

Inteligência artificial: raízes e escopo

Parafraseando Sancho Pança, tudo deve ter um início; e esse inicio deve estar ligado a algo que já existiu antes. Para os hindus, o mundo é sustentado por um elefante, mas o elefante se encontra apoiado em cima de uma tartaruga. Deve-se humildemente admitir que a invenção não consiste em se criar a partir do nada, mas sim a partir do caos; em primeiro lugar, deve-se dispor dos materiais necessários...

—MARY SHELLEY, *Frankenstein*

Inteligência artificial: uma tentativa de definição

A inteligência artificial (IA) pode ser definida como o ramo da ciência da computação que se ocupa da automação do comportamento inteligente. Essa definição é particularmente apropriada a este livro porque enfatiza nossa convicção de que a IA faz parte da ciência da computação e que, desse modo, deve ser baseada em princípios teóricos e aplicados sólidos nesse campo. Esses princípios incluem as estruturas de dados usadas na representação do conhecimento, os algoritmos necessários para aplicar esse conhecimento e as linguagens e técnicas de programação usadas em sua implementação.

Entretanto, essa definição sofre com o fato de que a própria inteligência não é muito bem definida ou compreendida. Embora a maioria das pessoas esteja certa de que reconhece o comportamento inteligente quando o vê, não é certo que alguém possa chegar perto de definir a inteligência de um modo que seria específico o suficiente para ajudar na avaliação de um programa de computador supostamente inteligente, enquanto ainda captura a vitalidade e a complexidade da mente humana.

Como resultado da assombrosa tarefa de criar uma inteligência genérica, os pesquisadores da IA normalmente assumem o papel dos engenheiros ao moldarem artefatos inteligentes específicos. Estes geralmente vêm na forma de ferramentas de diagnóstico, prognóstico ou visualização, que permitem que seus usuários humanos realizem tarefas complexas. Alguns exemplos dessas ferramentas incluem os modelos ocultos de Markov para a compreensão da linguagem natural, sistemas de raciocínio automatizado para provar novos teoremas na matemática, redes Bayesianas dinâmicas para rastrear sinais através de redes corticais e visualização de padrões de dados de expressão de gene, conforme vistos nas aplicações da Seção 1.2.

O problema de *definir* o campo inteiro da inteligência artificial é semelhante ao de definir a própria inteligência: ela é uma única faculdade ou é apenas um nome para uma coleção de capacidades distintas e não relacionadas? Até que ponto a inteligência é aprendida e não existe desde o nascimento? O que acontece exatamente quando ocorre o aprendizado? O que é criatividade? O que é intuição? A inteligência pode ser deduzida do comportamento observável ou ela requer evidências de um mecanismo interno em particular? Como o conhecimento é representado

no tecido nervoso de um ser humano e que lições isso nos traz para o projeto de máquinas inteligentes? O que é autopercepção? Que papel ela desempenha na inteligência? Além disso, o conhecimento sobre a inteligência humana é necessário para construir um programa inteligente, ou uma técnica estritamente de “engenharia” é suficiente para tratar do problema? É possível conseguir inteligência em um computador, ou uma entidade inteligente requer a riqueza de sensações e experiências que só poderiam ser encontradas em uma existência biológica?

Essas são perguntas não respondidas que têm ajudado a modelar os problemas e as metodologias de solução que constituem o núcleo da IA moderna. De fato, parte da atração da inteligência artificial é que ela oferece uma ferramenta única e poderosa para explorar exatamente essas perguntas. A IA oferece um meio e um banco de ensaio para teorias da inteligência: essas teorias podem ser indicadas na linguagem de programação de computadores e podem, consequentemente, ser testadas e verificadas pela execução desses programas em um computador real.

Por esses motivos, nossa definição inicial de inteligência artificial é um pouco ambígua. No máximo, ela só tem levado a mais perguntas e à noção paradoxal de um campo de estudos cujas principais metas incluem sua própria definição. No entanto, essa dificuldade em chegar a uma definição exata de IA é totalmente compreensível. A inteligência artificial ainda é uma disciplina jovem, e sua estrutura, suas considerações e seus métodos não estão definidos tão claramente quanto aqueles de uma ciência mais madura, como a física.

A inteligência artificial sempre esteve mais preocupada com a expansão das capacidades da ciência da computação do que com a definição de seus limites. Manter essa exploração baseada em princípios teóricos sólidos é um dos desafios que os pesquisadores de IA, em geral, e este livro, em particular, enfrentam.

Em virtude do seu escopo e da sua ambição, a inteligência artificial não tem uma definição simples. Até o momento, simplesmente a definimos como *a coleção de problemas e metodologias estudada pelos pesquisadores de inteligência artificial*. Essa definição pode parecer tola e sem sentido, mas ela reforça um argumento importante: a inteligência artificial, como toda ciência, é um empreendimento humano e talvez seja mais bem entendida nesse contexto.

Há várias razões para que qualquer ciência, incluindo a IA, se preocupe com um certo conjunto de problemas e desenvolva um conjunto de técnicas específicas para enfrentar esses problemas. No Capítulo 1, uma breve história da inteligência artificial e das pessoas e hipóteses que a têm modelado explicará por que certos grupos de perguntas chegaram a dominar esse campo e por que os métodos discutidos neste livro foram escolhidos para a sua solução.

IA: história e aplicações

Por natureza, todos os homens desejam conhecer...

—ARISTÓTELES, frase inicial de *Metaphysics*

Ouvi o resto, e ainda mais admirareis o valor das artes e indústrias que dei aos mortais. Antes de mim — e este foi meu maior benefício —, quando atacados por qualquer enfermidade, nenhum socorro para eles havia, quer em alimento, quer em poções, bálsamos ou medicamentos: eles pereciam. Hoje, graças às salutares composições que lhes ensinei, todos os males são curáveis.

Fui eu quem tornou visíveis aos olhos dos homens os sinais flamejantes do céu que havia antes do escurecer. Chega deste assunto. Abaixo da terra, as bênçãos escondidas do homem, cobre, ferro, prata e ouro — haverá alguém para alegar que as descobriu antes de mim? Ninguém, com certeza, que esteja disposto a dizer a verdade sobre isso. Uma breve frase explicará a história: toda a arte que os mortais possuem vem de Prometeu.

—ÉSQUILO, *Prometheus Bound*

1.1 Do Éden ao ENIAC: posicionamentos em relação à inteligência, ao conhecimento e à astúcia humana

Prometeu fala dos frutos de sua transgressão contra os deuses do Olimpo: sua finalidade não foi simplesmente roubar o fogo para a raça humana, mas também iluminar a humanidade através do dom da inteligência ou intelecto: a *mente racional*. Essa inteligência forma a base para toda a tecnologia humana e, em última instância, toda a civilização humana. O trabalho de Ésquilo, o dramaturgo grego clássico, ilustra uma consciência profunda e antiga do extraordinário poder do conhecimento. A inteligência artificial, em seu interesse muito direto no dom de Prometeu, tem sido aplicada a todas as áreas de seu legado — medicina, psicologia, biologia, astronomia, geologia — e a muitas áreas de empreendimento científico que nem Ésquilo poderia ter imaginado.

Embora a ação de Prometeu livrasse a humanidade da doença da ignorância, ela também lhe rendeu a ira de Zeus. Indignado por esse roubo do conhecimento que anteriormente pertencia apenas aos deuses do Olimpo, Zeus ordenou que Prometeu fosse acorrentado a uma rocha bruta para sofrer as devastações dos elementos pela eternidade. A noção de que os esforços humanos para ganhar conhecimento constituem uma transgressão contra as leis de Deus ou da natureza está profundamente arraigada ao pensamento ocidental. Ela é a base da história do Éden e aparece nos livros de Dante e Milton. Tanto Shakespeare quanto as antigas tragédias gregas representaram a ambição intelectual como causa de desastre. A crença de que o desejo por conhecimento por fim deve levar ao

desastre persistiu por toda a história, perdurando pela Renascença, pela Era do Iluminismo e até mesmo pelos avanços científicos e filosóficos dos séculos XIX e XX. Assim, não devemos ficar surpresos com o fato de a inteligência artificial inspirar tanta controvérsia em círculos acadêmicos e populares.

Na realidade, em vez de dissipar esse antigo medo das consequências da ambição intelectual, a tecnologia moderna apenas fez essas consequências parecerem mais prováveis, até mesmo iminentes. As lendas de Prometeu, de Eva e de Fausto foram recontadas na linguagem da sociedade tecnológica. Na introdução de *Frankenstein*, com o interessante subtítulo de *O moderno Prometeu*, Mary Shelley escreve:

Muitas e longas foram as conversas entre Lord Byron e Shelley, às quais fui uma ouvinte devota e silenciosa. Durante uma delas, diversas doutrinas filosóficas foram discutidas, e entre outras, a natureza do princípio da vida, e se havia qualquer probabilidade de que sequer fosse descoberta e comunicada. Elas falavam dos experimentos do Dr. Darwin (não me refiro ao que o doutor realmente fez ou disse que fez, mas, conforme meu propósito, no que se falava que ele teria feito), que preservaram um pedaço de vidro até que, por algum meio extraordinário, ele começou a se mover voluntariamente. Afinal, não é dessa forma que a vida devia ser criada. Talvez um corpo fosse reanimado; o galvanismo deu sinal disso: talvez as partes componentes de uma criatura possam ser fabricadas, reunidas e dotadas de calor vital (Butler, 1998).

Mary Shelley nos mostra em que grau avanços científicos como a obra de Darwin e a descoberta da eletricidade têm convencido até mesmo leigos de que o funcionamento da natureza não era segredo divino, mas poderia ser desmembrado e compreendido sistematicamente. O monstro de Frankenstein não é produto de encantamentos xamanistas ou transações inaudíveis com o submundo: é montado de componentes separadamente “fabricados” e infundidos com a força vital da eletricidade. Embora a ciência do século XIX fosse inadequada para estabelecer o objetivo de compreender e criar um agente totalmente inteligente, ela afirmou a noção de que os mistérios da vida e do intelecto poderiam ser trazidos à luz da análise científica.

1.1.1 Uma breve história dos fundamentos da IA

Quando Mary Shelley finalmente, e talvez irrevogavelmente, uniu a ciência moderna ao mito de Prometeu, os fundamentos filosóficos do trabalho moderno em inteligência artificial já haviam sido desenvolvidos durante milhares de anos. Embora as questões morais e culturais levantadas pela inteligência artificial sejam interessantes e importantes, nossa introdução se preocupa mais com a herança intelectual da IA. O ponto de partida lógico para essa história é o gênio Aristóteles ou, como a *Divina Comédia* de Dante se refere a ele, “o mestre dos que sabem”. Aristóteles reuniu as descobertas, as maravilhas e os temores da antiga tradição grega com a análise cuidadosa e o pensamento disciplinado que se tornaram o padrão para a ciência mais moderna.

Para Aristóteles, o aspecto mais fascinante da natureza era a mudança. Em *Física*, ele definiu sua “filosofia da natureza” como o “estudo das coisas que mudam”. Ele distinguiu a *matéria* e a *forma* das coisas: uma escultura é modelada a partir do *material* bronze e tem a *forma* de um ser humano. A mudança ocorre quando o bronze é moldado em uma nova forma. A distinção matéria/forma fornece uma base filosófica para noções modernas como computação simbólica e abstração de dados. Na computação (até mesmo com números), manipulamos padrões que são as formas de material eletromagnético, com as mudanças de forma desses padrões representando aspectos do processo de solução. A abstração da forma a partir do meio de sua representação não apenas permite que essas formas sejam manipuladas computacionalmente, mas também oferece a promessa de uma teoria de estruturas de dados, núcleo da moderna ciência da computação. Ela também dá suporte à criação de uma inteligência “artificial”.

Em sua obra *Metafísica*, começando com as palavras “Todos os homens, por natureza, desejam conhecer”, Aristóteles desenvolveu uma ciência de coisas que nunca mudam, incluindo sua cosmologia e sua teologia. Mais relevante à inteligência artificial, porém, foi a epistemologia de Aristóteles, ou a análise de como os humanos “conhecem” seu mundo, discutida em sua obra *Lógica*. Aristóteles se referia à lógica como o “instrumento” (*organon*), porque percebeu que o estudo do próprio pensamento era a base de todo o conhecimento. Em *Lógica*, ele investiga se certas proposições podem ser “verdadeiras” porque estão relacionadas com outras coisas que são

sabidamente “verdadeiras”. Assim, se sabemos que “todos os homens são mortais” e que “Sócrates é um homem”, então podemos concluir que “Sócrates é mortal”. Esse argumento é um exemplo do que Aristóteles se referia como um silogismo, usando a forma dedutiva *modus ponens*. Embora a axiomatização formal do raciocínio precisasse de outros dois mil anos para o seu amadurecimento completo nos trabalhos de Gottlob Frege, Bertrand Russell, Kurt Gödel, Alan Turing, Alfred Tarski e outros, as suas raízes remontam a Aristóteles.

O pensamento do Renascimento, construído sobre a tradição grega, iniciou a evolução de um modo diferente e poderoso de pensar a respeito da humanidade e sua relação com o mundo natural. A ciência começou a substituir o misticismo como um meio de entender a natureza. Os relógios e, eventualmente, horários de fábricas, sobrepujaram os ritmos da natureza para milhares de moradores urbanos. A maioria das ciências sociais e físicas tem sua origem na noção de que processos, quer sejam eles naturais quer sejam artificiais, poderiam ser analisados e entendidos matematicamente. Em particular, os cientistas e filósofos perceberam que o próprio pensamento, o modo como o conhecimento era representado e manipulado na mente humana, era um tema difícil, porém essencial para o estudo científico.

Talvez o principal evento no desenvolvimento da visão do mundo moderno foi a revolução de Copérnico, a substituição do antigo modelo do universo centralizado na Terra com a ideia de que a Terra e outros planetas estão na realidade em órbitas em torno do Sol. Depois de séculos de uma ordem “óbvia”, em que a explicação científica da natureza do cosmos foi condizente com os ensinos da religião e do bom senso, um modelo drasticamente diferente e não tão óbvio foi proposto para explicar os movimentos dos corpos celestes. Talvez, pela primeira vez, *nossas ideias sobre o mundo foram vistas como fundamentalmente distintas de sua aparência*. Essa separação entre a mente humana e sua realidade ao redor, entre as ideias sobre as coisas e as próprias coisas, é essencial para o estudo moderno da mente e de sua organização. Essa brecha foi ampliada pelos escritos de Galileu, cujas observações científicas contradisseram ainda mais as verdades “óbvias” sobre o mundo natural e cujo desenvolvimento da matemática como uma ferramenta para descrever esse mundo enfatizaram a distinção entre o mundo e nossas ideias a respeito dele. É dessa brecha que surgiu a noção moderna de mente: a introspecção se tornou um motivo comum na literatura, os filósofos começaram a estudar a epistemologia e a matemática, e a aplicação sistemática do método científico competiu com os cinco sentidos como ferramenta para o conhecimento do mundo.

Em 1620, *Novum Organum*, de Francis Bacon, ofereceu um conjunto de técnicas de busca para essa metodologia científica emergente. Com base na ideia Aristotélica e Platônica de que a “forma” de uma entidade era equivalente à soma de suas “características” necessárias e suficientes, Bacon articulou um algoritmo para determinar a essência de uma entidade. Primeiro, ele criava uma coleção organizada de todas as ocorrências da entidade, enumerando as características de cada uma em uma tabela. Depois, ele coletava uma lista semelhante de ocorrências negativas da entidade, focalizando especialmente instâncias próximas da entidade, ou seja, aquelas que se desviavam da “forma” da entidade por características únicas. Depois, Bacon tenta — essa etapa não fica totalmente clara — criar uma lista sistemática de todas as características essenciais à entidade, ou seja, aquelas que são comuns a todas as ocorrências positivas da entidade e que faltam nas ocorrências negativas.

É interessante ver uma forma da abordagem de Francis Bacon para o aprendizado do conceito refletida nos algoritmos de IA modernos em “Busca em espaço de versões”, na Seção 10.2. Uma extensão dos algoritmos de Bacon também fazia parte de um programa de IA para o aprendizado por descoberta, adequadamente denominado *Bacon* (Langley et al., 1981). Esse programa foi capaz de induzir muitas leis físicas a partir de coleções de dados relacionados aos fenômenos. Também é interessante observar que a questão de se um algoritmo de uso geral era possível para produzir provas científicas aguardou os desafios do matemático do princípio do século XX Hilbert (sua obra *Entscheidungsproblem*) e a resposta do gênio moderno Alan Turing (sua *Máquina de Turing* e provas de *computabilidade* e do *problema da parada*); ver Davis et al. (1976).

Embora a primeira máquina de calcular, o ábaco, tenha sido criada pelos chineses no século XXVI a.C., a maior mecanização dos processos algébricos esperou as habilidades dos europeus do século XVII. Em 1614, o matemático escocês, John Napier, criou logaritmos, as transformações matemáticas que permitiram que a multiplicação e o uso de expoentes fossem reduzidos à adição e à multiplicação. Napier também criou o dispositivo chamado “ossos de Napier”, que foi usado para representar valores de estouro para operações aritméticas. O dispositivo foi mais tarde usado por Wilhelm Schickard (1592-1635), um matemático e sacerdote alemão de Tübingen, que em 1623 inventou um *Relógio Calculador* para realizar adição e subtração. Essa máquina registrava o estouro de seus cálculos pelo soar de um relógio.

Outra máquina de cálculo famosa foi a *Pascaline* que Blaise Pascal, filósofo e matemático francês, criou em 1642. Embora os mecanismos de Schickard e Pascal fossem limitados à adição e à subtração — incluindo estouros e empréstimos —, eles mostraram que processos que, segundo se imaginava, exigiam pensamento e habilidade humana poderiam ser totalmente automatizados. Conforme Pascal citou posteriormente em *Pensées* (Pensamentos) (1670), “A máquina aritmética produz efeitos que chegam mais perto do pensamento que todas as ações dos animais”.

Os sucessos de Pascal com máquinas de cálculo inspirou Gottfried Wilhelm von Leibniz, em 1694, a concluir uma máquina funcional que se tornou conhecida como a *Roda de Leibniz*. Ela integrava um tambor móvel e uma manivela para impulsionar as rodas e os cilindros que realizavam as operações mais complexas de multiplicação e divisão. Leibniz também era fascinado pela possibilidade de uma lógica automatizada para provas de proposições. Retornando ao algoritmo de especificação de entidade de Bacon, onde conceitos eram caracterizados como a coleção de suas características necessárias e suficientes, Leibniz imaginou uma máquina que poderia usar com essas características para produzir conclusões logicamente corretas. Leibniz (1887) também idealizou uma máquina que refletia ideias modernas de inferência dedutiva e prova, pela qual a produção de conhecimento científico poderia se tornar automatizada, um cálculo para raciocínio.

Os séculos XVII e XVIII também viram muita discussão de questões epistemológicas; talvez a mais influente tenha sido o trabalho de René Descartes, uma figura central no desenvolvimento dos conceitos modernos do pensamento e teorias da mente. Em sua obra *Meditações*, Descartes (1680) tentou achar uma base para a realidade puramente por meio da introspecção. Rejeitando sistematicamente a entrada de seus sentidos como não confiáveis, Descartes foi forçado a duvidar até mesmo da existência do mundo físico e foi deixado apenas com a realidade do pensamento; até mesmo sua própria existência teve de ser justificada em termos de pensamento: “*Cogito ergo sum*” (Penso, logo existo). Depois que estabeleceu sua própria existência puramente como uma entidade de pensamento, Descartes deduziu a existência de Deus como um criador essencial e por fim reafirmou a realidade do universo físico como a criação necessária de um Deus generoso.

Podemos fazer duas observações: primeiro, a divisão entre a mente e o mundo físico se tornou tão completa que o processo de pensamento podia ser discutido isoladamente de qualquer entrada sensorial ou questão material específica; segundo, a conexão entre a mente e o mundo físico era tão tênue que exigia a intervenção de um Deus generoso para dar suporte ao conhecimento confiável do mundo físico! Essa visão da dualidade entre a mente e o mundo físico forma a base de todo o pensamento de Descartes, incluindo seu desenvolvimento da geometria analítica. De que outra forma ele poderia ter unificado um ramo da matemática aparentemente tão mundano como a geometria com uma estrutura matemática tão abstrata como a álgebra?

Por que incluímos essa discussão entre mente/corpo em um livro sobre inteligência artificial? Há duas consequências dessa análise que são essenciais para um empreendimento da IA:

1. Ao tentar separar mente e mundo físico, Descartes e outros pensadores relacionados estabeleceram que a estrutura das ideias sobre o mundo não foi necessariamente a mesma que a estrutura de seu tema. Isso está por trás da metodologia da IA, juntamente com os campos da epistemologia, da psicologia, de grande parte da matemática mais avançada e da maior parte da literatura moderna: os processos mentais têm uma existência própria, obedecem às suas próprias leis e podem ser estudados por si próprios.
2. Uma vez que a mente e o corpo são separados, os filósofos acharam necessário encontrar um meio de reconectar os dois, pois a interação entre os planos mental, *res cogitans*, e físico, *res extensa*, de Descartes é essencial para a existência humana.

Embora milhões de palavras tenham sido escritas sobre esse *problema mente-corpo*, e diversas soluções propostas, ninguém explicou com sucesso as interações óbvias entre os estados mentais e as ações físicas, podendo afirmar uma diferença fundamental entre eles. A resposta mais bem aceita para esse problema, e aquela que oferece uma base essencial para o estudo da IA, afirma que a mente e o corpo não são, de modo nenhum, entidades fundamentalmente diferentes. Nessa visão, os processos mentais são realmente alcançados por sistemas físicos como os cérebros (ou computadores). Os processos mentais, como os processos físicos, por fim podem ser caracterizados por meio da matemática formal. Ou então, como reconhecido em *Leviatā*, do filósofo inglês do século XVII Thomas Hobbes (1651), “por raciocínio, quero dizer cálculo”.

1.1.2 IA e as tradições racionalista e empirista

Questões de pesquisa modernas em inteligência artificial, como em outras disciplinas científicas, são formadas e evoluem a partir de uma combinação de pressões históricas, sociais e culturais. Duas das pressões mais proeminentes para a evolução da IA são as tradições empirista e racionalista na filosofia.

A tradição racionalista, como vimos na seção anterior, teve um proponente antigo em Platão e foi continuada pelos escritos de Pascal, Descartes e Leibniz. Para o racionalista, o mundo exterior é reconstruído a partir de ideias claras e distintas da matemática. Uma crítica a esse enfoque dualista é o desligamento forçado dos sistemas representativos de seu campo de referência. A questão é se o significado atribuído a uma representação pode ser definido independentemente de suas condições de aplicação. Se o mundo é diferente de nossas crenças sobre ele, nossos conceitos e símbolos criados ainda podem ter significado?

Muitos programas de IA têm muita coisa desse aspecto racionalista. Os primeiros planejadores de robôs, por exemplo, descreviam seu domínio de aplicação ou “mundo” como conjuntos de declarações de cálculo de predicados, e depois um “plano” para ação era criado provendo teoremas sobre esse “mundo” (Fikes et al., 1972; ver também a Seção 8.4). O *Physical Symbol System Hypothesis*, de Newell e Simon (introdução à Parte II e ao Capítulo 16) é visto por muitos como o arquétipo desse enfoque na IA moderna. Duras críticas elegem esse viés racionalista como parte da falha da IA em resolver tarefas complexas, como o entendimento das linguagens humanas (Searle, 1980; Winograd e Flores, 1986; Brooks, 1991a).

Em vez de afirmar como “real” o mundo de ideias claras e distintas, os empiristas continuam a nos lembrar que “nada entra na mente senão por meio dos sentidos”. Essa restrição leva a mais questionamentos de como os seres humanos possivelmente podem perceber conceitos gerais ou as formas puras da caverna de Platão (Platão, 1961). Aristóteles foi um antigo empirista, enfatizando, em *De Anima*, as limitações do sistema perceptivo humano. Empiristas mais modernos, especialmente Hobbes, Locke e Hume, enfatizam que o conhecimento deve ser explicado por meio de uma psicologia introspectiva, porém empírica. Eles distinguem dois tipos de fenômenos mentais: a percepção, de um lado, e o pensamento, a memória e a imaginação, de outro. O filósofo escocês David Hume, por exemplo, distingue *impressões* de *ideias*. As impressões são dinâmicas e vívidas, refletindo a presença e a existência de um objeto externo e não sujeito ao controle voluntário, a *qualia* de Dennett (2005). As ideias, por outro lado, são menos vívidas e detalhadas e mais sujeitas ao controle voluntário da pessoa.

Dada essa distinção entre impressões e ideias, como surge o conhecimento? Para Hobbes, Locke e Hume, o mecanismo explanatório fundamental é a *associação*. Determinadas propriedades perceptivas são associadas por meio da experiência repetitiva. Essa associação repetitiva cria uma disposição na mente para associar as ideias correspondentes, um precursor do enfoque comportamentalista do século XX. Uma propriedade fundamental dessa explicação é apresentada com o ceticismo de Hume. A explicação puramente descritiva de Hume para as origens das ideias não pode, segundo ele, dar suporte à crença na causalidade. Até mesmo o uso da lógica e da indução não pode ser apoiado racionalmente nessa epistemologia empirista radical.

Em *An Inquiry Concerning Human Understanding* (Uma investigação sobre o entendimento humano) (1748), o ceticismo de Hume se estendeu para a análise dos milagres. Embora Hume não tratasse da natureza dos milagres diretamente, ele questionou a crença no miraculoso baseada em testemunhos. Esse ceticismo, é claro, foi visto como uma ameaça direta pelos crentes na Bíblia, bem como por muitos outros perpetuadores de tradições religiosas. O Reverendo Thomas Bayes era um matemático e um ministro religioso. Um de seus trabalhos, denominado *Essay towards Solving a Problem in the Doctrine of Chances* (Ensaio para a solução de um problema na doutrina das probabilidades) (1763), abordou os questionamentos de Hume matematicamente. O teorema de Bayes demonstra formalmente como podemos, por meio do aprendizado das correlações dos efeitos de ações, determinar a probabilidade de suas causas.

A explicação associativa do conhecimento desempenha um papel significativo no desenvolvimento das estruturas e dos programas representativos da IA, por exemplo, na organização de memória com *redes semânticas* e *MOPS* e no trabalho na compreensão de linguagem natural (ver seções 7.0, 7.1 e Capítulo 15). As explicações associativas possuem influências importantes do aprendizado de máquina, especialmente com redes conexionistas (ver seções 10.6, 10.7 e Capítulo 11). O associonismo também desempenha um papel importante na psicologia

cognitiva, incluindo os *esquemas* de Bartlett e Piaget, bem como na inteira confiança da tradição comportamentalista (Luger, 1994). Por fim, com ferramentas de IA para análise estocástica, incluindo a *rede Bayesiana de crença* (RBC; ou BBN, da sigla em inglês Bayesian belief network) e suas extensões atuais aos sistemas completos de Turing de primeira ordem para modelagem estocástica, as teorias associativas encontraram uma base matemática sólida e um poder de expressão maduro. As ferramentas Bayesianas são importantes para a pesquisa, incluindo diagnósticos, aprendizado de máquina e compreensão de linguagem natural (ver capítulos 5 e 13).

Immanuel Kant, filósofo alemão treinado na tradição racionalista, foi fortemente influenciado pelos escritos de Hume. Como resultado, ele começou a síntese moderna dessas duas tradições. Para Kant, o conhecimento contém duas energias colaborativas, um componente *a priori* vindo da razão do sujeito, juntamente com um componente *a posteriori* vindo da experiência ativa. A experiência só é significativa a partir da contribuição do sujeito. Sem uma forma de organização ativa proposta pelo sujeito, o mundo não seria nada mais que sensações transitórias passageiras. Por fim, no nível do discernimento, Kant afirma que imagens ou representações passageiras são ligadas pelo sujeito ativo e tomadas como as diversas aparições de uma identidade, de um “objeto”. O realismo de Kant iniciou o empreendimento moderno de psicólogos como Bartlett, Brunner e Piaget. O trabalho de Kant influencia a iniciativa moderna de IA do aprendizado de máquina (Seção IV), bem como o desenvolvimento contínuo de uma epistemologia construtivista (ver Capítulo 16).

1.1.3 Desenvolvimento da lógica formal

Uma vez que o pensamento começou a ser considerado como uma forma de cálculo, sua formalização e eventual mecanização foram as etapas seguintes óbvias. Como observamos na Seção 1.1.1, Gottfried Wilhelm von Leibniz, com *Calculus Philosophicus*, introduziu o primeiro sistema de lógica formal, além de propor uma máquina para automatizar suas tarefas (Leibniz, 1887). Além disso, as etapas e os estágios dessa solução mecânica podem ser representadas como movimento pelos estados de uma árvore ou grafo. Leonhard Euler, no século XVIII, com sua análise da “conectividade” das pontes unindo as margens de rios e ilhas da cidade de Königsberg (veja a introdução do Capítulo 3), introduziu o estudo das representações que podem capturar de forma abstrata a estrutura dos relacionamentos no mundo, bem como as etapas distintas dentro de um cálculo sobre esses relacionamentos (Euler, 1735).

A formalização da teoria dos grafos também possibilitou a *busca no espaço de estados*, uma ferramenta conceitual importante da inteligência artificial. Podemos usar grafos para modelar a estrutura mais profunda de um problema. Os nós de um *grafo de espaço de estados* representam estágios possíveis da solução de um problema; os arcos do grafo representam inferências, movimentos em um jogo ou outros passos na solução de um problema. Resolver o problema é um processo de buscar, no grafo de espaço de estados, um caminho para uma solução (introdução à Parte II e Capítulo 3). Por descreverem o espaço completo de soluções do problema, os grafos de espaço de estado são uma ferramenta poderosa para se medir a estrutura e a complexidade de problemas e analisar a eficiência, a correção e a generalidade de estratégias de solução.

Como um dos criadores da ciência da pesquisa operacional e projetista das primeiras máquinas computacionais mecânicas programáveis, Charles Babbage, matemático do século XIX, pode ser também considerado um dos pioneiros da inteligência artificial (Morrison e Morrison, 1961). O *motor diferencial* de Babbage era uma máquina de propósito específico para calcular os valores de certas funções polinomiais e foi o precursor de seu *motor analítico*. Esse motor, concebido, mas não construído com sucesso durante a sua vida, era uma máquina computacional programável de propósito geral que anteviu muitas das suposições arquitetônicas que formam a base do computador moderno.

Descrevendo o motor analítico, Ada Lovelace (1961), amiga, incentivadora e colaboradora de Babbage, disse:

Podemos afirmar muito apropriadamente que o motor analítico tece padrões algébricos do mesmo modo que o tear de Jacquard tece flores e folhas. Aqui, nos parece, reside muito mais originalidade do que se poderia atribuir ao motor diferencial.

A inspiração de Babbage era o desejo de aplicar a tecnologia de seu tempo para libertar o homem da árdua tarefa dos cálculos aritméticos. Nesse sentido, bem como dentro da sua concepção de computadores como dispositivos mecânicos, Babbage pensava puramente em termos do século XIX. O seu motor analítico, entretanto, também incluía muitas noções modernas, como a separação entre memória e processador, o *depósito* e o *moinho*, em termos de Babbage, o conceito de uma máquina digital em vez de uma máquina analógica e a noção de programabilidade baseada na execução de uma série de operações codificadas em cartões de cartolina. A característica mais impressionante na descrição de Ada Lovelace, e em toda a obra de Babbage, é o seu tratamento dos “padrões” de relacionamentos algébricos como entidades que podem ser estudadas, caracterizadas e, enfim, implementadas e manipuladas mecanicamente sem a preocupação com valores particulares que são, por fim, passados pelo “moinho” da máquina de calcular. Esse é um exemplo da implementação “da abstração e da manipulação da forma” que foi descrita pela primeira vez por Aristóteles e Liebniz.

O objetivo de se criar uma linguagem formal para o pensamento também aparece na obra de George Boole, outro matemático do século XIX, cujo trabalho deve ser incluído em qualquer discussão sobre as raízes da inteligência artificial (Boole, 1847, 1854). Embora ele tenha realizado contribuições em várias áreas da matemática, seu trabalho mais conhecido é aquele na formalização das leis da lógica, uma realização que representa o núcleo da ciência da computação moderna. Embora o papel da álgebra booleana no projeto de circuitos lógicos seja bem conhecido, os objetivos pessoais de Boole em desenvolver seu sistema parecem mais próximos dos objetivos dos pesquisadores contemporâneos da IA. No primeiro capítulo de *Uma Investigação das leis do pensamento, sobre as quais estão fundamentadas as teorias matemáticas da lógica e da probabilidade*, Boole (1854) descreveu os seus objetivos como:

investigar as leis fundamentais das operações mentais pelas quais se realiza o raciocínio: expressá-las na linguagem simbólica de um cálculo e, sobre essa base, estabelecer a ciência da lógica e instruir o seu método; ... e, por fim, coletar de vários elementos verdadeiros, revelados no decorrer destas consultas, algumas indicações prováveis relativas à natureza e à constituição da mente humana.

A importância da realização de Boole está no poder extraordinário e na simplicidade do sistema que ele concebeu: três operações, “E” (representada por * ou \wedge), “OU” (representada por + ou \vee) e “NÃO” (representada por \neg), formam o núcleo do seu cálculo lógico. Essas operações foram a base para todos os desenvolvimentos subsequentes da lógica formal, incluindo a concepção dos computadores modernos. Ao mesmo tempo que mantém o significado desses símbolos praticamente idêntico às operações algébricas correspondentes, Boole observou que “os símbolos da lógica estão adicionalmente sujeitos a uma lei especial, à qual os símbolos de quantidades, como tais, não estão sujeitos”. Essa lei diz que, para qualquer X , que é um elemento da álgebra, $X * X = X$ (ou seja, se algo é sabidamente verdadeiro, então a sua repetição não pode aumentar esse conhecimento). Essa noção levou à restrição característica dos valores booleanos a apenas dois números que podem satisfazer essa equação: 1 e 0. As definições padrão da multiplicação booleana (E) e da adição (OU) advêm desse critério.

O sistema de Boole não apenas forneceu a base da aritmética binária, mas também demonstrou que um sistema formal extremamente simples era adequado para captar todo o poder da lógica. Essa suposição e o sistema que Boole desenvolveu para demonstrá-la formam a base de todos os esforços modernos para formalizar a lógica, desde o *Principia Mathematica* (Princípios matemáticos) (Whitehead e Russell, 1950), passando pela obra de Turing e Gödel, até os sistemas modernos de raciocínio automatizado.

Gottlob Frege, em sua obra *Fundamentos da Aritmética* (Frege, 1879, 1884), criou uma linguagem de especificação matemática para descrever a base da aritmética de forma clara e precisa. Com essa linguagem, Frege formalizou muitas das questões que foram tratadas inicialmente pela Lógica de Aristóteles. A linguagem de Frege, agora conhecida como *cálculo de predicados de primeira ordem*, oferece uma ferramenta para descrever as proposições e atribuições de valores verdade que compõem os elementos do raciocínio matemático e descreve a base axiomática do “significado” para essas expressões. O sistema formal do cálculo de predicados, que inclui símbolos de predicados, uma teoria de funções e variáveis quantificadas, foi concebido para ser uma linguagem que descreva a matemática e suas bases filosóficas. Ele tem também um papel importante na criação de uma teoria de representações para a inteligência artificial (Capítulo 2). O cálculo de predicados de primeira ordem fornece as

ferramentas necessárias para o raciocínio automatizado: uma linguagem para expressões, uma teoria para suposições relacionadas com o significado das expressões e um cálculo logicamente correto para inferir novas expressões verdadeiras.

O trabalho de Whitehead e Russell (1950) é particularmente importante para a fundamentação da IA, uma vez que o seu objetivo declarado era derivar a matemática como um todo a partir de operações formais sobre uma coleção de axiomas. Embora muitos sistemas matemáticos tenham sido construídos a partir de axiomas básicos, o interessante aqui é o comprometimento de Russell e Whitehead com a matemática como um sistema puramente formal. Isso significa que os axiomas e teoremas seriam tratados apenas como sequências de caracteres: as provas se processariam apenas pela aplicação de regras bem definidas para manipular essas sequências. Não se poderia confiar na intuição ou no significado de teoremas como base para provas. Cada passo de uma prova é resultado da aplicação estrita de regras formais (sintáticas) a axiomas ou teoremas provados previamente, mesmo quando provas tradicionais possam considerar esse passo como “óbvio”. O “significado” que os teoremas e axiomas do sistema podem ter em relação ao mundo seria independente das suas derivações lógicas. Esse tratamento do raciocínio matemático em termos puramente formais (e consequentemente mecânicos) proporciona uma base essencial para sua automação em computadores físicos. A sintaxe lógica e as regras formais de inferência desenvolvidas por Russell e Whitehead fornecem ainda a base para sistemas de prova automática de teoremas, apresentados no Capítulo 14, bem como para os fundamentos teóricos da inteligência artificial.

Alfred Tarski é outro matemático cujo trabalho é essencial para o embasamento da IA. Tarski criou uma *teoria de referência*, por meio da qual se pode dizer que as *fórmulas bem formadas* de Frege ou Russell e Whitehead se referem, de um modo preciso, ao mundo físico (Tarski, 1944, 1956; ver Capítulo 2). Essa visão serve de embasamento para a maioria das teorias de semântica formal. No seu artigo “*A concepção semântica da verdade e os fundamentos da semântica*”, Tarski descreve sua teoria de referência e relacionamentos de valores verdade. Vários cientistas da computação moderna, especialmente Scott, Strachey, Burstall (Burstall e Darlington, 1977) e Plotkin, relacionaram essa teoria com linguagens de programação e outras especificações para computação.

Embora nos séculos XVIII, XIX e no início do XX a formulação da ciência e da matemática tenha criado o pré-requisito intelectual para o estudo da inteligência artificial, somente com a introdução do computador digital no século XX é que a IA se tornou uma disciplina científicamente viável. No final dos anos 1940, os computadores eletrônicos digitais demonstraram seu potencial para disponibilizar memória e poder de processamento requeridos pelos programas inteligentes. Era possível, agora, implementar sistemas de raciocínio formal em um computador e testar empiricamente a sua aptidão para exibir inteligência. Um componente essencial da ciência da inteligência artificial é esse compromisso com os computadores digitais como veículo para criar e testar teorias sobre a inteligência.

Os computadores digitais não são apenas um veículo para testar teorias sobre a inteligência. A sua arquitetura sugere também um paradigma específico para essas teorias: inteligência é uma forma de processamento de informação. A noção de busca como uma metodologia para solução de problemas, por exemplo, está mais ligada à natureza sequencial da operação do computador do que a qualquer modelo biológico de inteligência. A maioria dos programas de IA representa o conhecimento em alguma linguagem formal que é, então, manipulada por algoritmos, respeitando a separação entre dados e programa, que é fundamental ao estilo de computação de von Neumann. A lógica formal emergiu como uma ferramenta representacional importante para as pesquisas em IA, assim como a teoria dos grafos desempenha um papel essencial na análise de espaços de problemas e fornece a base para as redes semânticas e outros modelos similares de sentido semântico. Essas técnicas e formalismos são discutidos detalhadamente ao longo de todo o livro e são mencionados aqui para enfatizar a relação simbiótica entre o computador digital e os alicerces da inteligência artificial.

Esquecemos frequentemente que as ferramentas que criamos para os nossos propósitos tendem a moldar a nossa concepção do mundo a partir de sua estrutura e suas limitações. Embora aparentemente restritiva, essa interação é um aspecto essencial da evolução do conhecimento humano: uma ferramenta (e as teorias científicas são, no final das contas, apenas ferramentas) é desenvolvida para solucionar um problema em particular. Conforme essa ferramenta é usada e refinada, ela própria parece sugerir outras aplicações, levando a outras questões e, por fim, ao desenvolvimento de novas ferramentas.

1.1.4 Teste de Turing

Um dos primeiros artigos a tratar da questão da inteligência de máquina, especificamente em relação ao computador digital moderno, foi escrito em 1950 pelo matemático britânico Alan Turing. A obra *Maquinismo Computacional e Inteligência* (Turing, 1950) permanece atual tanto em relação à sua ponderação dos argumentos contra a possibilidade de se criar uma máquina computacional inteligente quanto por suas respostas a esses argumentos. Turing, que era conhecido principalmente por suas contribuições à teoria da computabilidade, abordou a questão se seria possível ou não fazer uma máquina pensar. Ao notar que as ambiguidades fundamentais contidas na própria questão (o que é pensar? o que é uma máquina?) obstruam qualquer resposta racional, ele propôs que a questão sobre a inteligência fosse substituída por um teste empírico mais claramente definido.

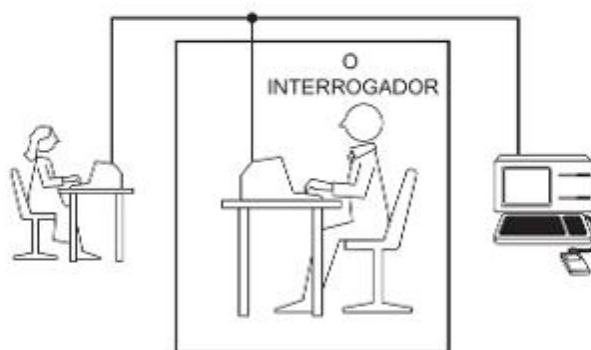
O teste de Turing mede o desempenho de uma máquina, aparentemente inteligente, em relação ao desempenho de um ser humano, indiscutivelmente o melhor e único padrão de comportamento inteligente. O teste, que foi chamado de *jogo de imitação* por Turing, coloca a máquina e seu correspondente humano em salas separadas de um segundo ser humano, referido como o *interrogador* (Figura 1.1). O interrogador não é capaz de ver nenhum dos dois participantes ou de falar diretamente com eles. Ele também não sabe qual entidade é a máquina e só pode se comunicar com eles por um dispositivo textual, como um terminal. A tarefa do interrogador é distinguir o computador do ser humano utilizando apenas as respostas de ambos a perguntas formuladas por meio desse dispositivo. Se o interrogador não puder distinguir a máquina do ser humano, então, argumenta Turing, pode-se supor que a máquina seja inteligente.

Isolando-se o interrogador tanto da máquina quanto do outro participante humano, o teste assegura que ele não será influenciado pela aparência da máquina ou por qualquer outra propriedade mecânica de sua voz. Entretanto, o interrogador é livre para fazer qualquer pergunta, não importando quão intrincada ou indireta ela seja, em um esforço para descobrir a identidade do computador. O interrogador pode, por exemplo, solicitar que os dois participantes realizem um cálculo aritmético bastante complicado, presumindo que seja mais provável que o computador, e não o ser humano, realize o cálculo corretamente; para neutralizar essa estratégia, o computador precisa saber quando ele deve deixar de obter a resposta correta a tais problemas, de modo a se parecer com um ser humano. Para descobrir a identidade humana a partir de sua natureza emocional, o interrogador poderia pedir para os dois participantes reagirem a um poema ou a uma obra de arte; essa estratégia requer que o computador tenha conhecimento acerca da constituição emocional dos seres humanos.

As características importantes do teste de Turing são:

1. Ele tenta nos dar uma noção objetiva de inteligência, isto é, o comportamento de um ser sabidamente inteligente em resposta a um conjunto particular de questões. Isso nos fornece um padrão para determinar a inteligência, evitando os debates inevitáveis sobre a sua “verdadeira” natureza.

Figura 1.1 Teste de Turing.



2. Ele evita que sejamos desviados por essas questões confusas e atualmente irresponsáveis, como, por exemplo, se um computador usa ou não os processos internos adequados ou se a máquina tem ou não consciência de suas ações.
3. Ele elimina qualquer viés em favor dos organismos vivos, forçando o interrogador a focar apenas no conteúdo das respostas às questões.

Por causa dessas vantagens, o teste de Turing fornece uma base para vários esquemas realmente utilizados para avaliar programas de IA modernos. Um programa que potencialmente tenha adquirido inteligência em alguma área de especialidade pode ser avaliado comparando seu desempenho a um especialista humano, diante de certo conjunto de problemas. Essa técnica de avaliação é apenas uma variação do teste de Turing; pede-se que um grupo de pessoas compare, às cegas, o desempenho de um computador e de um ser humano em relação a um conjunto de problemas. Como veremos, essa metodologia se tornou uma ferramenta essencial tanto no desenvolvimento quanto na verificação dos sistemas especialistas modernos.

O teste de Turing, apesar do seu apelo intuitivo, é vulnerável a várias críticas justificáveis. Uma das mais importantes se refere ao seu viés direcionado para as tarefas de solução de problemas puramente simbólicos. Ele não testa as habilidades necessárias para a percepção ou para a destreza manual, muito embora elas sejam componentes importantes da inteligência humana. Por outro lado, existem algumas críticas que sugerem que o teste de Turing restringe desnecessariamente a inteligência de máquina a se encaixar no molde humano. Talvez a inteligência de máquina seja simplesmente diferente da inteligência humana, e tentar avaliá-la em termos humanos seja um erro fundamental. Realmente desejamos uma máquina que realize operações matemáticas de modo tão lento e inexato como o faz um ser humano? Uma máquina inteligente não deveria tirar proveito dos seus próprios recursos, como o fato de dispor de uma memória grande, rápida e confiável, em vez de tentar emular a cognição humana? De fato, vários praticantes da IA moderna (por exemplo, Ford e Hayes, 1995) consideram que responder totalmente ao desafio do teste de Turing é um erro e um grande desvirtuamento da tarefa mais importante que se apresenta: o desenvolvimento de teorias gerais para explicar os mecanismos da inteligência nos seres humanos e em máquinas, com a aplicação dessas teorias no desenvolvimento de ferramentas para solucionar problemas práticos, específicos. Embora concordemos com as preocupações de Ford e Hayes, consideramos ainda o teste de Turing como um componente importante na verificação e na validação do software de IA moderno.

Turing também tratou da real possibilidade de se construir um programa inteligente em um computador digital. Pensando em termos de um modelo específico de computação (uma máquina de computação eletrônica de estados discretos), ele formulou algumas conjecturas bem fundamentadas com relação à capacidade de armazenamento, à complexidade do programa e à filosofia básica de projeto, necessárias para esse tipo de sistema. Por fim, ele discutiu uma série de objeções morais, filosóficas e científicas sobre a possibilidade de se construir um programa assim em termos de uma tecnologia real. O leitor deve consultar o artigo de Turing para ter um quadro resumido ainda atual e relevante do debate sobre as possibilidades das máquinas inteligentes.

Duas das objeções citadas por Turing merecem uma consideração mais detalhada. A *Objeção de Lady Lovelace*, formulada originalmente por Ada Lovelace, argumenta que os computadores podem fazer apenas aquilo que lhes foi anteriormente dito, em consequência não podendo realizar ações originais (e, portanto, inteligentes). Essa objeção se tornou uma parte reconfortante, ainda que um tanto dúbia, do folclore tecnológico contemporâneo. Os sistemas especialistas (Seção 1.2.3 e Capítulo 8), especialmente na área de diagnósticos, têm chegado a conclusões não previstas pelos seus projetistas. Na verdade, vários pesquisadores consideram que a criatividade humana pode ser expressa em um programa de computador.

A outra objeção relacionada, o *Argumento da informalidade do comportamento*, defende a impossibilidade de se criar um conjunto de regras que digam a um indivíduo exatamente o que fazer em todas as circunstâncias possíveis. Certamente, a flexibilidade que possibilita a uma inteligência biológica responder a um conjunto quase infinito de situações de uma forma razoável e, muitas vezes, ótima, é uma característica distintiva do comportamento inteligente. Apesar de ser verdadeira a afirmação de que a estrutura de controle utilizada nos programas de computador mais tradicionais não demonstra grande flexibilidade ou originalidade, não é verdade que todos os programas devam ser escritos dessa forma. Na realidade, grande parte dos esforços em IA nos últimos 25 anos tem se concentrado no desenvolvimento de linguagens de programação e de modelos, tais como sistemas de pro-

dução, sistemas baseados em objetos, representações por redes neurais e outros que são discutidos neste livro, que procuram superar essa deficiência.

Muitos programas de IA modernos consistem, geralmente, em uma coleção de componentes modulares, ou regras de comportamento, que não é executada segundo uma ordenação rígida, mas que é invocada conforme a necessidade em resposta à estrutura de um caso particular de um problema. Os comparadores de padrões permitem a aplicação de regras gerais sobre certo conjunto de casos. Esses sistemas têm uma flexibilidade extrema que possibilita que programas relativamente pequenos exibam um vasto domínio de comportamentos possíveis em resposta a problemas e situações diferentes.

Entretanto, se esses sistemas, no final das contas, podem ou não exibir a flexibilidade que um organismo vivo possui ainda é um tema de muito debate. O ganhador do Prêmio Nobel Herbert Simon argumentou que muito da originalidade e da variabilidade de comportamento demonstradas pelas criaturas vivas é devido à riqueza do seu ambiente e não à complexidade dos seus próprios programas internos. Em *As Ciências do Artificial*, Simon (1981) descreve a progressão tortuosa de uma formiga ao longo de uma extensão de solo irregular e atravancada. Embora a trilha da formiga pareça ser bastante complexa, Simon argumenta que o objetivo da formiga é muito simples: retornar à sua colônia o mais rápido possível. Os giros e voltas na sua trilha são causados pelos obstáculos que ela encontra no seu caminho. Simon conclui:

Uma formiga, vista como um sistema comportamental, é muito simples. A aparente complexidade do seu comportamento ao longo do tempo é, em grande parte, um reflexo da complexidade do ambiente no qual ela se encontra.

Se, no final das contas, for provado que essa ideia se aplica não só às criaturas simples como os insetos, mas também aos organismos com maior inteligência, então ela constitui um argumento poderoso de que, esses sistemas são relativamente simples e, consequentemente, compreensíveis. É interessante notar que, se alguém aplicar essa ideia aos seres humanos, isso se torna um forte argumento para a importância da cultura na formação da inteligência. Em vez de crescer no escuro, como os cogumelos, a inteligência aparentemente depende de uma interação com um ambiente adequadamente rico. Cultura é tão importante na geração de seres humanos quanto os seres humanos são na geração de cultura. Em vez de denegrir nossos intelectos, essa ideia enfatiza a riqueza milagrosa e a coerência das culturas que se formaram a partir das vidas de seres humanos separados. De fato, a ideia de que a inteligência emerge das interações de elementos individuais de uma sociedade é uma das intuições que dão suporte à abordagem para a tecnologia de IA apresentada na próxima seção.

1.1.5 Modelos biológicos e sociais da inteligência: teoria de agentes

Até agora, abordamos o problema da construção de máquinas inteligentes do ponto de vista da matemática, com a crença implícita de que o raciocínio lógico é o paradigma da inteligência em si, bem como com o compromisso com a fundamentação “objetiva” do raciocínio lógico. Essa forma de se considerar o conhecimento, a linguagem e o pensamento reflete a tradição racionalista da filosofia ocidental, a qual evolui através de Platão, Galileu, Descartes, Leibniz e muitos dos outros filósofos que discutimos anteriormente neste capítulo. Ela também reflete as suposições que fundamentam o teste de Turing, particularmente a sua ênfase no raciocínio simbólico como teste de inteligência, e a crença de que uma comparação direta com um ser humano é adequada para confirmar a inteligência de máquina.

A confiança na lógica como um meio de representar conhecimento e na inferência lógica como o mecanismo primordial para o raciocínio inteligente é tão dominante na filosofia ocidental que a sua “verdade” frequentemente parece óvia e incontestável. Portanto, não é de surpreender que abordagens baseadas nessas suposições têm dominado a ciência da inteligência artificial desde o seu início até os dias de hoje.

Entretanto, a última metade do século XX presenciou numerosos desafios à filosofia racionalista. Várias formas de relativismo filosófico questionaram a base objetiva da linguagem, da ciência, da sociedade e do próprio pensamento. A filosofia póstuma de Ludwig Wittgenstein (Wittgenstein, 1953) forçou-nos a reconsiderar a base do significado, tanto da linguagem natural quanto da formal. Os trabalhos de Gödel (Nagel e Newman, 1958) e de Turing puseram em dúvida os fundamentos da própria matemática. O pensamento pós-moderno mudou o

nosso entendimento sobre significado e valor nas artes e na sociedade. A inteligência artificial não ficou imune a essas críticas; de fato, as dificuldades que a IA encontrou para alcançar seus objetivos são frequentemente consideradas como evidência da deficiência do ponto de vista racionalista (Winograd e Flores, 1986; Lakoff e Johnson, 1999; Dennett, 2005).

Duas tradições filosóficas, a de Wittgenstein (1953), bem como a de Husserl (1970, 1972) e Heidegger (1962), são centrais a essa reconsideração da tradição filosófica ocidental. Na sua obra póstuma, Wittgenstein questionou muitas das suposições da tradição racionalista, incluindo os fundamentos da linguagem, da ciência e do próprio conhecimento. A linguagem natural foi um foco central da análise de Wittgenstein: ele desafiou a noção de que a linguagem humana tenha derivado o seu significado a partir de qualquer tipo de fundamentação objetiva.

Para Wittgenstein, bem como pela teoria dos atos da fala desenvolvida por Austin (1962) e seus seguidores (Grice, 1975; Searle, 1969), o significado de qualquer expressão vocal depende de onde ela se situa em um contexto humano e cultural. O nosso entendimento da palavra “cadeira”, por exemplo, depende da existência de um corpo físico que corresponda a uma postura sentada e às convenções culturais sobre o uso de cadeiras. Quando, por exemplo, uma pedra grande e plana seria uma cadeira? Por que é estranho se referir ao trono da Inglaterra como uma cadeira? Qual a diferença entre o entendimento humano sobre uma cadeira e o de um cachorro ou gato, que são incapazes de sentar, no sentido humano? Com base nos seus ataques aos fundamentos do significado, Wittgenstein argumentou que deveríamos considerar o uso da linguagem em termos de escolhas feitas e ações realizadas em um contexto cultural variável. Wittgenstein estendeu, inclusive, as suas críticas à ciência e à matemática, argumentando que elas eram construções sociais, assim como o uso da linguagem.

Husserl (1970, 1972), o pai da fenomenologia, estava comprometido com abstrações enraizadas no *mundo da vida* (*Lebenswelt*) concreto: um modelo racionalista é secundário em relação ao mundo concreto que o suporta. Para Husserl, bem como para seu estudante Heidegger (1962) e seu defensor Merleau-Ponty (1962), a inteligência não é saber o que é verdadeiro, mas saber como lidar com um mundo em constante modificação e evolução. Gadamer (1976) também contribuiu para essa tradição. Assim, pelas tradições existencialista e fenomenologista, a inteligência é vista como sobrevivência no mundo, em vez de um conjunto de proposições lógicas acerca do mundo (combinada com um esquema de inferência).

Muitos autores, por exemplo Dreyfus e Dreyfus (1985) e Winograd e Flores (1986), se inspiraram no trabalho de Wittgenstein e de Husserl/Heidegger para suas críticas à IA. Embora muitos praticantes da IA continuem desenvolvendo a agenda racional/lógica (também conhecida como GOFAI, do inglês *Good Old Fashioned AI* — “a boa e velha IA”), um número crescente de pesquisadores da área incorporaram essas críticas em novos e excitantes modelos de inteligência. Mantendo a ênfase de Wittgenstein sobre as raízes antropológicas e culturais do conhecimento, eles se voltaram, como inspiração, para modelos sociais de comportamento inteligente, algumas vezes chamados de *baseados em agente* ou *situados*.

Como exemplo de alternativa para uma abordagem baseada em lógica, os pesquisadores da área de aprendizado conexionalista (Seção 1.2.9 e Capítulo 11) retiraram a ênfase na lógica e no funcionamento da mente racional, em um esforço para alcançar a inteligência modelando a arquitetura do cérebro físico. Os modelos neurais da inteligência enfatizam a habilidade do cérebro em se adaptar ao mundo no qual ele está inserido pela modificação dos relacionamentos entre neurônios individuais. Em vez de representar o conhecimento por sentenças lógicas explícitas, eles capturam o conhecimento implicitamente, como uma propriedade de padrões de relacionamentos.

Outro modelo de inteligência baseado na biologia busca sua inspiração nos processos pelos quais espécies inteiras se adaptam ao seu ambiente. Os trabalhos em vida artificial e em algoritmos genéticos (Capítulo 12) aplicam os princípios da evolução biológica aos problemas de encontrar soluções para problemas difíceis. Esses programas não resolvem problemas raciocinando logicamente sobre eles; ao contrário, eles geram populações de soluções candidatas competidoras e as forçam a evoluir para soluções melhores por um processo que imita a evolução biológica: possíveis soluções ruins tendem a se extinguir, enquanto aquelas que se mostram mais promissoras em resolver um problema sobrevivem e se reproduzem, construindo novas soluções a partir de componentes dos seus pais bem-sucedidos.

Os sistemas sociais fornecem outra metáfora para a inteligência, uma vez que eles exibem comportamentos globais que permitem resolver problemas que confundiriam qualquer um de seus membros individuais. Por exemplo, embora nenhum indivíduo possa prever precisamente o número de países que serão consumidos na

cidade de Nova York em um determinado dia, o sistema de padarias de Nova York como um todo realiza muito bem a tarefa de manter a cidade abastecida com pães, com o mínimo de desperdício. O mercado de ações faz um excelente trabalho em fixar os valores relativos de centenas de empresas, muito embora cada investidor individual tenha somente um conhecimento limitado sobre poucas empresas. Um último exemplo vem da ciência moderna. Indivíduos estabelecidos em universidades, na indústria ou em instituições governamentais enfrentam problemas comuns. Problemas importantes à sociedade como um todo são atacados e solucionados por indivíduos agindo quase que independentemente, tendo conferências e revistas científicas como principais meios de comunicação, embora o progresso, em muitos casos, também seja conduzido por agências de fomento.

Esses exemplos compartilham dois temas: primeiro, a visão da inteligência como enraizada na cultura e na sociedade, e, como consequência, emergente. O segundo, que a inteligência é o reflexo dos comportamentos coletivos de um grande número de indivíduos semiautônomos muito simples que interagem entre si, ou *agentes*. Quer esses agentes sejam células neurais, quer sejam membros individuais de uma espécie, ou ainda uma única pessoa de uma sociedade, as suas interações produzem inteligência.

Quais são os grandes temas que justificam uma visão de inteligência emergente orientada a agentes? Entre eles se incluem:

1. Agentes são autônomos ou semiautônomos, isto é, cada agente tem certas responsabilidades na solução de problemas, com pouco ou nenhum conhecimento do que outros agentes fazem ou de como o fazem. Cada agente realiza a sua parte independentemente da solução do problema: produz um resultado (faz alguma coisa), ou relata seus resultados para outros indivíduos da comunidade (agente comunicante).
2. Os agentes são “situados”. Cada agente é sensível ao seu ambiente particular e (normalmente) não tem conhecimento do domínio completo de todos os agentes. Assim, o conhecimento de um agente é limitado às tarefas atuais: “o arquivo que estou processando” ou “a parede próxima a mim” sem qualquer conhecimento do conjunto total de arquivos ou das restrições físicas da tarefa de solucionar um problema.
3. Agentes são interativos. Isso significa que eles formam uma coleção de indivíduos que cooperam em uma tarefa particular. Nesse sentido, eles podem ser vistos como uma “sociedade” e, como na sociedade humana o conhecimento, as habilidades e as responsabilidades, mesmo quando vistas como coletivos, estão distribuídos entre os indivíduos de uma população.
4. A sociedade de agentes é estruturada. Na maioria das visões de solução de problemas orientada a agentes, cada indivíduo, embora tendo o seu ambiente e seu conjunto de habilidades particulares e únicas, precisa se coordenar com outros agentes para a solução global do problema. Assim, uma solução final não será vista apenas como coletiva, mas também como cooperativa.
5. Finalmente, o fenômeno da inteligência nesse ambiente é “emergente”. Embora agentes individuais sejam vistos como se possuissem conjuntos de habilidades e responsabilidades, o resultado cooperativo global da sociedade de agentes pode ser visto como maior que a soma de suas contribuições individuais. A inteligência é vista como um fenômeno que reside e emerge da sociedade, e não apenas uma propriedade de um agente individual.

Com base nessas observações, definimos um agente como um elemento de uma sociedade que pode perceber aspectos (frequentemente limitados) de seu ambiente e afetá-lo, quer diretamente, quer através da cooperação com outros agentes. A maioria das soluções inteligentes requer uma variedade de agentes. Aí se incluem agentes rotineiros, que simplesmente capturam e comunicam partes da informação; agentes de coordenação, que dão suporte às interações entre outros agentes; agentes de busca, que examinam conjuntos de informações e retornam pedaços importantes deles; agentes aprendizes, que examinam coleções de informações e formam conceitos ou generalizações, e agentes de decisão, que podem tanto disparar tarefas como chegar a conclusões com base em informação e processamento limitados. Retomando uma definição mais antiga de inteligência, agentes podem ser vistos como mecanismos que suportam a tomada de decisão no contexto de recursos computacionais limitados.

Os requisitos principais para projetar e construir uma sociedade de agentes são:

1. estruturas para a representação da informação;
2. estratégias para busca por meio de soluções alternativas;
3. criação de arquiteturas que possam suportar a interação de agentes.

Os capítulos restantes deste livro, especialmente a Seção 7.4, incluem prescrições para a construção de ferramentas de suporte para essa sociedade de agentes, além de muitos exemplos de solução de problemas baseada em agente.

A nossa discussão preliminar sobre a possibilidade de uma teoria da inteligência automática não teve a intenção de exagerar o progresso atingido na atualidade ou minimizar o trabalho que se encontra pela frente. Como enfatizamos ao longo de todo o texto, é importante que tenhamos consciência das nossas limitações e que sejamos honestos acerca de nossos sucessos. Houve, por exemplo, resultados apenas limitados com programas dos quais se pode dizer, em qualquer sentido interessante, que “aprendem”. Nossas realizações na modelagem das complexidades semânticas de uma linguagem natural, como o inglês, também têm sido modestas. Mesmo questões fundamentais, como a organização do conhecimento ou o completo gerenciamento da complexidade e da correção de programas de computadores muito grandes (como grandes bases de conhecimento), requerem pesquisas adicionais consideráveis. Os sistemas baseados em conhecimento, embora tenham atingido um sucesso comercial do ponto de vista da engenharia, ainda têm muitas limitações na qualidade e na generalidade do seu raciocínio. Aqui se incluem a sua incapacidade de realizar *raciocínio de senso comum* ou em exibir conhecimento de uma realidade física rudimentar, por exemplo, como as coisas mudam com o tempo.

Mas devemos manter uma perspectiva razoável. É fácil subestimar as realizações da inteligência artificial quando encaramos honestamente o trabalho que nos resta. Na próxima seção, estabelecemos essa perspectiva a partir de um panorama de várias áreas da pesquisa e do desenvolvimento em inteligência artificial.

1.2 Uma visão geral das áreas de aplicação da IA

O Motor Analítico não tem pretensões de criar algo original. Ele pode fazer qualquer coisa, desde que saibamos como lhe dar a ordem.

—ADA BYRON, Condessa de Lovelace

Desculpe-me, Dave; não posso deixar que faça isso.

—HAL 9000 em 2001: *Uma odisseia no espaço*, de Arthur C. Clarke

Retornamos agora ao nosso objetivo de definir inteligência artificial por meio de um exame das ambições e das realizações dos pesquisadores da área. As duas preocupações fundamentais dos pesquisadores em IA são a *representação de conhecimento* e a *busca*. A primeira trata do problema de capturar em uma linguagem formal, isto é, em uma linguagem adequada para ser manipulada em computador, toda a extensão de conhecimento necessário para um comportamento inteligente. O Capítulo 2 introduz o cálculo de predicados como uma linguagem para descrever as propriedades e os relacionamentos entre objetos em domínios de problemas que, para a sua solução, necessitam de raciocínio qualitativo, em vez de cálculos aritméticos. Depois, a Seção III discute as ferramentas que a inteligência artificial desenvolveu para representar as ambiguidades e as complexidades de áreas como raciocínio de senso comum e compreensão de linguagem natural.

A busca é uma técnica de solução de problemas que explora sistematicamente o espaço de *estados do problema*, isto é, os estágios sucessivos e alternativos no processo de solução do problema. Exemplos de estados de um problema incluem as diferentes configurações do tabuleiro em um jogo ou os passos intermediários em um processo de raciocínio. Esse espaço de soluções alternativas é, então, explorado para se encontrar uma resposta final. Newell e Simon (1976) propuseram que essa é a base essencial da solução de problemas por seres humanos. De fato, quando um jogador de xadrez examina os efeitos de diferentes movimentos ou um médico considera uma série de diagnósticos alternativos, eles estão realizando uma busca entre diferentes alternativas. As implicações

desse modelo e dessas técnicas para sua implementação são discutidas nos capítulos 3, 4, 6 e 16. A Sala Virtual deste livro oferece implementações desses algoritmos em Lisp, Prolog e Java.

Como a maioria das ciências, a IA é composta de uma série de disciplinas que, ao mesmo tempo em que compartilha uma abordagem essencial para a solução de problemas, está preocupada com aplicações diferentes. Nesta seção, apresentamos algumas das mais importantes áreas de aplicação e suas contribuições para a inteligência artificial como um todo.

1.2.1 Jogos

Muitas das pesquisas iniciais sobre busca em espaço de estados foram realizadas utilizando jogos de tabuleiro comuns, como xadrez, damas e o quebra-cabeça dos 15 (*15-puzzle*). Além do seu apelo intelectual inerente, os jogos de tabuleiro têm certas propriedades que os tornaram objetos de estudo ideais para esses trabalhos iniciais. A maioria dos jogos utiliza um conjunto bem definido de regras: isso faz com que seja fácil gerar o espaço de busca e libera o pesquisador de muitas das ambiguidades e complexidades inerentes a problemas menos estruturados. As configurações do tabuleiro usadas nesses jogos são facilmente representáveis em um computador, dispensando o formalismo complexo necessário para capturar as sutilezas semânticas de domínios de problemas mais complexos. Como os jogos podem ser praticados facilmente, o teste de programas de jogos não apresenta nenhum ônus financeiro ou ético. Nos capítulos 3 e 4, é apresentada a busca em espaço de estados, o paradigma que serve de base na maioria das pesquisas sobre jogos.

Os jogos podem gerar espaços de busca extremamente grandes. Eles podem ser grandes e complexos o suficiente para necessitarem de técnicas poderosas para determinar quais alternativas devem ser exploradas no espaço do problema. Essas técnicas são chamadas de *heurísticas* e constituem uma área importante da pesquisa em IA. Uma heurística é uma estratégia de solução útil, mas potencialmente falível, como, por exemplo, checar para assegurar que um equipamento que não está respondendo esteja ligado na tomada antes de admitir que ele esteja quebrado, ou ainda, rocar, a fim de proteger o próprio rei da captura em uma partida de xadrez. Muito do que chamamos de inteligência parece estar relacionado com as heurísticas usadas pelos seres humanos para resolver problemas.

Como a maioria das pessoas tem alguma experiência com esses jogos simples, é possível projetar e testar a eficácia de nossas próprias heurísticas. Não precisamos consultar um especialista como em algumas áreas restritas, por exemplo, a medicina e a matemática (xadrez é uma exceção óbvia a essa regra). Por essas razões, os jogos proporcionam um domínio rico para o estudo da busca heurística. O Capítulo 4 introduz a noção de heurística usando esses jogos simples. Os programas de jogos, apesar da sua simplicidade, apresentam seus próprios desafios, incluindo um adversário cujos movimentos não podem ser antecipados deterministicamente (capítulos 5 e 9), e a necessidade de se considerar fatores tanto psicológicos como táticos na estratégia do jogo.

Os recentes sucessos em jogos baseados em computador incluem campeonatos mundiais de gamão e xadrez. Também é interessante notar que, em 2007, o espaço de estados completo para o jogo de damas foi mapeado, permitindo que ele seja, desde o primeiro movimento, determinístico!

1.2.2 Raciocínio automatizado e prova de teoremas

Podemos afirmar que a prova automática de teoremas é o ramo mais antigo da inteligência artificial, sendo possível traçar as suas raízes passando pelo *Teorista Lógico*, de Newell e Simon (Newell e Simon, 1963a), bem como por seu *Resolvedor Geral de Problemas* (Newell e Simon, 1963b), passando pelos esforços de Russell e Whitehead para tratar toda a matemática como uma derivação puramente formal de teoremas, a partir de axiomas básicos, até as suas origens nos escritos de Babbage e Leibniz. De qualquer forma, esse tem sido certamente um dos ramos mais frutíferos da IA. A pesquisa em prova automática de teoremas foi responsável por muitos trabalhos iniciais na formalização de algoritmos de busca e no desenvolvimento de linguagens de representação formais como o cálculo de predicados (Capítulo 2) e a linguagem de programação em lógica Prolog.

Muito do apelo da prova automática de teoremas reside no rigor e na generalidade da lógica. Por ser um sistema formal, a lógica se presta bem a ser automatizada. Uma ampla variedade de problemas pode ser abordada representando a descrição do problema e os conhecimentos básicos relevantes como axiomas lógicos e tratando instâncias do problema como teoremas a serem provados. Essa abordagem é a base dos trabalhos em prova automática de teoremas e em sistemas de raciocínio matemático (Capítulo 14).

Infelizmente, os esforços iniciais para criar provadores de teoremas fracassaram em desenvolver um sistema que pudesse solucionar, de modo consistente, problemas complicados. Isso se deveu à habilidade de qualquer sistema lógico razoavelmente complexo de gerar um número infinito de teoremas prováveis: sem técnicas poderosas (heurísticas) para guiar a sua busca, os provadores automáticos de teoremas provavam um grande número de teoremas irrelevantes antes de encontrar o teorema correto. Em resposta a essa ineficiência, muitos pesquisadores afirmam que métodos sintáticos, puramente formais para guiar a busca, são inherentemente incapazes de lidar com um espaço tão grande e que a única alternativa é confiar nas estratégias *ad hoc* informais que os seres humanos parecem utilizar para resolver problemas. Essa é a estratégia implícita no desenvolvimento de sistemas especialistas (Capítulo 8) e que provou ser adequada.

Apesar disso, o apelo do raciocínio baseado em lógica matemática formal é forte demais para ser ignorado. Muitos problemas importantes, como o projeto e a verificação de circuitos lógicos, a verificação da correção de programas computacionais e o controle de sistemas complexos, parecem corresponder a essa abordagem. De fato, a comunidade de prova de teoremas tem obtido sucesso em conceber heurísticas poderosas para a solução de problemas que dependem apenas de uma estimativa da forma sintática de uma expressão lógica, e, como resultado, capazes de reduzir a complexidade do espaço de busca sem recorrer a técnicas *ad hoc* usadas pela maioria dos peritos humanos.

Outra razão para o interesse continuado em provadores automáticos de teoremas é a compreensão de que tais sistemas não precisam ser capazes de resolver independentemente problemas extremamente complexos sem assistência humana. Muitos provadores de teoremas modernos funcionam como assistentes inteligentes, liberando os seres humanos para realizar as tarefas mais exigentes de decomposição de um problema grande em subproblemas e para conceber heurísticas de busca no espaço de provas possíveis. O provador de teoremas, então, realiza a tarefa mais simples, mas ainda assim exigente, de provar lemas, verificar conjecturas menores e completar os aspectos formais de uma prova delineada por seu associado humano (Boyer e Moore, 1979; Bundy, 1988; Veroff, 1997; Veroff e Spinks, 2006).

1.2.3 Sistemas especialistas

Uma das mais importantes conclusões que foi tirada dos trabalhos iniciais em solução de problemas foi a importância do conhecimento específico do domínio. Um médico, por exemplo, não é efetivo em diagnosticar uma doença apenas porque ele possui uma habilidade inata em resolver problemas genéricos; ele é eficaz porque sabe muito sobre medicina. Da mesma forma, um geólogo é eficaz em descobrir depósitos de minérios porque ele é capaz de aplicar uma grande quantidade de conhecimento teórico e empírico sobre geologia ao problema específico. O conhecimento especialista é uma combinação de um entendimento teórico do problema com uma coleção de regras heurísticas para resolver problemas, que a experiência demonstrou ser efetiva no domínio. Os sistemas especialistas são construídos a partir da extração desse conhecimento de um especialista humano, codificando-o de uma forma que um computador possa aplicá-lo a problemas similares.

Uma característica fundamental dos sistemas especialistas é que a sua estratégia para resolver problemas é dependente do conhecimento de um especialista humano no domínio. Embora existam alguns programas em que o projetista é também a fonte do conhecimento do domínio, geralmente é muito mais provável que esses programas sejam um produto da colaboração entre um especialista do domínio, como um médico, um químico, um geólogo ou um engenheiro, e um especialista em inteligência artificial. O especialista no domínio fornece o conhecimento necessário do domínio do problema, tanto por meio de uma discussão geral dos seus métodos de resolução de problema quanto pela demonstração dessas habilidades em um conjunto cuidadosamente escolhido de exemplos de problemas. O especialista em IA, ou *engenheiro do conhecimento*, como frequentemente são co-

nhecidos os projetistas de sistemas especialistas, é responsável por implementar esse conhecimento em um programa que seja tanto efetivo como aparentemente inteligente do ponto de vista de seu comportamento. Uma vez que esse programa esteja escrito, é necessário refiná-lo a partir da apresentação de exemplos de problemas a resolver, sob a supervisão crítica do especialista no domínio, e realizar quaisquer alterações necessárias no conhecimento do programa. Esse processo é repetido até que o programa atinja o nível desejado de desempenho.

Um dos sistemas mais antigos a explorar o conhecimento específico do domínio para a solução de problemas foi o DENDRAL, desenvolvido em Stanford no final dos anos 1960 (Lindsay et al., 1980). Esse sistema foi projetado para inferir a estrutura de moléculas orgânicas a partir de suas fórmulas químicas e de informações de espectrografia de massa das ligações químicas presentes nas moléculas. Como as moléculas orgânicas normalmente são muito grandes, o número de estruturas possíveis para essas moléculas tende a ser enorme. O DENDRAL trata o problema desse grande espaço de busca aplicando o conhecimento heurístico de especialistas em química no problema de elucidação da estrutura. O método utilizado mostrou-se muito efetivo, obtendo rotineiramente a estrutura correta entre milhões de possibilidades após algumas tentativas. A abordagem se mostrou tão bem-sucedida que, hoje em dia, são usados programas descendentes e extensões daquele sistema em laboratórios químicos e farmacêuticos espalhados por todo o mundo.

Enquanto o DENDRAL foi um dos primeiros programas a usar efetivamente o conhecimento específico do domínio para alcançar o desempenho de especialistas na resolução do problema, o MYCIN estabeleceu a metodologia dos sistemas especialistas contemporâneos (Buchanan e Shortliffe, 1984). O MYCIN utiliza conhecimento de especialistas médicos para diagnosticar e prescrever tratamentos para meningite espinhal e infecções bacterianas do sangue. Esse programa, desenvolvido em Stanford em meados dos anos 1970, foi um dos primeiros a tratar os problemas do raciocínio com informações incertas ou incompletas. Ele fornecia explanações lógicas claras do seu próprio raciocínio, utilizava uma estrutura de controle apropriada para o domínio específico do problema e identificava critérios para avaliar o desempenho obtido de forma confiável. Muitas das técnicas de desenvolvimento de sistemas especialistas em uso atualmente foram desenvolvidas originalmente no projeto MYCIN (Capítulo 8).

Outros sistemas especialistas clássicos são o programa PROSPECTOR, para determinar a localização e o tipo prováveis de depósitos de minério com base em informação geológica sobre um sítio (Duda et al., 1979a, 1979b), o programa INTERNIST, para realizar diagnósticos na área da medicina interna, o Dipmeter Advisor, para interpretar os resultados de registros de perfuração de poços petrolíferos (Smith e Baker, 1983), e o XCON, para configurar computadores VAX. O XCON foi desenvolvido em 1981 e, naquele tempo, todo computador VAX vendido pela Digital Equipment Corporation era configurado por esse software. Vários outros sistemas especialistas são atualmente aplicados para resolver problemas em áreas como medicina, educação, negócios, projeto e ciências (Waterman, 1986; Durkin, 1994). Veja também os anais das conferências Innovative Applications of Artificial Intelligence (IAAI).

É interessante notar que a maioria dos sistemas especialistas foi escrita para domínios com nível de pericia relativamente especializado. Esses domínios são geralmente bem estudados e têm estratégias de solução de problemas claramente definidas. Problemas que dependem de uma noção de “senso comum”, definida de forma menos rígida, são muito mais difíceis de serem resolvidos por esses meios. Apesar das promessas dos sistemas especialistas, seria um erro superestimar a habilidade dessa tecnologia. Entre as deficiências mais comuns encontradas atualmente, estão:

1. Dificuldade em capturar conhecimento “profundo” do domínio do problema. Ao MYCIN, por exemplo, falta conhecimento real da fisiologia humana. Ele não sabe o que ocorre com o sangue ou qual é a função da medula espinhal. Conta-se que, ao selecionar uma droga para o tratamento de meningite, o MYCIN perguntou se o paciente estava “grávida”, mesmo sabendo que ele era do sexo masculino. Mesmo havendo dúvidas sobre o ocorrido, se é verdade ou apenas folclore, isso ilustra a limitação potencial do conhecimento em sistemas especialistas.
2. Falta de robustez e flexibilidade. Se a um ser humano for apresentado um problema que ele não consiga resolver imediatamente, ele geralmente poderá realizar um exame dos princípios fundamentais e chegar a uma estratégia para atacá-lo. Os sistemas especialistas, em geral, não possuem essa habilidade.

3. Incapacidade de fornecer explanações aprofundadas. Como os sistemas especialistas não têm um conhecimento profundo de seus domínios, as suas explanações são geralmente restritas a uma descrição dos passos que eles realizaram para encontrar a solução. Normalmente, por exemplo, eles são incapazes de explicar “por que” adotaram certa abordagem.
4. Dificuldades na verificação. Embora a correção de qualquer sistema computacional extenso seja difícil de ser provada, os sistemas especialistas são particularmente difíceis de serem verificados. Esse é um problema sério, uma vez que a tecnologia dos sistemas especialistas vem sendo utilizada em aplicações críticas, como o controle de tráfego aéreo, a operação de reatores nucleares e os sistemas de armamentos.
5. Pouco aprendizado por experiência. Os sistemas especialistas atuais são artesanais; quando o sistema já se encontra desenvolvido, o seu desempenho não irá melhorar sem a ajuda de seus programadores. Isso levanta severas dúvidas sobre a inteligência desses sistemas.

Apesar dessas limitações, os sistemas especialistas têm provado o seu valor em várias aplicações importantes e constituem um dos principais tópicos deste livro, sendo discutidos nos capítulos 7 e 8. Veja também os anais das conferências Innovative Applications of Artificial Intelligence (IAAI) para novas aplicações.

1.2.4 Compreensão da linguagem natural e modelagem semântica

Um dos objetivos da inteligência artificial, que vem sendo perseguido há muito tempo, é a criação de programas que sejam capazes de entender e gerar a linguagem humana. A habilidade em utilizar e compreender a linguagem natural não apenas parece ser um aspecto fundamental da inteligência humana, mas também a sua automação teria um impacto inacreditável sobre a facilidade de utilização e eficácia dos próprios computadores. Foram feitos grandes esforços no desenvolvimento de programas que comprehendem a linguagem natural. Embora esses programas tenham alcançado sucesso em contextos restritos, sistemas que possam usar linguagem natural com a flexibilidade e a generalidade que caracterizam a fala humana ainda estão além das metodologias atuais.

Compreender a linguagem natural envolve muito mais que a simples análise de sentenças, da separação de suas partes individuais e da procura dessas palavras em um dicionário. A compreensão real depende de um extenso conhecimento do domínio do discurso e das expressões idiomáticas utilizadas naquele domínio, bem como da habilidade em aplicar conhecimento contextual genérico para resolver omissões e ambiguidades que são parte usual da fala humana.

Considere, por exemplo, as dificuldades de se estabelecer uma conversação sobre beisebol com um indivíduo que entende o inglês, mas não sabe nada sobre as regras, os jogadores e a história do jogo. Será que essa pessoa poderia compreender o significado da sentença: “sem ninguém no topo da nona e com uma corrida na segunda, o treinador tirou o seu reserva do curral”? Muito embora cada uma das palavras na sentença possa ser compreendida individualmente, a sentença soaria incompreensível mesmo para uma pessoa muito inteligente que não fosse fã de beisebol.

A tarefa de coletar e organizar esse conhecimento de fundo, de forma que ele possa ser aplicado à compreensão da linguagem, constitui o problema fundamental na automação da compreensão da linguagem natural. Em resposta a essa necessidade, os pesquisadores desenvolveram muitas das técnicas para estruturar o significado semântico usado em toda a inteligência artificial (capítulos 7 e 15).

Devido à enorme quantidade de conhecimento necessário para a compreensão da linguagem natural, a maioria dos trabalhos é realizada em áreas de problemas especializadas e bem conhecidas. Um dos primeiros programas a explorar essa metodologia de “micromundo” foi o SHRDLU, de Winograd, um sistema de linguagem natural que podia “conversar” sobre uma configuração simples de blocos de diferentes formas e cores (Winograd, 1973). O SHRDLU podia responder a perguntas como “qual é a cor do bloco que está sobre o cubo azul?”, bem como planejar ações como “coloque a pirâmide vermelha sobre o tijolo verde”. Problemas desse tipo, que envolvem descrição e manipulação de arranjos simples de blocos, apareceram com uma frequência surpreendente em pesquisas iniciais de IA e são conhecidos como problemas de “mundo de blocos”.

Apesar do sucesso do SHRDLU em conversar sobre arranjos de blocos, os seus métodos não podem ser generalizados para além desse domínio. As técnicas representacionais usadas no programa eram simples demais para capturar, de uma maneira útil, a organização semântica de domínios mais ricos e complexos. Grande parte do trabalho atual em compreensão da linguagem natural é devotada a encontrar formalismos representacionais que sejam gerais o suficiente para serem usados em um amplo espectro de aplicações, mas também que se adaptem bem à estrutura específica de um dado domínio. Várias técnicas diferentes (a maioria das quais são extensões ou modificações de *redes semânticas*) são exploradas para esse propósito e são usadas no desenvolvimento de programas que podem compreender a linguagem natural em domínios de conhecimento restritos, mas interessantes. Por fim, em pesquisas recentes (Marcus, 1980; Manning e Schutze, 1999; Jurafsky e Martin, 2009), modelos estocásticos, descrevendo como conjuntos de palavras “ocorrem” no uso da linguagem, são empregados para caracterizar tanto a sintaxe como a semântica. Entretanto, a compreensão total da linguagem permanece além do estado da arte atual.

1.2.5 Modelando o desempenho humano

Embora grande parte da discussão anterior utilize a inteligência humana como um ponto de referência ao considerar a inteligência artificial, os programas discutidos não seguem o modelo de organização da mente humana. De fato, muitos programas de IA são concebidos para resolver algum problema útil sem levar em consideração as suas similaridades com a arquitetura mental humana. Mesmo os sistemas especialistas, que extraem muito do seu conhecimento de especialistas humanos, não procuram realmente simular os processos mentais internos humanos de solução de problemas. Se o desempenho é o único critério pelo qual um sistema é julgado, pode ser que não haja muitos motivos para simular os métodos humanos de solução de problemas; de fato, os programas que utilizam abordagens não humanas para resolver problemas são, frequentemente, mais bem-sucedidos que os seus correspondentes humanos. Ainda assim, o projeto de sistemas que modelam explicitamente algum aspecto do desempenho humano tem sido uma área fértil de pesquisa, tanto em inteligência artificial quanto em psicologia.

A modelagem do desempenho humano, além de proporcionar à IA grande parte de sua metodologia básica, tem se mostrado uma ferramenta poderosa para formular e testar teorias da cognição humana. As metodologias para solução de problemas, desenvolvidas por cientistas da computação, forneceram aos psicólogos uma nova metáfora para explorar a mente humana. Em vez de formular teorias da cognição na linguagem vaga usada nas pesquisas iniciais, ou desistir de descrever completamente os processos internos da mente humana (como sugerido pelos comportamentalistas), muitos psicólogos adotaram a linguagem e a teoria da ciência da computação para formular modelos de inteligência humana. Essas técnicas não apenas fornecem um novo vocabulário para descrever a inteligência humana, mas também as implementações computacionais dessas teorias oferecem aos psicólogos uma oportunidade de testar empiricamente, criticar e refinar as suas ideias (Luger, 1994; ver conferências e periódicos da *Cognitive Science Society*). Uma relação entre a inteligência artificial e inteligência humana é resumida no Capítulo 16.

1.2.6 Planejamento e robótica

A pesquisa em planejamento começou como um esforço para projetar robôs que pudessem realizar as suas tarefas com algum grau de flexibilidade e resposta em relação ao mundo externo. Em suma, planejamento pressupõe um robô que seja capaz de realizar certas ações atômicas. Ele procura encontrar uma sequência dessas ações que realize uma tarefa de alto nível, como se mover através de uma sala com vários obstáculos.

Planejamento é um problema difícil por uma série de razões, entre as quais o tamanho do espaço de sequências possíveis de movimentos. Mesmo um robô extremamente simples é capaz de gerar um vasto número de sequências de movimentos potenciais. Imagine, por exemplo, um robô que pode se mover para a frente, para trás, para a direita ou para a esquerda, e considere de quantas maneiras diferentes o robô pode se mover em uma sala. Suponha, também, que haja obstáculos na sala e que o robô deva selecionar um caminho que desvie desses obstáculos de uma forma eficiente. Desenvolver um programa que possa descobrir de modo inteligente o melhor cami-

nho sob essas circunstâncias, sem ser sobrepujado pelo número enorme de possibilidades, requer técnicas sofisticadas para representar conhecimento espacial e controlar a busca através de ambientes possíveis.

Um dos métodos que os seres humanos usam para planejamento é a *decomposição hierárquica do problema*. Ao planejar uma viagem a Londres, você geralmente tratará separadamente dos problemas de organizar o voo, chegar ao aeroporto, fazer as conexões aéreas e o traslado em Londres, apesar de eles serem parte de um plano global maior. Cada um desses problemas pode ser ainda decomposto em subproblemas menores, como encontrar um mapa da cidade, entender o sistema de metrô e encontrar um pub adequado. Essa abordagem não apenas res-tringe efetivamente o tamanho do espaço que deve ser buscado, mas também permite armazenar os subplanos frequentemente usados para serem reutilizados no futuro.

Diferentemente dos seres humanos que planejam sem grande esforço, o desenvolvimento de programas de computador que façam o mesmo é um grande desafio. Uma tarefa aparentemente simples, como decompor um problema em subproblemas independentes, na verdade requer heurísticas sofisticadas e conhecimento extensivo sobre o domínio do planejamento. Determinar quais subplanos devem ser armazenados e como eles podem ser generalizados para uso futuro é, também, um problema difícil.

Um robô dificilmente seria considerado inteligente se executasse cegamente uma sequência de ações sem responder a alterações no ambiente ou se não fosse capaz de detectar e corrigir erros no seu plano original. Um robô pode não possuir sensores adequados para localizar todos os obstáculos que se encontram no caminho planejado. Esse robô deve começar movendo-se pelo ambiente com base no que ele “percebeu”, corrigindo o seu caminho, conforme outros obstáculos são detectados. Organizar planos, de modo que seja possível responder a condições variáveis do ambiente, é um dos principais problemas do planejamento (Lewis e Luger, 2000; Thrun et al., 2007).

Por fim, a robótica foi uma das áreas da IA que produziu muitas das ideias que dão suporte à solução de problemas orientada a agentes (Seção 1.1.4). Frustrados tanto com a complexidade de se manter um grande espaço representacional, quanto com a concepção de algoritmos de busca adequados ao planejamento tradicional, pesquisadores como Agre e Chapman (1987), Brooks (1991a) e Thrun et al. (2007) reformularam o problema global em termos da interação de múltiplos agentes semiautônomos. Cada agente é responsável por uma porção individual do problema e, por meio da coordenação entre eles, a solução global emergirá.

A pesquisa em planejamento agora se estende para além dos domínios da robótica, incluindo a coordenação de qualquer conjunto complexo de tarefas e objetivos. Planejadores modernos são aplicados tanto a agentes (Nilsson, 1994) como para controlar aceleradores de feixe de partículas (Klein et al., 1999, 2000).

1.2.7 Linguagens e ambientes para IA

Alguns dos produtos secundários mais importantes da pesquisa em inteligência artificial foram avanços em linguagens de programação e em ambientes de desenvolvimento de software. Por uma série de razões, incluindo aí o tamanho de muitos programas de aplicação de IA, a importância de uma metodologia de “prototipação”, a tendência dos algoritmos de busca em gerar espaços enormes e a dificuldade de prever o comportamento de programas guiados por heurísticas, os programadores em IA foram forçados a desenvolver um conjunto poderoso de metodologias de programação.

Esses ambientes de programação incluem técnicas de estruturação do conhecimento, como a programação orientada a objetos. Linguagens de alto nível, como Lisp e Prolog, que suportam desenvolvimento modular, ajudam a gerenciar o tamanho e a complexidade do sistema. Pacotes de rastreamento permitem que um programador reconstrua a execução de um algoritmo complexo e tornam possível esclarecer as complexidades da busca guiada por heurísticas. Sem essas ferramentas e técnicas, é improvável que muitos dos sistemas de IA significativos tivessem sido construídos.

Muitas dessas técnicas são agora ferramentas-padrão para a engenharia de software e têm pouca relação com a teoria de IA. Outras, como a programação orientada a objetos, são de elevado interesse teórico e prático. Finalmente, muitos dos algoritmos de IA agora são construídos também em linguagens de programação mais tradicionais, como C++ e Java.

As linguagens desenvolvidas para a programação de inteligência artificial estão intimamente ligadas à estrutura teórica da área. Criamos muitas das estruturas representativas neste livro em Prolog, Lisp e Java, colocando-as à

disposição em Luger e Stubblefield (2009) e na Internet. Neste livro, não tomamos partido nas discussões ideológicas sobre seus méritos relativos. Em vez disso, seguimos o ditado “um bom trabalhador conhece todas as ferramentas”.

1.2.8 Aprendizado de máquina

O aprendizado permanece sendo uma área desafiadora para a IA. A importância do aprendizado, entretanto, é inquestionável, particularmente porque essa habilidade é um dos componentes mais importantes do comportamento inteligente. Um sistema especialista pode executar cálculos extensivos e custosos para resolver um problema. Entretanto, diferentemente de um ser humano, se em uma outra vez lhe for apresentado o mesmo problema ou outro similar, ele normalmente não se lembrará da solução. Ele realizará a mesma sequência de cálculos novamente. Isso é verdade para a segunda vez, bem como para a terceira, para a quarta e para qualquer outra oportunidade em que ele resolva o problema — o que não é o comportamento esperado de um sistema inteligente para resolver problemas. A solução óbvia é permitir que esses sistemas aprendam por conta própria, seja por sua própria experiência, por analogia, por exemplos, por um professor que lhes “diga” o que fazer, ou por recompensa ou punição, dependendo dos resultados.

Embora o aprendizado seja uma área difícil, há diversos programas que sugerem que isso não é impossível. Um programa pioneiro é o AM (*Automated Mathematician* — matemático automatizado), concebido para descobrir leis matemáticas (Lenat, 1977, 1982). A partir de conceitos e axiomas da teoria dos conjuntos inicialmente fornecidos, o AM foi capaz de induzir conceitos matemáticos importantes como cardinalidade, aritmética inteira e muitos resultados da teoria dos números. O AM formulou novos teoremas a partir da modificação da sua base de conhecimento atual e usou heurísticas para encontrar o “melhor” teorema entre uma série de teoremas possíveis. Mais recentemente, Cotton et al. (2000) concebeu um programa que inventa automaticamente sequências “interessantes” de inteiros.

Entre os trabalhos pioneiros influentes em aprendizado, está a pesquisa de Winston sobre a indução de conceitos estruturais como “arco” a partir de um conjunto de exemplos do mundo de blocos (Winston, 1975a). O algoritmo ID3 obteve sucesso em aprender padrões gerais a partir de exemplos (Quinlan, 1986a). O MetaDENDRAL aprende regras para interpretar dados de espectrografia de massa em química orgânica a partir de exemplos de dados extraídos de compostos de estrutura conhecida. *Tiresias*, uma interface para sistemas especialistas, converte recomendações descritas em alto nível em regras novas para a sua base de regras (Davis, 1982). *Hacker* realiza o planejamento de manipulações no mundo de blocos a partir de um processo iterativo de conceber um plano, testá-lo e corrigir qualquer falha descoberta no plano candidato (Sussman, 1975). Os trabalhos em aprendizado baseado em explanação mostraram a eficácia do conhecimento prévio sobre o aprendizado (Mitchell et al., 1986; DeJong e Mooney, 1986). Atualmente, existem também muitos modelos biológicos e sociológicos importantes de aprendizado; apresentaremos uma revisão desses modelos nos capítulos sobre aprendizado conexionista e aprendizado emergente.

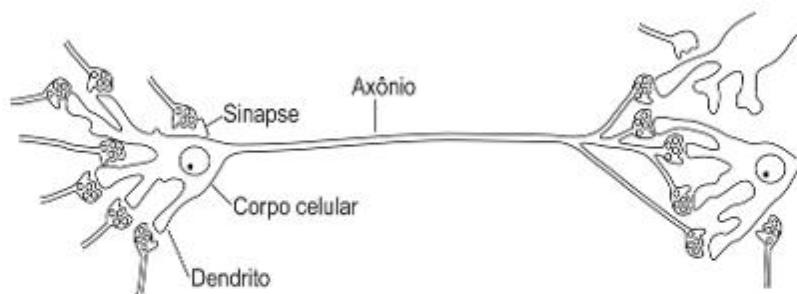
O sucesso dos programas de aprendizado de máquina sugere a existência de um conjunto de princípios gerais do aprendizado que permitirá a construção de programas com capacidade de aprender em domínios realistas. Apresentamos várias abordagens ao aprendizado na Seção IV.

1.2.9 Representações alternativas: redes neurais e algoritmos genéticos

A maioria das técnicas apresentadas neste livro sobre IA utiliza o conhecimento representado explicitamente e algoritmos de busca cuidadosamente concebidos para implementar a inteligência. Uma abordagem bem diferente procura construir programas inteligentes utilizando modelos que imitam a estrutura dos neurônios no cérebro humano ou padrões evolutivos encontrados nos algoritmos genéticos e em vida artificial.

Um esquema simples de um neurônio (Figura 1.2) consiste em um corpo celular com várias protuberâncias ramificadas, chamadas de *dendritos*, e um único ramo chamado de *axônio*. Os dendritos recebem sinais de outros neurônios. Quando esses impulsos combinados excedem um determinado limiar, o neurônio dispara e um impulso é

Figura 1.2 Um diagrama simplificado de um neurônio a partir de Crick e Asanuma (1986).



propagado ao longo do axônio. Os ramos nas terminações do axônio formam *sinapses* com os dendritos de outros neurônios. A sinapse é o ponto de contato entre os neurônios e pode ser *excitatória* ou *inibitória*, dependendo se elas contribuem, respectivamente, para aumentar o sinal global ou, então, para diminuí-lo.

Essa descrição de um neurônio é excessivamente simples, mas capta as características que são relevantes para os modelos neurais de computação. Em particular, cada unidade computacional calcula uma função específica de suas entradas e passa o resultado para outras unidades da rede que estão conectadas a ela: os resultados finais são produzidos pelo processamento paralelo e distribuído dessa rede de conexões neurais e seus limiares de peso.

As arquiteturas neurais são mecanismos que possuem grande apelo para implementar a inteligência por várias razões. Os programas tradicionais de IA podem ser frágeis e excessivamente suscetíveis a ruído. A inteligência humana é muito mais flexível e consegue interpretar muito bem sinais ruidosos, como reconhecer um rosto em uma sala escura ou manter uma conversa durante uma festa barulhenta. As arquiteturas neurais, por capturarem o conhecimento por meio de um grande número de unidades distribuídas em uma rede, têm um potencial maior para reconhecer dados ruidosos e incompletos.

Com os algoritmos genéticos e a vida artificial, desenvolvemos novas soluções para o problema a partir de elementos de soluções anteriores. Os operadores genéticos, como *combinação* e *mutação*, à semelhança de seus equivalentes genéticos do mundo natural, produzem a cada nova geração soluções potenciais cada vez melhores para o problema. A vida artificial produz uma nova geração como uma função da “qualidade” de seus vizinhos em gerações anteriores.

Tanto as arquiteturas neurais como os algoritmos genéticos fornecem um modelo natural para o paralelismo, porque cada neurônio, ou segmento de uma solução, é uma unidade independente. Hillis (1985) observou o fato de que os seres humanos avançam mais rapidamente em uma tarefa à medida que adquirem mais conhecimento, enquanto os computadores tendem a diminuir o ritmo. Essa desaceleração se deve ao custo da busca sequencial em uma base de conhecimento; uma arquitetura maciçamente paralela como o cérebro humano não deveria sofrer desse problema. Por fim, existe algo intrinsecamente atraente ao abordar os problemas de inteligência do ponto de vista neural ou genético. Afinal, o cérebro evoluiu até alcançar a inteligência, utilizando, para isso, uma arquitetura neural. Apresentamos as redes neurais, os algoritmos genéticos e a vida artificial nos capítulos 10 e 11.

1.2.10 IA e filosofia

Na Seção 1.1, apresentamos as raízes filosóficas, matemáticas e sociais da inteligência artificial. É importante compreender que a moderna IA não é apenas um produto dessa rica tradição intelectual, mas ela também contribui para isso.

Por exemplo, as questões que Turing propôs sobre programas inteligentes, por exemplo, são refletidas também no nosso entendimento da própria inteligência. O que é inteligência e como ela pode ser descrita? Qual é a natureza do conhecimento? O conhecimento pode ser representado? Como o conhecimento em uma área de aplicação se

relaciona com a habilidade de resolver problemas nesse domínio? Como é que *saber o que* é verdadeiro — a *teoria* de Aristóteles — relaciona-se com *saber como* realizar — a sua *praxis*?

As respostas propostas para essas questões constituem uma parte importante das atividades dos pesquisadores em IA. Em termos científicos, os programas de IA podem ser vistos como experimentos. Uma concepção se torna concreta em um programa e o programa é executado como um experimento. Os projetistas desses programas observam os resultados e, então, modificam o projeto e realizam novamente o experimento. Dessa forma, podemos determinar se as nossas representações e os nossos algoritmos são modelos suficientes de comportamento inteligente. Newell e Simon (1976) propuseram essa abordagem para o entendimento científico na sua palestra do *Turing Award* de 1976, e um modelo mais forte para a inteligência por meio da sua hipótese de sistema simbólico físico: *a condição necessária e suficiente para um sistema físico exibir inteligência é que ele seja um sistema simbólico físico*. Na Seção VI, analisaremos o que essa hipótese significa na prática e como ela tem sido criticada por vários pensadores modernos.

Várias aplicações de IA suscitam, também, questões filosóficas profundas. Em que medida podemos dizer que um computador pode entender expressões em linguagem natural? Para produzir ou compreender uma linguagem, é necessária a interpretação de símbolos. Não é suficiente ser capaz de dizer que uma cadeia de símbolos está bem formada. Um mecanismo para a compreensão deve ser capaz de atribuir significado ou interpretar símbolos dentro de um contexto. O que é significado? O que é interpretação? Em que medida a interpretação requer responsabilidade?

Questões filosóficas similares emergem de muitas áreas de aplicação da IA, desde a construção de sistemas especialistas para cooperar com peritos humanos até o projeto de sistemas de visão computacional, ou ainda, o projeto de algoritmos para aprendizado de máquina. Abordaremos várias dessas questões quando elas surgirem nos diversos capítulos deste livro e retornaremos à questão geral da relevância para a filosofia na Seção VI.

1.3 Inteligência artificial — um resumo

Tentamos definir inteligência artificial a partir da discussão de suas áreas mais importantes de pesquisa e aplicação. Essa visão geral revela um campo de estudo jovem e promissor, cujo interesse principal é encontrar um modo efetivo de entender e aplicar técnicas inteligentes para a solução de problemas, para o planejamento e as habilidades de comunicação em uma ampla gama de problemas práticos. Apesar da variedade de problemas tratados pela inteligência artificial, emerge uma série de características importantes que parecem ser comuns a todas as divisões da área; entre elas se incluem:

1. O uso do computador para executar raciocínio, reconhecimento de padrões, aprendizado ou outras formas de inferência.
2. Um foco em problemas que não respondem a soluções algorítmicas. Isso implica a utilização de busca heurística como uma técnica de IA para a solução de problemas.
3. Um interesse na solução de problemas utilizando informação inexata, faltante ou insuficientemente definida, e o uso de formalismos representacionais que possibilitem ao programador compensar esses problemas.
4. Raciocínio que utiliza as características qualitativas significativas de uma situação.
5. Uma tentativa de tratar de questões que envolvem tanto significado semântico como forma sintática.
6. Respostas que não são nem exatas nem ótimas, mas que são “suficientes” em um certo sentido. Isso é resultado de se utilizar essencialmente métodos de solução de problemas baseados em heurísticas em situações em que resultados ótimos, ou exatos, são caros demais ou mesmo impossíveis.
7. O uso de grandes quantidades de conhecimento específico de um domínio para resolver problemas. Essa é a base dos sistemas especialistas.
8. O uso de metaconhecimento para produzir um controle mais sofisticado sobre as estratégias de resolução de problemas. Embora esse seja um problema muito difícil, tratado por relativamente poucos sistemas, ele vem emergindo como uma área essencial de pesquisa.

Esperamos que essa introdução tenha fornecido uma impressão geral da estrutura e do significado da área de inteligência artificial. Esperamos, também, que as breves discussões sobre questões técnicas, como busca e representação, não tenham sido excessivamente enigmáticas e obscuras; elas serão desenvolvidas em uma profundidade mais adequada ao longo deste texto e foram incluídas aqui para demonstrar a sua importância na organização geral da área.

Como mencionamos na discussão sobre solução de problemas orientada a agentes, os objetos adquirem significado a partir da sua relação com outros objetos. Isso é igualmente válido para fatos, teorias e técnicas que constituem uma área de estudo científico. A intenção é fornecer uma ideia geral desses inter-relacionamentos, de modo que, quando os temas técnicos da inteligência artificial forem apresentados separadamente, eles encontrarão o seu lugar em uma compreensão crescente da substância e em direções globais da área. Somos guiados por uma observação feita por Gregory Bateson, o psicólogo e teórico de sistemas (Bateson, 1979):

Se você quebrar o padrão que conecta os itens do aprendizado, você necessariamente destruirá toda a qualidade.

1.4 Epílogo e referências

A área da IA reflete alguns dos interesses mais antigos da civilização ocidental à luz do modelo computacional moderno. As noções de racionalidade, representação e raciocínio estão agora, talvez mais do que nunca, sob uma observação minuciosa, porque nós pesquisadores da IA necessitamos compreendê-las de modo algorítmico! Ao mesmo tempo, a situação política, econômica e ética da nossa espécie nos força a refletir sobre a nossa responsabilidade pelos efeitos das nossas conquistas.

Várias fontes excelentes disponíveis sobre os tópicos levantados neste capítulo são *Mind Design* (Haugeland, 1997), *Artificial Intelligence: The Very Idea* (Haugeland, 1985), *Brainstorms* (Dennett, 1978), *Mental Models* (Johnson-Laird, 1983), *Elbow Room* (Dennett, 1984), *The Body in the Mind* (Johnson, 1987), *Consciousness Explained* (Dennett, 1991) e *Darwin's Dangerous Idea* (Dennett, 1995), *Prehistory of Android Epistemology* (Glymour, Ford e Hayes, 1995a) e *Sweet Dreams* (Dennett, 2006).

Várias das fontes clássicas também estão disponíveis, incluindo as obras *Física*, *Metafísica* e *Lógica*, de Aristóteles; os artigos de Frege; os escritos de Babbage, Boole, e Russell, e Whitehead. Os artigos de Turing também são muito interessantes, especialmente as suas discussões sobre a natureza da inteligência e a possibilidade de construção de programas inteligentes (Turing, 1950). O famoso artigo de Turing (1937), *On Computable Numbers, with an Application to the Entscheidungsproblem*, elabora a teoria de máquinas de Turing e a definição de computabilidade. A biografia de Turing, *Alan Turing: The Enigma* (Hodges, 1983), também proporciona uma excelente leitura. *Pandemonium* (1959), de Selfridge, é um exemplo pioneiro de aprendizado. Uma importante coleção de artigos pioneiros em IA pode ser encontrada em Webber e Nilsson (1981).

Computer Power and Human Reason (Weizenbaum, 1976) e *Understanding Computers and Cognition* (Winograd e Flores, 1986) oferecem comentários soberbos sobre as limitações e questões éticas em IA. *The Sciences of the Artificial* (Simon, 1981) é uma afirmação positiva sobre as possibilidades da inteligência artificial e o seu papel na sociedade.

As aplicações da IA mencionadas na Seção 1.2 visam introduzir o leitor aos variados interesses dos pesquisadores de IA e procuram delinear muitas das questões importantes que estão sendo investigadas. Cada uma dessas subseções se refere às principais áreas deste livro, onde esses tópicos são apresentados. *The Handbook of Artificial Intelligence* (Barr e Feigenbaum, 1989) oferece uma introdução a cada uma dessas áreas. *Encyclopedia of Artificial Intelligence* (Shapiro, 1992) oferece uma visão clara e comprehensível do campo da inteligência artificial.

A compreensão da linguagem natural é um campo de estudo muito dinâmico; alguns pontos de vista importantes se encontram em *Natural Language Understanding* (Allen, 1995), *Language as a Cognitive Process* (Winograd, 1983), *Computer Models of Thought and Language* (Schank e Colby, 1973), *Grammar, Meaning and the Machine Analysis of Language* (Wilks, 1972), *The Language Instinct* (Pinker, 1994), *Philosophy in the Flesh*

(Lakoff e Johnson, 1999) e *Speech and Language Processing* (Jurafsky e Martin, 2009); uma introdução à área se encontra nos nossos capítulos 7 e 15.

O uso de computadores para modelar a capacidade humana, que tratamos rapidamente no Capítulo 17, é discutido com certa profundidade em *Human Problem Solving* (Newell e Simon, 1972), *Computation and Cognition* (Pylyshyn, 1984), *Arguments Concerning Representations for Mental Imagery* (Anderson, 1978) e *Cognitive Science: the Science of Intelligent Systems* (Luger, 1994), *Problem Solving as Model Refinement: Towards a Constructivist Epistemology* (Luger et al., 2002) e *Bayesian Brain* (Doya et al., 2007).

O aprendizado de máquina é discutido na Parte IV; são importantes fontes para esse assunto o conjunto de vários volumes, *Machine Learning* (Michalski et al., 1983, 1986; Kodratoff e Michalski, 1990), o *Journal of Artificial Intelligence* e o *Journal of Machine Learning*. Outras referências poderão ser encontradas nos quatro capítulos da Parte IV.

Por fim, o Capítulo 12 apresenta uma visão de inteligência que enfatiza a sua estrutura modular e a adaptação a um contexto social e natural. O *Society of Mind*, de Minsky (1985), é uma das primeiras e mais provocantes articulações desse ponto de vista. Veja também *Android Epistemology* (Ford et al., 1995b) e *Artificial Life* (Langton, 1995).

1.5 Exercícios

1. Crie e justifique sua definição para inteligência artificial.
2. Dê vários outros exemplos da distinção aristotélica entre “matéria” e “forma”. Você pode mostrar como os seus exemplos poderiam se encaixar em uma teoria da abstração?
3. Muito do pensamento ocidental tradicional se fundamenta na relação mente-corpo. A mente e o corpo são:
 - a. entidades distintas que interagem de alguma forma; ou
 - b. a mente é uma expressão de “processos físicos”; ou
 - c. o corpo é apenas uma ilusão da mente racional?Discuta suas ideias a respeito do problema mente-corpo e a sua importância para uma teoria da inteligência artificial.
4. Critique os critérios de Turing para que um sistema computacional seja “inteligente”.
5. Descreva seus próprios critérios para que um sistema computacional seja considerado “inteligente”.
6. Embora a computação seja uma disciplina relativamente nova, filósofos e matemáticos têm pensado por milhares de anos sobre as questões envolvidas na solução automática de problemas. Qual é a sua opinião sobre a relevância dessas questões filosóficas para a concepção de um dispositivo para a resolução inteligente de problemas? Justifique a sua resposta.
7. Dadas as diferenças entre as arquiteturas dos computadores modernos e a do cérebro humano, qual é a relevância da pesquisa sobre a estrutura fisiológica e a função de sistemas biológicos para a engenharia de programas de IA? Justifique a sua resposta.
8. Escolha uma área de aplicação na qual você considere que se justifique o trabalho necessário para implementar uma solução por sistema especialista. Explique o problema escolhido detalhadamente. Com base na sua própria intuição, quais aspectos dessa solução seriam mais difíceis de serem automatizados?
9. Cite dois benefícios dos sistemas especialistas, além daqueles já listados no texto. Discuta-os em termos de resultados intelectuais, sociais ou financeiros.
10. Discuta por que você acha que o problema do aprendizado de máquina é tão difícil.
11. Discuta se você acha possível, ou não, que um computador comprehenda e use uma linguagem natural (humana).
12. Liste e discuta dois efeitos potencialmente negativos para a sociedade do desenvolvimento de técnicas de inteligência artificial.