http://doi.org/10.7213/1980-5934.32.055.DS06
ISSN 1980-5934

Artefatos biológicos artificiais: do modelo Imitativo de Inteligência Artificial ao Advento de organismos vivos programados

Artificial biological artifacts: from the Imitative model of Artificial Intelligence to the Advent of programmed living organisms

KLEBER BEZ BIROLO CANDIOTTO[a]

Resumo

O programa de pesquisa em Inteligência Artificial (IA), desde sua origem, tem a inteligência humana como modelo, e sua reprodução (ou superação) como escopo. Com o objetivo de sustentar a necessidade de um novo modelo de investigação sobre inteligência, a presente pesquisa destaca inicialmente o caráter imitativo da IA clássica mediante as críticas de Searle e Dreyfus ao projeto da IA. Procuramos sustentar que uma abordagem mais ampla de inteligência, como a sugerida por Bickhard, que tem como referência os sistemas físicos autossustentáveis recursivamente, partindo da investigação do funcionamento de inteligências primitivas, é a perspectiva mais produtiva para a IA. Para isso, apresentamos o advento dos "*xenobots*", os organismos vivos programados, como um marco nesta perspectiva de pesquisa em IA que se desprende do modelo imitativo de inteligência, sem desconsiderar a dimensão interacionista dos organismos vivos, para promover uma "inteligência biológica artificializada".

[a] Pontifícia Universidade Católica do Paraná (PUCPR), Curitiba, PR, Brasil. Doutor em Filosofia, e-mail: kleber.candiotto@pucpr.br

Palavras-chave: Inteligência Artificial. Modelo Imitativo. Abordagem Interacionista.

Abstract

The research program of Artificial Intelligence has, since its origin, held human intelligence as a model, and its reproduction as its scope. In order to support the need for a new model of research in intelligence, the present research reduces the imitative characters of classic AI using criticisms of Searle and Dreyfus to the AI project. We seek to support a more comprehensive approach to intelligence, as suggested by Bickhard. The latter holds recursively self-sustaining physical systems as a reference, starting from the investigation of the functioning of primitive intelligences, and a more productive perspective for an AI. For this, we submit the arrival of xenobots, the living programs of organisms, as a milestone in this particular perspective of research within AI. It shows the imitative model of intelligence, without disregarding the interactionist dimension of living beings, to promote "artificilized biological intelligence".

Keywords: *Artificial Intelligence. Imitative model. Interactionist Approach.*

Considerações Iniciais

A publicação de Sam Kriegman, Douglas Blackiston, Michael Levin e Josh Bongard, intitulada *A scalable pipeline for designing reconfigurable organisms*, de 13 de janeiro de 2020, na revista *Proceedings of the National Academy of Sciences* (PNAS), anuncia perspectivas revolucionárias para a pesquisa em Inteligência Artificial com o presságio de uma nova classe de artefatos. O resultado fundamental desta pesquisa é viabilização de novas máquinas vivas, que não são nem robôs tradicionais e nem alguma forma de vida conhecida. Trata-se de organismos vivos e programáveis que é resultado do potencial de simulação de protótipos vivos pela máquina computacional e sua incorporação para formar novos organismos vivos, porém, programados biologicamente desde o início.

O presente artigo procura analisar os possíveis avanços para o projeto da Inteligência Artificial[1] promovidos pela confecção desta nova categoria de artefatos, a qual poderá se tornar o marco de uma nova perspectiva em IA. Para tanto, será

[1] Doravante IA, para se referir ao programa de pesquisa científica que tem como escopo a artificialização da inteligência humana.

elaborado inicialmente uma caracterização do programa de pesquisa em IA que, em que pese seu intenso e rápido desenvolvimento, mantém constante sua referência à inteligência humana. Aspetos das críticas de Searle e Dreyfus à IA serão os fundamentos para esta caracterização. Em seguida, sugere-se uma perspectiva mais ampla de inteligência, que tem como referência os sistemas físicos autossustentáveis recursivamente, partindo da investigação do funcionamento de inteligências primitivas, conforme estudos de Bickhard. Por fim, ao abordarmos a nova categoria de artefatos viabilizada por Kriegman, Blackiston, Levin e Bongard, os organismos vivos e programáveis, trataremos sobre a mudança de perspectiva deste invento para o âmbito da IA, em especial, sua contribuição para aumento de inteligência nas funções biológicas existentes, ou seja, uma "inteligência biológica artificializada".

Modelo imitativo da IA

O otimismo promovido pelos resultados científicos da primeira metade do século XX sobre computação e cognição ensejou projetos de replicação, ou até de ampliação da capacidade cognitiva humana. Surge, assim, um programa de investigação científica denominado de Inteligência Artificial, termo cunhado por John McCarthy no seminário de Darthmouth em 1956, com a presença de Marvin Minsky, Claude Shannon, Allen Newell, Herbert Simon, entre outros pioneiros da ciência da computação (HANDERSON, 2007, p. 44).

Esse programa de investigação tinha como escopo identificar as condições estruturais e materiais para artefatos artificiais realizarem comportamentos considerados inteligentes. Com isso, a Inteligência Artificial suscitou implicações muito além de seu propósito de engenharia advindo das ciências da computação. No âmbito da filosofia, o maior desafio tem sido discutir sobre a possibilidade de as máquinas produzirem pensamento e, assim, viabilizar a construção de uma mente artificial. É o que sugere Allan Turing, em seu revolucionário artigo de 1950, *Computing machinery and intelligence*, que buscou provar a possibilidade de uma máquina exibir comportamento inteligente equivalente aos humanos, no teste que recebe seu

nome[2]. Para Turing, "a nova formulação do problema pode ser descrita em termos de um jogo a que nós chamamos 'jogo da imitação'" (1950, p. 434). Mais do que imitar, a máquina poderia ampliar a capacidade cognitiva humana, uma vez que, como a força física humana se expandiu em grande escala com o surgimento das máquinas a vapor da Revolução Industrial, também a atividade cognitiva de pensar racionalmente, capacidade até então restrita aos humanos, poderia ser ampliada pelas máquinas computacionais.

Do comportamento da máquina se deduz sua possível inteligência. Todavia, vale destacar que a dedução só é possível desde que o comportamento seja similar ao humano, ou com base nos resultados obtidos por humanos. Por outro lado, o pensamento é concebido sob o aspecto computacional, ou seja, sob uma perspectiva de processamento de dados representados simbólica e binariamente.

O teste de Turing parece ser simplista, porém, sua tese do pensamento como computacional foi determinante para o surgimento e aprimoramento da ciência cognitiva e do programa de pesquisa da Inteligência Artificial. O incipiente computador da década de 1950, com o pioneiro Alan Turing, não possuía tecnologia suficiente para imitar o comportamento humano, resultante de um processo inteligente, capaz de passar no seu teste[3]. Apesar disso, sua concepção adiantou vários desafios das décadas seguintes, entre eles, os pressupostos

[2] Essencialmente, o Teste de Turing consistia na tentativa de um grupo de avaliadores, numa sala, descobrir se as respostas dadas às suas perguntas a dois monitores eletrônicos na parede (um monitor conectado a um computador e outro, a um humano) vinham de um computador ou de um humano. Se os avaliadores não soubessem distinguir com evidências se a resposta fora dada por um computador ou por uma pessoa, significaria que a máquina passara no teste, logo, seria possível afirmar que esta máquina exibiria comportamento inteligente.

[3] O primeiro a passar no desafio do "jogo da imitação" proposto por Turing em 1950 foi um supercomputador desenvolvido na Universidade de Reading em Londres, em 2014, ao convencer os juízes que a máquina era um garoto de 13 anos chamado Eugene Goostman. Fonte: <https://www.theguardian.com/technology/2014/jun/08/super-computer-simulates-13-year-old-boy-passes-turing-test>. Acesso: 12 de agosto de 2019. Nesta matéria, há um importante relato do professor Kevin Warwick, da Universidade de Reading: "No campo da inteligência artificial, não há marco mais icônico e controverso do que o teste de Turing. É apropriado que um marco tão importante tenha sido alcançado na Royal Society em Londres, o lar da ciência britânica e o cenário de muitos grandes avanços na compreensão humana ao longo dos séculos. Esse marco será registrado na história como um dos mais emocionantes".

ontológicos e as opções metodológicas de um campo de investigação próprio denominado *ciências cognitivas*[4].

Mesmo considerando todo o avanço científico que a ciência cognitiva proporcionou desde Turing até o final do século XX, a IA como um projeto de imitação da inteligência humana foi objeto de consistentes críticas. Entre os filósofos contrários às pretensões de reprodução da inteligência humana destacam-se John Searle e Hubert Dreyfus.

Em filosofia, já é tradicional a classificação de Searle quanto ao projeto de Inteligência Artificial: um projeto *fraco*, que considera o computador uma referência metafórica ou um instrumento que simula situações aparentemente racionais; e outro *forte*, a IA Forte, que toma o computador como a base para a construção de uma mente artificial. Neste último sentido, as inteligências humana e artificial são tomadas igualmente como processamento de informação.

A crítica de Searle (1980; 1999) é dirigida à IA Forte. Na sua concepção, a linguagem computacional está adstrita aos processos sintáticos e, por isso, não atinge os processos semânticos da linguagem genuinamente humana. Searle argumentou, com o célebre experimento mental da "sala chinesa"[5], que a imitação proposta pelo teste de Turing é insuficiente para sustentar a tese de que as máquinas pensam. Em suma, o experimento leva em conta uma pessoa numa sala, de posse de um dicionário chinês-inglês contendo todas suas regras gramaticais. Sua tarefa é traduzir uma folha do chinês para o inglês. Esta pessoa, no entanto, não fala e não entende chinês, somente faz a substituição dos ideogramas para o inglês. Após concluir a tradução, o indivíduo dentro da sala a envia por uma janela a outra pessoa do lado de fora, que não sabe nada a respeito do que se passa no interior da sala. A

[4] Gardner (2003, p. 20) sistematizou os fundamentos epistemológicos das ciências cognitivas: 1) atividades cognitivas humanas devem ser tomadas como representações mentais, com a criação de um nível de análise distinto do biológico e do cultural; 2) o computador é o modelo de compreensão do funcionamento da mente humana; 3) no nível de análise cognitivo, fatores não cognitivos, tais como emocionais, culturais ou históricos, devem ser isolados; 4) as pesquisas em ciências cognitivas devem ser interdisciplinares, dada a complexidade de seu entendimento; 5) antigas questões da agenda filosófica são retomadas, mas com um aporte empírico.

[5] Esse experimento mental foi introduzido por Searle em 1980, com seu artigo *Minds, brains and programs* e aprofundado em suas publicações subsequentes.

questão, portanto, é: a pessoa fora da sala que recebeu a folha com o conteúdo em inglês pode afirmar categoricamente que há no interior da sala um falante da língua chinesa? Para Searle, a resposta é não. O que acontece no interior da sala é apenas uma manipulação de símbolos com base em regras ajustadas, sem conhecimento de seu significado. Esta é a mesma condição dos computadores que, segundo Searle, apenas manipulam símbolos, sem qualquer domínio de seu significado. O projeto forte da Inteligência Artificial, com base neste argumento do quarto chinês, estaria, por princípio, equivocado, visto que a máquina computacional se restringe tão somente a seguir regras formais previamente programadas para manipular símbolos, sem qualquer compreensão dos seus significados.

A versão fraca da Inteligência Artificial, por sua vez, fornece comparações produtivas para o entendimento dos processos mentais, muito embora não seja uma representação fiel da capacidade mental humana, que é dotada de consciência. Por essa razão, a consciência passou a ser entendida por Searle como um fenômeno natural exclusivo dos humanos, passível de compreensão com base em seu *status* ontológico subjetivo.

Na perspectiva de Searle, os programas computacionais jamais podem ser tomados como idênticos a processos mentais, uma vez que são puramente formais e sintáticos. Assim, "as mentes são mais do que sintáticas. As mentes são semânticas, no sentido de que possuem mais do que uma estrutura formal, têm um conteúdo" (SEARLE, 1997, p. 39). Além das habilidades sintáticas, a inteligência humana é composta também por habilidades semânticas, que não são derivadas das sintáticas, uma vez que pressupõem intencionalidade[6], conceito central na agenda de pesquisa de Searle.

Searle dirigiu sua crítica à pesquisa sobre IA da década de 1980. Naquele contexto, ainda não havia a Rede Mundial de Internet (WWW) e Turing era a referência principal para a computação. É importante notar que o modelo de computação de Turing era de caráter imitativo: a máquina seria "pensante" se

[6] Para Searle, intencionalidade é "a propriedade de muitos estados e eventos mentais pela qual eles são dirigidos para ou acerca de objetos e estados de coisas no mundo" (SEARLE, 1995, p. 1).

passasse no teste, ou seja, se um testador não conseguisse apontar categoricamente a identidade (máquina ou humano) da entidade emissora da resposta.

Antes de Searle, Dreyfus já havia advertido sobre os limites da IA, em seu memorável artigo *What Computers Can't Do* [7] . Sua crítica direciona-se fundamentalmente ao pressuposto de igualdade entre humanos e computadores para a simulação da mente, uma vez que o modo de pensar e agir dos humanos não segue regras rígidas, inviabilizando, assim, sua formalização e operacionalização por uma máquina digital. A principal diferença elucidada por Dreyfus é a constatação de que, em suas ações, humanos aplicam senso comum e distinções pragmáticas que nem sempre são traduzíveis para uma linguagem formal. Por isso, a principal limitação da IA clássica era estar limitada a regras restritas para realização de tarefas específicas, o que levaria ao problema da regressão ao infinito: regras de procedimento para aplicação de regras, que precisam também de outras regras, e mais regras para estas, num processo ao infinito. Por esta razão, torna-se infecunda a tese de que humanos possuem um algoritmo de procedimento inconsciente.

Na visão computacional, "'saber que' é entendido como fundamental, e todas as outras habilidades e comportamentos inteligentes — tudo, desde a compreensão da linguagem até o reconhecimento de rostos — são considerados simplesmente 'problemas de complexidade'" (DREYFUS, 1992, p. 55). Para Dreyfus, portanto, a visão computacional clássica não é meramente falsa, mas inversa. O conhecimento procedural (saber-como) é o que fundamentalmente permite os humanos resolverem seus problemas contextuais cotidianos, e não apenas mediante raciocínio formal. É o primeiro que torna possível o segundo.

O maior desafio do projeto da IA provavelmente seja incluir noções de senso comum na máquina computacional, uma vez que tais noções não são obtidas pelo mero acúmulo de informação e domínio de regras (programação). O desafio está alicerçado no fato de que nem todo conhecimento pode ser transferível pela linguagem: como uma pessoa pode aprender a dirigir um carro apenas com a leitura

[7] Sua primeira versão ocorreu em 1972 e foi revisado em 1979. Em 1992, seu texto foi reeditado pelo MIT Press sob o título *What Computers Still Can't Do: A Critique of Artificial Reason*. Na "Introdução à edição revisada", Dreyfus ressalta que o livro permaneceu praticamente intacto desde sua primeira versão de 1972, com apenas pequenas alterações e novas introduções a cada nova edição.

das instruções e informações de um manual, sem a interação com o veículo e o ambiente? Contudo, ao que parece, não é a capacidade de raciocínio de senso comum que faz surgir a inteligência, mas sim o mecanismo inteligente que consegue acumular o conhecimento responsável pelo senso comum.

O argumento central de Dreyfus é que, contrariamente às pretensões formalistas[8], o significado é intrinsicamente dependente do contexto, sendo que a dependência do contexto, em princípio, não pode ser formalizada, posto que os contextos são inerentemente indeterminados. Para ser formalizada, a experiência humana deveria ser organizada a partir de critérios de relevância e significação previamente aplicados, o que requer interpretação do contexto em tantos outros subcontextos, em uma nova regressão infinita, inviabilizando a construção de uma hierarquia de situações e regras gerais para a ação. Esse problema é conhecido na linguística e na pesquisa em IA como o *frame problem*. Trata-se de determinar relevância das regras aplicáveis ao contexto, algo como "enquadrar" a regra. "Enquadrar", neste caso, significa determinar o contexto apropriado para compreendê-lo e fazer isso equivale a determinar o que é e o que não é relevante para o seu significado.

Para determinar o contexto mais apropriado para a compreensão de um fenômeno, exige-se apelo a outro contexto maior, e assim por diante. Por isso, o tratamento formal do problema incorre à regressão infinita. Por exemplo, o som da campainha do apartamento leva em conta uma sequência de contextos, como o tipo do som (é uma campainha ou um som da TV?), quem atende (eu atendo? Sou o proprietário? Há alguém próximo da porta?), conveniência de atender (estou esperando alguém? Qual o horário? Há riscos?), forma de atendimento (pergunto antes quem é? Como reconhecer?), o atendimento mais adequado (quais padrões culturais devem ser levados em conta?), e assim por diante. Determinar o contexto apropriado para a compreensão de um fenômeno, inevitavelmente exige apelo a

[8] A tese fundamental dos formalistas é que a cognição necessita fundamentalmente de representações. Por isso, os formalistas também são comumente considerados representacionistas. É importante destacar que críticos como Dreyfus não se preocupam em negar a existência de representações mentais, mas rechaçam a tese de que estas sejam os componentes fundamentais da cognição.

outro contexto maior. Um exemplo desta tentativa equivocada dos formalistas, argumenta Dreyfus (1993, p. 57), vem da ciência cognitiva ao tentar construir leis psicológicas, ou intencionais, com a aplicação de cláusulas *ceteris paribus*[9] para amenizar a rigidez do sistema. No entanto, a cláusula *ceteris paribus* seria também uma regra formal pois necessita também da representação do conhecimento humano, o que significa que em qualquer situação específica nunca pode ser totalmente explicitado sem regressão.

Por isso, tratar formalmente o problema leva sempre à regressão infinita de contextos. A existência do *frame problem* leva a concluir que os formalistas entenderam mal a natureza da inteligência ao conceberem que criaturas inteligentes são confrontadas *com* situações, sendo que, ao certo, os humanos estão *nas* situações, em que mundo e agentes inteligentes estão imbricados. Este mundo, na perspectiva heideggeriana[10] de Dreyfus, não é composto apenas de fatos, mas também de comportamentos, intenções, tendências, práticas e habilidades, formando o saber-como, que funciona como um pano-de-fundo da ação humana. Por isso, fragmentar em contextos para classificá-los e ordená-los, como pretendiam os formalistas, incorre no afastamento da maneira como a inteligência humana funciona. Esta leva em conta a condição do corpo no mundo, que seleciona pré-reflexivamente o que é relevante entender para, consequentemente, agir. Humanos estão imersos no mundo em uma rede de conexões e práticas que o tornam familiar.

O enfrentamento[11] com o mundo requer mais construir senso comum do que empregar teorias e analisá-las. A teorização ocorre posteriormente e tem a função de guiar algumas formas de ação deliberada, quando o emprego automático

---

[9] Segundo Jerry Fodor (1991, p. 20), o objetivo destas leis psicológicas, ou intencionais, é definir mecanismos computacionais (isto é, regras ou algoritmos formais) que expliquem as leis intencionais. Todas essas leis, no entanto, devem conter condições *ceteris paribus*, ou seja, elas são necessariamente "não rígidas" ou aplicam se "todo o resto é igual".

[10] Para uma explanação de sua perspectiva heideggeriana, ver Dreyfus (2007, p. 247-268).

[11] Três anos antes de seu falecimento, Dreyfus publicou *Skillful Coping: Essays on the Phenomenology of Everyday Perception and Action*. Editada por Mark Wrathall, esta obra compila ensaios de Dreyfus ao longo de sua carreira acadêmica, atualizando suas argumentações, mas reafirmando a insuficiência da IA para reproduzir ou imitar a inteligência humana. Nesta obra, é fundamental a noção de "enfrentamento hábil" (*Skillful Coping*), como sugere o título de capa, que é uma maneira de ser e de agir, na qual a pessoa está imersa em suas ações, de modo que não está pensando ou refletindo.

do senso comum pode ser ineficaz ou insuficiente, como um cálculo matemático puro, por exemplo. Por outro lado, um exímio jogador de xadrez humano não decide a melhor jogada analisando milhares de jogadas possíveis e examinando as consequências de todas as ações identificadas. Em vez disso, o especialista em xadrez[12] pode levar apenas alguns segundos para decidir o que fazer, valendo-se da intuição como uma forma altamente desenvolvida de bom senso para descartar certos movimentos desde o início.

As objeções de Dreyfus à IA apenas se fortaleceram ao longo de seu programa de pesquisa em filosofia, levando-o a considerar superada a questão, conforme sua declaração em entrevista a Nicholas Fearn, em 2006: "Eu não penso mais em computadores. Eu acho que venci e acabou: eles desistiram" (DREYFUS in FEARN, 2006, p.52). A posição de que mentes humanas e computadores funcionam de maneiras muito diferentes perdura até o fim de sua vida.

[12] É clássico referenciar o xadrez para ilustrar o desafio da capacidade de cálculo entre humanos e computadores, dada suas infinitas formas de jogadas, como atesta Diego Rasskin-Gutman (apud SILVER, 2012, p. 217): "Existem mais jogos de xadrez possíveis do que o número de átomos no universo". A memorável vitória do Deep Blue, o supercomputador da IBM, sobre o campeão russo Garry Gasparov, em 1997, pareceu mostrar a supremacia da máquina na arte do jogo de xadrez. Contudo, conforme Nate Silver, houve um erro no funcionamento do programa Deep Blue: "o erro surgiu na quadragésima quarta jogada de seu primeiro jogo contra Kasparov; incapaz de selecionar uma jogada, o programa adotou o padrão de segurança à prova de falhas de último recurso, no qual ele selecionava uma jogada completamente aleatória. O bug foi inconsequente, chegando no final do jogo em uma posição que já havia sido perdida. De fato, o bug não foi nada lamentável para o Deep Blue: foi provavelmente o que permitiu ao computador vencer Kasparov. Na recontagem popular da partida de Kasparov contra o Deep Blue, foi o segundo jogo em que seus problemas se originaram - quando ele cometera o erro quase sem precedentes de perder uma posição que ele provavelmente poderia ter empatado. Mas o que havia inspirado Kasparov a cometer esse erro? Sua ansiedade em relação ao quadragésimo quarto movimento de Deep Blue no primeiro jogo - o movimento em que o computador mudou sua torre sem nenhum objetivo aparente. Kasparov concluiu que o jogo contra-intuitivo deve ser um sinal de inteligência superior. Ele nunca considerou que era simplesmente um bug. Por mais que confie na tecnologia do século XXI, ainda temos os pontos cegos de Edgar Allan Poe sobre o papel que essas máquinas desempenham em nossas vidas. O computador fez Kasparov piscar, mas apenas por causa de uma falha de design. [...] Os computadores são muito, muito rápidos em fazer cálculos. Além disso, pode-se contar com eles para calcular fielmente - sem se cansar ou se emocionar ou mudar seu modo de análise no meio do caminho. [...] Enquanto isso, os computadores não são muito bons em tarefas que exigem criatividade e imaginação, como planejar estratégias ou desenvolver teorias sobre o funcionamento do mundo" (SILVER, 2012, p. 236-237).

As críticas de Searle e Dreyfus são dirigidas a um projeto de IA dependente da noção de inteligência humana. Bostrom (2014, p. 3) lembra que desde a década de 1940 já havia o propósito de criar máquinas equivalentes a seres humanos em inteligência que associa senso comum[13] e habilidades para aprender e raciocinar em uma ampla gama de domínios naturais e abstratos. Nesta época, Turing considerava que computadores no futuro poderiam aprender mediante sua interação com o meio. Contudo, o modelo de "máquina pensante" existente até a década de 1990 ainda era imitativo, centrado num modelo digital de inteligência humana, o que tem gerado críticas filosóficas como as de Searle e Dreyfus.

Abordagem Interacionista de Bickhard

Uma abordagem mais geral da inteligência se mostra como cada vez mais necessária para avanços teóricos eficazes em IA. É o que sugere a abordagem interacionista de Mark Bickhard (2004, 2009a, 2009b) em suas pesquisas sobre robótica cognitiva. Em sua perspectiva, a inteligência passa a ser compreendida com base em uma acepção mais ampla, para além da corporeidade e inteligência humanas. Com fundamento na teoria das estruturas dissipavas de Prigogine, ou seja, de sistemas termodinamicamente abertos, que operam em condições de não equilíbrio (ou também chamados de distante equilíbrio — *far-from-equilibrium system*), Bickhard desenvolveu uma teoria da cognição e da inteligência a partir de estruturas e funções de processos e sistemas físicos, sendo os seres humanos um destes sistemas (SUSSER, 2013, p.282).

Bickhard (2009b, p. 551) fundamenta sua argumentação numa metafísica de processos [14], a qual considera diretamente estabilidades do processo e sua

[13] Para uma compreensão dos resultados recentes sobre a realização de senso comum em IA, ver SHANAHAN (2016) com a noção de *lei da inércia do senso comum*, entendendo que o ser humano é capaz de determinar o que é ou não relevante na tomada de decisão para sua ação, mas ainda há dificuldade para saber como é possível *modelá-la*. Para evitar a longa busca de infinitas regras, o senso comum parece incluir uma cláusula *ceteris paribus* cuja aplicação varia conforme o contexto.

[14] Sobre a importância de uma metafísica de processos, Bickhard (2009b, p. 565) argumenta: "A mudança para uma metafísica de processo, no entanto, induz grandes

persistência no tempo. São basicamente duas classes muito amplas de estabilidades de organização do processo:

Estável: a instância da organização permanece estável, desde que uma quantidade acima do limite de energia colide com ela. É o que ocorre com grande parte de objetos do mundo. Uma rocha, por exemplo, sem forças externas, vai se manter estável e pode persistir por durações cosmológicas. É importante destacar que estas instâncias de organização podem ser isoladas de seus ambientes sem perturbar essa estabilidade, pois ficarão em seu equilíbrio termodinâmico.

Longe do equilíbrio termodinâmico: estas instâncias de organizações, se forem isoladas, entram em equilíbrio e deixam de existir. Por isso, para se manterem, dependem de sua manutenção em condições longe do equilíbrio. Essa manutenção pode se dar de duas formas: b.1) *A partir de fontes externas ao processo*, como bombas que fornecem um fluxo contínuo de produtos químicos para um tanque; ou, b.2) *Com automanutenção*: mediante auto-organização, com processos contribuem para sua própria estabilidade.

Entre os sistemas mencionados, uma classe específica para os propósitos de entendimento de cognição e inteligência é daqueles que contribuem para a manutenção de suas próprias condições longe de equilíbrio, como ocorre com a chama da vela, por exemplo. A chama mantém a temperatura acima do limiar de combustão, o que faz com que a cera derreta, possibilitando a chama adentrar no pavio, que leva à vaporização da cera no pavio queimando, o que, em condições atmosféricas e gravitacionais normais, induz convecção térmica (uma forma de

---

mudanças em nossa estrutura geral de suposições: - Primeiro, a mudança se torna o padrão explicativo, e é a estabilidade que requer explicação. Da mesma forma, processos, diferentemente dos átomos ou do 'material' de substâncias, não têm limites inerentes, e também os limites, portanto, devem ser explicados, não assumidos. - Segundo, os processos têm seus poderes causais em virtude de sua organização. A organização não pode ser deslegitimada como um possível *locus* de poder causal sem eliminar toda causalidade do universo. Mas, se a organização é um local potencial de poder causal, o mesmo ocorre com a organização de nível superior. Em particular, não há bloqueio metafísico para a possibilidade de poder causal emergente na nova organização. - Terceiro, se a emergência é uma possibilidade metafísica [...], então a porta está aberta para a possibilidade de que a normatividade e a mente sejam emergentes. Isso desfaria a divisão metafísica de dois domínios que persiste por mais de dois milênios."

propagação de calor nos líquidos), que traz novo oxigênio e elimina o desperdício. A chama de vela apresenta uma situação de automanutenção. Contudo, só lhe resta queimar, pois não possui opções e não pode selecionar entre as opções. Se ficar sem cera ou se não houver oxigênio, por exemplo, não há como a chama corrigir essa ameaça à sua existência contínua. "Sistemas de manutenção automática mais sofisticados, no entanto, têm opções e podem fazer seleções de acordo com as condições variáveis de seus ambientes, a fim de corrigir ou compensar essas condições variáveis" (BICKHARD, 2009b, p. 562).

Um sistema longe de equilíbrio, como uma chama de vela, quando isolado, fica sem oxigênio ou cera e alcança o equilíbrio, deixando, assim, de existir. Bactérias, por exemplo, que se deslocam em um gradiente de sacarose, buscam regiões com maior concentração de açúcar para sua alimentação. O consumo do açúcar leva à diminuição da concentração e isso poderia levar ao equilíbrio do sistema com o consequente desaparecimento das bactérias. Todavia, diferentemente da chama da vela, as bactérias podem identificar a diferença e ir em busca de novas regiões concentradas, mantendo sua existência. É desta automanutenção que se caracterizam os sistemas dependentes das condições termodinâmicas longe do equilíbrio, em que seus subsistemas estão em constante relação, um utilizando o estado do outro para promover comportamentos que viabilizem a manutenção de suas próprias condições longe de equilíbrio em dada situação.

Há, portanto, uma função, que é manter as condições de um sistema para que fique longe de equilíbrio. Neste caso, a função é a manutenção do sistema. "Este é um modelo de função como utilidade, e não como design (evolutivo). Esse modelo de função é de uma propriedade causalmente eficaz: a persistência ou cessação do processo longe do equilíbrio faz uma diferença causal para o mundo" (BICKHARD, 2004, p. 11). Tal propriedade se caracteriza por ser normativa e relacional, como define Bickhard: "é uma propriedade normativa, na medida em que tal contribuição pode ser positiva ou negativa, adequada ou inadequada. É uma propriedade relacional: o coração de um parasita é funcional para o parasita, mas é disfuncional para o hospedeiro" (2004, p. 12). A normatividade, portanto, surge em um sentido termodinâmico, mais especificamente, em sistemas longe do equilíbrio termodinâmico, onde há interação entre as partes do sistema. Estas interações são

normativas, isto é, podem ser bem-sucedidas ou falhar, pois pertencem a um sistema aberto. Para Bickhard (2004, p. 8), a função é o sentido mais próprio de normatividade, como o coração com sua função de bombear o sangue para o corpo como seu caráter normativo, uma vez que o coração pode ser disfuncional se não bombear ou bombear inadequadamente o sangue. A função normativa é, assim, a base de uma cadeia hierárquica de emergências normativas. Fenômenos como atividade cognitiva são fundamentalmente normativos, emergentes de uma cadeia hierárquica com a normatividade funcional biológica como base. "Alguns outros locais e níveis na hierarquia incluem representação, percepção, memória, aprendizado, emoções, socialização, linguagem, valores, racionalidade e ética" (BICKHARD, 2004, p. 13-14).

A complexidade decorrente desta automanutenção, em que um sistema físico mantém suas próprias condições de existência ao ser bem-sucedido para discriminar ambientes viáveis de inviáveis, é a base do comportamento inteligente. Neste sentido, a tese de Bickhard amplia o entendimento de inteligência com o argumento dos sistemas físicos autossustentáveis recursivamente, mesmo naqueles mais primitivos. Aliás, é a partir do entendimento de como esta inteligência primitiva funciona que é possível entender o surgimento de inteligências mais complexas (BICKHARD, 2009a, p. 352).

Os computadores, de acordo com este raciocínio, são compostos por materiais que não estão longe do equilíbrio, por isso tendem à estabilidade termodinâmica. Uma pedra, por exemplo, quando isolada, entra em estabilidade termodinâmica e seu equilíbrio perdurará por milhões ou até bilhões de anos. Neste mesmo sentido, os computadores ou robôs, constituídos por componentes metálicos, se isolados termodinamicamente, não irão deixar de existir, como ocorre com os sistemas longe do equilíbrio termodinâmico. Além disso, não são entidades que se atualizam, uma vez que não possuem antecipação normativa da interação para suas ações no mundo, como os sistemas longe do equilíbrio de automanutenção.

Os argumentos de Bickhard contribuem para uma mudança de perspectiva em IA. Conforme já apresentado aqui, as críticas de Searle e Dreyfus elucidam uma

IA com base em engenharia reversa, de cima para baixo, que tem como objetivo a artificialização da inteligência humana. As observações de Bickhard mostram que o foco da IA deveria ser em Inteligência Artificial, e não em inteligência "humana" artificial. Com base nesta perspectiva, a IA pode ampliar seus resultados se tiver em vista a confecção de sistemas simples artificialmente inteligentes, adotando uma concepção de inteligência em sentido geral e não adstrito ao modo humano.

Ao entender a inteligência a partir de escalas menores e construir pequenos sistemas físicos, dotados de habilidades com flexibilidade para se adaptar e aprender (sistemas inteligentes simples), a IA pode avançar em perspectivas mais eficazes para formas complexas de inteligência que exerçam interação corpo-mundo. Isso requer que a IA enfatize não somente *como* sistemas são inteligentes (a crítica à IA Forte de Searle), mas sobretudo *porque* são inteligentes. As argumentações de Dreyfus e Bickhard apresentam caminhos para uma IA que pense este *porquê* a partir de criaturas inteligentes, integradas e incorporadas em um mundo (Dreyfus), na busca de cumprir funções, especialmente de se manter nele de forma autossustentável e funcionando com sucesso (Bickhard).

Xenobots: organismos vivos e programáveis

A limitação da IA, de acordo com o exposto, está no fato de que a confecção das tecnologias já realizadas pela humanidade se dá fundamentalmente a partir de materiais sintéticos não vivos, dada sua relativa simplicidade, facilidade e previsibilidade para se manter. Por outro lado, os sistemas vivos, por serem resultantes de processos evolutivos, possuem funções e estrutura complexas, de maneira pré-configurada, por isso resistem a alterações para novos comportamentos. Assim, a criação de novas formas de vida limita-se atualmente a organismos simples já existentes[15] ou a organoides de bioengenharia *in vitro*, que não resistem a contextos longe do equilíbrio termodinâmico.

[15] Construção de sistemas bio-híbridos, inspirados em modelos morfológicos relativamente simples, mostram evolução na tentativa de manipular as funcionalidades biológicas de organismos vivos, mas continuam esbarrando nas limitações da IA imitativa, pois não são

Um caminho provavelmente revolucionário para este desafio parece surgir. É o que mostra a pesquisa publicada por Sam Kriegman, Douglas Blackiston, Michael Levin e Josh Bongard, no artigo *A scalable pipeline for designing reconfigurable organisms*, de 13 de janeiro de 2020, na revista *Proceedings of the National Academy of Sciences* (PNAS). Simulações em IA desde a origem possibilitam projetar e criar uma nova classe de artefatos: as novas máquinas vivas, que não são nem robôs tradicionais e nem alguma forma de vida conhecida. O estudo teve como escopo apresentar um procedimento escalável para a criação de novas formas de vida funcionais. Para isso, partiu-se de métodos de IA para projetar computacionalmente diversas formas de vida possíveis para desempenhar alguma função desejada. Em seguida, os projetos viáveis são "criados usando um kit de ferramentas de construção baseado em células para realizar sistemas vivos com os comportamentos previstos" (KRIEGMAN, BLACKISTON, LEVIN e BONGARD, 2020, p. 1853).

Esta pesquisa se desenvolveu em dois momentos. Inicialmente, na Universidade de Vermont (EUA), com a aplicação de algoritmos evolutivos em um supercomputador, pesquisadores identificaram *in silico* maneiras possíveis de se construir sistemas vivos pretendidos para desempenharem funções como o transporte de determinado medicamento para células específicas do organismo humano. No segundo momento, na Universidade de Tuffs (EUA), biólogos reaproveitaram células vivas para reagruparem de acordo com as projeções computacionais, viabilizando, assim, formas inéditas de vida. Foram colhidas células-tronco dos embriões de uma espécie de sapos africana chamada *Xenopus laevis*: por isso a expressão "*xenobots*" para estes novos artefatos. Separadas em células únicas (da pele e do coração) e deixadas para incubar, as células foram cortadas e unidas microscopicamente para ficarem o mais fielmente possível ao projeto modelado computacionalmente. Em seguida, reunidas em formas corporais

capazes de prever situações aleatórias e arbitrárias. Por isso, estas máquinas biológicas se assemelham a organismos existentes, dependentes de suas formas de design. O exemplo de imitações do modelo morfológico relativamente simples dos peixes batóides (PARK, 2016, p. 158), como arraias, sugere um animal artificial para nadar e seguir fototaticamente uma orientação clara. Mas, conforme apresentado no estudo, os movimentos eram previamente controlados, nas condições anatômicas das partes dos seres vivos utilizados no empreendimento (PARK, 2016, p. 158-162).

nunca criadas naturalmente, as células demonstraram trabalhar de forma conjunta. A junção de células da pele (mais passivas) com células do músculo cardíaco (mais ativas, com contrações aleatórias) possibilitou criar movimentos ordenados, guiados pelo design do computador e auxiliados por padrões espontâneos de auto-organização. Estes organismos apresentam condições para se autolocomover e explorar um ambiente aquoso com nutrientes suficientes para "viverem" dias ou semanas.

Para cumprir a função desejada, os organismos reconfiguráveis necessitaram desempenhar os comportamentos de locomoção, manipulação de objetos, transporte de objetos e de coletividade (cooperação entre células para criar anatomias funcionais). Propriedades geométricas e emergentes, simuladas e refinadas computacionalmente, são transferidas para processos bioelétricos, bioquímicos e biomecânicos, ou seja, incorporadas dos métodos de pesquisa evolutiva para descobrir projetos que podem ser instanciados em materiais biológicos não artificiais.

Os *xenobots* não são robôs tradicionais, nem se caracterizam como alguma espécie conhecida de ser vivo existente. São, portanto, organismos vivos e programáveis, máquinas completamente biológicas desde o início, consideradas como uma nova classe de artefatos. Embora o material biológico seja totalmente de *Xenopus laevis*, o resultado final não é um sapo ou qualquer outro ser vivo projetado pela natureza ou manipulado geneticamente, mas uma máquina computacional biológica.

A capacidade própria dos seres vivos de resistirem à entropia, mantida nestes novos artefatos, permite, em tese, superar uma série de limitações das tecnologias estáticas existentes, especialmente as dificuldades para automanutenção. Além disso, quando houver necessidade de descarte após o cumprimento da função almejada, a biodegradação destes artefatos possibilita diversas vantagens ambientais. A possibilidade de aplicação é muito ampla e promissora, como sugerem os pesquisadores, porém, a contribuição teórica para a IA pode ser revolucionária, por ser tratar de uma possível superação concreta das limitações conceituais identificadas pelos críticos supracitados.

Considerações finais

O caso dos *xenobots* é um exemplo da aplicação de IA para além de seus modelos clássicos, uma vez que concebe "inteligência" a partir de uma acepção geral com base nas funcionalidades de sistemas simples automantidos, conforme advoga Bickhard. Pesquisas em IA, que se baseiam no ser humano como modelo de inteligência com o objetivo de reproduzi-la, incorrem em comprometedoras limitações, como o problema da intencionalidade (e consequentemente da consciência), mencionada por Searle, e a inviabilidade da formalização dos contextos, argumentada por Dreyfus.

Considerando uma possível complementariedade na argumentação de Searle, Dreyfus e Bickhard sobre da IA, bem como as novas perspectivas promovidas pelo surgimento dos *xenobots,* é plausível prospectar, por um lado, a inviabilidade da IA, forte com seus projetos de reprodução da inteligência humana e, por outro, um caminho de profícuo desenvolvimento da IA, com as devidas prevenções de consequências não intencionais temerárias decorrente de tentativas de criação de uma IA superior à capacidade humana. Esta nova classe de artefatos, que são organismos vivos e programáveis, máquinas completamente biológicas desde o início, pressupõe o entendimento das regras mais simples da inteligência de organismos elementares, o que abre oportunidade para uma nova forma de investigação para alcançar formatos mais complexos de interação entre o organismo e o contexto de realização da função, os quais são inerentemente integrados. Ao invés de partir da inteligência humana, a IA abre caminho para artificializar a inteligência em sua forma biológica geral, a saber, uma inteligência biológica artificializada.

Referências

BICKHARD, M. H. Interactivism. In: J. Symons, P. Calvo. (Eds.) *The Routledge Companion to Philosophy of Psychology*. London: Routledg, 2009a. p. 346-359.

BICKHARD, M. H. The interactivist model. *Synthese*, v. 166, n. 3, p. 547-591, 2009b.

BICKHARD, M. Part II: Applications of Process-Based Theories: Process and Emergence: Normative Function and Representation. *Axiomathes*, v. 14, n. 1, p. 121-155, 2004.

BOSTROM, N. *Superintelligence*: paths, dangers, strategies. New York: Oxford University Press, 2014.

DREYFUS, H. *What computers can't do*: a critique of artificial reason. New York: Harper & Row, 1972.

DREYFUS, H. *What Computers Still Can`t Do*: a critique of artificial reason. Cambridge: MIT Press, 1992.

DREYFUS, H. L. Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Artificial Intelligence*, Elsevier, v. 171p. 1137-1160, 2007.

DREYFUS, H. *Skillful Coping*: Essays on the Phenomenology of Everyday Perception and Action. Oxford University Press, 2014.

FEARN, N. *The latest answers to the oldest questions:* a philosophical adventure with the world's greatest thinkers. New York: Grove Press, 2005.

FODOR, J., You Can Fool Some of the People All of the Time, Everything Else Being Equal. "Hedged Laws and Psychological Explanations". *Mind*, v. 100, n. 397, p. 19-34, 1991.

GARDNER, H. *A nova ciência da mente*. 3. ed. São Paulo: EDUSP, 2003.

HANDERSON, H. *Artificial Intelligence*: mirrors for the mind. New York: Chelsea House Publishers, 2007.

KRIEGMAN, S., BLACKISTON, D., LEVIN, M., E BONGARD, J. A scalable pipeline for designing reconfigurable organisms. Proceedings of the National Academy of Sciences (PNAS), v. 117, n. 4, p. 1853-1859. Doi: https://doi.org/10.1073/pnas.1910837117.

PARK, S. J. et al. Phototactic guidance of a tissue-engineered soft-robotic ray. *Science*, v. 353, n. 6295, p.158-162, 2016.

SEARLE, J. *A redescoberta da mente.* São Paulo: Martins Fontes, 1997.

SEARLE, J. *Intencionalidade.* São Paulo: Martins Fontes, 1995.

SEARLE, J. R. Minds, brains, and programs. *Behavioral and Brain Sciences*, Cambridge, v. 3 n. 3, p. 417-457, 1980.

SEARLE, J. *The Chinese Room*. The MIT Encyclopedia of the Cognitive Sciences. Cambridge: MIT Press, 1999. p. 115-116

SHANAHAN, M. The Frame Problem. In: ZALTA, E. N. (Ed.). *The Stanford Ency- clopedia of Philosophy*. Spring 2016 ed. [s.l.]. Metaphysics Research Lab, Stanford University, 2016. Disponível em: <https://plato.stanford.edu/archives/spr2016/entries/frame-problem/>.

SILVER, N. *The Signal and the Noise*: Why Most Predictions Fail – but Some Don't. New York: The Penguin Press, 2012.

SUSSER, D. Artificial Intelligence and the Body: Dreyfus, Bickhard, and the Future of AI. In: MULLER, V. *Philosophy and Theory of Artificial Intelligence*. Berlin: Springer, 2013.

TURING, A. M. Computing machinery and intelligence. *Mind*, Oxford, n. 59, p. 433-460, 1950.

RECEBIDO: 02/11/2019
APROVADO: 02/12/2019

RECEIVED: 11/02/2019
APPROVED: 12/02/2019