

Uma cartografia comum aproximando Inteligência Artificial, Filosofia e Psicologia

Luciano Frontino de Medeiros¹

Alvino Moser²

Marilene S. S. Garcia³

Resumo: O artigo relaciona quatro problemas epistemológicos a partir de uma perspectiva em que a Inteligência Artificial compartilha um domínio de conhecimento comum às áreas da Filosofia e da Psicologia: i) o problema clássico do *frame* ou quadro de referência, surgido a partir das pesquisas em IA sobre a limitação da representação em lógica de primeira ordem; ii) o problema de Hume, exposto por Daniel Dennett, abordando representações que raciocinam sobre representações; iii) na direção indicada por William Frawley, é apresentado o problema de Platão, expondo sobre a eficácia do conhecimento a partir de evidências escassas e fragmentadas do mundo; iv) o problema de Wittgenstein sobre a compatibilidade entre a linguagem determinística computada e a linguagem probabilística real. A ideia chave é mostrar que a Inteligência Artificial não é apenas engenharia de robôs ou sistemas inteligentes, mas também uma área que suscita questionamentos mais profundos, contendo objetos de interesse próximos ou mesmo comuns com os campos de atuação da Filosofia e da Psicologia.

Palavras-chave: Inteligência Artificial. Filosofia da Mente. Psicologia Cognitiva. Representações de conhecimento. Epistemologia da Inteligência Artificial.

Abstract: This paper presents a discussion of four philosophical problems from perspectives in which Artificial Intelligence, Philosophy and Psychology have a common interest. The issues are: the classical frame problem, which originated from AI research concerning the restriction of representations in first order logic; David Hume's view on representation and on reasoning about representations. In a line of argumentation presented by William Frawley, the paper focuses on two philosophical problems, (1) Plato's question of how world knowledge can grow from the mere fragments of which our cognition of facts is made up and (2) Wittgenstein's question concerning the compatibility between natural and computational languages. The main purpose of the paper is to show how certain fundamental philosophical premises may exclude the

¹ Doutor em Engenharia e Gestão do Conhecimento pela Universidade Federal de Santa Catarina (2010). Professor Titular do Programa de Mestrado Profissional em Educação e Novas Tecnologias do Centro Universitário Internacional UNINTER. E-mail: luciano.me@uninter.com.

² Doutor em Filosofia e Ética pela Université Catholique de Louvain, Louvain-la-Neuve, Bélgica (1973). Professor Titular do Programa de Mestrado Profissional em Educação e Novas Tecnologias do Centro Universitário Internacional UNINTER. E-mail: alvino.m@uninter.com.

³ Marilene S. S. Garcia concluiu pós-doutorado pela PUC-SP – TIDD. É autora dos livros: *Mobilidade tecnológica e planejamento didático* (2017); *Avaliação e validação de projetos* (2018), ambos pela editora Senac-SP. Possui doutorado pela USP e Mestrado pela UNICAMP, com estágio de pesquisa pelas Universidades de Freiburg e Oldenburg, na Alemanha. É professora do Mestrado Profissional em Educação e Novas Tecnologias da UNINTER- PR, em que ministra a disciplina de Aprendizagens ativas, Metodologias ativas e ensino híbrido. Coordena pesquisa em design de aplicativo para a inclusão de analfabetos funcionais. E-mail: marilene@uol.com.br.

arguments of specialists from other domains of research, such as, philosophy, psychology and AI.

Keywords: Artificial Intelligence. Philosophy of Mind. Cognitive Psychology. Knowledge representations. Epistemology of AI.

Introdução

A Inteligência Artificial (IA) como área consolidada de conhecimento científico, tem alcançado nos dias atuais uma popularidade nunca experimentada em sua recente história. O desenvolvimento de tecnologias como o Aprendizado de Máquina (*machine learning*) e o advento do paradigma de Aprendizagem Profunda (*deep learning*) demonstram a franca vitalidade atual desse campo de pesquisa (GOODFELLOW; BENGIO; COURVILLE, 2016; HINTON; OSINDERO; TEH, 2006; MOHAMED; DAHL; HINTON, 2009). Alia-se a isso o sucesso e a proliferação de plataformas de automação de baixo custo, que tem auxiliado na disseminação sem par das tecnologias robóticas. Em decorrência, constata-se que as tecnologias consideradas “inteligentes” estão mais protagonistas no dia a dia da Sociedade do Conhecimento, seja para a melhoria da qualidade de vida alicerçada na dimensão construtiva do ser humano, seja para motivações de exploração de mercado (e no limite, bélicas) ou ainda permitir as alienações sociais excludentes.

Apesar de a Inteligência Artificial estabelecer como foco das suas investigações a criação de artefatos inteligentes (RUSSELL; NORVIG, 2004), seus estudos e desenvolvimentos também tiveram influência significativa (assim como foi influenciada) sobre as áreas disciplinares da Filosofia e da Psicologia. Por ter também o homem e a sua inteligência como objeto de estudo e ponto de partida, encontra-se nos seus fundamentos todo o edifício filosófico envolvendo as questões profundas e existenciais sobre o que é ser “humano”, bem como o conhecimento relativo aos aspectos biológicos e psicológicos da mente humana. O filósofo Daniel Dennett explorou diversas possibilidades em um trabalho de quase três décadas atrás sobre essa cartografia comum em que transitam filósofos da mente, psicólogos experimentais e os engenheiros da Inteligência Artificial (DENNETT, 2006, p. 163-183). Dennett destacou a importância do estudo da IA por parte dos filósofos da mente, ressaltando a grande proximidade com o campo de estudo dos psicólogos.

Em outro ramo dessa árvore interdisciplinar, a Psicologia Cognitiva manifesta hoje uma série de aproximações com a Inteligência Artificial. Podem-se citar obras como Sternberg (2014) e Lefrançois (2013), que apresentam capítulos específicos sobre IA junto à discussão sobre os temas gerais da Psicologia Cognitiva e teorias de aprendizagem. Sternberg alça ainda as simulações de Inteligência Artificial ao status de componentes do método de pesquisa em Psicologia Cognitiva. “Os psicólogos cognitivos usam experimentos controlados, pesquisa psicobiológica, autoavaliações, observação naturalista e simulações por computador e IA para estudar os fenômenos cognitivos” (STERNBERG, 2014, p. 16-17).

Sternberg considera ainda a Psicologia Cognitiva, a IA, a Psicobiologia, a Filosofia, a Linguística e a Antropologia como campos que fornecem métodos e ideias para um campo mais geral e multidisciplinar, a Ciência Cognitiva (NICKERSON, 2005 apud STERNBERG, 2014).

Ao exposto dessas considerações preliminares, a proposta deste artigo é expor quatro problemas epistemológicos, ampliando as percepções consolidadas por Dennett com outras abordagens como as exploradas por Frawley com respeito às aproximações entre as ciências cognitivas e o sociointeracionismo (FRAWLEY, 2000). Esses problemas têm origem em uma área de conhecimento que se sobressai a partir de um campo comum e pode ensejar respostas em uma ou outra dessas áreas. Dennett destaca acertadamente o problema de Hume, originário da Filosofia, o qual tem uma resposta peculiar e única dada pela Inteligência Artificial; e o problema do *frame*, originário do campo da Inteligência Artificial, que tem impacto sobre as formas de representação, com base na lógica, das coisas na realidade. Por sua vez, Frawley discute também o problema de Platão, que pode ser abordado adequadamente pela Inteligência Artificial e por último o problema de Wittgenstein (assim denominado por Frawley), relativo ao compromisso da linguagem como forma computada e como uso. A ideia geral é reforçar a constatação do quanto essas áreas, a IA, a Filosofia e a Psicologia estão imbricadas entre si, na emergência de um campo de conhecimento interdisciplinar em que estudos particulares em uma área específica devem ser ampliados no contexto maior do mapa comum constituído por elas.

A categorização adotada aqui não tem a intenção de acondicionar o âmbito de toda a Filosofia pensada pelas personalidades que concedem a denominação, mas apenas para permitir uma proposta de estudos que possa pavimentar caminhos entre as diferentes áreas. Não se deseja racionalizar ou compactar toda a contribuição filosófica de um Platão ou de um Wittgenstein em poucas linhas ou conceitos, mas sim mostrar que certas questões, quando abordadas na pesquisa empírica da IA, já foram objeto de uma abordagem filosófica extensa, a qual pode fornecer a fundamentação teórica necessária para que essa pesquisa possa se desenvolver de forma ampliada. Como mencionado por Dennett, os pesquisadores de IA precisam estudar filosofia, “a menos que estejam contentes em reinventar a roda a cada poucos dias”.

Quando a IA reinventa a roda, ela é caracteristicamente quadrada, ou, no melhor dos casos, hexagonal, e pode fazer apenas algumas centenas de revoluções antes de parar. As rodas do filósofo, por sua vez, são círculos perfeitos, e não requerem, em princípio, nenhuma lubrificação; e podem ir pelo menos duas direções em duas direções ao mesmo tempo. Claramente, um encontro de mentes é necessário. (DENNETT, 2006, p. 183)

O problema de Platão

A concepção de sistemas inteligentes parte da premissa que estabelece a existência de processos internos ao agente, os quais têm inspiração na forma como os seres biológicos inteligentes pensam (ou como se imagina que pensam). A preocupação dos pesquisadores de IA se concentra nesse ponto em como dotar o agente inteligente de conhecimentos suficientes para que possa enfrentar a situação experimental de resolução de um problema. Portanto, há um lado interno, relativo ao agente, que precisa ser equipado com algum tipo de representação do mundo em que ele irá atuar e essa representação requer robustez para fazer frente a variações que um problema idealizado genericamente pode apresentar.

Platão apresentava a sua teoria das ideias como forma de dar um fundamento estável ao *logos*, para escapar da multiplicidade das interrelações contraditórias do sensível. A busca do inteligível é a busca pelo lugar da não contradição, enquanto que o sensível é o lugar da contradição, em que a identidade é una e múltipla. “O sensível tira sua pouca realidade da sua participação nessa realidade suprema que as Ideias representam” (ROGUE, 2011, p. 86-111).

Embasado na concepção desse dualismo ontológico de Platão, Frawley ressalta que possuímos um conhecimento rico e eficaz, ainda que as evidências e fatos apresentados a nós, a partir dos sentidos, sejam escassos, fragmentados e inconclusivos. Dessa forma, põe-se em uma interrogação o problema de Platão “de que forma sabemos tanto a partir de tão pouco? O que faz com que o lado interno de nossas mentes seja tão pleno e sistemático enquanto nossos lados externos sejam tão heterogêneos?” (FRAWLEY, 2000, p. 42).

Platão também dá uma resposta no âmbito de sua teoria das ideias: “Possuímos um conjunto de verdades em nossa razão que se origina da exposição à realidade das formas puras e ideais”. Frawley comenta, entretanto, que respostas ao problema de Platão não precisam necessariamente ser da categoria de uma resposta de Platão. Por exemplo, a posição do behaviorismo radical contradiz o fato de que o que sabemos não é tão pouco, como posto no enunciado. Mas de maneira geral, nas ciências cognitivas, as respostas ao problema de Platão têm genuinamente uma resposta de Platão (FRAWLEY, 2000, p. 43).

As abordagens correlatas para essa resposta podem ser exploradas a partir dos dois principais paradigmas existentes dentro da Inteligência Artificial: o representacionalismo (ou computacionalismo, ou ainda simbolismo) e o conexionismo. O representacionalismo identifica a mente operando por meio de símbolos, enquanto que o conexionismo vê o cérebro, as redes neurais e seu potencial para associação de padrões (DINSMORE, 1992, p. 1; FRAWLEY, 2000, p. 73). A corrente do representacionalismo tem à frente alguns expoentes como Jerry Fodor, Noam Chomsky e Alan Newell. O conexionismo tem, dentre várias personalidades, pesquisadores como David Rumelhart, Frank Rosenblatt e Geoffrey Hinton.

O representacionalismo, ou computacionalismo, é a tese da metáfora do computador na essência do funcionamento da mente, tal como um sistema baseado em fatos e regras que produz inferências a partir da manipulação dessas estruturas, objetivando preservar significado e condições de verdade (DINSMORE, 1992, p. 2). Dessa forma, o pensamento e outros processos ligados à mente efetuam cálculos operacionais utilizando a sintaxe de representações que compõem a linguagem do pensamento. Na concepção de que a mente possui um código interno simbólico, ela

deve ser capaz de representar as hipóteses sobre o mundo e permitir que a sua validade seja verificada. O código interno é uma espécie de linguagem proposicional do pensamento (FRAWLEY, 2000, p. 74).

Uma linguagem proposicional compõe-se de fatos e regras. Fatos são representações tidas como verdadeiras sobre um dado contexto sendo representado e tem característica estática. As regras, por sua vez, permitem que se derivem novos fatos a partir de fatos existentes, agregando um aspecto ampliador e dinâmico ao raciocínio proposicional. Esses sistemas também são conhecidos como sistemas de produção do tipo “se... então”, que englobam premissas e conclusões (NEWELL; SIMON, 1972 apud STERNBERG, 2014, p. 285).

Dessa forma, essa é uma primeira resposta ao problema de Platão: a partir de um sistema rico de verdades necessárias e suficientes, o ser humano é capaz de produzir novos conhecimentos na exposição e validação das formas presentes em um código interno da mente ao mundo. A partir do conhecimento proveniente de fatos estáticos, as regras permitem a ampliação dos limites do nosso próprio conhecimento.

Nossas mentes vêm ao mundo equipadas com um sistema rico e predeterminado de verdades necessárias: formas atemporais, idealizadas, imutáveis, absolutas e autoevidentes levadas à consciência pela introspecção, não desenvolvidas através de imitação ou de outro tipo de código do mundo externo. (FRAWLEY, 2000, p. 43)

Entretanto, a corrente do conexionismo também permite obter uma resposta ao problema de Platão. O paradigma conexionista tem como pilar a concepção de que o cérebro consiste de um número massivo de elementos interconectados, os quais enviam sinais excitatórios ou inibitórios entre si, atualizando suas excitações a partir de mensagens simples (McCLELLAND; RUMELHART; HINTON, 1986, p. 10). Com base nessa premissa, redes neurais artificiais podem ser construídas, criando-se uma arquitetura de unidades (neurônios) interconectadas (por sinapses) que podem aprender tarefas como reconhecimento de padrões a partir de algoritmos que utilizam processos de tentativa e erro (CHURCHLAND, 2004, p. 244-252).

Um dos tipos clássicos que aparece no estudo de aprendizado em redes neurais artificiais é o perceptron, criado pelo psicólogo Frank Rosenblatt (RUSSELL; NORVIG, 2004, p. 732). O perceptron possui unidades sensoriais, associativas e geradoras de

respostas. As sinapses ou pesos são descritos em termos de valores que irão ponderar os sinais de entrada, permitindo a ativação de neurônios específicos conforme a atividade de aprendizado desejada. Uma rede neural artificial aprende a partir de um conjunto de amostras representativas do contexto do problema de classificação ou reconhecimento de padrões. Após o algoritmo ser executado várias vezes, a rede neural pode alcançar um nível de aprendizado tal que faça inferências de forma muito precisa sobre as amostras oferecidas no treinamento. “O padrão de conexões determina o que o sistema conhece e como vai responder” (LEFRANÇOIS, 2013, p. 290).

Entretanto, uma das características que fazem as redes neurais artificiais um dos tipos de IA mais utilizados para semelhantes tarefas é a sua capacidade de generalização (HAYKIN, 2001). Caso sejam apresentadas amostras que não fizeram parte do conjunto de treinamento, a rede neural artificial pode inferir de maneira aproximada, ou seja, fornecer uma resposta de tal forma que seja uma interpolação das amostras em questão, em comparação com o conjunto de amostras do conjunto de treinamento.

Portanto, o conexionismo também fornece uma resposta ao problema de Platão, pois as redes neurais podem armazenar em sua estrutura simples um conhecimento que pode extrapolar e generalizar para além do aprendizado retido a partir de treinamento prévio, permitindo a ampliação do conhecimento com o qual a rede pode lidar. Portanto, o código interno, que permite a extrapolação a partir dos seus limites, seja no caso das regras para o representacionalismo, ou a capacidade de generalização no caso do conexionismo, tem condições de preencher a demanda do problema epistemológico de Platão, cada uma com suas peculiaridades e distinções.

Para a ciência cognitiva, o código interno é coincidente com a mente. Em uma explicação representacionista ele é explícito, local, inflexível e dedicado; em uma explicação conexionista ele é emergente, distribuído, flexível e ajustável. (FRAWLEY, 2000, p. 78)

Ainda com relação à aprendizagem, Frawley aponta que os cientistas cognitivos entram em concordância quanto à existência de um código interno para acomodar a aprendizagem. Entretanto, enquanto os representacionistas a minimizam, os conexionistas a maximizam (FRAWLEY, 2000, p. 77).

O problema de Hume

Dennett (2006) foi o primeiro a formalizar a possibilidade, por meio da abordagem da Inteligência Artificial, de resolução do que foi denominado por ele de “problema de Hume” que, por mais de duzentos anos, foi alvo de estudo por filósofos e psicólogos. Se a Inteligência Artificial assume o papel de uma investigação mais abstrata da possibilidade do conhecimento, as representações de conhecimento, que sempre estiveram na ordem do dia das pesquisas da IA, abriram uma trilha na direção de uma Psicologia Cognitiva bem sucedida na explicação dos fenômenos complexos da mente. Na abertura da “caixa preta”, preterida pelos behavioristas radicais por explicações comportamentais enraizadas no ambiente, as representações internas da mente foram bem respaldadas nas metáforas construídas para o raciocínio mecânico feito por sistemas inteligentes.

A única psicologia que poderia possivelmente ser bem sucedida na explicação das complexidades da atividade humana deve postular representações internas. [...] Para os empiristas britânicos, as representações internas eram chamadas ideias, sensações, impressões; mais recentemente, os psicólogos têm falado de hipóteses, mapas, esquemas, imagens, proposições, engramas, sinais neuronais e mesmo hologramas e teorias inatas inteiras. (DENNETT, 2006, p. 178)

Entretanto, se uma representação qualquer existe é porque houve uma intencionalidade prévia de alguém para a atividade de elaborar essa representação. É necessário também que exista um sujeito que desempenhe o papel de interpretar a representação, de dar significado a ela.

Nada é intrinsecamente uma representação de nada; algo é uma representação apenas para alguém; qualquer representação ou sistema de representações assim requer um usuário ou interpretador da representação, que é externo a ela. Qualquer interpretador como tal deve possuir vários traços psicológicos ou intencionais; ele deve ser capaz de várias compreensões, e deve possuir crenças e objetivos [...]. Tal interpretador é um tipo de homúnculo. (DENNETT, 2006, p. 178)

Hume fundamenta a sua teoria das ideias abstratas, originárias das ideias concretas e das impressões do mundo exterior sobre os sentidos, estabelecendo a possibilidade de conexões associativas entre elas (MORRIS; BROWN, 2017, p. 14). No entanto, sua teoria das ideias não converge para qualquer *locus* de significação. Na

esteira dessa proposição, filósofos como Kant e Husserl apresentaram posições contrárias à teoria das ideias de Hume. Kant, no âmbito do idealismo transcendental, criticou Hume por pressupor que conceitos provenientes dos sentidos prevalecessem sobre o raciocínio causal *a priori*; a generalização não poderia chegar jamais por meio dos sentidos e também não poderia estar contida ao mesmo tempo na forma pura da intuição sensível, pois ela é um ato do poder de representação, e o “eu penso”, a presunção de um sujeito, tem que estar por trás de todas as representações (KANT, 2015, p. 127-129). E, na gênese da sua fenomenologia, Husserl investe nos modos de consciência e nas vivências intencionais para preencher o vazio deixado pelas conexões associativas de Hume que parecem manifestar, de certa forma, uma espécie de intencionalidade (HUSSERL, 2012, p. 157).

As representações internas de Hume eram impressões e ideias, e ele sabiamente se esquivou da noção de um *eu* interno que manipularia de maneira inteligente esses elementos; mas isso o deixou na necessidade de fazer as ideias e impressões “pensarem por si mesmas”. O resultado foi sua teoria do “eu” como um “feixe” [...] de impressões e ideias. Ele tentou colocar essas impressões e ideias em interação dinâmica, postulando diversas ligações associacionistas, de forma que cada ideia sucessiva na corrente de consciência arrastava sua sucessora para o palco de acordo com um ou outro princípio, todos sem o benefício de uma supervisão inteligente. (DENNETT, 2006, p. 179)

Contudo, a IA respondeu ao problema de Hume a partir da existência de estruturas de dados que integram os sistemas inteligentes, as quais têm condições de serem pensadas como representações que se entendem por si mesmas e prescindindo, portanto, de um núcleo significador. Um sistema concebido em alto nível pode ser decomposto sucessivamente em agentes cada vez menores, que irão desempenhar a tarefa em questão executando apenas uma pequena parte do todo; o sistema seria uma organização de diversos pequenos agentes que assumem papéis variados para a emergência da tarefa inteligente em alto nível.

Se pudermos ter uma equipe ou comitê de homúnculos relativamente ignorantes, de mente estreita, cegos, para produzir o comportamento inteligente do todo, isso é um progresso. Um fluxograma é tipicamente o esquema organizacional de um comitê de homúnculos [...]; cada quadro especifica um homúnculo ao prescrever uma função sem dizer como ela é realizada. (DENNETT, 2006, p. 180)

Entretanto, mesmo com a decomposição sucessiva, Dennett complementa que não será possível atingir uma representação que se autocompreenda. No limite, todos os homúnculos são eliminados e há um sistema que manifesta comportamento inteligente de alto nível, composto por subsistemas que executam em partes mínimas a tarefa global (DENNETT, 2006, p. 181).

A moderna teoria da IA tem por base a construção de agentes inteligentes. Na esteira da noção de homúnculo dada por Dennett, um agente possui uma intencionalidade e uma racionalidade, percebendo o ambiente, tomando decisões conforme a sequência das percepções e agindo sobre ele, atualizando seus estados internos, as representações que permitem a execução de alguma tarefa inteligente.

Para cada sequência de percepções possível, um agente racional deve selecionar uma ação que se espera venha (sic) a maximizar sua medida de desempenho, dada a evidência fornecida pela sequência de percepções e por qualquer conhecimento interno do agente. (RUSSELL; NORVIG, 2004, p. 36)

É interessante ressaltar, sobre a resposta da IA na perspectiva de Dennett, dada ao problema de Hume que, ainda que considerada para sistemas muito limitados, é até o momento o único modo conhecido de resolver o problema (DENNETT, 2006, p. 182).

O problema do *Frame*

Originado de uma situação primeiramente posta por McCarthy e Hayes (1969), no âmbito da Inteligência Artificial, o problema do *frame*⁴ (também conhecido como problema do quadro, ou ainda o problema da estrutura)⁵ é a tarefa desafiadora de representar os efeitos de uma ação em lógica sem ser necessária a representação explícita de uma enorme número de não efeitos óbvios.⁶ Seria possível a limitação do escopo do raciocínio que é requerido para derivar as consequências de uma ação em determinado contexto? De maneira geral, o que faz o ser humano ter uma habilidade

⁴ Termo inspirado na Física Mecânica, significando o quadro inercial de referência, necessário para se fazer a análise de movimentos de objetos.

⁵ Dennett sugere ainda que o problema do *frame* fosse denominado também de problema de Kant (DENNETT, 2006, p. 183).

⁶ Se um agente inteligente precisa raciocinar, por exemplo, sobre um mundo de blocos para dizer o que é um cubo, uma pirâmide ou uma esfera, ele precisa dizer também que um cubo não é uma pirâmide e uma esfera, uma pirâmide não é um cubo ou uma esfera, e uma esfera não é um cubo ou uma pirâmide. Esse conhecimento precisa ser representado de forma explícita. Pequenos conjuntos são passíveis de uma representação adequada, mas, como se pode intuir, à medida em que o número de elementos cresce, as combinações possíveis crescem exponencialmente. Ao abordar problemas reais, não se tem memória (ou tempo) suficiente para acondicionar todas as possibilidades de representação.

peculiar para a tomada de decisões, baseando-se no que é relevante em uma situação em andamento, sem ter que explicitamente considerar tudo o que não é relevante (SHANAHAN, 2016, p. 1)?

Russel e Norvig (2004, p. 320-321) afirmam que o problema do *frame* é a representação de todas as coisas que permanecem iguais, em uma situação em que se encontra um agente que tem a possibilidade de executar determinadas ações a partir de um mecanismo de raciocínio, baseado em lógica de primeira ordem e contendo regras (axiomas) explícitas de efeito, alterando certos elementos e deixando outros inalterados. Portanto, uma grande parte da dinâmica qualitativa que é necessária para o planejamento de uma situação consiste na inferência de elementos que não irão se modificar. O problema do *frame*, de forma sumária, é o problema em como formalizar o raciocínio inercial necessário (THOMASON, 2016, p. 36).

O problema do *frame* puramente lógico pode ser resolvido utilizando-se a lógica monotônica, simplesmente escrevendo axiomas explícitos que dizem o que não pode ser modificado enquanto uma ação é executada por um agente. Entretanto, soluções baseadas em lógica não-monotônica têm sido amplamente pesquisadas (THOMASON, 2016, p. 37). Em suma, o desafio está em encontrar uma forma de capturar os não-efeitos de ações de uma forma sucinta em lógica formal. A premissa fundamental é denominada de *lei de inércia do senso comum* (SHANAHAN, 2016, p. 3).

Dennett extrapola o problema do *frame* para o patamar de um problema epistemológico abstrato, descoberto no escopo da IA. Se um agente cognitivo, que possui diversas crenças sobre o mundo que o rodeia, realiza uma ação, o próprio mundo em que o agente existe se modifica e, dessa forma, as crenças desse agente precisariam ser atualizadas ou mesmo revistas.

Se supõe, como fizeram tradicionalmente os filósofos, que as crenças de alguém são um conjunto de proposições, e que o raciocínio é uma inferência ou dedução a partir de membros desse conjunto, fica-se em dificuldades, pois é inteiramente claro (embora ainda controvertido) que os sistemas que se baseiam apenas em tais processos fiquem atolados por explosões combinatórias em seu esforço de atualização. (DENNETT, 2006, p. 182-183)

Dennett comenta ainda sobre a inegável capacidade dos seres humanos de manter as suas crenças de maneira aproximada em consonância com a realidade em que vivem.

Frawley (2000, p. 24) apresenta o problema do *frame* como um enigma relativo à afirmação clara e integral das condições que restringem de forma global as decisões tomadas por um agente inteligente. E menciona que a psicologia de Vygotsky pode ter um papel importante para mostrar não apenas uma solução, mas um método para abordar o problema do *frame*. Frawley apresenta a teoria sociointeracionista de Vygotsky e Luria não como uma explicação do que é “externo”, mas sim como uma teoria da internalização do “externo”. O problema do *frame* é mais uma questão de como o raciocínio lida com a adição de informações que influencia a situação de comprovação das conclusões, sendo capaz de inibir ou descartar uma diversidade de opções inferenciais.

A teoria sociointeracionista propõe uma linguagem para o pensamento, uma fala privada (VYGOTSKY, 1978, p. 24) que é acionada na resolução de problemas e que “codifica a internalização individual única da significação social e cultural externa” (FRAWLEY, 2000, p. 38). A estruturação dessa linguagem de inferências do pensamento de onde são derivadas as soluções possui uma função inibidora que descarta alternativas e dá uma orientação ao pensamento representacional.

Para os vygotskianos, um dos métodos padrão de experimento é interromper o comportamento introduzindo uma tarefa difícil e, então, observar como o sujeito recupera o controle cognitivo. No endireitar do self, o sujeito utiliza a fala privada como uma orientação simbólica externa e o que emerge são pistas linguísticas para a luta do indivíduo com a estrutura (frame) cognitivo. (FRAWLEY, 2000, p. 38)

Sternberg (2014, p. 67) expõe a comparação do ser humano com os robôs. Ainda que o ser humano possua uma visão limitada e que esteja sujeito a ilusões, possui uma capacidade extraordinária superior a dos robôs para codificar as representações visuais e dar significado a elas.

Dada a sofisticação dos robôs de hoje em dia, qual é a razão da superioridade humana? Possivelmente várias, mas o conhecimento é uma delas. O ser humano simplesmente conhece muito mais acerca do ambiente e das fontes de regularidade no ambiente que os robôs. O conhecimento humano é uma

grande vantagem que os robôs – ao menos os atuais – não são capazes de suplantar. (STERNBERG, 2014, p. 67-68)

Sternberg, na abordagem dos processos da atenção, destaca também o limite da mente humana quanto ao volume de informação na qual pode se concentrar. Os fenômenos psicológicos da atenção possibilitam o uso de recursos mentais que são limitados de forma sensata.

Ao diminuir a atenção sobre muitos estímulos externos (sensações) e internos (lembranças), podemos focar os estímulos que mais nos interessam. Esse foco acentuado aumenta a probabilidade de resposta rápida e precisa aos estímulos que interessam. (STERNBERG, 2014, p. 108)

Portanto, parece que o ser humano é naturalmente e biologicamente muito bem equipado em relação à abordagem aproximada do problema do *frame*. Enquanto que nos agentes da IA é necessária uma grande quantidade de conhecimento para acomodar inclusive os não-efeitos, o ser humano parece suplantar o problema do *frame* de uma maneira muito natural e eficaz. A corrente representacionista da mente também se defronta com essa dificuldade para explicar a mente operando de acordo com um código simbólico, lidando de forma natural com o problema do *frame*.

O problema de Wittgenstein

Como último problema epistemológico a ser abordado, Frawley (2000, p. 55) apresenta o problema de Wittgenstein a partir do questionamento sobre como o ser humano consegue coordenar a experiência idealizada do mundo interior, virtual e a experiência vivenciada sujeita às circunstâncias contingentes. Concernente aos aspectos da linguagem, esse problema se posta de forma importante para a ciência cognitiva para que se alcance um “compatibilismo” entre os padrões legítimos e determinísticos presentes na mente e a visão das pessoas como agentes com livre arbítrio e liberdade de escolha (FRAWLEY, 2000, p. 55).

É necessário então invocar a unidade dialética do filósofo Ludwig Wittgenstein autor de duas obras emblemáticas que influenciaram o estilo de pensar contemporâneo: *O Tractatus Logico-Philosophicus* (publicado na sua primeira edição em 1921), assinalado como o “primeiro Wittgenstein”, e *Investigações Filosóficas* (lançado

em 1953), relativo ao “segundo Wittgenstein”. O primeiro Wittgenstein segue a corrente da lógica adotando um estilo de demonstração e buscando a forma e o significado das proposições lógicas associadas à linguagem sob as condições de verdade determinada. Já o segundo Wittgenstein é um estudioso da história natural que aborda, seguindo um estilo discursivo, a respeito dos usos de declarações sob condições indeterminadas em relação à verdade e ao significado.

Wittgenstein sempre trabalhou na fronteira entre a determinação e a indeterminação. A partir de um único ponto de vista teórico, ele explorou os dois lados, concentrando-se, no início, na lógica e, posteriormente, no uso. Por esse motivo, alguns atribuíram uma unidade dialética a Wittgenstein. (FRAWLEY, 2000, p. 55)

O primeiro Wittgenstein se insere na discussão a partir dos trabalhos sobre o logicismo de Frege, cuja influência se constata de forma significativa no decorrer do *Tractatus*. Apresenta uma solução para a antinomia de Zermelo-Russell, desconsiderando a teoria de tipos que Bertrand Russell havia proposto para a abordagem ao problema (BUCHHOLZ, 2009; WITTGENSTEIN, 2010, p. 159). Visando confrontar na época a abordagem do psicologismo na fundamentação da teoria do conhecimento, a elucidação da inteligência humana não deveria acontecer por meio de investigações psicológicas, mas sim por uma teoria do significado proposicional. Essa teoria procederia adequadamente apenas se fosse identificado o significado linguístico a partir do emprego lógico dos sinais linguísticos (BUCHHOLZ, 2009, p. 41).

Em *Investigações Filosóficas*, temos um segundo Wittgenstein preocupado com jogos de linguagem, envolvendo a totalidade em que consiste a linguagem e as atividades que as englobam (WITTGENSTEIN, 2012, p. 19). De acordo com Fann (2013), Wittgenstein se dá conta de que as doutrinas do *Tractatus* se baseavam em uma imagem particular da essência da linguagem humana. Seria a teoria do significado-correspondência, cuja essência eram as palavras individuais que nomeavam objetos, o objeto que representava a palavra era seu significado. “Wittgenstein ahora se da cuenta de que su anterior parecer sobre las proposiciones no era el resultado de una investigación, era un requisito [...] Su concepción del lenguaje requería que toda proposición tuviera un sentido definido.” (FANN, 2013, p. 78-79).

Portanto, conforme o segundo Wittgenstein, podemos chegar apenas a uma análise correta por meio do que se poderia chamar de investigação lógica dos próprios fenômenos, ou seja, em certo sentido *a posteriori*, e não por meio de conjecturas sobre as possibilidades *a priori* (FANN, 2013, p. 62).

Um Wittgenstein único apresentaria “uma tensão perpétua entre a representação e a atuação, entre o tempo formalmente computável e o decorrido no mundo real” (FRAWLEY, 2000, p. 56). Portanto, o problema de Wittgenstein se resume em “como a linguagem se enquadra na relação entre máquinas (fatos determinados em forma computável) e pessoas (escolha de valores sob regras indeterminadas)” (FRAWLEY, 2000, p. 57).

O problema de Wittgenstein assume ascendência direta sobre as teorias representacionistas. Se a mente humana opera segundo uma linguagem da mente, operando sobre cálculos operacionais, chega-se a uma condição de limite sobre como se dá a relação de compromisso entre representações determinísticas operando sobre a sintaxe formal da mente com o que é indeterminado e ocorrido na realidade. Da perspectiva da IA, esse é um problema fundamental: como um robô, programado com um sistema de regras determinísticas, irá lidar com as situações indeterminadas provenientes da interação com o ambiente externo? Como deverá ser elaborado (ou se é possível de elaborar) um sistema de aprendizado de máquina que torne o robô apto a compatibilizar o problema de Wittgenstein, assim como um ser humano habilmente o faz?

Frawley sugere que o tratamento do problema de Wittgenstein deva consistir em duas soluções: para o primeiro Wittgenstein, a abordagem é muito semelhante com aquela feita com relação ao problema e a resposta de Platão, com a Psicologia Cognitiva internalista subsidiando por meio de um conjunto de resultados tanto teóricos quanto empíricos. Já, para o segundo Wittgenstein, é necessário que se aborde o problema por meio de uma psicologia social da ação, a partir do programa de pesquisa de Vygotsky, Luria e Leontiev sobre a mente inserida no contexto sociocultural (FRAWLEY, 2000, p. 61).

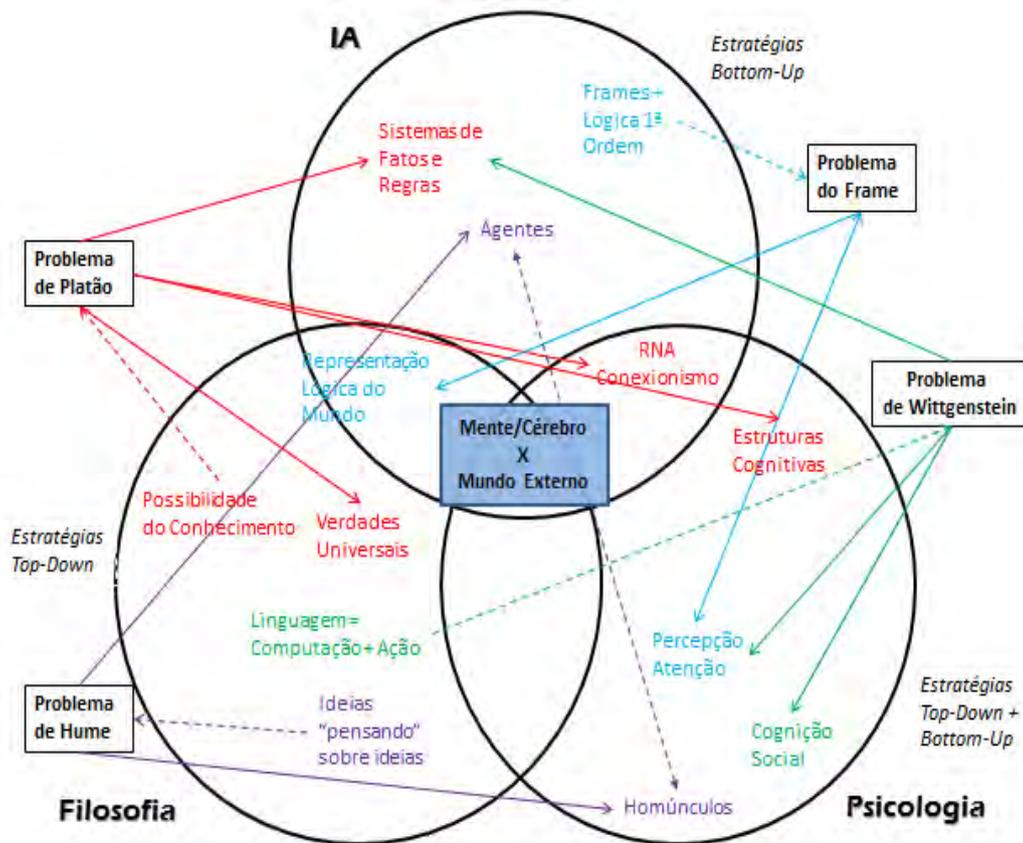


Figura 1. Uma cartografia sintetizando os quatro problemas sobre as áreas da Filosofia, Psicologia e Inteligência Artificial.

Considerações finais

A explanação dos quatro problemas filosóficos apresentados em detalhe teve como intuito demonstrar algumas das aproximações existentes entre as áreas do conhecimento da Filosofia e da Psicologia com a Inteligência Artificial e mesmo a existência de uma área interdisciplinar, uma cartografia comum indicando um domínio em que as áreas manifestam a convergência dos objetos de estudo relacionados à fenomenologia da inteligência. A figura 1 é uma tentativa de sintetizar e ligar os temas desenvolvidos para cada problema em uma representação comum. No centro, a ideia de que as três áreas exploram as questões relacionadas aos fenômenos do sistema cérebro/mente e suas relações com o mundo externo.

É natural pensar as pesquisas em IA, sobre a reprodução de raciocínio automatizado ou comportamentos inteligentes, como algo embasado fortemente na tecnologia que torna possível a construção desses artefatos, sem compromisso com

uma abordagem mais reflexiva sobre aspectos do funcionamento da mente. Porém, a IA pode ser pensada como uma área de pesquisa que emerge a partir da atividade de escrutínio incessante da mente humana, iniciada na Grécia Antiga e que perpassa por séculos de história, quando inserida no tempo presente com a manifestação da alta tecnologia. Neste século XXI, a humanidade é capaz de reproduzir de forma artificial os processos naturais da mente humana, exteriorizando-se a inteligência. Talvez esse seja o aspecto mais peculiar da IA. Enquanto que na Filosofia e na Psicologia se busca entender os aspectos da mente inteligente enquanto manifestações internas e encerradas nos limites do ser, a IA, por sua vez, as externaliza.

A confluência entre as áreas da Filosofia, da Psicologia e da IA não se esgota com esses quatro problemas epistemológicos. Há diversos temas que ensejam aproximações como, por exemplo, o problema mente-cérebro, o teste de Turing, o argumento da sala chinesa de Searle e também os aspectos éticos que estão surgindo em função da inserção da IA de forma intensiva na sociedade. Porém, é preponderante afirmar que na investigação ou aquisição de conhecimentos em uma área específica, o necessário aporte às outras áreas, na condição de correlatas, não pode deixar de ser considerado. É possível filosofar com IA ou entender processos psicológicos complexos com IA, assim como é legítimo tomar por empréstimo acadêmico os conhecimentos e os avanços da IA para se possa fazer Filosofia ou Psicologia.

Enviado: 12 março 2018

Aprovado: 15 abril 2018

Referências

BUCHHOLZ, K. *Compreender Wittgenstein*. 2a. ed. Petrópolis, RJ: Vozes, 2009.

CHURCHLAND, P. M. *Matéria e consciência*. São Paulo: Editora UNESP, 2004.

DENNETT, D. C. *Brainstorms: Escritos filosóficos sobre a Mente e a Psicologia*. São Paulo: UNESP, 2006.

DINSMORE, J. Thunder in the gap. In: DINSMORE, J. (Org.). *The symbolic and connectionist paradigms: closing the gap*. Hillsdale, NJ: Lawrence Erlbaum, p. 1-23, 1992.

FANN, K. T. *El concepto de filosofía en Wittgenstein*. 3ª. ed. Madrid: Tecnos, 2013.

FRAWLEY, W. *Vygotsky e a ciência cognitivas: linguagem e interação das mentes social e computacional*. Porto Alegre: Artes Médias Sul, 2000.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. Cambridge, MA: MIT Press, 2016.

HAYKIN, S. *Redes neurais: princípios e prática*. Porto Alegre: Bookman, 2001.

HINTON, G. E.; OSINDERO, S.; TEH, Y. W. A fast learning algorithm for deep belief nets. In: *Neural computation*, v. 18, n. 7, p. 1527-1554, 2006.

HUSSERL, E. *Investigações lógicas: investigações para a fenomenologia e a teoria do conhecimento*. Rio de Janeiro: Forense, 2012.

KANT, I. *Crítica da razão pura*. 4a. ed. Petrópolis, RJ: Vozes, 2015.

LEFRANÇOIS, G. R. *Teorias da aprendizagem*. São Paulo: Cengage Learning, 2013.

McCARTHY, J.; HAYES, P. J. Some philosophical problems from the standpoint of artificial intelligence. In: GINSBERG, Matthew (Org.). *Readings in nonmonotonic reasoning*. San Francisco, CA: Kaufmann, p. 26-45, 1987.

McCLELLAND, J. L.; RUMELHART, D.; HINTON, G. E.; PARALLEL DISTRIBUTED GROUP. *Parallel distributed processing: exploration of the microstructure of cognition*. Cambridge, MA: MIT Press, 1986. p. 3-44.

MOHAMED, A.; SAINATH, T.N.; DAHL, G.; RAMABHADRAN, B.; HINTON, G.; PICHENY, M.A. Deep belief networks for phone recognition. In: *Proceedings of the IEEE international conference on acoustics, speech and signal processing*, p. 5060-5063, 2011.

MORRIS, W. E.; BROWN, C. R. David Hume. In: ZALTA, E. N. (Org.). *Stanford encyclopedia of philosophy*. 2017. Disponível em: <<https://plato.stanford.edu/archives/spr2017/entries/hume>>. Acesso em: 30 jun. 2018.

ROGUE, C. *Comprender Platão*. 6a. ed. Petrópolis, RJ: Vozes, 2011.

RUSSELL, S.; NORVIG, P. *Inteligência Artificial*. Rio de Janeiro: Campus, 2004.

SHANAHAN, M. The frame problem. In: ZALTA, E. N. (Org.). *Stanford encyclopedia of philosophy*. 2016. Disponível em: <<https://plato.stanford.edu/archives/spr2016/entries/frame-problem>>. Acesso em: 30 jun. 2018.

STERNBERG, R. J. *Psicologia Cognitiva*. São Paulo: Cengage Learning, 2014.

THOMASON, R. Logic and Artificial Intelligence. In: ZALTA, E. N. (Org.). *Stanford encyclopedia of philosophy*. 2016. Disponível em: <<https://plato.stanford.edu/archives/win2016/entries/logic-ai>>. Acesso em: 30 jun. 2018.

VYGOTSKY, L. S. *Mind in society: the development of higher psychological processes*. Cambridge, MA: Harvard University Press, 1978.

WITTGENSTEIN, L. *Tractatus logico-philosophicus*. São Paulo: EDUSP, 2010.

_____. *Investigações filosóficas*. 7a. ed. Petrópolis, RJ: Vozes, 2012.