Inteligência artificial e consciência

Arthur Araújo/UFG

Resumo

Este artigo trata de alguns recentes avanços epistemológicos em Filosofia da mente. Particularmente enfoca os chamados modelos de processos mentais e as teorias filosóficas subjacentes.
(Palavras-chave: inteligência, consciência, modelo, filosofia.)

Abstract

This article deals with recent epistemological developments in the Philosophy of Mind. Particularly, it focuses on the so-called modelsof mental processes and their underlying philosophical theories.
(Key-words: intelligence, conscience, model, Philosophy.)

Eu, desgraçado, que inventei para os homens tais engenhos, para mim mesmo não descubro uma artimanha com que do presente suplício me liberte. (Ésquilo, *Prometeu Acorrentado*)

A inteligência artificial (IA) vai aparecer, como disciplina científica, mais ou menos nos anos 50 com o objetivo de estudar a representação de comportamentos cognitivos e computacionais de tipo "inteligente", por meio de símbolos e de regras, com a finalidade de simulação das atividades mentais. Por outro lado, deve-se dizer que, desde os anos 70, a chamada filosofia da mente (ou filosofia dos processos mentais) tem sido um campo de interface entre problemas filosóficos e estudos computacionais da mente. Deve-se dizer ainda que o estudo filosófico da IA não significa uma redução da reflexão filosófica a uma crítica da pesquisa científica, mas, ao contrário, pode ser o caso de uma reformulação teórica do materialismo a partir da implementação de modelos computacionais dos processos mentais. Temas como representação, consciência, auto-reflexão, intencionalidade etc, são tratados do ponto de vista das condições materiais dos processos mentais e a partir de modelos computacionais. É uma perspectiva filosófica que visa reformular as bases do materialismo e integrar os temas sacros da tradição filosófica a um processo de laicização epistemológica.

Considera-se, como um problema típico, o seguinte: como é possível a coordenação entre o sistema representacional de um organismo ou de

uma máquina e os chamados itens do mundo exterior ? Trata-se de um problema de *comunicação* que envolve a capacidade de representação e ação efetivas de um indivíduo (animal ou máquina) no meio externo. Compreende-se a representação de uma ação como parte integrante de um processo de desenvolvimento mecânico e cognitivo. Aqui nasce o paradigma da IA: um computador digital poderia simular e coordenar as relações entre mente e ação no meio. A mente vai ser tratada como um conjunto de estados de consciência. Resta saber se tal conjunto é discreto ou contínuo. Caso seja discreto, como explicar a passagem de um estado para outro e o chamado fluxo de consciência? Haveria um algoritmo para tal fim ? Aliás, A. Turing (1950) publica um artigo consagrado à seguinte questão: as máquinas podem pensar?[1] Turing partia do que se poderia chamar a distinção entre dois níveis (ou estados) de consciência:

a - consciência operacional (mecânica): *awareness.*

b - consciência reflexiva (auto-consciência): *self-consciousness.*

A conclusão de Turing é a seguinte: uma máquina poderia realizar tudo o que um ser humano faz, porém não teria consciência do que faz.[2] É a chamada "objeção de consciência" (*consciousness objection* — cf. Turing, 1950, *p.* 59-61). Nós poderíamos dizer que a motivação filosófica de Turing remonta a R. Descartes (*Discours de la Méthode* — 5ª Parte). Descartes concebeu uma classificação categórica dos autômatos:

a - autômatos de 1º grau: máquinas e animais.

b - autômatos de 2º grau: homem.

A diferença entre ambos é, segundo Descartes, a capacidade de auto-reflexão. Faltaria aos autômatos de 1º grau o espírito. Por outro lado, as teorias recentes da consciência interpretam o espírito de um modo distinto da tradição clássica desde Descartes. O termo espírito significa *mind* (mente), um correlato variante do sentido clássico. Mente representa um conjunto de estados de consciência não-transcendentais e não-reflexivos: a mente corresponde a um conjunto formado por processos mecânicos (característicos dos autômatos de 1º grau) e por processos reflexivos (característicos dos autômatos de 2º grau).

Pode-se dizer que as teorias recentes da consciência vão reformular as bases do problema cartesiano da comunicação mente—corpo e laicizar a tradição espiritualista da auto-consciência. Com efeito, vamos observar o aparecimento de três momentos da IA

relacionados a fases subjacentes das teorias filosóficas da consciência:

1º momento: máquinas mecânicas: teoria mecanicista da consciência

{ consciência entendida como processo mecânico

2º momento: analogia cérebro—máquina, e
3º momento: auto-organização

{ teoria materialista da consciência (consciência entendida como processo material

Entre várias perspectivas, algumas teorias da consciência abandonam a clássica distinção entre métodos de espírito e métodos de máquina. Aliás, uma distinção já abandonada na fase embrionária da IA baseada nas máquinas mecânicas (modelo da máquina de Turing).

Máquinas mecânicas e teoria mecanicista da consciência

Aqui entende-se a consciência como um processo mecânico. O modelo histórico e característico da consciência são as máquinas de calcular. Descreve-se os estados de consciência por seqüências lógico-algébricas. A idéia básica é aquela formulada por Leibniz e desenvolvida por D. Boole (1854): uma linguagem universal capaz de estruturar os processos de pensamento independente de conteúdo mental ou de objeto. Os processos de pensamento representam estados de consciência expressos por sinais e estruturados como linguagem simbólica. Mais recentemente, as chamadas "Teorias computacionais da mente" vão seguir de modo geral o princípio da "Álgebra booleana": os estados mentais podem ser interpretados em termo de 'leis lógicas'.

Boole desenvolve uma Álgebra da Lógica como modelo interpretante da lógica formal aristotélica associando formas lógicas e expressões algébricas; por exemplo: '0' e '1' correspondem, respectivamente, aos valores lógicos falso e verdadeiro. Posteriormente, a álgebra de Boole vai ser a base lógica de operação das máquinas eletrônicas de 1ª geração. Trata-se do desenvolvimento histórico do projeto de Leibniz de interpretar os processos de pensamento através de um formalismo universal adequado; por exemplo, em termos de máquinas eletrônicas, associam-se estados da corrente elétrica (negativo e positivo) a símbolos lógico-algébricos (0 e 1). Aliás, Leibniz foi um dos pioneiros na construção de máquinas de calcular, ao lado de Pascal e Descartes.

Resumidamente, diríamos que o princípio característico das máquinas

mecânicas vai expressar uma teoria da consciência baseada na manipulação de símbolos e de regras pré-fixadas. É um tratamento reducionista da consciência. Parece, por outro lado, um anacronismo. Atualmente ainda sentimos um certo refluxo da perspectiva mecanicista. Trata-se do debate entre Penrose e Minsky acerca da descrição da consciência por meio de estruturas algorítmicas. Penrose defende um ponto de vista cartesiano, ou seja, os estados de consciência são mais complexos e superiores do que as seqüências lógico-algébricas. Ele defende, portanto, uma incompatibilidade entre processo de máquina e processo de espírito. Por sua vez, Minsky defende a possibilidade de uma descrição algorítmica da consciência desde que seja possível compatibilizar processos mecânicos e a complexidade dos processos da mente. Minsky sustenta uma perspectiva bastante próxima àquela de Leibniz e Boole.

Redes neurais e teoria materialista da consciência.

Analogia cérebro/máquina

Os estados de consciência são descritos como parte de processos materiais. O objetivo é superar o abismo entre níveis operacionais e reflexivos de consciência a partir da analogia cérebro/máquina. Amplia-se o modelo computacional da máquina de Turing com a finalidade de descrever propriedades específicas da atividade neural. O modelo pioneiro de McColluch & Pitts (1943) interpreta o cérebro como o equivalente material da máquina de Turing.

Considera-se a estrutura do cérebro como um complexo de redes neurais conectadas umas às outras, responsáveis pelo processamento de informação. Teríamos assim dois estados básicos das redes neurais: inibição (0) e excitação (1). Os chamados 'objetivos mentais' e os estados mentais são interpretados em termos de quantidade processada de informação. É a primeira vez que vai aparecer a idéia de uma realização física da máquina de Turing associando símbolos matemáticos (operacionais) e conteúdos mentais. O modelo McColluch & Pitts é pioneiro na interpretação dos objetos e estados mentais como configurações variáveis das redes neurais.

Pode-se dizer que a estrutura e as condições materiais do cérebro (conjunto das redes neurais) têm prioridade ontológica sobre os elementos de ordem mental (por exemplo, elementos intencionais).[3] Os conteúdos mentais não dependem mais de uma explicação extra-neural do tipo *ad hoc*; por exemplo, uma entidade transcendental a partir da qual decidiría-

mos sobre nossa atividade mental. O modelo McColluch & Pitts parte da idéia de simulação de processos mentais por meio da implementação dos chamados neurônios formais (*formal neurons*); unidades que constituem uma máquina lógico-realizável da máquina de Turing. Busca-se representar as mesmas propriedades e características dos neurônios naturais. Com efeito, o cérebro é um processador de informação que conta basicamente com sua estrutura e propriedades para a orientação de ações do organismo no meio a partir de representações informacionais.[4] Claro, a função principal do cérebro não é representar o mundo exterior, mas ao mesmo tempo ele torna possível coordenar a quantidade de informação, o equivalente às representações dos eventos e dos objetos do mundo e as ações correspondentes do organismo no meio.

Auto-organização

O que marca o aparecimento da auto-organização (AO) como orientação epistemológica são principalmente os trabalhos do biofísico H. von Foerster a partir dos anos 60. AO é um princípio operacional que isenta o organismo ou a máquina de recorrer a um elemento exterior como fonte de organização do sistema representacional. O sistema não depende de instrução do mundo exterior (elementos sensíveis, mentais, transcendentais etc). AO faz referência às próprias propriedades e características do conjunto de redes neurais (naturais ou formais) que constituem o sistema representacional do organismo ou da máquina.

Obviamente, não é o caso de um solipsismo do sistema neural, mas, ao contrário, é mantida uma permanente e constante interação do organismo com o meio.[5] Todavia, não são as variações ou as instruções do meio que estruturam a capacidade de ação e de reflexão do organismo ou da máquina. O modelo de AO apresenta uma explicação dos estados de consciência que parte exatamente da implicação recíproca entre fenômenos de ordem mecânica (*awareness*) e fenômenos de ordem reflexiva (*self-consciousness*): o organismo deve ser capaz de refletir sua ação no meio; a consciência representa um momento (intervalo) entre o processamento de informação e a ação correspondente. Quanto maior o intervalo, maior o grau do estado de consciência.

Abandona-se, por outro lado, a oposição percepção-cognição. O cérebro recebe uma enorme quantidade de informação, via nervo ótico, que as redes neurais processam de modo paralelo e distribuído, ao contrário da computação tradicional, de tipo seqüencial. A percepção já é um estado

de consciência reflexiva. Por exemplo, agora eu sei que estou aqui e não tenho que refletir sobre meu estado. O chamado 'eu' do cérebro é muito mais o efeito do sistema representacional neural e menos a expressão de uma unidade reflexiva da consciência (Cf. Eccles & Popper, 1991). As orientações filosóficas materialistas são também designadas monistas, diferentemente do chamado dualismo, exatamente porque não aceitam a distinção mente-cérebro, ou ainda a idéia de um "eu" separado da estrutura física cerebral. Talvez o modelo que melhor ilustra o paradigma da AO seja o sistema imunológico. Este constitui parte do sistema representacional do organismo: o sistema cria as defesas (representações) imunológicas indiferente ao referente externo (por exemplo, condições do meio ou um vírus). Analogamente, a máquina poderia criar as representações sem determinação de instrução ou de intervenção direta do meio.

Assim, por exemplo, as chamadas representações mentais correspondem a certas quantidades de informação que o cérebro processa de modo paralelo e distribuído. As representações significam níveis complexos de articulação entre os itens do sistema representacional (o conjunto das redes neurais) e os itens do meio (o conjunto dos elementos perceptivos): o processamento de informação torna possível uma coordenação entre a informação processada e a ação (consciente) do organismo (ação coordenada). Por outro lado, as configurações variáveis das redes neurais vão permitir a identificação dos objetos mentais e a respectiva ação. Com efeito, o cérebro realiza um certo tipo de *scanning* perceptivo:

1º momento: representação operacional (mecânica) dos dados perceptivos.

2º momento: identificação dos dados (por exemplo., reconhecimento de formas sonoras ou visuais).

Há, portanto, uma coordenação entre processos de ordem mecânica e processos de ordem reflexiva. Trata-se de compreender os estados de consciência como parte integrante da vida biológica, o que nos leva certamente a uma verticalização do fenômeno da consciência, i.é, ao mapeamento das raízes biológicas dos processos cognitivos e interativos (sociais). Podemos ainda redefinir as bases da teoria materialista da consciência com o seguinte problema de implementação em máquinas: como um sistema representacional auto-organizado é capaz de coordenar representação operacional de informação e ação reflexiva no meio ? Propomos duas considerações finais:

Primeira: é possível a realização de um *scanning* perceptivo em redes neurais ? O cérebro deve ser capaz de representar a realidade, via processo mecânico, como um processo reflexivo de representação da ação do organismo no meio.

E segunda: um processo material e mecânico é capaz de fornecer as condições de desenvolvimento de um processo auto-reflexivo representando uma ação consciente no meio.

São dois pontos que caracterizam problemas de implementação, ou seja, modelos capazes de estruturar as bases de uma teoria materialista da consciência. Isso inclui certamente o debate intelectual franco com outras perspectivas teóricas; por exemplo, a robótica e o desenvolvimento de máquinas capazes de reflexão, comportamento inteligente no meio, 'vida interior'.

Notas

1. É comum aparecer a expressão "máquina de Turing"em ciência cognitiva e IA. Trata-se de uma modelo formal que se tornou a base dos computadores digitais. Basicamente a "máquina de Turing" requer: um conjunto de símbolos, regras pré-fixadas e um procedimento efetivo, como elementos básicos dos chamados "comportamentos inteligentes".

2. Aqui nós teríamos a idéia do computador como modelo da consciência. Cf. Putnam, 1988, p.162: "(...) a consciência tem um <<programa>> ou conjunto de regras, análogas às regras que governam um computador, e o pensamento implica a manipulação de palavras e de outros símbolos (nem toda esta manipulação é <<consciente>>, no sentido de ser susceptível de ser verbalizada pelo computador)." Cf. também Searle (1990): a manipulação de símbolos não é condição suficiente ao desenvolvimento de auto-consciência.

3. A identidade mente=cérebro aparece em muitas teorias como base ontológica dos comportamentos cognitivos. Trata-se de uma identidade que visa, por outro lado, compatibilizar os 'estados de máquina' (*hardware*) e os 'estados de programa' (*software*).

4. Em IA e ciência cognitiva, o termo 'representação' refere-se à informação internalizada como conteúdo das estruturas neurais.

5. Há uma controvérsia entre alguns autores da AO quanto ao grau de interação sistema-meio. Particularmente, Maturana e von Foerster (1978) defendem um grau '0' de interação no desenvolvimento dos sistemas (estes são 'fechados' organizacionalmente). Por sua vez, Atlan (1978) defende um desenvolvimento 'aberto' dos sistemas auto-organizados.

Referências Bibliográficas

ATLAN, H. Consciência e desejo em sistemas cognitivos. In: MORIN, E. *et alii.* (Eds.). *A unidade do homem*: invariantes biológicos e culturais. Trad. de Heloísa de Lima Dantas. São Paulo: Cultrix-Edusp, 1978.

ECCLES, J. C. & POPPER, K. *O eu e seu cérebro.* Trad. de Sílvio M. Garcia, Helena Cristina F. Arantes e Aurélio Osmar C. de Oliveira. Campinas: Papirus, Ed. da UnB, 1991.

MATURANA, H. Estratégias cognitivas. In: MORIN, E., op. cit.

PUTNAM, H. Formalização. In: *Enciclopédia Einaudi.* v. 13. Porto: Imprensa Nacional - Casa da Moeda, 1988.

SEARLE, J. L'esprit est-il un programme d'ordinateur. *Pour la science.* Paris, n. 149, p. 38-44, mars, 1990.

TURING, A. M. . Computing machinery and intelligence. In: HOFSTADTER, D.; DENNETT, D. (Eds.). *The Mind's I.* New York: Basic Books, Inc., Publishers, 1981.

VON FOERSTER, H. Notas sobre a epistemologia dos objetos vivos. In: Morin, E., op. cit.