugerido por Domingos (2015)?

Ainda não há resposta. De todo modo, a ausência de um norte claro fez com que a IA, inadvertidamente ou não, explorasse diferentes aspectos da nossa compreensão pré-teórica de inteligência ao longo de sua trajetória. A seguir, veremos alguns casos em que isso diferentes formas de pensar a inteligência humana tiveram efeito sobre o tipo de pesquisa produzida em IA.

Da inteligência humana para a IA

Historicamente, a IA não deu muita atenção à questão da inteligência porque ela demorou a constituir um problema prático. Em particular, a ampla adesão ao famoso *Teste de Turing* (Turing,

1950) permitiu aos pesquisadores se concentrar nas capacidades associadas à falar a coisa certa, na hora certa.[15] A ideia é sedutora: se e quando um computador for capaz de produzir *outputs* linguísticos tal qual um ser humano, a ponto de fazer um interlocutor humano pensar que está conversando com um igual, então ele já simulará tudo o que importa simular. O apelo à capacidade linguística é especialmente interessante porque a linguagem se acomoda facilmente em todos os contextos de atividade humana. Se alguém é capaz de se manifestar linguisticamente em um número aberto de situações, não é de todo implausível supor que esse alguém partilha das nossas capacidades cognitivas.

Contudo, a pertinência do teste de Turing está longe de ser clara. Quais seriam as inferências adequadas a fazer diante de alguém que não se deu conta de estar conversando com um *chatbot*? Faz sentido mensurar os avanços da empreitada a partir da capacidade de enganar um interlocutor? Apelar para a semelhança comportamental pode ser válido para alguns fins, em alguns contextos, mas certamente não todos. A inferência de (i) "P exibe comportamentos idênticos aos nossos" para (ii) "P possui mecanismos funcionalmente idênticos aos nossos" não é justificável sem um arcabouço empírico robusto, e há casos em que (i) é verdadeiro e (ii) é falso. Os LLMs são exemplos claros de que é possível a um sistema emular o comportamento verbal humano sem emular a inteligência humana, e isso sugere um limite para o teste de Turing enquanto norte para a IA, pois ele não joga luz sobre casos desse tipo.

[15] Mais precisamente, o que influenciou os pesquisadores foi uma certa interpretação dele como constituindo um teste empiricamente plausível, o que é polêmico. Para discussão, ver Gonçalves, 2024.

O isomorfismo de *outputs* comportamentais linguísticos, contudo, não é o único modo pelo qual uma concepção da inteligência "natural" influencia a concepção de inteligência de máquina que guia a IA. Há pelo menos dois outros movimentos que merecem destaque: 1) o modo como os modelos computacionais clássicos descritos na seção anterior foram substituídos por modelos neurais treinados via algoritmos de aprendizado (McClelland et al., 1987; Rumelhart et al., 1986); e 2) o abandono do cognitivismo clássico em virtude uma concepção *situada* da inteligência.

No caso (1), temos a adoção de arquiteturas computacionais razoavelmente distintas, vagamente inspiradas na estrutura neuronal do cérebro. Trata-se de uma abordagem ainda compatível com a teoria computacional da mente, mas que leva a uma mudança substancial no modo como os modelos são gerados. Se antes era praxe fazer uma análise funcional cuidadosa e elaborar modelos computacionais dessas funções, as arquiteturas neurais normalizaram a delegação desse trabalho para algoritmos de aprendizado. Tais algoritmos são expostos a um grande volume de dados e buscam mapear os padrões estatísticos e as diversas relações ali presentes. Essas relações entre diferentes objetos ou propriedades podem ser muito mais fracas do que as utilizadas nos modelos clássicos. Se antes tínhamos relativamente poucas relações "fortes" como "A implica B", agora podemos ter um número gigantesco de relações fracas como "A aumenta em 0,3% a chance de B". Além disso, as categorias sobre os quais essas relações são estabelecidas não são (necessariamente) dadas pelo desenvolvedor. Elas podem ser inferidas pelo próprio algoritmo de aprendizado, levando ao uso de categorias que nos são desconhecidas e mesmo opacas.

Por sua vez, no caso (2) temos uma rejeição de tudo o que é tipicamente associado ao cognitivismo. Trata-se de uma crescente influência oriunda das ciências cognitivas, em particular as *frameworks* que trabalham com cognição 4EA, segundo a qual a mente é incorporada, integrada ao ambiente, estendida, enativa e afetiva.[16] Embora a influência oriunda das ciências cognitivas seja relativamente recente, cognição situada e IA são velhas conhecidas. Elas tiveram seu primeiro contato muito antes de a cognição situada se estabelecer entre os cientistas cognitivos, a partir do trabalho de Dreyfus (1972), que se utilizou dessa ideia para criticar as pretensões da IA clássica ainda nos anos 1970.

Inspirado por autores da tradição fenomenológica como Heidegger e Merleau-Ponty, Dreyfus argumentava que ser inteligente é ser capaz de habitar um mundo. A ideia de mundo a que Dreyfus alude é fenomenológica: o que ele tem em mente não é uma porção do espaço-tempo, mas uma estrutura de significados. Essa estrutura tem um arranjo utensiliar, isto é, ela não articula objetos, propriedades e relações num quadro teórico, mas sim os usos que fazemos dos utensílios de que dispomos, conforme nossas necessidades e objetivos. Num exemplo clássico, um martelo aparece como algo que serve para pregar, que aparece como algo que serve para (por exemplo), construir uma cerca, que aparece como algo que serve para demarcar, que aparece como algo associado ao nosso anseio por segurança. A ideia central é a de que somos lançados num mundo assim estruturado e aprendemos a navegá-lo. Agir de forma inteligente, portanto, não é exercitar habilidades racionais, mas sim exercitar habilidades práticas. Esse *know-how*, e não nosso intelecto, é o que caracteriza

[16] Do inglês, *embodied*, *embedded*, *extended*, *enactive* e *affective*.

nossa relação primordial com o mundo e nossa inteligência. Trata-se, portanto, de uma rejeição radical de várias ideias caras ao cognitivismo.

Contemporaneamente, as ciências cognitivas incorporaram varias ideias inspiradas na tradição fenomenológica. Mas nem todos vão tão longe quanto Dreyfus, e certamente não é preciso rejeitar o uso de mecanismos computacionais na explicação da cognição situada (Clark, 1997; Piccinini, 2021 e 2022). Isso permite que essas ideias exerçam alguma influência sobre a concepção de inteligência que norteia pesquisas em IA. A despeito da pluralidade de concepções de inteligência a nortear os trabalhos atuais – bem como a poluição oriunda dos interesses econômicos, conforme discutido há pouco – alguns nichos de pesquisa na IA, como a robótica e os estudos de vida artificial, incorporam aspectos da tradição situada com maior ênfase.

Por sua vez, hoje (1) praticamente se confunde com a empreitada, a ponto de muitos só caracterizarem sistemas computacionais como IAs se estes forem modelos neurais gerados a partir de algoritmos de aprendizado). Tanto (1) quanto (2), portanto, são exemplos de mudanças no modo como pensamos a inteligência humana que levaram, em alguma medida, à revisão do modo como produzimos sistemas de IA. Mas e quanto à direção contrária? Podemos encontrar algo na IA que nos leva a refletir sobre a inteligência tal como se manifesta no humano? Embora nos encontremos num cenário compatível com uma profunda disjunção (exclusiva) entre o que é inteligência humana e inteligência "de máquina", nossa única referência pré-teórica é a inteligência enquanto um fenômeno que se manifesta no humano. Por um lado, isso faz com que a IA seja uma empreitada um tanto difícil de analisar. Por outro, a existência dessa referência pré-teórica permite, mesmo hoje, extrair

da IA lições sobre a inteligência humana, ainda que por uma via indireta.

Da IA para a inteligência humana

Os primeiros sistemas de IA foram concebidos e modelados utilizando formalismos lógicos. Buscava-se, por meio deles, modelar os estados de coisas no mundo (e.g. um objeto sobre a mesa) e as possíveis permutações (e.g. o objeto cai ao chão, o objeto é colocado de cabeça pra baixo, etc.). Esses modelos articulavam, invariavelmente, objetos, propriedades e relações, e todos esses elementos eram delineados em termos de condições necessárias e suficientes. Nesse cenário inicial, "calcular" o que deve ser feito para que se migre de um estado de coisas A até um estado de coisas B tinha um sentido muito próximo do empregado nos formalismos lógicos: o processo de concluir que B passaria a ser o caso após algum evento envolvia verificar se é possível provar um teorema que descreve B a partir de A, em conjunção com as demais premissas.[17] Nesse sentido, a inteligência era tomada como a capacidade de *resolver* problemas formais, e resolver um problema de modo adequado era equivalente a articular os objetos, as propriedade e as relações relevantes na forma de premissas, para então calcular os passos inferenciais necessários até a conclusão.

[17] Ver, por exemplo, Hayes, 1977 e McCarthy, 1980.

As limitações dessa abordagem logo começaram a aparecer, afinal, quanto mais próximo do real, menos claras e distintas as coisas se mostram. Qual exatamente o teorema que caracterizaria, por exemplo, a tarefa de determinar se um dado comportamento observado num aeroporto é suspeito? Diante de dificuldades como essa, as arquiteturas computacionais que usavam formalismos lógicos foram deixadas de lado em função de alternativas como mapas semânticos, *frames* (Minsky, 1997) e sistemas de produção (Newell, 1994). Contudo, nenhuma delas se mostrou suficientemente radical. Embora trouxessem algumas vantagens práticas para os desenvolvedores (menor curva de aprendizado, a possibilidade de expressar ideias de forma mais enxuta, etc.), nenhuma delas foi ao cerne da questão. Continuava necessário delinear cuidadosamente os objetos, as propriedades e as relações que caracterizavam as situações ou os problemas com os quais se lidaria, e o que tornava um processo mais (ou menos) inteligente não era a forma como essa modelagem era feita, mas sim a eficiência com que o sistema produzia inferências.

Dito de outro modo, a inteligência continuava a ser concebida como o processo de *resolver problemas* no interior de situações modeladas pelos desenvolvedores.

Isso levou ao encontro recorrente de duas dificuldades relacionadas: primeiro, a dificuldade de especificar estratégias "gerais" para realização de inferências, isto é, estratégias que guiem adequadamente o raciocínio em quaisquer situações. Segundo, a consequente necessidade de se limitar a soluções locais. Elas permitem ao sistema lidar adequadamente com um conjunto restrito de contextos, mas têm limitações severas na sua capacidade de extrapolação diante de novas circunstâncias.

O aspecto mais desafiador do primeiro problema é que, em qualquer dada situação, o conjunto de inferências *possíveis* é potencialmente ilimitado. Se o sistema tiver que considerar exaustivamente todos os caminhos inferenciais logicamente disponíveis, este será um processo sem fim ou, no mínimo, demandará uma quantidade de tempo que inviabilizaria o projeto. Ele precisa, portanto, se ater ao subconjunto de inferências que são pertinentes na situação presente. Mas como decidir quais inferências, dentre as possíveis, são pertinentes? Se for necessário verificar uma a uma, volta-se ao problema inicial, pois será preciso considerar exaustivamente todos os caminhos inferenciais possíveis a fim de delimitar os que são pertinentes. É preciso, portanto, aplicar critérios que permitam ao sistema selecionar os caminhos inferenciais com maiores chances de ser pertinentes, mas estes critérios são invariavelmente *locais* e não generalizáveis.

Para que possamos compreender melhor o que isso significa, suponhamos um caso em que é preciso encontrar um livro numa biblioteca gigantesca. Como podemos evitar a necessidade de checar todos os livros, um por um, até encontrar o livro desejado? Toda estratégia disponível depende de como os livros estão organizados. Podemos delimitar a busca observando o sobrenome do autor em ordem alfabética, mas se os livros estiverem organizados pela cor da capa, o resultado é um retorno à necessidade de verificar todos os livros, um por um. Além disso, há um número potencialmente ilimitado de estratégias a utilizar. Podemos classificar os livros por diferentes características da capa, pelo estado de conservação, pelas iniciais do título, pelas iniciais do autor, e assim por diante. O corolário é que, para evitar a necessidade de checar todos os livros, precisamos ter algum conhecimento prévio sobre como estão organizados.

Esse problema se generaliza: para qualquer tipo de situação, em qualquer domínio, a única forma de evitarmos a necessidade de uma consideração exaustiva de todas as possibilidades lógicas é fazendo uso de alguma informação prévia sobre o domínio. O preço a pagar, contudo, é que a estratégia adotada vale apenas para a situação específica em que a informação previamente disponível calhe de ser verdadeira. Por isso, quando um sistema é aplicado a situações não previstas, ele é incapaz de se adaptar, e isso aparece ao seu desenvolvedor como a necessidade de redesenhá-lo ou retreiná-lo.

Mesmo nos sistemas contemporâneos mais avançados (inclusive LLMs), esse desafio é mitigado não pelo desenvolvimento de estratégias cada vez mais gerais que permitiriam ao sistema exibir uma adaptatividade cada vez maior a situações novas. Em vez disso, o que temos é uma tentativa de acumular cada vez mais informações sobre um número cada vez maior de situações com as quais o sistema estará previamente familiarizado e tentando minimizar as chances de que ele encontre situações imprevistas nas quais sua dificuldade de adaptação venha à tona. Esta é a razão pela qual os sistemas mais avançados exigem *data centers* inteiros para si. O gigantismo desses sistemas se deve, em larga medida, à necessidade de acumular uma enorme quantidade de dados que permitam capturar o maior número possível de situações às quais o sistema poderá ser exposto.

Para um exemplo mais concreto, podemos considerar o caso dos veículos autônomos. Em larga medida, a razão de ainda não dispormos de veículos completamente autônomos está na dificuldade de ampliar o conjunto de contextos em que eles se mostram funcionais. Não por acaso, há uma tendência em "resolver" essa dificuldade constrangendo artificialmente a possibilidade de variações nos ambientes em que veículos autônomos operam, seja impedindo a presença de pedestres

ou ciclistas, seja atribuindo a eles ruas exclusivas. Quando um sistema é incapaz de se adaptar à complexidade do seu ambiente de operação, simplificar o ambiente é sempre uma opção.

Essas dificuldades nos levam a retomar o contraste com a inteligência humana. O ser humano consegue se adaptar forma fluida a um número aberto de situações, e fazemos isso com uma fração do poder de processamento de um *data center*. Como isso é possível? Esse é o cenário que tornou saliente a importância do *insight* na inteligência humana.

Em busca do *insight*

Uso o termo "insight" no mesmo sentido de Kaplan; Simon (1990). Não se trata de um súbito lampejo que coloca diante de nós a trajetória inferencial adequada para resolver um problema. Encontrar uma trajetória complexa que permita solucionar um problema igualmente complexo é tipicamente experienciado como o ápice de um período de grande concentração e esforço, como quando concluímos todos os passos que levam à solução de um problema matemático. Em contraste, o *insight* envolve a adoção de uma nova perspectiva, de um novo ponto de vista. Adotar essa perspectiva reduz a necessidade de grandes cadeias inferenciais, pois faz com que os elementos mais relevantes para a solução de um problema se tornem salientes. Nesse sentido, ter um *insight* é *inteligir* uma situação de um modo que nos permite explorar novas trajetórias inferenciais. Não se trata, portanto, de encontrar uma nova forma de solucionar um problema, mas sim uma nova forma de concebê-lo.

Um exemplo pode ajudar. Cukier; Mayer-Schoenberger; Vericourt (2021) contam a história de como Regina Barzilay, professora de IA no MIT, participou da descoberta de um novo antibiótico. Nessa pesquisa, os cientistas testaram mais de 2500 compostos químicos para verificar se algum deles inibia o crescimento da bactéria *E. coli*. Os compostos que se mostraram capazes de inibi-la, foram usados para treinar uma IA, que aprendeu a mapear e associar propriedades estruturais desses compostos à capacidade de combater a bacteria. Uma vez treinada, essa IA foi utilizada para realizar buscas em diferentes bases de dados de moléculas, a fim de encontrar aquelas que partilhavam das características estruturais previamente mapeadas. O resultado foi a identificação de um novo antibiótico capaz de tratar bactérias resistentes aos medicamentos até então disponíveis. [18]

O que nos interessa nessa história, contudo, não é o feito da IA, mas sim o de Barzilay. Sua descoberta só foi possível porque o problema foi concebido de uma nova forma. O desafio de encontrar novos medicamentos é tipicamente pensado como o problema de encontrar substâncias com moléculas similares à dos medicamentos existentes. No caso dos antibióticos, os resultados costumam ser limitados, tanto porque a maior parte das substâncias de composição similar à dos antibióticos conhecidos já foi analisada, quanto porque essa similaridade facilita o desenvolvimento de resistência por parte das bactérias. Barzilay reformulou a pergunta que guiava a pesquisa. Em vez de explorar o espaço de possíveis medicamentos a partir da pergunta "quais compostos têm estrutura semelhante à dos

[18] Ver Trafton, 2020.

antibióticos conhecidos?", esse espaço foi explorado a partir da pergunta "quais compostos se mostraram capazes de inibir a *E. coli*?"

Por trás de um aparente sucesso da IA, está um feito da inteligência humana: a capacidade de inteligir uma mesma situação, de múltiplas formas. A IA foi capaz de automatizar a busca pela solução, mas apenas porque um ser humano estruturou a situação e delimitou cuidadosamente o espaço de trajetórias inferenciais potencialmente relevantes para a investigação.

Apesar de o exemplo utilizado pertencer ao contexto científico, esse aspecto da inteligência está sempre operante em nossa lida com o mundo. Todo comportamento inteligente se dá no interior de uma situação, tal qual inteligida e estrutura por nós, de modo conforme a nossos interesses e expectativas. O que pode passar despercebido, contudo, é que mesmo situações repetitivas do cotidiano demandam essa abertura à reformulação contínua. A razão para isso é que nenhuma situação concreta é idêntica à outra. Podemos nos deslocar todos os dias para o mesmo ambiente de trabalho, realizar o mesmo tipo de tarefa, mas ainda assim estamos sempre atentos e abertos a mudanças no modo como compreendemos a situação ocorrente. Um colega de trabalho pode resolver desabafar conosco sobre um problema, e de repente precisamos articular e sopesar aspectos da vida profissional e pessoal. Devo dar ouvidos? É adequado continuar essa conversa em outro ambiente? Esse movimento é percebido como falta de profissionalismo, ou como o início de uma potencial amizade? Aquilo que consideraremos adequado numa situação como essa depende de como ela é compreendida por nós. Ao inteligi-la de uma forma ou outra, desdobramos certos cursos de ação e certas trajetórias inferenciais, ao mesmo tempo em que fechamos outras. Essa é a razão pela qual não precisamos do poder de processamento de um *data*

center inteiro para exibirmos inteligência em um número indefinidamente multiplicável de contextos.

Dizer que a inteligência humana envolve a capacidade de ter *insights* está longe de ser uma novidade, e certamente não precisaríamos recorrer à história da IA para afirmar isso. O que ela nos ajuda a enxergar, contudo, é a ubiquidade desse fenômeno. Longe de ser um feito incomum ou reservado a momentos de grande genialidade, o *insight* é um dos fundamentos do comportamento inteligente.

Não obstante, essa capacidade de revisar nossa compreensão e reestruturar o mundo de forma fluida e adaptativa é um ponto cego das pesquisas em IA. Via de regra, os pesquisadores ainda tratam da estrutura do mundo como uma condição de possibilidade para o exercício da inteligência, mas o modo sistemático e reiterado com que as dificuldades acima mencionadas se apresentam (influenciando as decisões de escopo de projeto, por exemplo), sugerem fortemente que essa estruturação não é uma condição *para*, mas uma capacidade *da* inteligência. Nesse sentido, a inteligência opera num espaço anterior ao tipo de raciocínio e inferência a que os pesquisadores normalmente se atentam, pois o que está em jogo não é o que fazer num mundo previamente estruturado, mas sim decidir como estruturá-lo de modo adaptativo.

Esse ponto cego não é insuperável. Se mais pesquisadores de IA dedicarem mais tempo e recursos à compreensão do *insight*, a IA pode vir a ser ainda mais informativa. Salvo por mudanças bruscas de cenário, ela provavelmente não irá modelar mecanismos que sirvam como hipótese para explicar o modo como o *insight* opera na inteligência humana. O que ela pode fazer, porém, é mostrar se ou como é possível simular o *insight* em sistemas computacionais.

A IA já fez isso antes, ao gerar sistemas capazes de jogar xadrez com maestria, e ao gerar LLMs capazes de produzir *output* linguístico adequado. Ambos, xadrez e linguagem, já foram tomados como desafios intransponíveis, pois acreditava-se que emular a inteligência *tal como se manifesta no humano* seria um requisito necessário. A IA nos ensinou muito sobre xadrez, e tem nos ensinado também sobre a linguagem. Talvez ela possa contribuir para nossa compreensão do *insight*.

Considerações finais

A história da IA é (também) a história de como tentamos inferir nossas capacidades a partir de nossos feitos, e de como seus pontos cegos são reveladores e informativos. Separar o mundo em objetos, propriedades e relações para então articulá-los de modo conforme a nossos interesses é um feito da inteligência humana, e não o arcabouço a partir do qual ela opera. Podemos perceber isso tanto na IA clássica quanto na IA contemporânea.

Na primeira, isso se dava pela suposição de que poderíamos tomar uma ontologia formal como dada. Inadvertidamente, partia-se do princípio de que a inteligência humana inteligia e estruturava o mundo inteiro de uma única forma.

Na segunda, isso se mostra na ideia de que podemos gerar essa mesma estrutura a partir de rastros dos nossos comportamentos passados. A diferença está apenas no fato de que não estamos mais restritos a ontologias formais, uma vez que o elenco de categorias, bem como sua articulação, fica a cargo dos algoritmos de aprendizado.

Esse cenário sugere que nossa busca pela compreensão do *insight* mal começou. Por isso, antes de finalizar, vale a pena compreender por que o *insight* é um desafio explicativo, tanto no âmbito dos modelos computacionais clássicos, quanto no âmbito dos modelos neurais.

A principal razão é que sua expressão algorítmica precisa escapar de uma circularidade: o que é pertinente para nós depende de como a situação em que nos encontramos é inteligida, mas como inteligimos a situação em que nos encontramos depende do que é pertinente para nós.[19]

E quanto à cognição situada? Ao rejeitar modelos computacionais, não estaria ela evitando esse problema? Infelizmente, não. Isso porque a abordagem da cognição situada, ao menos na forma defendida por Dreyfus, se ancora essencialmente numa familiaridade prévia com o mundo. Ela enfatiza a capacidade de navegar uma estrutura dada, mas o que está em jogo é a capacidade de operar com permutações dessa estrutura. Como vimos, o *insight* se caracteriza pela capacidade de ir além do que nos é familiar e lidar com as particularidades de situações nunca antes encontradas. Isso fez com que o próprio Dreyfus reconhecesse o *insight* como um desafio ainda fora do alcance, tanto do cognitivismo, quanto da abordagem situada (1987), e pouco ou nada mudou a esse respeito desde então.

[19] A manifestação mais clara dessa circularidade na IA se dá no *frame problem*. Para um tratamento extenso do problema, ver Barth, 2024.

Referências bibliográficas

Barth, C. (2024) *Representational Cognitive Pluralism: Towards a Cognitive-Science of Relevance-Sensitivity*. Tese de Doutorado – Belo Horizonte: Faculdade de Filosofia e Ciências Humanas, Universidade Federal de Minas Gerais.

Bishop, T. (2025) *Microsoft CEO Satya Nadella has a Formula to Gauge the Long-Term Success of AI Investments*. Disponível em: <https://www.geekwire.com/2025/microsoft-ceo-satya-nadella-has-a-formula-to-gauge-the-long-term-success-of-ai investments/>. Acesso em: 16 fev. 2026.

Chollet, F. (2019) On the Measure of Intelligence. *arXiv :1911.01547v2*.

Clark, A. (1997) *Being There: Putting Brain, Body, and World Together Again*. Cambridge: MIT Press.

Cobbe, K. et al. (2021) Training Verifiers to Solve Math Word Problems. *arXiv : 2110.14168*.

Cukier, K.; Mayer-Schoenberger, V.; Vericourt, F. (2021) *Framers: Human Advantage in an Age of Technology and Turmoil. WH Allen.*

Deng, J. et al. (2009) *ImageNet: A Large-Scale Hierarchical Image Database*. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Domingos, P. (2015) *The Master Algorithm: How the Quest for the Ultimate Learning Machine will Wemake our World*. New York: Basic Books.

Dreyfus, H. (1972) *What Computers Can't do: a Critique of Artificial Reason*. New York: Harper & Row.

Dreyfus, H. L.; Dreyfus, S. E. (1987) How to stop worrying about the frame problem even though it's computationally insoluble. In:

Pylyshyn, Z. W. (Ed.). *The Robot's Dillemma: The Frame Problem in Artificial Intelligence*. Ablex, pp. 95-111.

Gibson, J. J. (1979) *The Ecological Approach to Visual Perception*. Houghton Mifflin.

Gonçalves, B. (2024) *The Turing Test Argument*. Taylor & Francis.

Haugeland, J. (1997) *Mind Design II: Philosophy, Psychology, Artificial Intelligence*. MIT Press.

Haugeland, J. (1998) The Nature and Plausibility of Cognitivism. In: Haugeland, J. (Ed.). *Having Thought*. Cambridge: Harvard University Press. pp. 9-45.

Hayes, P. J. (1977) In Defence of Logic. *Proceedings of the IJCAI*.

Hutter, M.; Legg, S. (2007) A Collection of Definitions of Intelligence. *arXiv : 0706.3639*.

Kaplan, C. A.; Simon, H. A. (1990) In Search of Insight. *Cognitive Psychology*, v. 22, n. 3, pp. 374-419, jul.

Mccarthy, J. (1980) Circumscription: a Form of Non-Nonotonic Reasoning. *Artificial Intelligence*, v. 13, n. 1, pp. 27-39.

Mcclelland, J. L. (1987) et al. *Parallel Distributed Processing, Vol. 2: Psychological and Biological Models*. Cambridge, MA: MIT Press.

Minsky, M. (1997) A Framework for Representing Knowledge. In: Haugeland, J. (Ed.). *Mind Design II: Phylosophy, Psychology, Artificial Intelligence*. MIT Press. pp. 111-142.

Newell, A. (1994) *Unified Theories of Cognition*. Harvard University Press.

Newell, H. A.; Allen, S. (1976) Computer Science as Empirical Inquiry: Symbolsand Search. *Communications of the ACM*, v. 19, 1 mar.

Norvig, P.; Arcas, B. A. Y. (2023) Artificial General Intelligence is Already Here. *Noema*. Disponível em:

<https://www.noemamag.com/artificial-general-intelligence-is-already-here/>. Acesso em: 24 mar. 2025.

OpenAI. OpenAI Charter. Disponível em: <https://openai.com/charter/>. Acesso em: 16 fev. 2026.

Piccinini, G. (2021) *Neurocognitive Mechanisms: Explaining Biological Cognition*. Oxford University Press.

Piccinini, G. (2022) Situated Neural Representations: Solving the Problems of Content. *Frontiers in Neurorobotics*, v. 16, abr.

Rolla, G. (2021) *A mente enativa*. Editora Fi.

Rumelhart, D. E. et al. (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations I.* Cambridge, MA: MIT press.

Smith, B. C. (2019) *The Promise of Artificial Intelligence: Reckoning and Judgment*. MIT Press.

Talmor, A. et al. (2019) *Commonsense QA: A Question Answering Challenge Targeting Commonsense Knowledge. Proceedings of the 2019 Conference of the North*. Association for Computational Linguistics. Disponível em: http://dx.doi.org/10.18653/v1/N19-1421.

Trafton, A. *Artificial Intelligence Yields New Antibiotic*. Disponível em: <https://news.mit.edu/2020/artificial-intelligence-identifies-new-antibiotic-0220>. Acesso em: 16 fev. 2026.

Turing, A. (1950) Computing Machinery and Intelligence. *Mind*, v. LIX, n. 236, pp. 433-460.

Van Gelder, T. (1995) What Might Cognition Be, If Not Computation? *The Journal of Philosophy*, v. 92, jul.

Varela, F. J.; Rosch, E.; Thompson, E. T. (1991) *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.