

# O QUE É INTELIGÊNCIA ARTIFICIAL

**João de Fernandes Teixeira**

## INTRODUÇÃO

Conta uma velha anedota que uma vez um famoso teólogo da idade Média foi visitar o rei Alberto, o Grande. Quando chegou ao palácio real foi recebido por um boneco mecânico que se encarregou de abrir-lhe a porta e fazer-lhe as medidas de um autêntico mordomo. Indignado, o teólogo não resistiu aos seus impulsos e estraçalhou o boneco mecânico.

Esta anedota – talvez verdadeira, quem sabe – pode nos ajudar a dar um primeiro passo para entendermos o que seja a Inteligência Artificial. A expressão “Inteligência Artificial” soa de maneira assustadora, e talvez muitos de nós reagíssemos como o teólogo em sua indignação ao ver que uma máquina pode fazer aquilo que achamos ser uma exclusividade do gênero humano: pensar e agir racionalmente, executando tarefas para as quais se supõe que a inteligência seja necessária. Talvez muitos já tenham ouvido falar de projetos mirabolantes, como a construção de “cérebros eletrônicos” que seriam ligados na tomada e teriam “pensamentos” iguais aos nossos, ou até mesmo poderiam comunicar-se conosco, falando normalmente.

O que muitos de nós não sabem, entretanto, é que a idéia de se criar algo parecido com “máquinas pensantes” ou uma inteligência artificial paralela à nossa é hoje um projeto no qual trabalham cientistas de várias partes do mundo. Esses cientistas trabalham em várias áreas do conhecimento humano: lingüística, psicologia, filosofia, ciência da computação etc. O que os reúne é, entretanto uma característica comum: a idéia de que é possível criar “máquinas pensantes” e que o caminho para isso é o estudo e a elaboração de sofisticados programas de computador.

Para os pesquisadores da Inteligência Artificial (que daqui por diante abreviaremos IA) a mente humana funciona como um computador, e por isso o estudo dos programas computacionais é a chave para se compreender alguma coisa acerca de nossas atividades mentais. Podemos construir programas que imitem nossa capacidade de raciocinar, de perceber o mundo e identificar objetos que estão à nossa volta, e até mesmo de falar e de compreender nossa linguagem.

É esta grande novidade da IA, que a distingue de ciências afins como a cibernética e a computação, englobando-as num projeto muito mais ambicioso: a produção de comportamento inteligente. Tarefas para as quais se requer alguma inteligência já são executadas por algumas máquinas de que dispomos e que utilizamos para telefonar ou mesmo para lavar roupa. Contudo, esse tipo de máquinas não tem interesse para a IA de que falaremos aqui, cuja preocupação é não só aliviar o trabalho humano, mas também desvendar alguma coisa acerca da natureza da nossa mente. Para isso é preciso que essas máquinas realizem tarefas que requerem inteligência, e de uma maneira muito similar e próxima do modo como nós, seres humanos, as realizamos.

Assim, por exemplo, não basta simplesmente projetar e criar uma máquina de calcular (como nós a temos e usamos todos os dias, carregando-a no bolso) para dizermos que estamos fazendo IA. É preciso que essa máquina imite nossa atividade mental quando estamos fazendo uma operação aritmética.

Claro que esta imitação não poderá ser absolutamente perfeita. Afinal, somos constituídos de matéria viva e até agora ainda não pudemos construir um mecanismo totalmente semelhante ao nosso cérebro, com suas células e as ligações nervosas que existem entre elas. Mas, assim como existem mecanismos que imitam nossa capacidade de andar, como por exemplo, o automóvel, da mesma maneira o teórico da IA propõe-se a criar mecanismos que, embora não sejam idênticos a nós, possam imitar nossas atividades mentais. A imitação que se pode obter é, portanto, apenas aproximada, e é por isso que em IA fala-se da elaboração de programas de computador que são *modelos* de nossa capacidade de raciocinar, de enxergar, de falar etc.

A IA como projeto efetivo só se tornou possível após o aparecimento dos computadores modernos, ou seja, após a Segunda Guerra Mundial (de 1945 em diante). Até

então havia dificuldades técnicas que precisavam ser superadas para que o projeto dessas máquinas mais modernas pudesse sair do papel.

Quando apareceu pela primeira vez uma máquina dita “pensante” – uma máquina dotada de um programa que demonstrava automaticamente teoremas de matemática -, o impacto sobre as ciências do homem foi tremendo. Subitamente a comunidade científica percebeu que uma verdadeira revolução havia se iniciado. Uma revolução com profundas influências na psicologia, na lingüística e na filosofia. A ciência da computação deixava de ser uma disciplina puramente técnica, e suas realizações passaram a estender-se para outros campos. A idéia de estudar a mente humana à semelhança de um programa de computador parecia despontar como uma nova etapa para as ciências humanas.

Para a psicologia, a IA trouxe uma revolução, na medida em que o novo modelo apontava para uma alternativa à turbulência teórica que os psicólogos estavam atravessando. Havia muita discussão sobre a própria natureza do estudo a que se propunham os cientistas dessa área: se o seu estudo devia apenas se concentrar no comportamento dos organismos, ou se a psicologia deveria ser um estudo de nossas atividades mentais mediante um auto-exame, através do qual fecharíamos os olhos e tentaríamos perscrutar o que se passa em nossa cabeça quando raciocinamos. Ambas possibilidades deixavam muito a desejar.

Para a lingüística, a IA significava uma revolução: agora seria possível criar um programa de computador onde estivessem representadas as estruturas gramaticais das diversas línguas humanas. Se isto pudesse ser feito, um sonho muito antigo seria realizado: teríamos descoberto a raiz comum de todas as línguas humanas e uma máquina universal de tradução tornar-se-ia possível.

Mas foi realmente sobre a filosofia que o impacto da IA foi maior: criar uma máquina pensante significa desafiar uma velha tradição que coloca o homem e sua capacidade racional como algo único e original do universo. Mais do que isto, criar uma máquina pensante significa dizer que o pensamento pode ser recriado artificialmente sem que para isto precisemos de algo como uma “alma” ou outra marca divina.

Algumas questões que tradicionalmente atormentaram os filósofos ao longo dos séculos passaram a receber um novo enfoque a partir da IA: por exemplo, o problema das

relações entre a mente e o corpo, que se arrasta há milênios. Durante muitos anos os filósofos discutiram entre si se os nossos estados mentais (raciocínio, sonhos, imagens mentais etc.) seriam apenas manifestações de nossa atividade cerebral (materialismo) ou se eles não seriam reveladores da existência de algo imaterial como, por exemplo, uma alma imortal (dualismo). A IA oferece uma nova perspectiva para situarmos este problema para além das soluções existentes, que pendem seja para o materialismo seja para o dualismo.

A recepção dada pelos filósofos ao novo modelo da mente humana proposta pela IA foi, entretanto, muito ambígua. Alguns filósofos viram na IA uma alternativa para a filosofia tradicional, que deixaria de ser apenas um conjunto de discussões acadêmicas às vezes consideradas estereis e inconclusivas. As propostas filosóficas poderiam agora ser testadas em laboratórios, criando-se modelos computacionais para nossa maneira de raciocinar, de perceber o mundo e de formar pensamentos e idéias a partir dos objetos que estão à nossa volta. Esta seria a verdadeira filosofia “experimental” que muitos filósofos do passado gostariam de ter visto.

Outros filósofos, contudo, reagiram com a mesma indignação que parece ter sido experimentada pelo teólogo medieval em visita ao palácio real. Reduzir o ser humano e o pensamento às atividades de uma máquina seria uma proposta no mínimo ultrajante.

Se o pensamento humano pode ou não ser mecanizado, como pretendem os teóricos da IA, é uma questão que ainda permanece em aberto. Tudo dependerá ainda de realizações futuras e de algum tipo de consenso a que os filósofos ainda hesitam em chegar. Para se ter uma noção mais precisa do que a IA propõe como programa de pesquisa, é preciso saber um pouco de sua história, de suas realizações até agora, e saber, em linhas gerais, como funciona um computador, o que para muitos é ainda um mistério. É também a estes temas que dedico os capítulos a seguir.

## **UM POUCO DE HISTÓRIA. E DE ESTÓRIA**

Fazer uma história precisa do desenvolvimento da IA não é tarefa fácil. Isto porque, embora seu aparecimento como disciplina científica só tenha ocorrido a partir da Segunda Guerra Mundial (1939-1945), a idéia de construir uma máquina pensante ou uma criatura artificial que imitasse as habilidades humanas é muito antiga. Os primeiros registros de criaturas artificiais com habilidades humanas têm uma forma mítica ou por vezes lendária, tornando difícil uma separação nítida entre imaginação e realidade. A idéia é de fato muito antiga, mas as condições técnicas para a sua realização são coisa recente. É esta confusão entre mito e realidade e, por vezes, a impossibilidade de distinguí-los que faz com que a IA possa ser considerada uma disciplina com um extenso passado, mas com uma história relativamente curta.

Um dos episódios mais interessantes do passado mítico da Ia é a lenda do Golém. Joseph Golém era um homem artificial que teria sido criado no fim do século XVI por um rabino de Praga, na Tchecoslováquia, que resolvera construir uma criatura inteligente, capaz de espionar os inimigos dos judeus – então confinados no gueto de Praga. O Golém era de fato um ser inteligente, mas que um dia se revoltou contra seu criador, o qual então lhe tirou a inteligência e o devolveu ao mundo do inanimado.

Alguns registros mais recentes mostram que nos séculos XVII e XVIII proliferam mais mitos e lendas acerca de criaturas artificiais. Fala-se de um flautista mecânico que teria sido capaz de tocar seu instrumento com grande perfeição, e que teria sido construído lá pelos fins do século XVII. Há registros também do célebre “pato de Vaucanson”, que teria sido construído por um artífice homônimo. A grande novidade dessa criatura teria sido sua capacidade de bater as asas, andar, grasnar, comer grãos e expeli-los após a digestão – uma perfeita imitação das funções biológicas.

A existência passada dessas criaturas artificiais até hoje não está definitivamente comprovada. Sabe-se apenas que seus projetos estão registrados em alguns museus da Europa e que sua arquitetura interna teria sido extremamente complexa.

Os séculos XVII e XVIII conheceram também pela primeira vez uma preocupação filosófica com algumas implicações teóricas envolvidas na construção dos primeiros autômatos de que se tem notícia. Descartes (1596-1650), filósofo racionalista do século XVII e oficialmente considerado o criador da filosofia moderna, expressou este tipo de

preocupação em várias passagens de sua obra, argumentando que os autômatos, por mais bem construídos que fossem, jamais se igualariam aos seres humanos em termos de suas habilidades mentais. Isto porque os autômatos nunca viriam a ter uma alma imortal, igual à nossa, que lhes permitisse agir livremente e encadear sentenças de modo a expressar pensamentos como nós, humanos, o fazemos.

Mesmo que se construísse um autômato com cordas vocais e boca semelhantes às de um ser humano, ele jamais seria capaz de falar. No máximo, seria um repetidor de palavras, como um papagaio, mas isso não significa falar: significa apenas pronunciar palavras de uma maneira vazia, pois atrás destas não haveria pensamentos. Esse tipo de argumento formulado por Descartes foi ressuscitado por um filósofo norte-americano contemporâneo, John Searle, na forma de uma forte objeção às pretensões da IA.

Descartes não foi o único a se preocupar com os problemas filosóficos envolvidos na construção de autômatos e com a discussão de suas possíveis habilidades mentais. Um século depois outro filósofo francês, La Mettrie (1709-1751), escreveu um livro criticando os pontos de vista de Descartes. Ele sustentava uma posição oposta à de Descartes, argumentando em favor da idéia de que o pensamento não é o resultado da atividade de uma alma imortal e imaterial que teríamos dentro de nós. As faculdades de pensar e de falar, dizia La Mettrie, aparecem naquelas criaturas que têm um organismo mais complexo, com um cérebro mais desenvolvido. Se os animais ou os autômatos ainda não desenvolveram essas faculdades, isso se deve apenas a imperfeições fisiológicas ou mecânicas dessas criaturas, cujos organismos ou mecanismos têm um grau de complexidade inferior ao nosso.

La Mettrie defendeu suas idéias com grande paixão, e quando hoje em dia lemos sua obra, que infelizmente ainda não foi traduzida para o português, temos a impressão de estarmos diante de um verdadeiro manifesto em favor da IA. Nos seus trabalhos, este autor declara, entusiasticamente, que da mesma maneira que a humanidade teve, no começo dos tempos, um herói grego como Prometeu, que roubou o fogo dos deuses para dá-lo aos homens, um dia teríamos um segundo Prometeu que construiria um homem mecânico capaz de falar.

O século XIX não foi muito fértil no que diz respeito a discussões filosóficas acerca

das habilidades e possibilidades dos autômatos. Tudo se passou como se momentaneamente o assunto tivesse sido esquecido e as preocupações tivessem se voltado para outros temas que então inquietavam a humanidade: a questão social e o agigantamento do industrialismo, que parecia prometer uma completa modificação na imagem do mundo.

Foi somente na literatura – e na literatura do movimento romântico – que o tema do surgimento de criaturas artificiais parece ter sido lembrado. Nessa época é publicado o famoso romance *Frankenstein*, que explora o mito de um ser criado a partir de membros e órgãos de outras criaturas artificialmente reunidos. O sopro vital era substituído por um choque elétrico que fez com que a criatura adquirisse vida. Mas Frankenstein era um monstro. Um monstro que logo em seguida se revoltou contra seu criador.

Mas se o século XIX não produziu uma literatura onde o tema dos autômatos aparecia de forma explícita, nem por isso se pode dizer que nele faltaram descobertas que mais tarde seriam úteis para a IA. Invenções matemáticas e avanços tecnológicos, resultantes da descoberta da eletricidade, foram registrados – avanços que propiciariam, mais tarde, a construção dos modernos computadores.

A IA só aparecerá no século XX, e num contexto bem diferente do que poderíamos ter imaginado. O advento da Segunda Guerra Mundial trouxe pressões decisivas para a comunidade científica dos países aliados. Os bombardeios aéreos feitos pelos nazistas sobre as cidades européias pressionaram o desenvolvimento de canhões anti-aéreos dotados de um sistema de pontaria que corrigisse os eventuais desvios causados pelo deslocamento do alvo e do próprio canhão no momento do disparo. Esse tipo de mecanismo de autocorreção começou a ser visto como uma incipiente imitação de um comportamento humano. Para um observador leigo, tudo se passava como se o comportamento do canhão, ao perseguir seu alvo com precisão, estivesse sendo guiado por propósitos ou intenções semelhantes ao do ser humano.

Paralelamente a estes desenvolvimentos na indústria bélica, o surgimento dos campos de concentração abriu uma nova possibilidade: a realização de experimentos e estudos do cérebro em seres humanos. Esses estudos eram feitos não só com cobaias humanas (prisioneiros de guerra) como também com soldados que tivessem sofrido lesões cerebrais durante os combates. Através desses estudos alguns cientistas tentavam aprender

alguma coisa acerca dos princípios gerais do funcionamento do cérebro. Isso incluía a localização das áreas cerebrais responsáveis pelas diversas atividades humanas. Assim, por exemplo, para se tentar localizar a área cerebral que corresponderia à função da fala, removiam-se sucessivamente partes do cérebro do paciente, até que este se visse impossibilitado de falar.

No fim da Segunda Guerra Mundial, os cientistas já tinham registrado importantes invenções na área eletrônica, além de pesquisas sobre mecanismos que imitavam ações humanas e estudos sobre cérebro humano desenvolvidos por médicos e por psicólogos. Isso os levou a programarem um encontro nos Estados Unidos, onde pesquisadores dessas áreas apresentavam suas descobertas, numa primeira tentativa de reuni-las e compor algo parecido com uma ciência geral do funcionamento da mente humana. Esse encontro ficou conhecido como o Simpósio de Hixon, e aconteceu em 1948.

Quando os cientistas se reuniram no Simpósio de Hixon, já sabiam que a construção de um computador eletrônico tinha se tornado uma realidade. Naquela época, os escritos do matemático inglês Alan Turing, que continham os princípios de funcionalidade dos computadores modernos, começavam a sair nas revistas especializadas. Alan Turing era um jovem preocupado com questões matemáticas, cuja adolescência havia sido marcada pelo interesse por problemas filosóficos, como, por exemplo, as relações entre a alma e o corpo e a reencarnação. Quando se dedicou ao estudo de certas questões matemáticas, ele provavelmente não sabia que suas descobertas tornariam possível a construção dos computadores modernos. Ele descobriu o princípio fundamental do funcionamento dessas máquinas – um princípio que até hoje norteia a construção de computadores, por mais sofisticados que sejam.

Os resultados do Simpósio de Hixon não teriam sido tão surpreendentes se não levassem, através de uma intuição verdadeiramente criadora, a se estabelecer uma analogia entre o cérebro humano e os computadores. Essa analogia certamente foi produto do encontro entre psicólogos, neurofisiólogos e engenheiros eletrônicos que perceberam que o modo como estão dispostas as células do nosso cérebro (neurônios), ligadas através de fios nervosos minúsculos, é semelhante ao circuito elétrico de um computador eletrônico. Estava aberto o caminho para se dizer que a mente humana pode ser imitada por um

computador. Com isto nascia a nova disciplina, a Inteligência Artificial.

As décadas seguintes foram marcadas por novas invenções e descobertas surpreendentes. Na década de 50, dois cientistas americanos desenvolveram um programa de computador capaz de demonstrar automaticamente teoremas matemáticos. Esse programa foi chamado de “O Teórico da Lógica”, e sua inovação estava no fato de ele poder realmente gerar demonstrações de teoremas, e não simplesmente apresentá-las através de um artifício de memória. Os dois cientistas norte americanos estavam realmente convencidos de que sua máquina era uma autêntica simulação do modo como os seres humanos resolvem seus problemas matemáticos, e escreveram vários artigos a esse respeito.

Mas a maior novidade ainda estava por vir. No final da década de 60, aparece um programa de computador capaz de imitar um psicanalista. Esse psicanalista mecânico, que foi chamado “DOCTOR”, era na verdade uma variação de um outro programa batizado de “ELIZA”. ELIZA foi um programa originalmente desenvolvido para simular diálogos e conversas. O programa analisava as frases e devolvia as respostas utilizando a máquina de escrever.

O princípio de funcionamento desse tipo de programa era simples: a sentença enviada pelo parceiro humano era decomposta, e suas partes enviadas para um *script* armazenado no interior do computador. O *script* era um conjunto de regras semelhantes àquelas que são dadas para um ator quando se requer que ele improvise acerca de um tema qualquer. ELIZA podia receber vários tipos de *script*, e, dependendo do conteúdo destes, desenvolver conversas acerca de vários temas.

Quando Eliza trabalhava com um *script* especial chamado Doctor, ele se transformava num psicanalista mecânico. O *script* era cuidadosamente elaborado para que as respostas simulassem o comportamento verbal de um psicanalista ao receber um paciente pela primeira vez.

Muitas pessoas, na época, chegaram a afirmar que ELIZA era apenas um truque de memória, e que suas respostas sempre obedeciam a padrões preestabelecidos. Se ELIZA é ou não um autêntico psicanalista artificial é uma questão que pode ser discutida. Qualquer que seja a conclusão, uma coisa porém é certa: Eliza é uma imitação muito convincente do

comportamento verbal de um psicanalista humano.

Por volta de 1970 foi inventado, no Massachusetts Institute of Technology (o famoso MIT), nos Estados Unidos, um sistema chamado “SHRDLU”. O Shrdlu simulava uma espécie de robô fechado num ambiente artificial onde estavam colocados blocos de madeira coloridos, do tipo dos usados em jogos infantis. O Shrdlu era capaz de obedecer a instruções e falar sobre a posição dos blocos que movia e de seus “braços”.

Essas realizações trouxeram a consolidação da IA como disciplina científica. A partir delas, o entusiasmo por imitar as atividades da mente humana tem crescido cada vez mais. As pesquisas hoje em dia concentram-se em várias áreas: a tradução automática, a criação de máquinas capazes de perceber e identificar objetos que estão à sua volta, o aperfeiçoamento de programas para jogar xadrez ou damas e assim por diante. Uma pesquisa que tem atraído muito a atenção dos cientistas é a criação de uma máquina que reconheça vozes humanas. Quando essa máquina for criada, teremos a possibilidade de dispensar os teclados, e passaremos a nos comunicar com os computadores simplesmente falando. Mas isso ainda requer um progresso técnico muito grande. Atualmente só podemos nos comunicar com os computadores usando o teclado e em linguagens especiais: as linguagens nas quais são escritos os programas, muito diferentes da linguagem que falamos normalmente.

As realizações da IA não deixaram de chamar a atenção dos filósofos, que perceberam que muitos de seus conceitos e idéias teriam de ser revistos. Os progressos da IA tiveram um impacto muito grande nas concepções habituais que temos da mente humana, e esse fato foi imediatamente notado por aqueles que tem preocupações filosóficas e religiosas. Mas antes de expor as implicações filosóficas da IA, falarei um pouco dos princípios que norteiam a construção dessas máquinas e de onde vem seu extraordinário poder, que faz com que elas cada vez mais participem da nossa vida.

## **A INVENÇÃO DE TURING**

Para muitas pessoas o computador é ainda uma máquina misteriosa e às vezes assustadora. Quando olhamos para um microcomputador e deparamos com sua tela, seu teclado, e sua impressora, temos às vezes a sensação de estarmos diante de uma máquina mágica. Mesmo a grande maioria daqueles que o utilizam para realizar operações complexas muitas vezes não sabe o que ocorre dentro dessas máquinas. A situação é semelhante à de um motorista que dirige um carro sem, entretanto saber os princípios de funcionamento do motor.

O princípio de funcionamento dos computadores é relativamente simples, mas foram precisos anos para que se pudesse descobri-lo. Essa descoberta deveu-se a Alan Turing (1912–1954), um matemático inglês que, apesar de ter sido brilhante na sua época, teve uma vida particularmente atribulada: não pertencia à aristocracia da Inglaterra, o que lhe criava dificuldades em certos meios acadêmicos, e., ademais, era homossexual, o que escandalizava a sociedade britânica. Embora tenha morrido prematuramente (ele se suicidou, provavelmente por motivos que tinham a ver com sua homossexualidade), Turing deixou uma vasta produção de trabalhos e invenções matemáticas.

Foi na tentativa de resolver um problema matemático muito complexo que estava sendo discutido na década de 30 que ele criou a chamada máquina de Turing. Para termos uma idéia do que seja uma máquina de Turing basta que imaginemos uma longa fita de papel com símbolos e marcas a intervalos regulares, formando pequenos quadrados. Imaginemos agora que podemos estipular uma espécie de marcador ou um ponto fixo em relação ao qual pudéssemos mover a fita de papel para a esquerda ou para a direita. A situação de que falamos pode ser representada assim:

	S1	E2 s2	D1 s3	R s4	A s5	
--	----	-------	-------	------	------	--

|

fig. 1

marcador

Suponhamos agora que o nosso marcador tenha também um dispositivo que permita

reconhecer se num determinado quadrado há um símbolo ou ainda não, imprimir e apagar símbolos que aparecem na fita e ainda movê-la para a esquerda ou para a direita, dependendo do símbolo que aparece impresso. Na fita que aparece na figura, os quadrados têm dois tipos de símbolos: letras minúsculas e letras maiúsculas. Mover a fita para a esquerda ou para a direita (e num número determinado de quadrados) dependerá do símbolo em maiúsculas e é identificado pelo marcador. Além de mover a fita em determinadas direções, o símbolo em maiúsculas pode significar que o marcador deve imprimir ou apagar um símbolo num certo quadrado.

Em outras palavras, os símbolos A, B, C, D, E etc. representam as instruções que devem ser seguidas pela máquina, movendo a fita ou apagando os outros símbolos  $s_1$ ,  $s_2$ ,  $s_3$  etc. No caso da fita que representamos na figura anterior, podemos convencionar que os símbolos E e D significam mover a fita para a esquerda ou para a direita, e que o número que está junto de E ou de D representa o número de casas que se quer que a fita mova, seja numa direção ou outra. R significa “imprima o símbolo em minúsculas que está ao lado”, A significa “apague”, e assim por diante.

Vamos agora fazer uma outra suposição: a de que alguém queira usar uma máquina deste tipo para efetuar uma operação aritmética simples, como, por exemplo, uma soma. Para que isto possa ser feito, cada número terá de ser representado por uma sucessão de I, ocupando cada um deles um quadrado. Assim, o número 2 será representado por dois quadrados, sendo que em cada um deles deve figurar o símbolo I. O número 3 ocupará três quadrados, cada um com o símbolo I, e assim por diante. Na nossa máquina isto aparecerá assim:

	I	I	+	I	I	I	
--	---	---	---	---	---	---	--

|

fig. 2

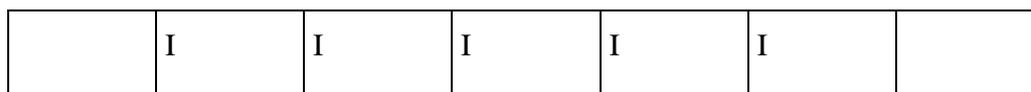
marcador

Que tipo de instruções teremos de dar à máquina para que ela efetue a operação

2+3, isto é, para que ela venha a representar o número 5? Para isto temos de fazer com que ela obedeça às seguintes instruções:

- a) Apague o sinal +.
- b) Imprima o sinal I na mesma casa.
- c) Mova a fita duas casas para a esquerda do marcador, isto é, puxe a fita duas casas para a direita.
- d) Apague o símbolo I.

Quando efetuamos a última operação, a fita estará assim:



|

fig. 3

marcador

e isto corresponde à representação do número 5, isto é, à soma desejada. É desta maneira que a máquina de Turing procede para efetuar uma soma. Que novidade há nisto? Aparentemente nenhuma. Ao contrário, parece que conseguimos complicar um processo simples, cotidiano.

Contudo, as coisas não são bem assim. O que Turing inovou com a invenção de sua máquina foi à descoberta de uma espécie de princípio geral para a construção de computadores.

Este princípio geral tem como ponto de partida a noção matemática de *procedimento efetivo*. As instruções que damos para a máquina têm de ser executadas passo a passo, formando uma sucessão. Cada vez que uma instrução é executada, a máquina passa de um estado para outro. A mudança de um estado para outro corresponde a uma mudança de configuração. Para se mudar de uma configuração para outra existem certas instruções (como, por exemplo, no nosso caso: mova a fita para a direita, apague um símbolo etc.) que

estabelecem exatamente aquilo que deve ser feito.

Quando existe esse tipo de receita que diz exatamente o que deve ser feito para se passar de um estado para outro num processo, temos um procedimento efetivo, ou seja, um conjunto finito de instruções não ambíguas que nos dizem o que fazer passo a passo, e que nos garantem a obtenção de um resultado no final.

Ora, a descoberta de Turing consiste no fato de ele ter demonstrado, através da invenção de sua máquina, que toda e qualquer tarefa que possa ser representada na forma de um procedimento efetivo pode ser mecanizada, ou seja, pode ser realizada por um computador. Com sua invenção ele demonstrou ademais, que todo e qualquer tipo de computador pode, em última análise, ser reduzido a uma máquina de Turing, pois, embora os computadores possam diferir entre si quanto à sua finalidade e até mesmo ao material de que são compostos, eles podem ser imitados por sua máquina. E isso sem dúvida torna a máquina de Turing um verdadeiro princípio universal.

A máquina de Turing que descrevemos há pouco e que faz adições pode ser vista como uma imitação das várias máquinas de calcular de que dispomos. Na fita podemos representar os números (cada unidade será um quadrado com um símbolo), e há procedimentos efetivos para realizar operações aritméticas, ou seja, as operações aritméticas podem ser descritas através de um conjunto preciso de instruções: puxe a fita para a direita, para a esquerda etc.

Podemos dizer que nesta máquina rudimentar a fita contém uma representação dos números, e que o conjunto de instruções corresponde ao *programa* da máquina, da mesma maneira que dizemos que os computadores têm um programa. Se temos uma máquina que efetua adições e subtrações (uma maneira de imaginar a subtração seria pelo processo de apagar uma certa quantidade de símbolos da fita), podemos dizer que em princípio, nossa máquina será capaz de efetuar qualquer operação aritmética. Pois afinal, o que são multiplicações senão repetições de somas, e divisões a repetição de subtração? Dividir 16 por 4 significa quantas vezes o número 4 pode ser subtraído de 16. Claro que no caso da nossa máquina com fita e marcador, o programa ficaria bastante complicado, e se os números fossem grandes, a fita teria de ser extraordinariamente longa.

Máquinas de calcular constituem um grande e primeiro passo para mecanizar parte

de nossas atividades mentais. Claro que máquinas de calcular já existiam antes da invenção da máquina de Turing. Mas o que torna a invenção de Turing realmente interessante é a possibilidade de mecanizar tarefas executadas pela nossa mente, desde que elas possam se representadas por símbolos e na forma de procedimentos efetivos.

Imaginemos agora que em vez de trabalharmos com uma máquina rudimentar, com fita e marcador, tenhamos uma máquina bem mais sofisticada, mas cujo princípio seja o mesmo de uma máquina de Turing. Uma das diferenças seria que em vez de termos uma fita onde os quadrados teriam vários símbolos, teríamos apenas dois símbolos básicos, 0 e 1. Representar números e instruções na fita desse tipo de máquina se torna muito mais complicado: é preciso usar uma série de artifícios quando se dispõe de apenas dois símbolos. Mas certamente há aqui uma vantagem: se os símbolos a serem utilizados são apenas 0 e 1, podemos traçar uma correspondência entre estes e um circuito elétrico, com uma série de interruptores do tipo daqueles que usamos para apagar ou acender a luz de uma sala.

Nesses interruptores só há dois estados possíveis: quando eles estão ligados, passa a corrente, acende-se a lâmpada. Quando estão desligados a situação é inversa: não passa a corrente, a lâmpada fica apagada. Tudo se passa como se pudéssemos imaginar que estes estados de cada interruptor correspondessem aos símbolos que estão nos quadrados da nossa fita de papel: 0 quando não passa corrente, e 1 quando a corrente passa.

Ora, é exatamente este princípio que nos permite chegar a algo como uma representação elétrica do pensamento. Nossa máquina mais sofisticada não terá fita, mas um complexo circuito com interruptores. Em vez de ser operada por um movimento da fita de um lado para outro que apaga ou imprime símbolos, ela terá uma forma mais sofisticada de transmitir as instruções desejadas. Poderemos também conceber um tipo de “marcador” mais sofisticado. Tudo isso pode ser feito hoje em dia ocupando-se em espaço cada vez menor: os progressos da eletrônica permitem a construção de circuitos cada vez mais complexos e mais minúsculos. Foram estes progressos que permitiram reduzir tanto o tamanho dos circuitos e das máquinas que hoje em dia uma calculadora é tão pequena quanto um relógio de pulso.

Mas o que dissemos até agora serve apenas para mostrar como uma pequena parte

de nossas atividades mentais – aquelas relacionadas com operações aritméticas ou matemáticas – pode ser mecanizada. Mas e quanto ao resto de nossos pensamentos? Nem todas as nossas atividades mentais são dirigidas para realizar operações com números, e é precisamente a possibilidade de se mecanizar este outro tipo de atividades que constitui a grande novidade introduzida pelos computadores modernos – que nada mais são do que complexas máquinas de Turing que operam apenas com os símbolos 0 e 1.

E como podemos representar outros tipos de pensamentos além de números usando apenas os símbolos 0 e 1? Para isto os pesquisadores da IA e aqueles que começaram a construir computadores mais sofisticados precisaram, inicialmente, usar um artifício. O ponto de partida de tudo é a idéia de que nossos pensamentos são expressos em linguagem – não apenas em linguagem falada, mas em linguagem escrita. Ora a linguagem escrita nada mais é do que um sistema de símbolos construído a partir dos elementos básicos que compõem nosso alfabeto. O que precisamos então é arranjar um meio de representar todas as letras do alfabeto em termos de 0 e 1.

Sabemos que a totalidade das letras do alfabeto que usamos mais os outros caracteres normalmente empregados por nós, tais como números, vírgulas, pontos, espaço entre palavras, sinais de adição, subtração, multiplicação etc. totalizam 256 caracteres. Se se acha este número muito grande, basta dar uma olhada no teclado de uma máquina de escrever elétrica para se ver que na realidade usamos muito mais caracteres para expressar informação do que as letras do alfabeto. Com estes caracteres podemos expressar praticamente todo e qualquer pensamento, contar história da Revolução Francesa, a história da filosofia, realizar operações matemáticas e até escrever um livro sobre IA.

Vamos agora supor que temos um baralho com 256 cartas e que em cada uma delas está impresso um dos 256 caracteres de que falamos. Alguém seleciona uma carta ao acaso e a entrega para mim, virada com a face impressa para baixo, de maneira que eu não possa ver o caractere que está impresso nela. Essa pessoa pede que eu diga qual é a caractere que está impresso na carta.

Uma maneira de adivinhar o caractere é pegar a lista com os 256 caracteres e ir perguntando, um por um, até eu saber qual é aquele que está na carta. Ora, este é um método muito trabalhoso e demorado, pois corremos o risco de ter que repetir a mesma

questão 256 vezes.

Mas certamente existe uma outra saída para este problema: em vez de fazer 256 perguntas, posso começar perguntando: está o caractere da carta na primeira metade da lista? Qualquer que seja a resposta, teremos dividido a lista em duas partes com 128 caracteres cada uma – e numa dessas metades o caractere terá de estar. Tendo isolado uma das metades da lista onde o caractere em questão se encontra, posso repetir a estratégia mais de uma vez e dividir os 128 caracteres em duas metades de 64. Repito a questão e isolo uma lista com 64 caracteres. O processo deve ser repetido sucessivamente até que tenhamos um lista com apenas 2 caracteres e o caractere a ser identificado será necessariamente um deles. É fácil ver que com esta estratégia reduzi o número de perguntas a apenas 8, pois partindo de 256 caracteres, para se chegar a apenas 2 é preciso dividir a lista inicial oito vezes: na primeira vez obtivemos duas sublistas de 128 caracteres, na segunda vez duas sublistas de 64 caracteres, até que na oitava vez restarão apenas dois caracteres.

Foi utilizando esta estratégia que os pesquisadores da IA tornaram possível a representação de letras de nosso alfabeto e de outros caracteres que usamos para expressar informação em termos de apenas dois símbolos, 0 e 1. Se quisermos que isto seja feito pelo próprio computador, temos de fornecer-lhe a lista com os 256 caracteres e instruções para identificar um caractere qualquer, por exemplo, a letra *a*, dividindo a lista em metades e metades de metades. Quando o caractere não se encontra na primeira metade da lista dos 256 isto é representado por um 0. Caso contrário, ele será representado por 1. Após a repetição do processo por oito vezes sucessivas, teremos uma seqüência de oito 0 e 1. A letra *a*, por exemplo, é representada neste processo por 0100001. Mas a seqüência de 0 e 1 que correspondem a uma letra precisam, por sua vez, ser combinadas com outras seqüências de 0 e 1 para formar palavras, o que nos dá uma seqüência ainda maior. A representação de uma sentença no computador pode construir uma seqüência extraordinariamente longa de 0 e 1.

A transformação de palavras e sentenças em seqüências de 0 e 1 significa que uma representação elétrica de pensamentos, em termos de circuitos com interruptores abertos ou fechados, tornou-se possível. Isto ficou ainda mais fascinante quando se passou a projetar

circuitos capazes de representar as várias maneiras através das quais podemos combinar sentenças – o que equivale a ter representações elétricas de algumas de nossas formas básicas de raciocinar.

Quando raciocinamos, o que passa pela nossa cabeça são proposições diversas. Uma proposição é um pensamento expresso nessa sentença. Por exemplo, quando dizemos “A tarde está bela”, isto é um pensamento expresso nessa sentença. Uma proposição tem sempre uma característica específica: ela só pode ser verdadeira ou falsa. No caso da proposição “A tarde está bela”, temos meios de decidir se ela é verdadeira ou falsa: Podemos olhar pela janela ou mesmo ir dar um passeio.

Raciocinar é formular proposições e encadeá-las. O modo de encadear proposições para formar raciocínios corretos tem sido objeto de estudo desde a Antiguidade, quando surgiu uma disciplina com esta finalidade, a lógica. A partir do século passado a lógica conheceu avanços substanciais. Um destes avanços constituiu em se estabelecer uma representação simbólica para as proposições. Passou-se a usar as letras A, B, C etc para designar proposições, e logo se fez um estudo das maneiras passíveis de encadeá-las.

Estudos preliminares apontaram para a existência de quatro maneiras básicas de combinar proposições: a conjunção, a disjunção, a implicação e a bi-implicação. Podemos também negar uma proposição qualquer, o que neste sistema de representação será feito colocando-se o sinal  $\sim$  na sua frente. A negação de A será representada por  $\sim A$  a conjunção de proposições significa ligá-las pela partícula *e*. A conjunção é simbolizada em lógica pelo sinal  $\wedge$ . Se tivermos duas proposições, A e B, a conjunção será representada assim:  $A \wedge B$ . Se a proposição A for “O sol está brilhando” e a proposição B for “A tarde está bela”,  $A \wedge B$  significará: “O sol está brilhando e a tarde está bela”. A disjunção significa ligar proposições pela partícula *ou*. Assim, “O carro está andando ou o carro está parado” é um exemplo de disjunção. A implicação é uma ligação do tipo: “Se está chovendo, então a terra está molhada”.

A lógica tratou não apenas do modo como podemos encadear proposições, como também estabeleceu regras para sabermos quando estas ligações resultariam em verdades ou em falsidades. Assim, suponhamos, por exemplo, que as proposições “O sol está brilhando” e “A tarde está bela” sejam falsas. Neste caso, a conjunção destas proposições

também será falsa. Se a primeira proposição for falsa, mas a segunda for verdadeira, a conjunção será falsa. Todas as possíveis combinações para todos os tipos de proposições são previstas pela lógica. As combinações são agrupadas de acordo com o tipo de ligação entre as proposições, e formam aquilo que chamamos de “tabelas de verdade”.

Foi usando este tipo de estratégia que os pesquisadores da IA encontraram um caminho para construir circuitos que imitam nossa maneira de raciocinar. Da mesma forma que uma proposição só pode ter dois valores de verdade possíveis, ou seja, ser verdadeira ou falsa, os interruptores de um circuito também só podem ter duas posições: ou estão abertos (a corrente não passa) ou estão fechados (a corrente passa). É com base nesta analogia que podemos conceber um processo de representação elétrica dos raciocínios humanos.

O processo se inicia quando o programador coloca no computador as proposições que comporão um determinado raciocínio. Em seguida ele precisará informar ao computador o que se deseja fazer, ou seja, se se quer ligar as proposições através de uma conjunção, de uma disjunção ou implicação. Em outras palavras, é preciso fornecer ao computador as instruções que deverão ser seguidas. A partir do momento em que as instruções são fornecidas, um determinado circuito será acionado. Dependendo do tipo de instrução que é fornecido, o circuito que é acionado terá interruptores que permanecerão abertos ou fechados com a passagem de corrente elétrica. A construção desses circuitos e o que os interruptores neles fazem quando passa corrente elétrica obedecerá ao esquema das tabelas de verdade.

Assim, por exemplo, se se convencionou que quando existe passagem de corrente estamos representando uma proposição verdadeira, um circuito encarregado de fazer a conjunção de duas proposições verdadeiras deve ser construído de modo que a passagem de corrente elétrica faça com que seus interruptores se fechem e façam a corrente passar, permitindo que uma lâmpada se acenda ou outro sinal se manifeste. Se uma das proposições for falsa, um interruptor permanecerá aberto, a corrente não passará pelo circuito e a lâmpada não se acenderá. Neste caso saberemos que o resultado da conjunção de proposições não é verdadeiro.

Este processo de mecanização dos raciocínios pode se tornar extremamente

complexo, sobretudo quando se tem várias proposições e várias maneiras de conectá-las entre si. Neste caso, várias instruções terão de ser fornecidas, e muitos circuitos com interruptores com posições e características diferentes serão acionados até que se tenha um resultado final.

Os computadores modernos contam ainda com uma série de recursos adicionais além dos circuitos de que falamos. Esses recursos incluem a memória, que permite à máquina estocar informações (proposições, expressões) e retirá-las desse registro sempre que for necessário. Nessas máquinas mais complexas, o fornecimento de instruções pode se tornar uma operação complicada, que deve ser feita em linguagens especiais ou linguagens de programação. Contudo, é preciso assinalar que mesmo máquinas mais sofisticadas abedecem ao mesmo princípio de que falamos aqui, ou seja, tudo é processado e reduzido a dois símbolos básicos, 0 e 1, a chamada linguagem de máquina, que indica fluxo de corrente ou a sua ausência.

Isto nos dá uma imagem da quantidade e da complexidade de operações que são executadas por um computador num intervalo de poucos segundos. Operações que levaríamos horas para executar podem ser executadas em poucos segundos, embora o modo como são efetuadas seja extremamente complexo. Hoje em dia, em muitas situações, querer competir com a velocidade de um computador seria tão ridículo como querer apostar corrida com um avião.

Mas alguém ainda poderia dizer: sem dúvida, este processo de mecanização do pensamento tem se revelado eficiente, mas será que é este o modo como a mente opera para podermos dizer que o computador é um autêntico modelo do funcionamento da mente humana? Se não se puder responder afirmativamente a esta pergunta, não tem sentido falarmos de IA. Em outras palavras, o que se está questionando é se este procedimento de mecanização pode ser identificado com nossos processos de pensamento, o que nos permite dizer que a afirmação de que máquina pode pensar tem sentido.

Alguns teóricos tentaram responder a esta questão apontando para uma analogia entre os circuitos elétricos de um computador e o modo de funcionamento de nosso cérebro. Em nosso cérebro existem bilhões de células, os neurônios, que estão ligados entre si por minúsculos fios nervosos. Sabemos que em nosso cérebro passam correntes elétricas,

apesar de estas terem uma voltagem baixíssima. Será que o modo como raciocinamos não é idêntico ao modo como o computador o faz? Será que ao raciocinarmos não estão passando correntes elétricas entre os neurônios, que funcionariam como minúsculos interruptores?

Levantar este tipo de questão significa, em outras palavras, perguntar qual é a natureza daquilo que normalmente chamamos de *pensar*. Indagar o que seja aquilo que chamamos de pensamento pode nos levar a formular outras questões ainda mais inquietantes: terá cabimento afirmar que quando um circuito de um computador é acionado e nos fornece o resultado de uma soma, essa máquina está efetivamente pensando? Até que ponto será legítimo afirmar que essas máquinas constituem uma autêntica mecanização dos processos mentais humanos? Não seriam os computadores simples imitações ou simples truques que permitem realizar certas operações – máquinas às quais não poderíamos atribuir todas as características e propriedades de um ser humano?

Turing parece ter percebido as dificuldades que este tipo de questão envolve quando escreveu os artigos em que descrevia os princípios de sua máquina. Nesses artigos ele sustentou que não há nenhuma razão para se supor que uma máquina não possa pensar. Suas respostas a esse tipo de questão foram por vezes sarcásticas. Ele dizia que mesmo que para se dizer que uma criatura pensa seja preciso que esta tenha alma, não há nenhum impedimento para que Deus ponha alma nos computadores e não apenas nos seres humanos. Sustentar que Deus não poderia fazer isto seria o mesmo que querer limitar seus poderes. Deus dá almas a quem ele quiser.

Mas alguém ainda poderia dizer que mesmo que aceitássemos essas afirmações de Turing – de que as máquinas pensam - poderíamos ainda colocar as seguintes questões: Como saberíamos se uma máquina pensa ou não? Não é o pensar um processo invisível? Que critérios devemos utilizar para podermos afirmar que uma máquina pensa ou não?

Turing respondeu a estas perguntas da seguinte maneira: suponhamos que construíssemos uma máquina e a disfarçássemos tão bem que ela tivesse a aparência de um ser humano. Suponhamos também que ao observar o comportamento desta máquina, constatássemos que ela faz tudo o que um ser humano faz – respondendo a perguntas, fazendo operações aritméticas etc. - e que não tivéssemos meios de distinguir entre o comportamento desta máquina e o de um ser humano qualquer. Neste caso, pergunta

Turing, haveria alguma razão para não supor que esta máquina pensa, da mesma maneira que nós, seres humanos, o fazemos? Afinal de contas, qual o critério que usamos habitualmente para supor que nossos semelhantes pensam, a não ser a observação de seu comportamento e suas reações cotidianas? E, mesmo que uma máquina tão perfeita ainda não tenha sido construída hoje em dia, o que impede que no futuro nossa tecnologia avance tanto que ela se torne possível? Não se dizia no século passado que o homem jamais seria capaz de ir à Lua?

As afirmações de Turing geraram muitas polêmicas na sua época, principalmente entre os filósofos. Mas mesmo aqueles que discordaram de Turing, por achar que o jovem matemático inglês estava querendo humilhar a espécie humana ao sugerir que o pensamento não é um privilégio dado a nós pelo Criador, reconheceram mais tarde que a IA traz importantes contribuições para a filosofia. Mesmo que a mente humana não possa ser retratada numa máquina, o surgimento dos computadores abre novas perspectivas para a compreensão de nossas próprias atividades mentais, e traz sugestões interessantes para reavaliar a natureza de problemas filosóficos tradicionais.

## **INTELIGÊNCIA ARTIFICIAL E FILOSOFIA**

Uma vez um célebre professor de filosofia australiano foi à Universidade de Londres fazer uma palestra. O tema da palestra era “A identidade mente-cérebro”. Depois de discorrer sobre o tema durante mais de uma hora, tentando defender sua teoria de que mente e cérebro são a mesma coisa, um estudante ergueu o braço e fez a seguinte pergunta: “Mas, professor, se mente e cérebro são a mesma coisa, e se sabemos que quase 50 por cento do cérebro é composto de água, então o senhor quer dizer que muitos de nossos pensamentos nada mais são do que água? E se assim for, por que a água que está no cérebro ‘pensa’ e a água que está no corpo ‘não pensa’?”

O célebre professor não perdeu a seriedade. Alisou a gravata, olhou para o teto, refletiu alguns segundos e respondeu: “E como você sabe que a água que está no corpo não pensa?”.

Alguém poderia dizer: mas que conversa de doidos! Mas nem tanto. Quando começamos a refletir sobre essa questão – as relações entre pensamento e cérebro - facilmente chegamos a paradoxos. E logo percebemos que sabemos muito pouco acerca da natureza do pensamento e do cérebro.

Uma coisa, porém, é certa: este tipo de questão parece atormentar muito as pessoas até hoje. Será que existe algo como uma alma ou espírito – algo imaterial e invisível de onde se originam nossos pensamentos - ou será que tudo é matéria? Quando falamos de algo como espírito, normalmente associamos esta noção a alguma idéia de imortalidade, de duração eterna. Por outro lado, quando falamos de matéria, parece que estamos lidando com alguma coisa visível, sólida, mas efêmera.

É este tipo de oposição que parece ter dividido as opiniões dos filósofos desde o século XVII. Nesta época, Descartes chegara à conclusão de que mente e corpo são coisas inteiramente distintas, ou seja, duas substâncias com características e propriedades diferentes. Esta doutrina foi chamada de *dualismo*.

Embora estejamos acostumados com a visão de que corpo e mente sejam coisas separadas e ainda sejamos educados de acordo com esta tradição, as coisas não são tão simples assim. Há muitas dúvidas acerca disso. Se a mente é imaterial e se pensamentos não são coisas, como podemos falar de uma localização dos pensamentos no espaço? Parece intuitivo supor que somente objetos têm uma localização no espaço, e que também não tem cabimento falar de abstrações como estando localizadas em algum lugar. Onde estariam os conceitos de linha reta, de triângulo e de círculo? Não há sentido em falar de entidades abstratas como se elas devessem estar aqui ou ali, atrás da cadeira ou em cima da mesa. Ora, se os pensamentos são imateriais, da mesma maneira que as entidades abstratas, como poderíamos supor que eles ocorrem na nossa cabeça?

Mas ainda, se pensamentos não são como as coisas materiais, isto é, são algo imaterial, como podem eles influir em algum tipo de processo ou estado corporal? Como pode algo imaterial, que não tem as propriedades de um objeto, ocasionar alguma alteração no domínio do mundo material, como por exemplo, um movimento do nosso corpo?

Se separarmos mente e corpo, supondo que pensamentos são imateriais, não podemos explicar as nossas ações como decorrência daquilo que se passa na nossa mente.

Em outras palavras, tenho que imaginar que meu pensamento de me levantar, andar até a janela e abrir a cortina não tem nenhum tipo de relação com o fato de meu corpo se levantar, andar até a janela e abrir a cortina. Ora, isto seria tornar inexplicável o porquê dos nossos comportamentos.

Pensar em levantar, andar até a janela e abrir a cortina, e o ato físico de levantar, andar até a janela e abrir a cortina ocorrem quase ao mesmo tempo, mas isso não passaria de uma mera coincidência.

A falta de uma resposta satisfatória para estes problemas levou alguns filósofos do século XVII e XVIII a propor outro tipo de solução para as dificuldades colocadas pelo dualismo, como conceber a existência de algo como uma harmonia preestabelecida no universo, de acordo com a qual fenômenos mentais e fenômenos físicos, embora tenham propriedades diferentes e pertençam a ordens distintas, caminham paralelamente no tempo e sempre se ajustam perfeitamente.

Tudo se passaria como se tivéssemos dois relógios: um que correspondesse à ordem de eventos físicos e outro à de eventos mentais. Damos corda nesses relógios exatamente ao mesmo tempo e quando eles estão marcando a mesma hora. Assim, embora não haja nenhuma ligação entre os dois relógios, saberemos que ambos sempre marcarão a mesma hora. Da mesma maneira os teóricos da harmonia preestabelecida concebiam as relações entre a mente e o corpo: uma correspondência que dispensaria qualquer tipo de ligação.

A idéia de uma harmonia preestabelecida seria uma resposta satisfatória para o problema colocado pelas relações mente e corpo, se ela não envolvesse a idéia de um Deus todo-poderoso que colocou as duas ordens, a do físico e a do mental, a se desenvolverem paralelamente, da mesma maneira que damos corda aos relógios exatamente ao mesmo tempo e quando eles estão marcando a mesma hora. Ocorre que a existência de um Deus nem sempre é uma hipótese inteiramente aceitável, sobretudo para aqueles que querem ter certeza de estarem fazendo ciência, ou uma filosofia “verdadeiramente científica”.

Mas o que dizer da doutrina oposta ao dualismo, ou seja, o materialismo, o qual estabelece uma identidade entre pensamento e matéria (cérebro)? A visão materialista não parece resolver as dificuldades colocadas pelo problema das relações entre mente e corpo. Dizer que o físico e o mental são a mesma coisa não ajuda muito, pois pouco sabemos

acerca da natureza da matéria.

Até hoje a física moderna ainda não encontrou uma teoria definitiva para explicar a natureza do mundo material. Esta sensação de indefinição se acentuou ainda mais após as descobertas feitas no início do século XX, que obrigaram os cientistas a rever todas as nossas concepções habituais acerca do que seja a matéria.

A teoria da relatividade e a teoria quântica propostas no século XX fizeram com que o conceito habitual de matéria, entendida como corpo sólido que se move num espaço vazio, fosse progressivamente sendo abandonado. Os trabalhos de Einstein mostraram que aquilo que chamamos de massa ou matéria nada mais é do que uma forma de energia.

Mais ainda, os estudos acerca da estrutura da matéria e do mundo subatômico que se iniciaram desde então revelaram que os átomos não são partículas sólidas e duras como se pensava. Ao contrário, neles tudo se passa como se houvesse uma imensa distância entre o núcleo e partículas extremamente pequenas, os elétrons, que se movem à sua volta. No mundo subatômico ocorre a transformação da matéria em energia e vice-versa, e aquilo que normalmente enxergamos como matéria ou corpo sólido nada mais é do que o resultado do movimento extremamente rápido das partículas em torno do núcleo. Tudo se passa como quando observamos uma hélice que gira tão rapidamente que temos a impressão de estar diante de um disco.

Ora, o filósofo materialista poderia até mesmo tirar partido destas descobertas e dizer que o pensamento nada mais é do que matéria transformada em energia. Assim, não haveria necessidade de se falar de uma separação entre mente e matéria. O dualismo deixaria de ter sentido. Mas o materialismo também não teria mais sentido, pois teríamos abandonado quase que inteiramente a nossa noção habitual de objeto físico entendido como algo sólido, visível e compacto.

Parece que o aprofundamento do exame da noção de matéria nos leva a um círculo vicioso: quanto mais analisamos a matéria, mais caminhamos em direção a abstrações. E para isto precisamos de algo distinto do mundo material, pois só uma mente pode formar as representações abstratas com as quais a teoria física trabalha. Chegamos a uma espécie de paradoxo do materialismo, pois ele estaria nos levando de volta para as mesmas dificuldades que encontramos na teoria dualista.

As dificuldades teóricas que enfrentamos quando adotamos uma destas posições filosóficas, seja a materialista ou a dualista, sugerem que uma solução para o problema das relações entre mente e corpo e da natureza dos estados mentais (pensamentos) talvez ainda esteja muito distante. É possível que uma solução definitiva nunca venha a ser encontrada e que tenhamos de nos contentar com algumas propostas de solução que visem pelo menos a tornar alguns aspectos do problema mais claros e compreensíveis. É neste sentido que o modelo computacional da mente proposto pela IA pode ajudar a estabelecer algumas analogias entre mentes e máquinas, que podem, por sua vez, esclarecer alguns aspectos do problema das relações entre mente e o corpo.

Se retomarmos a noção básica de uma máquina de Turing, veremos como a IA permite estabelecer uma nova perspectiva para o problema que viemos tratando até então. Vimos no capítulo anterior que uma máquina de Turing pode ser construída a partir de um equipamento relativamente simples, ou seja, uma fita de papel e um marcador. Para termos uma máquina de Turing é preciso estipular um conjunto de instruções que especifiquem quais as operações que serão realizadas pela máquina (no caso, mover a fita para a esquerda ou para a direita etc.)

Ora, a máquina de Turing incorpora um tipo de dualidade parecida com aquela que nos leva freqüentemente a opor mente e corpo: de um lado, um conjunto de regras abstratas (as instruções), e, de outro, a realização física dessas regras obtidas pelos diferentes estados da máquina. Foi precisamente esta dualidade presente na máquina de Turing que permitiu aos teóricos da IA propor uma perspectiva inovadora para o tradicional problema filosófico das relações mente e corpo: a idéia consiste em estabelecer uma analogia entre os estados mentais (pensamentos) e o *software* (conjunto de instruções da máquina ou programa do computador) de um lado, e entre estados cerebrais e o *hardware* ou os diferentes estados físicos pelos quais passa a máquina ao obedecer às instruções.

Trata-se, naturalmente, de se traçar uma *analogia* entre o modo como nossa mente se relaciona com o nosso cérebro e não de dizer que nós funcionamos de maneira idêntica a uma máquina de Turing. De qualquer forma, esta simples analogia já nos ajuda a esclarecer uma série de dificuldades conceituais. Em primeiro lugar, podemos conceber um paralelismo entre eventos mentais e eventos físicos ou cerebrais: as operações da máquina

de Turing estabelecem uma co-relação precisa entre instruções (*software* ou os estados mentais) e estados físicos da máquina (*hardware* ou estados cerebrais). E este tipo de co-relação ou harmonia preestabelecida não precisa postular a existência de um Deus para tornar-se possível.

O paralelismo entre eventos mentais/eventos cerebrais e o *software/hardware* de um computador sugeridos pelos teóricos da IA seria confirmado pelo fato de que programas escritos em linguagens computacionais diferentes podem ser rodados numa mesma máquina, com uma mesma e única configuração de *hardware*. De maneira inversa, um mesmo programa pode ser rodado em computadores diferentes, isto é, em *hardwares* diferentes. Isto significa que os programas têm uma grande autonomia em relação ao *hardware*, e que não teria sentido falarmos em reduzir um determinado programa ao *hardware* da máquina onde ele é rodado, embora tenhamos de falar de uma correlação entre programa e máquina. Uma coisa não seria possível sem a outra.

Ora, é precisamente este tipo de correlação – que embora estabeleça a existência de uma dependência entre duas partes (máquina e programa), ao mesmo tempo lhes assegura uma autonomia – que constitui um excelente modelo para conceber relações entre mente e corpo. Este modelo, contudo, não constitui uma solução definitiva para o problema.

Por exemplo, não podemos concluir, pela simples aplicação do modelo, algo acerca da própria natureza de nossos estados mentais, isto é, se eles são realmente compostos de uma substância diferente daquela que compõe a matéria física ou não. Não obtemos uma resposta definitiva para o problema da identidade mente-cérebro, mas talvez isso nem sequer seja necessário, se podemos pelo menos supor que existe algo que permite uma tradução ou uma ponte entre o físico e o mental. E essa ponte ou tradução seria dada pelo modelo da máquina de Turing.

A adoção deste tipo de analogia para tratar o problema das relações mente-corpo tem ainda uma consequência interessante: sabemos que o que caracteriza uma máquina de Turing não é o tipo de material ou de substância empregada na sua construção. A fita poderia ser de papel ou de material magnético, ou, em vez de termos uma fita e um marcador, poderíamos ter um sistema de relés ou de transistores e chips, como ocorre nos computadores modernos. Isto quer dizer que, em última análise, não é preciso ter um cérebro igual ao

nosso para que se tenha um sistema capaz de desenvolver algo parecido com estados mentais. A inteligência não depende da existência de algo como a matéria viva; ela é, antes de tudo, o resultado de maneiras específicas de organizar certas instruções ou preencher certas funções, o que pode ser feito seja por um cérebro, seja por uma máquina de Turing. E é neste sentido que a inteligência poderia ser recriada artificialmente. Eis aí o grande projeto – ou talvez o grande sonho – da Inteligência Artificial!

### **O ENIGMA DO SIGNIFICADO**

Uma nova maneira de conceber o problema das relações da mente com o corpo não foi a única novidade que a IA trouxe para a filosofia, mas também uma nova perspectiva para refletir sobre a natureza dos problemas da filosofia e da psicologia.

A construção de réplicas de algumas das atividades mentais humanas alterou profundamente nossas próprias concepções habituais acerca da natureza do conhecimento. Criou-se a possibilidade de construir e testar programas computacionais que simulem aspectos da percepção visual humana e aspectos de nossas atividades lingüísticas. Estes programas computacionais ainda não atingiram um grau de aperfeiçoamento que nos permita dizer que simulações perfeitas dessas atividades já tenham sido realizadas. Contudo, é através da tentativa de elaborá-los que passamos a compreender melhor o funcionamento e a natureza dos nossos próprios processos psicológicos envolvidos na percepção, na linguagem e em outras atividades mentais humanas.

O pressuposto desta nova perspectiva é que o estudo da possibilidade do conhecimento humano deve ser um estudo prático, experimental, e não apenas uma reflexão abstrata que antecederia o trabalho do cientista ou julgaria os resultados finais da ciência. É neste sentido que a IA tem gradualmente levado alguns filósofos a trocar a atividade puramente especulativa acerca dos fundamentos do conhecimento pelo estudo do modo como simulações de processos mentais possam ser construídos de maneira que resultados satisfatórios sejam obtidos.

Simulações funcionam como modelos, e é pela construção de modelos de nós

mesmos que algum dia poderemos vir a conhecer os processos mentais que nos permitem falar, perceber o mundo e efetuar raciocínios. Os próprios mecanismos com os quais produzimos conhecimento são as ferramentas que permitem desvendar sua natureza. A mente pode conhecer a si mesma sem que haja risco de cairmos em algum tipo de circularidade: eis o que propõem alguns filósofos da ciência contemporâneos e, juntamente com eles, os teóricos da IA. Contra a sentença que diz que o olho não pode enxergar a si mesmo, eles diriam: sim, isso era verdade. Até o dia em que foi inventado o espelho...

Mas nem tudo são flores. O entusiasmo com que foram recebidas as realizações iniciais da IA e as novas perspectivas de abordagem de problemas tradicionais da filosofia não foram partilhados por toda a comunidade filosófica. A idéia de se construir uma inteligência artificial, paralela à nossa, trouxe também muita indignação. A crítica, neste caso, tornou-se mais feroz do que nunca: afinal de contas, é entre os intelectuais que mais se valoriza o pensamento, e, assim, a possibilidade de que atividades mentais sejam executadas por uma máquina pode se tornar algo francamente indesejável.

Algumas críticas ao projeto da IA que apareceram nas décadas de 60 e 70 traduziram atitudes apenas emocionais. Chegou-se a afirmar que os computadores são máquinas ainda tão toscas e primitivas que usá-las para replicar a mente humana equivaleria a subir numa árvore empunhando uma vara de pescar e dizer que assim poderíamos tocar na lua. Outras críticas, entretanto, serviram para apontar para importantes falhas do projeto e tornaram-se valiosas contribuições de caráter construtivo. Elas evidenciaram a necessidade de um estudo mais profundo da natureza dos processos mentais e dos conceitos que usamos para descrevê-los.

A inquietação intelectual quanto às possibilidades de construção de máquinas pensantes teve início após um período de euforia provocado pelas realizações iniciais da IA. Após esse período, os pesquisadores começaram a perceber que algumas tarefas às quais eles se propunham encontravam grandes dificuldades para se realizar.

Isso aconteceu, sobretudo com as máquinas de tradução. A tentativa de construir uma máquina para traduzir vários idiomas esbarrava em dificuldades técnicas cuja superação parecia estar distante. Uma das dificuldades era o caráter ambíguo que assumem alguns termos da nossa linguagem, uma ambigüidade que nós, seres humanos, dissipamos

pela referência ao contexto no qual esses termos ocorrem.

Quando fazemos uma tradução de um texto do português para o inglês e encontramos, por exemplo, a palavra “linha”, sabemos pelo contexto se o termo se refere à linha de trem ou linha de costura. No primeiro caso a tradução correta será “rail”, e no segundo, “thread”. O termo, ou signo é o mesmo, mas o significado é completamente diferente. Se o significado não é apreendido corretamente, a tradução será errada. Mas como programar uma máquina de maneira a que ela venha a apreender algo parecido com um contexto? Para apreciar a verdadeira natureza desta dificuldade basta lembrar do princípio de funcionamento dos computadores: Como pode uma máquina, que representa palavras em termos de estados internos definidos como 0 e 1, apreender uma variação de significado motivada por fatores contextuais?

Estas dificuldades com que se defronta a construção de uma máquina para a tradução automática de idiomas parecem revelar a existência de um obstáculo muito grande para as pesquisas na área de processamento de linguagem feita em IA. A representação da linguagem humana feita pelos computadores é realizada em termos puramente sintáticos. Ou seja, um computador só é capaz de reconhecer se uma determinada sentença está bem construída do ponto de vista gramatical. Quanto ao significado de uma sentença, isto é, saber se ela tem sentido ou não, as máquinas de que dispomos até o momento ainda deixam muito a desejar.

Para se ter uma idéia mais precisa do tipo de problema com o qual a IA ainda se defronta, basta percebermos que a sentença “Idéias verdes dormem furiosamente” é perfeita do ponto de vista sintático, embora não tenha sentido. Para apreender o significado de sentenças da nossa linguagem, seria preciso um critério mais refinado do que o puramente sintático. A velha objeção de Descartes contra a possibilidade de construir um autômato que pudesse articular a linguagem da mesma maneira que nós, humanos, parece não ter sido superada.

Com o reconhecimento destas dificuldades, a euforia que predominava entre os pesquisadores da IA no pós-guerra começou a ser substituída por um conjunto de preocupações e dúvidas – muitas das quais foram levantadas por filósofos. Será que todos os nossos processos mentais podem ser mecanizados e reduzidos a procedimentos efetivos?

Seriam estes obstáculos reveladores de que todo o projeto da IA repousa sobre pés de barro?

Estas dúvidas levaram ainda os filósofos a questionar outros aspectos da IA. Por exemplo, a famosa proposição de Turing que diz que podemos considerar uma máquina como ser pensante na medida em que ela produz comportamentos indistinguíveis dos de um ser humano foi colocada em xeque. Será que o simples exame do comportamento externo, observável, de um organismo ou de uma máquina é suficiente para que possamos concluir que esse organismo ou máquina é capaz de ter pensamentos? Um CD player toca Bach e Beethoven da mesma maneira que um músico o faz. Mas nunca diríamos de um CD player que ele o faz intencionalmente. Nem tampouco aplaudiríamos um CD player no final de uma execução.

As inquietações não pararam por aí. No final da década de 70 o filósofo norte-americano John Searle fazia uma viagem de avião para uma cidade da Califórnia, na costa oeste dos Estados Unidos, quando teve uma intuição acerca de um dos problemas mais importantes que os teóricos da IA tinham deixado passar despercebido.

Searle preparava, a bordo do avião, uma palestra que seria apresentada num simpósio de IA. Ele estava muito impressionado com uma série de programas computacionais que estavam sendo desenvolvido em algumas universidades americanas. Esses programas tinham sido projetados com uma finalidade específica: compreender estórias.

Assim, por exemplo, se se fornece a um computador com esses programas o seguinte relato: um homem entra num restaurante, pede um sanduíche e sai sem pagar nem deixar gorjeta porque notou que o pão estava amanhecido – o programa era construído de tal maneira que o computador “respondia” coerentemente a questões elaboradas com base no texto da estória. Ou seja, tudo se passava como num exercício de interpretação de textos daqueles que costumam cair em exames vestibulares. A diferença era que a interpretação do texto era efetuada por um computador, convenientemente programado para fornecer respostas adequadas.

Tudo isso não teria nada demais se os autores deste tipo de programas computacionais não sustentassem que essas máquinas eram capazes de *compreender* as

estórias que lhes eram fornecidas, e que tais programas funcionam exatamente como seres humanos no que diz respeito à capacidade de compreender textos.

Searle não se conformou com esse tipo de informação e elaborou um argumento filosófico – o Chinese Room Argument (Argumento do Quarto Chinês) – para mostrar que a idéia de que tais programas simulavam a atividade humana de compreender estórias e textos era completamente equivocada. A idéia central do argumento é inverter a situação de simulação e imaginar a tarefa executada pelo computador sendo realizada por um ser humano.

Imaginemos então uma pessoa trancada num quarto que não tem portas nem janelas, apenas duas portinholas em paredes opostas. Esta pessoa fala apenas português, mas alguém lhe forneceu um texto em chinês e uma espécie de tabela com regras e truques (escritos em português) para que ela, a partir de sentenças escritas em chinês, gere novas sentenças em língua chinesa. De vez em quando, abre-se uma das portinholas do quarto e alguém fornece a essa pessoa um novo texto escrito em chinês.

O ocupante do quarto pega o texto inicial escrito em chinês, os novos textos que foram introduzidos e, usando as regras de transformação que estão na tabela, gera um terceiro texto em chinês. Como o processo é repetido regularmente, ele vai adquirindo uma habilidade muito grande no manejo das regras de transformação.

Ora, esta situação corresponde ao que ocorre no interior de um computador dotado de um programa para compreender estórias: o texto inicial, que está com a pessoa trancada no quarto, corresponde à estória que é fornecida ao computador. As novas sentenças que são geradas com base nas regras de transformação podem muito bem ser as respostas às perguntas que foram feitas com base no texto.

Ocorre que a pessoa que está no interior do quarto, manipulando a tabela com as regras de transformação, embora produza sentenças que são respostas adequadas às perguntas sobre o texto em chinês, *não compreende chinês*. A pessoa não compreende o texto inicial em chinês, nem tampouco as repostas que são geradas. Tudo se passa de maneira parecida ao incidente do macaco que penetrou numa fábrica de máquinas de escrever e, apertando teclas ao acaso, acabou produzindo o texto do **Hamlet** de Shakespeare. Diz-se que o macaco não tinha a menor idéia do texto que estava produzindo.

À diferença da tabela com as regras de transformação, o mesmo acontece na situação da pessoa trancada no quarto – uma situação imaginária que nada mais faz do que ilustrar, de forma mais didática, o que ocorre no interior dos computadores com seus programas.

Da mesma maneira que uma câmara de televisão *não vê nada*, mas apenas reproduz imagens às quais nós atribuímos interpretações, os programas elaborados para compreender histórias na verdade nada compreendem. Eles apenas manipulam símbolos - símbolos que não têm nenhum *significado* para a máquina. Trata-se de uma manipulação de símbolos inteiramente cega. Dizer que uma máquina compreende ou enxerga é, no entender de Searle, um grande equívoco. É o mesmo que dizer que um papagaio fala, quando ele na verdade apenas emite sons que são imitados após muitas repetições.

Mas o que faz com que nós, seres humanos – à diferença das máquinas – possamos compreender, enxergar e gerar significado para nossa linguagem, nossos pensamentos e nossas ações? Os filósofos chamaram a esta faculdade de *intencionalidade* – uma propriedade que caracteriza nossos estados mentais. A intencionalidade se manifesta na medida em que sabemos a que se referem nossos estados mentais. Quando falamos, não estamos apenas emitindo sons: sabemos do que estamos falando e que nossas palavras se referem a coisas que estão no mundo. Todos os nossos pensamentos – sejam expressos em palavras ou não – têm conteúdos que apontam para coisas ou situações do mundo. É impossível estar pensando sem estar pensando em alguma coisa. E quando estamos pensando, sabemos selecionar, entre nossos estados mentais, aqueles que apontam para objetos que estão à nossa volta e aqueles que são mais distantes, como, por exemplo, os conteúdos da nossa imaginação. De qualquer maneira, há sempre uma direcionalidade, algo como um apontar para fora de nós mesmos que faz com que nossos pensamentos adquiram significado ou sentido.

Dizer que o significado é um produto da intencionalidade não ajuda muito se não sabemos como e por que nossos pensamentos têm essa propriedade. Searle sustenta que a intencionalidade é o resultado das operações do nosso cérebro – operações químicas características dos seres vivos. Mas será que isso responde à nossa pergunta? Não estará ele tentando refazer o velho caminho do materialismo que procura identificar propriedades de coisas materiais com propriedades de estados mentais? E se nem a própria teoria proposta

por Searle explica definitivamente a natureza da intencionalidade, não estará ele apenas apresentando mais um quebra-cabeças construído a partir de conceitos obscuros?

Muito se falou dos argumentos de Searle, principalmente do Argumento do Quarto do Chinês. Filósofos favoráveis ao projeto da IA apresentaram vários contra-argumentos às objeções de Searle. Alguns deles apontavam para falhas no argumento principal, salientando que não sabemos se de fato os computadores podem ou não compreender alguma coisa. A situação seria semelhante a quando observamos um ser humano responder a perguntas acerca de um texto qualquer: como podemos estar certos de que essa pessoa compreende o que está fazendo? Por acaso muitos de nossos processos mentais cotidianos não são tão rotineiros que os fazemos por associação tão mecânica e cega como as do computador?

E como podemos saber se alguém está realmente compreendendo o que faz? Pelas suas declarações? “Sim, eu compreendo o que faço quando respondo a uma pergunta sobre o texto”. E uma máquina não poderia ser programada para fornecer esta declaração? O que nos resta é a observação do comportamento seja este o de um ser humano ou de uma máquina. A vida interior de um outro ser humano é algo a que temos um acesso muito limitado. Aliás, nossa própria vida interior é algo sobre o que pouco sabemos. Mesmo quando fechamos os olhos e tentamos examinar o fluxo de nossos pensamentos, isto não nos dá nenhuma informação acerca de como ocorrem as operações do nosso cérebro. Somos, em grande parte, opacos para nós mesmos, e quase que só temos acesso ao resultado de nossas operações mentais. Ora, será que não estamos na mesma situação de alguém que olha para os resultados das operações de um computador e, com base neles, quer sustentar a afirmação de que essa máquina nada compreende acerca das próprias operações que realiza?

Poderíamos dizer muito mais coisas ainda no sentido de refutar o Argumento do Quarto do Chinês. Contudo, mesmo com novos contra-argumentos dificilmente abriríamos mão da crença de que nossa mente tem características que ainda não foram reproduzidas por sistemas artificiais. Parece que sempre teremos a sensação de que temos algo a mais do que as máquinas. Esse tipo de sensação é que levou alguns filósofos a supor, como Searle, que a intencionalidade dos nossos pensamentos é algo de origem biológica e que a

inteligência seria uma característica exclusiva dos seres vivos.

Esses teóricos partem da suposição de que o significado dos estados mentais é algo que se forma através da interação dos organismos com seu meio ambiente. Uma interação que tem atrás de si toda uma história, ou melhor, toda uma trajetória de evolução das espécies que resultou na formação daquilo que chamamos de “mente” – ou um órgão com características muito especiais. Se esta suposição for correta e algum dia comprovada, os teóricos da IA terão de rever seu programa ou até abandoná-lo. Isto porque características de um ser vivo podem ser imitadas. Mas a vida, a história e a evolução, estas talvez nunca poderiam ser recriadas num laboratório.

## CONCLUSÃO

As críticas à IA de que falamos no capítulo anterior não significam que o projeto de construção de máquinas pensantes esteja fadado ao fracasso. Elas apontam, contudo, para a existência de um grande conjunto de obstáculos que precisam ainda ser superados para que tenhamos algo parecido com a simulação de todas as atividades mentais humanas.

Neste sentido a IA coloca um verdadeiro desafio à espécie humana: será a inteligência um fenômeno único e original no universo ou será que poderemos replicar a nós mesmos? Mas não significa isto equiparar nossa posição no universo à de um Criador?

Nos últimos anos tem se falado com um entusiasmo sempre crescente nas possibilidades abertas pela IA. Fala-se, inclusive, de computadores que, no futuro, serão capazes de reproduzir nossas próprias emoções. Mas, muitas vezes, o discurso entusiasta dos teóricos da IA parece servir apenas para ocultar uma enorme distância entre realidade e desejo.

Este mesmo discurso tem servido, entretanto, de alento para que as pesquisas nessa área continuem. E novos horizontes parecem estar surgindo continuamente, sobretudo agora, quando se passou a procurar novas arquiteturas internas para os computadores, quando novos modelos da mente humana, de inspiração biológica e neurofisiológica, têm

sido propostos, o que nos leva a crer que dentro em breve ocorrerá uma verdadeira revolução teórica na IA. Uma revolução profunda, que poderá trazer modificações e alternativas à própria invenção de Turing, que até agora tem sido o ponto de partida de todos os trabalhos e pesquisas feitos nesse campo.

Contudo, todo o esforço feito pelos pesquisadores da IA não teria sentido se com esse projeto só se objetivasse produzir criaturas semelhantes a nós. O verdadeiro espírito da pesquisa em IA consiste em usar seus métodos para ampliar o conhecimento que temos acerca de nossa própria mente. Pois para produzir criaturas idênticas ou semelhantes a nós existem métodos muito mais fáceis e prazerosos...

Apostar no sucesso ou no fracasso do projeto de IA nas próximas décadas é algo que fica por conta do grau de otimismo que queiramos ter. Aliás, para sermos criaturas realmente inteligentes, como pensamos que somos, é preciso que sejamos otimistas. Para que representarmos um mundo sombrio e pessimista? Uma vez, um filósofo alemão exprimiu seu otimismo na sentença: “Este é o melhor dos mundos possíveis”. E não é mesmo? Pois não é fantástico que usando apenas o pensamento e a linguagem, possamos nos situar no labirinto que quisermos? Que outro mundo poderia dar a nós a opção de escolher em que mundo queremos viver?

## SOBRE O AUTOR

**João de Fernandes Teixeira** bacharelou-se em Filosofia pela USP-SP e obteve o grau de mestre em lógica e filosofia da ciência na UNICAMP. Doutorou-se em filosofia da mente e ciência cognitiva na University of Essex, Inglaterra. Desde 1992 é professor no Departamento de Filosofia da Universidade Federal de São Carlos. Em 1995 e 1998 foi “visiting scholar” no Centro de Estudos Cognitivos da Tufts University, em Boston, a convite do Prof. Daniel Dennett. É bolsista de produtividade em pesquisa do CNPq. Publicou **O que é Inteligência Artificial** (Brasiliense, 1990), **O que é Filosofia da Mente** (Brasiliense, 1994), **Filosofia da Mente e Inteligência Artificial** (Edições CLE-UNICAMP, 1996), **Cérebros, Máquinas e Consciência** (EDUFSCar, 1996), **Mentes e Máquinas: uma introdução à Ciência Cognitiva** (Artes Médicas, 1998, finalista do Prêmio Jabuti) **Mente, Cérebro e Cognição** (Vozes, 2.000, segunda edição 2003) e **Filosofia e Ciência Cognitiva** (Vozes, 2004), além de **Filosofia da Mente: Neurociência, Cognição e Comportamento** (Editora Claraluz, 2005). Colabora com o curso de pós-graduação em Tecnologia da Inteligência e Design Digital da PUC-SP e no momento prepara um livro sobre a filosofia da mente de Daniel Dennett.

