

# 1 INTRODUÇÃO

*Tentamos aqui explicar porque consideramos a inteligência artificial um assunto merecedor de estudo e tentamos decidir o que ela é exatamente.*

A humanidade tem se dado o nome científico **homo sapiens** – homem sábio – pois sua capacidade mental é importante para nossa vida cotidiana e nosso senso de ego. O campo da Inteligência Artificial, ou IA, tenta entender entidades inteligentes. Deste modo, uma razão para estudá-la é aprender mais sobre nós mesmos. Mas, ao contrário da filosofia e da psicologia, que também se preocupam com a inteligência, a IA se esforça para construir entidades inteligentes assim como entendê-las. Outra razão para estudar IA é que estas entidades inteligentes construídas são interessantes e úteis em si. A IA tem produzido muitos produtos significativos e impressionantes mesmo neste seu estágio inicial de desenvolvimento. Embora ninguém possa prever o futuro em detalhes, está claro que computadores com um nível humano de inteligência (ou melhor) teria um enorme impacto em nossa vida cotidiana e no curso futuro da civilização.

A IA lança-se num de seus últimos desafios. Como é possível para um cérebro lento e pequeno, seja biológico ou eletrônico, perceber, entender, prever e manipular um mundo muito maior e mais complicado que ele próprio? Como poderemos fazer algo com essas propriedades? Estas são questões difíceis, mas ao contrário da busca da viagem mais rápida que a luz ou de um dispositivo antigravitacional, o pesquisador em IA possui uma forte evidência de que sua busca é possível. Tudo o que os pesquisadores têm que fazer é olhar no espelho e ver um exemplo de sistema inteligente.

A IA é uma das mais recentes disciplinas. Foi iniciada formalmente em 1956, quando o nome foi cunhado, embora naquele ponto certos trabalhos já tivessem sido discretamente realizados há cerca de cinco anos. Ao lado da Genética moderna, é regularmente citada como o “campo que eu mais iria gostar de atuar” por cientistas de outras disciplinas. Um estudante de Física pode ter o sentimento de que todas as boas idéias já foram abordadas por Galileu, Newton, Einstein, e o resto, e que toma-se muito tempo de estudo antes que alguém possa contribuir com novas idéias. A IA, por outro lado, ainda tem oportunidades para um Einstein em tempo integral.

O estudo da inteligência é também uma das mais antigas disciplinas. Por mais de 2000 anos, filósofos têm tentado entender como ver, aprender, lembrar, e raciocinar poderiam, ou deveriam, ser feitos<sup>1</sup>. O advento de computadores utilizáveis nos primórdios dos anos 50, mudaram a já aprendida mas antiquada especulação a respeito dessas faculdades mentais para uma disciplina real, experimental e teórica.

Muitos sentiam que os novos “Super-Cérebros eletrônicos” tinham potencial ilimitado para a inteligência. “Mais rápido do que Einstein” era uma manchete típica da época. Mas, assim como adaptar um veículo para criar entidades artificialmente inteligentes, o computador provê uma ferramenta para testar teorias de inteligência, e muitas teorias falham em especificar o teste, um caso de “pular da frigideira e cair no fogo”. A IA mostrou-se ser bem mais difícil do que muitos inicialmente imaginaram, e idéias modernas são mais ricas, mais sutis, e mais interessantes como resultado.

Atualmente a IA abrange uma alta variedade de subáreas, desde áreas de propósito geral como percepção e argumentação lógica, até tarefas específicas como jogar xadrez, provar teoremas matemáticos, escrever poesia e diagnosticar doenças. Frequentemente, cientistas de outras áreas migram gradualmente para a IA, onde encontram as ferramentas e o vocabulário para sistematizar e automatizar

---

<sup>1</sup> Um ramo mais recente da Filosofia preocupa-se em provar que a IA é impossível. Retornaremos a esse interessante ponto de vista no Capítulo 26.

as tarefas intelectuais nas quais têm trabalhado. Da mesma forma, os que trabalham em IA podem preferir aplicar seus métodos a qualquer área de esforço intelectual humano. Neste sentido, é um campo verdadeiramente universal.

## 1.1 O QUE É IA

Já explicamos porque IA é excitante, mas ainda não dissemos o que ela é. Poderíamos apenas dizer, “Bem, tem algo a haver com programas inteligentes, então vamos em frente e escrevamos alguns.” Mas a história da Ciência mostra que é útil estabelecer as metas certas. Os antigos alquimistas, buscando a porção da vida eterna e um método para transformar chumbo em ouro, provavelmente estavam provavelmente deslocados no caminho correto. Só quando a meta mudou, para aquela de encontrar teorias explícitas que davam prognósticos precisos do mundo terrestre, da mesma forma que a astronomia antiga predisse os movimentos aparentes das estrelas e planetas, é que o método científico emergiu e a ciência produtiva pôde ser realizada.

“O novo e excitante esforço para fazer os computadores pensar ... <i>máquinas com mentes</i> , no sentido amplo e literal” (Haugeland, 1985)	“O estudo das faculdades mentais através da utilização de modelos computacionais” (Charniak e McDermott, 1985)
“[A automação das] atividades que associamos ao pensamento humano, tais como tomada de decisão, resolução de problemas, aprendizado,...”(Bellman, 1978)	“O estudo das computações que tornam possível perceber, raciocinar e atuar” (Winstone, 1992)
“A arte de criar máquinas que executam funções que requerem inteligência quando executadas por pessoas” (Kurzweil, 1990)	“Um campo de estudo que pretende explicar e emular comportamento inteligente em termos de processos computacionais” (Schalkoff, 1990)
“O estudo de como fazer computadores realizarem coisas as quais, no momento, as pessoas as fazem melhor” (Rich e Knight, 1991)	“O ramo da Ciência da Computação preocupado com a automação do comportamento inteligente” (Luger e Stubblefield, 1993)
<b>Figura 1.1</b> Algumas definições de IA. Estão organizadas segundo as seguintes categorias	
Sistemas que pensam como humanos	Sistemas que pensam racionalmente
Sistemas que agem como humanos	Sistemas que agem racionalmente

Definições da IA de acordo com oito livros recentes são mostradas na figura 1.1. Estas definições variam ao longo de duas dimensões principais. As do topo preocupam-se com o processo do pensamento e raciocínio, enquanto as de baixo dirigem-se ao comportamento. Além disso, as definições da esquerda medem o sucesso em termos da desempenho humano, enquanto que as da direita o medem comparando com um conceito ideal de inteligência, que nós chamaremos de racionalidade. Um sistema é racional se ele faz a coisa certa. Isto dá-nos quatro possíveis metas para perseguir na inteligência artificial, como visto na legenda da figura 1.1.

Historicamente, todas as quatro abordagens têm sido seguidas. Como se poderia esperar, existe uma tensão entre abordagens centradas nos humanos e abordagens centradas na racionalidade<sup>2</sup>. Uma abordagem centrada nos humanos deve ser uma ciência empírica, envolvendo hipóteses e confirmação experimental. Uma abordagem racionalista envolve uma combinação de Matemática e Engenharia. Pessoas em cada grupo algumas vezes criticam trabalhos feitos por outros grupos, mas a verdade é que cada direção tem realizado valiosas contribuições. Vejamos cada uma mais detalhadamente.

<sup>2</sup> Devemos apontar que, por distinguir entre comportamento humano e racional, não estamos sugerindo que os humanos sejam necessariamente "irracionais" no sentido de "instabilidade emocional" ou insanidade. Alguém poderia notar que sempre cometemos erros; não somos todos grandes mestres de xadrez, muito embora possamos todos conhecer as regras desse jogo; e infelizmente nem todo mundo obtém nota máxima nos exames. Alguns erros sistemáticos no raciocínio humano são catalogados por Kahneman *et al.* (1982).

## Agindo humanamente: a abordagem do teste de Turing

O **teste de Turing**, proposto por Alan Turing (1950), foi desenvolvido para prover uma definição operacional satisfatória de inteligência. Turing definiu comportamento inteligente como a habilidade de alcançar desempenho humano em todas as tarefas cognitivas, o suficiente para ludibriar um interrogador. Falando mais claramente, pelo teste proposto, um computador deve ser interrogado por um humano, via um terminal, passando no teste se o interrogador não puder dizer se há um computador ou um humano do outro lado do terminal. O capítulo 26 discute sobre os detalhes do teste, e quando um computador é ou não realmente inteligente após passar no teste. Atualmente, programar um computador para passar no teste é uma tarefa árdua. O computador precisaria possuir as seguintes capacidades:

- **Processamento de linguagem natural** para que ele possa se comunicar com sucesso em Inglês (ou qualquer outro idioma);
- **Representação do conhecimento** para armazenar as informações fornecidas antes ou durante o interrogatório;
- **Raciocínio automático** para usar a informação armazenada para responder questões e para chegar a novas conclusões;
- **Capacidade de aprendizagem** para se adaptar às novas circunstâncias e para detectar e extrapolar padrões.

O teste de Turing evita deliberadamente a interação física direta entre o interrogador e o computador, porque a simulação física de uma pessoa é desnecessário para a inteligência. No entanto, o então chamado **teste total de Turing** inclui um sinal de vídeo, de maneira que o interrogador possa testar as habilidades de percepção visual do interrogado, assim como a oportunidade para o interrogador passar objetos físicos para o interrogado “pela escotilha”. Para passar no Teste total de Turing, o computador precisa:

- **Visão de computador** para perceber os objetos e
- **Robótica** para movê-los.

Dentro da IA, não tem sido um grande desafio tentar passar no teste de Turing. A capacidade de agir como humano se torna trivial quando programas com IA precisam interagir com pessoas, assim como quando um sistema especialista explica como ele chegou a seu diagnóstico, ou um sistema de processamento de linguagem natural tem um diálogo com um usuário. Estes programas precisam se comportar de acordo com certas convenções normais de interação humana de maneira a se fazerem entender. A representação subjetiva e o raciocínio em tal sistema pode ou não estar baseado em um modelo humano.

## Pensando humanamente: a abordagem da modelagem cognitiva

Se vamos dizer que um dado programa pensa com um humano, devemos ter alguma maneira de determinar como pensam os humanos. Precisamos nos inserir nos trabalhos reais da mente humana.

Existem duas maneiras de se fazer isso: através de introspecção - tentando capturar nossos próprios pensamentos quando eles ocorrem – ou através de experimentos psicológicos. Uma vez que tenhamos uma teoria suficientemente precisa sobre a mente, torna-se possível expressar essa teoria como um programa de computador. Se a entrada, saída e comportamento temporal do computador se igualam ao humano, isto é uma evidência de que alguns dos mecanismos do programa podem também estar operando em humanos. Por exemplo, Newell e Simon, que desenvolveram o GPS “Solucionador Geral de Problemas” (Newell e Simon, 1961), não se contentaram em ter seu programa solucionando corretamente os problemas. Estavam mais preocupados em comparar traços dos seus passos de raciocínio a traços de humanos resolvendo os mesmos problemas. Isto está em contraste com outros pesquisadores da mesma época (como Wang (1960)), preocupados em obter respostas corretas a respeito de como os humanos poderiam fazê-lo. O campo interdisciplinar da **ciência cognitiva** trás consigo modelos computacionais da IA e técnicas experimentais da Psicologia para tentar construir teorias precisas e testáveis sobre o funcionamento da mente humana.

Embora a ciência cognitiva seja um campo fascinante por si só, não falaremos muito sobre ela neste livro. Comentaremos ocasionalmente similaridades ou diferenças entre técnicas de IA e cognição humana. A ciência cognitiva real, no entanto, é necessariamente baseada em investigações experimentais dos humanos ou animais, e assumimos que o leitor tem acesso apenas a um computador para experimentação. Notaremos simplesmente que IA e ciência cognitiva continuam a fertilizar uma a outra, especialmente nas áreas de visão, linguagem natural e aprendizagem. A história das teorias psicológicas da cognição são brevemente cobertas na seção 1.2.

### **Pensando racionalmente: a abordagem das leis do pensamento**

O filósofo grego Aristóteles foi um dos primeiros a tentar codificar o pensamento correto, isto é, processos de raciocínio irrefutáveis. Seus famosos silogismos proveram padrões para estruturas de argumentação que sempre davam conclusões corretas a premissas corretas dadas. Por exemplo, “Sócrates é um homem; todo homem é mortal, portanto Sócrates é mortal”. Estas leis do pensamento eram supostamente entendidas como governantes da operação da mente e deram início ao campo da lógica.

O desenvolvimento da lógica formal no fim do século XIX e no início de século XX, que descrevemos com mais detalhes no Capítulo 6, provê uma notação precisa para declarações sobre todos os tipos de coisas no mundo e a relação entre eles. (Em contraste com a notação aritmética ordinal, que provê principalmente declarações de igualdade e desigualdade sobre os números.) Por volta de 1965, existiam programas que podiam, dados suficientes tempo e memória, tomar uma descrição de um problema em notação lógica e encontrar a solução para o mesmo, se ela existir. (Se não existir solução o programa pode nunca parar procurando por ela.) A então chamada tradição **logística** dentro da IA espera construir programas suficientes para criar sistemas inteligentes.

Há dois principais obstáculos para essa abordagem. Primeiro, não é fácil tomar conhecimento informal e estabelecê-lo em termos formais requeridos pela notação lógica, particularmente quando o conhecimento é menos de 100% certo. Segundo, há uma grande diferença entre estar apto a resolver um problema “a princípio” e fazê-lo na prática. Mesmo problemas com apenas poucas dúzias de fatos podem exaurir os recursos computacionais de qualquer computador, a menos que ele tenha alguma diretiva para o tratamento da ordem do passo de raciocínio. Embora ambos obstáculos se impliquem a qualquer tentativa de se construir sistemas de raciocínio computacional, eles apareceram primeiramente na tradição logística, porque o poder de representação dos sistemas de raciocínio são bem definidos, e muito bem entendidos.

### **Agindo racionalmente: a abordagem do agente racional**

Agir racionalmente significa agir de maneira a alcançar uma meta, dada uma crença. Um agente é apenas alguma coisa que percebe e age. (Isso pode ser um uso incomum da palavra, mas você se acostumará a ele.) Nessa abordagem, a IA é vista como o estudo e a construção de agentes racionais.

Na abordagem “lei do pensamento” para a IA, toda ênfase estava sobre as inferências corretas. Fazer inferências corretas é, às vezes, parte de ser um agente racional, porque uma maneira de agir racionalmente é raciocinar logicamente para a conclusão que uma dada ação alcançará uma dada meta, e então agir de acordo com essa conclusão. Por outro lado, inferência correta não é o todo da racionalidade, porque existem situações freqüentes onde não existe algo comprovadamente correto a se fazer, ainda que alguma coisa deva ainda ser feita. Existem também maneiras de agir racionalmente que não podem ser racionalmente ditas como envolvendo inferências. Por exemplo, puxar a mão de dentro de um forno quente é uma ação de reflexo que tem mais sucesso do que uma ação mais lenta tomada após uma ponderação cuidadosa.

Todas as “habilidades cognitivas” necessárias para o Teste de Turing estão lá para permitir ações racionais. Então, precisamos da habilidade de representar conhecimento e raciocinar com ele, porque isso nos habilita a alcançar boas decisões em uma ampla variedade de situações. Precisamos ser capazes de gerar sentenças compreensíveis em linguagem natural, porque dizendo essas sentenças ajuda-nos a sobressairmo-nos numa sociedade complexa. Precisamos aprender não apenas para erudição, mas porque ter uma idéia melhor de como o mundo trabalha, habilitamo-nos a gerar estratégias mais efetivas para

lidar com ele. Precisamos da percepção visual não só porque ver é divertido, mas de maneira a obter uma idéia melhor de que ação pode ser alcançada, por exemplo, ser capaz de ver um petisco gostoso ajuda-nos a mover-nos em direção a ele.

O estudo da IA como o desenvolvimento de agentes racionais tem duas vantagens. Primeiro, ele é mais geral do que a abordagem das “leis do pensamento”, porque inferência correta é apenas um mecanismo útil para alcançar racionalidade, mas não é um mecanismo necessário. Segundo, ele é mais aprazível para o desenvolvimento científico do que as abordagens baseadas em comportamento ou pensamento humanos, porque o padrão da racionalidade é claramente definido e completamente geral. Comportamento humano, por outro lado, é bem adaptado para um ambiente específico e é o produto, em parte, de um amplo e complicado e processo evolucionário desconhecido que ainda pode estar longe de alcançar a perfeição. *Esse livro se concentrará portanto nos princípios gerais dos agentes racionais, e nos componentes para construí-los.* Veremos que a despeito da aparente simplicidade com que os problemas possam ser estabelecidos, uma enorme variedade de questões surgem quando tentamos resolvê-los. O Capítulo 2 destaca alguns desses problemas com mais detalhes.

Um ponto importante para se lembrar: veremos antes de mais nada que alcançar a perfeita racionalidade - sempre fazendo a coisa certa - não é possível em ambientes complicados. As demandas computacionais são altas demais. Entretanto, na maior parte deste livro, adotaremos a hipótese de trabalho de que tomar decisões com perfeito entendimento é um bom começo. Isso simplifica o problema e provê a configuração apropriada para a maioria do material fundamental na área. Os capítulos 5 e 17 lidam explicitamente com o problema da **racionalidade limitada** - agindo apropriadamente quando não há tempo suficiente para fazer todas computações que deveriam ser feitas.

## 1.2 OS FUNDAMENTOS DA INTELIGÊNCIA ARTIFICIAL

---

Nesta sessão e na próxima, proveremos um breve histórico da IA. Apesar da IA em si ser uma área jovem, tem herdado muitas idéias, pontos de vista, e técnicas de outras disciplinas. Após 2000 anos de tradição em Filosofia, teorias do raciocínio e da aprendizagem têm emergido com o ponto de vista de que a mente é constituída pelas operações de um sistema físico. Depois de 400 anos de Matemática, temos as teorias formais da lógica, probabilidade, tomadas de decisões, e computação. Da Psicologia temos as ferramentas com as quais investigar a mente humana e uma linguagem científica com a qual expressamos as teorias resultantes. Da Lingüística, temos teorias da estrutura e do significado das linguagens. Finalmente, da Ciência da Computação temos as ferramentas para tornar a IA uma realidade.

Como qualquer história, esta é forçada a se concentrar em um pequeno número de pessoas e eventos e ignorar outras que foram também importantes. Escolhemos organizar eventos para contar a história de como vários componentes intelectuais da IA moderna vieram a se formar. Certamente não desejaríamos dar a impressão, entretanto, de que as disciplinas partir das quais vieram os componentes, têm todas sido trabalhadas almejando a IA como seu último fruto.

### Filosofia (428 A.C. - dias presentes)

*“A mais segura caracterização da tradição filosófica européia é aquela que a considera uma série de rodapés das obras de Platão.” - Alfred North Whitehead*

Começamos com o nascimento de Platão em 428 A.C. Suas escritas enveredaram pela Política, Matemática, Física, Astronomia e vários ramos da Filosofia. Juntos, Platão, seu professor Sócrates e seu aluno Aristóteles estabeleceram os fundamentos para a maioria dos pensamentos e culturas ocidentais. O filósofo Hubert Dreyfus (1979, p.67) diz que “A história da inteligência artificial pode bem ter começado por volta de 450 A.C.” quando Platão reportou um diálogo em que Sócrates pergunta à Euthyphro<sup>3</sup>, “Quero saber o que é característico da piedade que faz todas as ações piedosas ... que eu possa tê-la para mudá-lo, e para usá-lo como um padrão por onde eu possa julgar suas ações e aquelas dos outros

---

<sup>3</sup> O Euthyphro descreve os eventos imediatamente anteriores ao julgamento de Sócrates em 399 A.C. Deyfus claramente errou em colocá-lo 51 anos mais cedo.

homens.”<sup>4</sup> Em outras palavras, Sócrates estava pedindo por um *algoritmo* para distinguir piedade de não piedade. Aristóteles seguiu tentando formular mais precisamente as leis que governam a parte racional da mente. Ele desenvolveu um sistema informal de silogismos para raciocínio próprio, que em princípio permite gerar mecanicamente conclusões, dadas premissas iniciais. Aristóteles não acreditava que todas as partes da mente eram governadas por processos lógicos; ele também tinha uma noção de razão intuitiva.

Agora que temos a idéia de um conjunto de regras que podem descrever o trabalho (de pelo menos uma parte) da mente, o próximo passo é considerar a mente como um sistema físico. Tivemos que esperar até René Descartes (1596-1650) para uma clara discussão sobre a distinção entre a mente e a matéria e os problemas surgidos a partir de então. Um problema com a concepção física pura da mente é que isso aparenta deixar pouco espaço para o livre arbítrio. Se a mente é governada inteiramente por leis físicas, então ela não tem mais livre arbítrio do que uma rocha decidindo cair através do centro da terra. Embora um forte defensor do poder da racionalidade, Descartes foi também um proponente do **dualismo**. Ele assegura que existe uma parte da mente (ou alma ou espírito) que está fora da natureza, isento das leis físicas. Por outro lado, ele sentiu que os animais não possuíam esta qualidade dualista; eles podiam ser considerados como se fossem máquinas.

Uma alternativa ao dualismo é o **materialismo**, que assegura que todo o mundo (incluindo o cérebro e a mente) operam de acordo com as leis físicas<sup>5</sup>. Wilhelm Leibniz (1646-1716) foi provavelmente o primeiro a assumir uma posição materialista em suas conclusões lógicas e construir um dispositivo mecânico destinado à resolver operações mentais. Infelizmente, sua formulação de Lógica era tão fraca que seu gerador de conceito mecânico não pode produzir resultados interessantes.

É também possível adotar uma posição intermediária, na qual aceita-se que a mente tem uma base física, mas nega que ela possa ser explicada por uma redução aos processos físicos ordinários. Processos mentais e consciência são portanto parte do mundo físico, mas inerentemente mal instruídos; eles estão além do entendimento racional. Alguns filósofos críticos da IA têm adotado exatamente esta posição, como discutiremos no Capítulo 26.

Bloqueando estas possíveis objeções para os objetivos da IA, a Filosofia tem então estabelecido uma tradição na qual a mente era concebida como um dispositivo físico operando principalmente pelo raciocínio com o conhecimento que ela continha. O próximo problema é então estabelecer a fonte do conhecimento. O movimento **empiricista**, começando com o *Novum Organum*<sup>6</sup> de Francis Bacon (1561-1626), é caracterizado pela máxima de John Locke (1632-1704): “Nada está no entendimento, que não tenha antes estado no sentido.” O “*A Treatise of Human Nature*” (Hume, 1978) de David Hume (1711-1776) propôs o que é hoje conhecido como o princípio da indução: que regras gerais são adquiridas pela exposição de associações repetidas entre seus elementos. Foi dada à teoria um delineamento mais formal por Bertrand Russell (1872-1970), que introduziu o **positivismo lógico**. Esta doutrina sustenta que todo conhecimento pode ser caracterizado por teorias lógicas conectadas, em última instância, a **sentenças de observação** que correspondem a entradas sensoriais<sup>7</sup>. A **teoria da confirmação** de Rudolf Carnap e Carl Hempel tentou estabelecer a natureza da conexão entre as sentenças de observação e as teorias mais gerais - em outras palavras, entender como o conhecimento pode ser adquirido pela experiência.

O elemento final na imagem filosófica da mente é a conexão entre conhecimento e ação. Que forma poderia ter esta conexão, e como ações poderiam ser justificadas? Essas questões são vitais para IA, porque apenas pelo entendimento de como ações são justificadas pode ser entendido como construir um agente cujas ações são justificáveis, ou racionais. Aristóteles dá uma elegante resposta em *Nicomachean Ethics* (Livro III.3, 1112b):

---

<sup>4</sup> Note que outras traduções possuem “bondade/bom” no lugar de “piedade/piedoso”.

<sup>5</sup> Nesse ponto de vista, a percepção do “livre arbítrio” surge porque a geração determinística de comportamento é constituída pela operação da mente selecionando entre o que aparenta ser possíveis cursos da ação. Eles permanecem “possíveis” porque o cérebro não tem acesso aos seus próprios estados futuros.

<sup>6</sup> Uma atualização do *organon* de Aristóteles, ou instrumento de pensamento.

<sup>7</sup> Neste quadro, todas as afirmações significativas podem ser verificadas ou falsificadas tanto pela análise do significado das palavras quanto pela realização de experimentos. Pelo fato de extrapolar a maioria das metafísicas, como foi a intenção, o positivismo lógico foi impopular em alguns círculos.

Deliberamos não sobre fins, mas sobre meios, pois um doutor não delibera se ele deve curar, nem um orador se ele deve persuadir, nem a um homem de estado se ele deve produzir lei e ordem, nem a qualquer outro deliberar sobre seu fim. Eles assumem o fim e consideram como e por quais meios ele pode ser atingido, e se é mais fácil e melhor produzido deste modo; enquanto se isto for atingido por um meio apenas, eles consideram como ele deverá ser atingido por este, e por quais meios este será atingido, até que eles cheguem à primeira causa, que é, em ordem de descoberta, a última ... e o que é a última na ordem de análise, parece ser a primeira em ordem de adequação. E se nós chegamos a uma impossibilidade, desistimos da busca, por exemplo se precisamos de dinheiro e este não pode ser conseguido; mas se uma coisa aparenta-se possível tentaremos fazê-lo.

A abordagem de Aristóteles (com alguns poucos refinamentos secundários) foi implementado 2300 anos mais tarde por Newell e Simon em seu programa GPS, sobre o qual eles escreveram (Newell e Simon, 1972):

Os principais métodos do GPS encorpam juntamente a heurística da análise meios-fins. A análise meios-fins é tipada pelo seguinte tipo de argumento de senso comum:

Quero levar meu filho à escola de enfermagem. Qual a diferença entre o que eu tenho e o que eu quero? A distância. O que muda a distância? Meu automóvel. Meu automóvel não quer funcionar. O que é necessário para fazê-lo funcionar? Uma nova bateria. O que tem novas baterias? Uma oficina de automóveis. Eu quero que a oficina coloque uma nova bateria; mas a oficina não sabe que eu preciso de uma. Qual a dificuldade? A comunicação. O que permite comunicação? Um telefone ... e assim vai.

Esse tipo de análise, classificando coisas em termos de funções que eles oferecem e oscilações entre fins, funções requeridas, e meios que as executam, forma o sistema básico da heurística do GPS.

A análise meios-fins é útil, mas não diz o que fazer quando várias ações alcançarão a meta, ou quando nenhuma ação a alcançará completamente. Arnauld, um seguidor de Descartes, corretamente descreveu uma fórmula quantitativa para decidir qual ação tomar em casos como este (ver Capítulo 6). O livro *Utilitarianism* (Mill, 1863) de John Stuart Mill (1806-1873) amplifica esta idéia. Uma teoria mais formal das decisões é discutida na seção seguinte.

## Matemática (c. 800 - dias presentes)

Os filósofos estabeleceram a maioria das idéias importantes da IA, mas para fazer o salto para uma ciência formal, era necessário um nível de formalização matemático em 3 áreas principais: computação, lógica e probabilidade. A noção para se expressar uma computação como um **algoritmo** formal remonta de al-Khowarazmi, um matemático árabe do século IX, cujos escritos também introduziram os algarismos arábicos e a álgebra na Europa.

A lógica remonta pelo menos até Aristóteles, mas era uma disciplina mais filosófica que matemática, até que George Boole (1815-1864) introduziu a sua linguagem formal para fazer inferências lógicas, em 1847. A abordagem de Boole era incompleta, mas suficientemente boa para que outros preenchessem as lacunas. Em 1879, Gottlob Frege (1848-1925) produziu uma lógica que, exceto por algumas mudanças notacionais, formam a lógica de primeira ordem, que é usada como o mais básico sistema de representação de conhecimento<sup>8</sup>. Alfred Tarski (1902-1983) introduziu uma teoria de referência que mostra como 'relacionar os objetos em uma lógica' com os objetos do mundo real. O próximo passo foi determinar os limites do que poderia ser feito com a lógica e com a computação.

David Hilbert (1862-1943), um grande matemático por si só, é mais lembrado pelos problemas que ele não resolveu. Em 1900, ele apresentou uma lista de 23 problemas que ele corretamente previu que ocupariam os matemáticos pelo próximo século. O problema final indagava se existe um algoritmo para decidir a verdade sobre qualquer proposição lógica envolvendo os números naturais - o famoso *Entscheidungsproblem*, ou problema da decisão. Essencialmente, Hilbert foi perguntado se existiam limites fundamentais para o poder dos procedimentos de prova efetiva. Em 1930, Kurt Gödel (1906-1978) mostrou que existe um procedimento efetivo para provar qualquer declaração verdadeira na lógica de primeira ordem de Frege e Russel; mas a lógica de primeira ordem não poderia capturar o princípio da indução matemática necessária para caracterizar os números naturais. Em 1931, ele mostrou que o limite

---

<sup>8</sup> Para entender porque a notação de Frege não foi aceita universalmente, veja a capa deste livro.

real verdadeiramente existe. O seu **teorema da incompletude** mostrou que em qualquer linguagem expressiva o suficiente para descrever as propriedades dos números naturais, existem declarações verdadeiras que são indecidíveis: suas verdades não podem ser estabelecidos por qualquer algoritmo.

Esse resultado fundamental também pode ser interpretado como prova de que há funções nos inteiros que não podem ser representadas por algoritmo – isto é, elas não podem ser computadas. Isto motivou Alan Turing (1912-1954) a tentar caracterizar exatamente que funções são capazes de ser computadas. Essa noção atualmente ligeiramente problemática, pois a noção de uma computação ou de um procedimento efetivo realmente não podem ter uma definição formal. Contudo, a tese de Church-Turing, que estabelece que a Máquina de Turing (Turing 1936) é capaz de computar qualquer função computável, é geralmente aceita como provedora de uma definição suficiente. Turing também mostrou que existiam algumas funções que nenhuma Máquina de Turing poderia computar. Por exemplo, nenhuma máquina pode dizer *em geral* se um dado programa dará resposta para uma certa entrada ou se ele rodará indefinidamente.

Apesar da indecidibilidade e da não-computabilidade serem importantes para o entendimento da computação, a noção de **intratabilidade** tem um impacto bem maior. Falando grosseiramente, uma classe de problemas é chamada intratável se o tempo requerido para solucionar instâncias da classe cresce no mínimo exponencialmente com o tamanho das instâncias. A distinção entre o crescimento polinomial e exponencial na complexidade foi enfatizada pela primeira vez em meados dos anos 60 (Cobham, 1964; Edmonds, 1965). Isso é importante porque o crescimento exponencial implica que mesmo instâncias de tamanho moderado não podem ser solucionadas em tempo razoável. Entretanto, deve-se tentar dividir o problema de gerar um comportamento inteligente em subproblemas tratáveis como um todo, no lugar de problemas intratáveis. O segundo conceito importante na teoria da complexidade é a **redução**, que também surgiu nos anos 60 (Dantzig, 1960; Edmonds, 1962). Uma redução é uma transformação geral de uma classe de problemas ( $1^a$ ) para outra ( $2^a$ ), onde a solução para os problemas da primeira classe pode ser achada reduzindo-os para problemas da segunda classe e solucionando estes últimos problemas.

Como reconhecer um problema intratável? A Teoria da NP-completude, criado por Steven Cook (1971) e seguida por Richard Karp (1972), provê um método. Cook e Karp mostraram a existência de grandes classes de problemas de raciocínio canônicos de raciocínio e de busca combinatorial que são NP-completos.

Qualquer classe de problemas para o qual um problema NP-completo pode ser reduzido é igualmente intratável. (Embora ainda não tenha sido provado que problemas NP-completos sejam necessariamente intratáveis, poucos teóricos acreditam no contrário). Esses resultados contrastam severamente com o entusiasmo de alguns pela criação das "Super-Cérebros Eletrônicos" surgido com o advento dos computadores. Apesar da sempre crescente velocidade dos computadores, o uso cuidadoso e sutil dos recursos caracterizarão os sistemas inteligentes. Colocando de forma crua, o mundo é uma instância de problema *extremamente* grande!

Ao lado da lógica e da computação, a terceira grande contribuição dos matemáticos para a IA é a teoria da probabilidade. O italiano Gerolamo Cardano (1501-1576) traçou inicialmente a idéia da probabilidade, descrevendo-a em termos dos possíveis resultados dos jogos de azar. Antes dele, os resultados dos jogos eram vistos como a vontade dos deuses mais do que um capricho da sorte. A probabilidade rapidamente tornou-se parte válida de todas as ciências quantitativas, ajudando a lidar com medições incorretas e teorias incompletas. Pierre Fermat (1601-1665), Blaise Pascal (1623-1662), James Bernoulli (1654-1705), Pierre Laplace (1749- 1827) e outros avançaram a teoria e introduziram novos métodos estatísticos. Bernoulli também traçou uma visão alternativa de probabilidade, com o nome de "grau de confiança" ao invés de um resultado de razão objetiva. As probabilidades subjetivas, entretanto, podem ser atualizadas a medida que novas evidências são obtidas. Thomas Bayes (1702-1761) propôs uma regra para a atualização de probabilidades subjetivas à luz de novas evidências (publicado postumamente em 1763). A regra de Bayes, e o subsequente campo das análises bayesianas, formam as bases da moderna abordagem dos raciocínios incertos em sistemas de IA. Existem ainda debates entre os que apoiam as visões subjetiva e objetiva da probabilidade, mas não é claro se a diferença entre ambas tem grande significação para a IA. Ambas as versões obedecem ao mesmo conjunto axiomático. Os *Fundamentos de Estatística* de Savage (1954) dá uma boa introdução ao campo.



Assim como com a lógica, uma conexão deve ser estabelecida entre raciocínio probabilístico e ação. A **Teoria da Decisão**, desenvolvida por John von Neumann e Oskar Morgenstern (1944), combina a teoria da probabilidade com a teoria da utilidade (que dá a prova formal e completa para a especificação das preferências de um agente) para dar a primeira teoria geral que pode distinguir boas ações de más ações. A teoria da decisão é o sucessor matemático do utilitarismo, e fornece as bases teóricas para muitos dos projetos de agentes neste livro.

## Psicologia (1879 - dias presentes)

Pode-se dizer que a psicologia científica começou com o trabalho do físico alemão Hermann von Helmholtz (1821-1894) e seu aluno Wilhelm Wundt (1832-1920). Helmholtz utilizou os métodos científicos para estudar a visão humana e o seu *Manual de Fisiologia Ótica* é ainda hoje descrita como "o mais singular e importante tratado sobre física e fisiologia da visão humana" (Nalwa, 1993, p.15). Em 1879, o mesmo ano em que Frege lançou a lógica de primeira ordem, Wundt abriu o primeiro laboratório de psicologia experimental na Universidade de Leipzig. Wundt, ao contrário, insistia em experimentos cuidadosamente controlados, onde seus trabalhadores desempenhariam uma tarefa preceptiva ou associativa enquanto realizariam uma introspecção em seus processos mentais. O controle cuidadoso trilhou um grande caminho para fazer da psicologia uma ciência, mas, ao passo que a metodologia se expandia, um curioso fenômeno surgiu: cada laboratório reportaria dados introspectivos que apenas ocorreriam para corroborar as teorias que eram populares naquele laboratório. O movimento **comportamentalista** de John Watson (1878-1958) e de Edward Lee Thorndike (1874-1949) rebelou-se contra este subjetivismo, rejeitando qualquer teoria envolvendo processos mentais baseado no fato que a introspecção não poderia fornecer evidências confiáveis. O comportamentalismo insistia em estudar apenas medidas objetivas da percepção (ou estímulo) dada a um animal e suas ações resultantes (ou respostas). Construções mentais tais como conhecimento, crenças, metas e passos de raciocínio foram disseminadas como "psicologia popular" não-científica. O comportamentalismo descobriu muito sobre ratos e pombos, mas teve pouco sucesso ao tentar entender humanos. Não obstante, teve uma grande força na psicologia (especialmente nos Estados Unidos) entre 1920 e 1960.

A visão de que o cérebro possui e processa a informação, que é a principal característica da **psicologia cognitiva**, pode ser trilhada até os trabalhos de William James<sup>9</sup> (1842–1910). Helmholtz também insistia que a percepção envolvia uma forma de inferência lógica inconsciente. O ponto de vista cognitivo era largamente obscurecido pelo comportamentalismo até 1943, quando Kenneth Craik publicou *The Nature of Explanation*. Craik recolocou o passo mental esquecido entre o estímulo e a resposta. Ele afirma que crenças, metas e passos de raciocínio poderiam ser úteis como componentes válidos de uma teoria do comportamento humano, e são apenas tão científicos como, digamos, a utilização de pressão e temperatura para se falar sobre gases, a despeito de serem compostos de moléculas que não possuem nem uma coisa nem outra. Craik especificou os três passos chave de um agente baseado em conhecimento: (1) o estímulo deve ser traduzido em uma representação interna; (2) esta representação é manipulada por processos cognitivos para resultar em novas representações internas; (3) essas são, por sua vez, retraduzidas em ação. Ele claramente explicou porque isso era uma boa descrição para um agente:

*"Se o organismo carrega um 'modelo em escala reduzida' da realidade externa e de suas próprias ações possíveis dentro de sua cabeça, ele é capaz de tentar várias alternativas, concluir qual a melhor delas, reagir a situações futuras antes que elas surjam, utilizar o conhecimento de eventos passados para lidar com o presente e o futuro, e em todas as formas ele reage de maneira mais completa, segura e competente as emergências que ele venha a se deparar."* (Craik, 1943)

Um agente descrito desta forma pode, por exemplo, planejar uma longa viagem considerando várias rotas possíveis, comparando-as, e escolhendo a melhor, tudo antes de iniciar a jornada. Desde os anos 60, a idéia do processamento da informação tem dominado a psicologia. É agora quase tido como certo entre muitos psicólogos que 'uma teoria cognitiva deveria ser como um programa de computador. Por isso entende-se que a teoria deveria descrever cognição como um processo consistindo de transformações bem definidas operando num nível de informação levados pelos sinais de entrada.

---

<sup>9</sup> William James era o irmão do novelista Henry James. É dito que Henry escreveu ficção como se fosse psicologia e William escrevia psicologia como se fosse ficção.

Na maior parte da história recente da IA e da ciência cognitiva, nenhuma distinção significativa foi estabelecida entre os dois campos, e era comum ver programas de IA descritos como resultados psicológicos sem qualquer alusão quanto ao exato comportamento humano que eles estavam modelando. Na última década, entretanto, as distinções metodológicas têm se tornado cada vez mais claras, e a maioria dos trabalhos agora recaem em uma área ou na outra.

### **Engenharia da computação (1940 - dias presentes)**

Para a IA ter sucesso, precisamos de duas coisas: inteligência e um artefato. O computador foi unanimemente aclamado como o artefato com melhores chances de demonstrar inteligência. O computador eletrônico digital moderno foi inventado independentemente e quase simultaneamente por cientistas de três países envolvidos na Segunda Guerra Mundial. O primeiro computador operacional moderno foi o 'Heath Robinson'<sup>10</sup>, construído em 1940 pela equipe de Alan Turing, com o simples propósito de decifrar mensagens alemãs. Quando os alemães mudaram para um código mais sofisticado, a eletromecânica existente no 'Robinson' demonstrou ser demasiadamente lenta, e uma nova máquina chamada de 'Colossus' foi construída a partir de tubos de vácuo. Ele foi completado em 1943, e, na época do final da Guerra, 10 máquinas 'Colossus' estavam em uso diário.

O primeiro computador programável operacional foi o Z-3, inventado por Konrad Zuse, na Alemanha, em 1941. Zuse inventou números de ponto flutuante para o Z-3, e prosseguiu até o desenvolvimento, em 1945, do Plankalkul, a primeira linguagem de programação de alto nível. Embora Zuse tenha recebido algum apoio do terceiro Reich para aplicar a sua máquina ao projeto de aeronaves, a hierarquia militar não deu tanta importância para a computação assim como a Inglaterra havia feito.

Nos Estados Unidos, o primeiro computador eletrônico, o ABC, foi criado por John Atanasoff e o seu aluno de graduação Clifford Berry entre 1940 e 1942 na Universidade do Estado de Iowa. O projeto recebeu pouco apoio e foi abandonado depois que Atanasoff envolveu-se em pesquisa militar em Washington. Dois outros projetos de computadores foram iniciados como pesquisa militar secreta: o Mark I, II e III foram desenvolvidos em Harvard por uma equipe dirigida por Howard Aiken; e o ENIAC foi desenvolvido na Universidade da Pennsylvania por uma equipe que incluía John Mauchly e John Eckert. O ENIAC foi o primeiro computador de propósito geral, eletrônico e digital. Uma das suas primeiras aplicações foi computar tabelas de tiros de artilharia. Um sucessor, o EDVAC, que recebeu sugestões de Von Neumann para utilizar um programa armazenado, de maneira que os técnicos não teriam que 'tropeçar nos cabos' na tentativa de rodar um novo programa.

Mas, talvez, o avanço mais crítico foi o IBM 701, desenvolvido em 1952 por Nathaniel Rochester e seu grupo. Este foi o primeiro computador a render lucros ao seu fabricante. A IBM se tornou uma das maiores corporações do mundo, e a venda de computadores cresceu para 150 bilhões de dólares por ano. Nos Estados Unidos, a indústria de computadores (incluindo a de software e a de serviços) agora contabilizam quase 10% do produto interno bruto.

Cada geração de hardware de computador trouxe um aumento de velocidade e de capacidade, e uma redução de preço. A engenharia da computação tem alcançado um sucesso notável, dobrando regularmente o seu desempenho a cada dois anos, sem nenhum final imediato em vista desta taxa de crescimento. máquinas massivamente paralelas prometem adicionar vários zeros ao poder de processamento alcançável.

Certamente existiam dispositivos de cálculo antes dos computadores eletrônicos. O ábaco tem quase 7.000 anos de criação. Na metade do século XVII, Blaise Pascal construiu uma máquina mecânica de adicionar e subtrair chamada 'Pascaline'. Leibniz a aperfeiçoou em 1694, construindo um dispositivo mecânico que multiplicava fazendo repetidas adições. O progresso estagnou por mais de um século, até que Charles Babbage (1792-1871) sonhou que as tabelas logarítmicas poderiam ser computadas por máquinas. Ele projetou uma máquina para essa tarefa, mas nunca completou o projeto. Em vez disso, ele mudou para o projeto de uma máquina analítica, para a qual Babbage inventou as idéias de memória endereçável, programas armazenados e desvios condicionais. Embora a idéia de máquinas programáveis

---

<sup>10</sup> Heath Robinson foi um cartunista famoso por seus desenhos sobre as contraposições absurdamente complicadas das tarefas diárias como a de passar manteiga em uma torrada.

não ser nova - em 1805, Joseph Marie Jacquard inventou um tear que podia ser programado utilizando cartões perfurados - a máquina de Babbage era o primeiro artefato possuindo as características necessárias para a computação universal. Uma colega de Babbage, Ada Lovelace, filha do poeta Lord Byron, escreveu programas para a máquina analítica e até mesmo especulou que a máquina poderia jogar xadrez ou compor música. Lovelace foi a primeira programadora do mundo e a primeira de muitas a desistir em virtude de custos excessivos e a ter um ambicioso projeto inovador abandonado<sup>11</sup>. O projeto básico de Babbage foi provado ser viável por Doron Swade e seus colegas, que construíram um modelo operacional utilizando apenas as técnicas mecânicas disponíveis na época de Babbage (Swade, 1993). Babbage teve a idéia correta, mas lhe faltou habilidades organizacionais para ter a sua máquina construída.

IA também tem um débito com a área de software da ciência da computação, que tem suprido os sistemas operacionais, linguagens de programação e ferramentas necessárias para escrever programas modernos (e trabalhos sobre eles). Mas essa é uma área onde as dívidas têm sido saldadas: o trabalho de IA tem sido pioneiro de muitas idéias que têm feito seu caminho de volta para a corrente principal da ciência da computação, incluindo tempo compartilhado, interpretadores interativos, tipo de dados de listas ligadas, gerenciamento da memória automática e alguns dos conceitos chave da programação orientados a objeto e dos ambientes de desenvolvimento de programas integrados com interfaces gráficas de usuário.

### Linguística (1957 - dias presentes)

Em 1957, B. F. Skinner publicou *Verbal Behaviour*. Isso foi um relato detalhado e compreensivo da abordagem comportamental para a aprendizagem da linguagem, escrito pelo mais destacado especialista na área. Mas, curiosamente, uma revisão do livro tornou-se bem conhecida como o livro em si e serviu para quase eliminar o interesse no behaviorismo. O autor da revisão foi Noam Chomsky, que tinha apenas publicado um livro sobre sua própria teoria, *Syntactic Structures*. Chomsky mostrou como a teoria comportamental não enfocava a noção de criatividade da linguagem - não explicou como uma criança poderia entender e compor as sentenças que ele ou ela nunca tinham ouvido antes. A teoria de Chomsky - baseada nos modelos sintáticos que remontam ao linguística indiano Panini - poderia explicar isso e, ao contrário das teorias anteriores, foi suficientemente formal para ser, a princípio, programada.

Os desenvolvimentos posteriores da linguística mostraram que o problema era consideravelmente mais complexo do que pareceu em 1957. A linguagem é ambígua e deixa muito por dizer. Isso significa que o entendimento da linguagem requer uma compreensão do assunto e do contexto, não apenas uma compreensão da estrutura das sentenças. Isso pode parecer óbvio, mas não foi apreciado até o início dos anos 60. Muito dos trabalhos iniciais em **representação do conhecimento** (o estudo de como colocar o conhecimento numa forma que um computador possa raciocinar sobre ele) estava ligado à linguagem e era ensinado por pesquisadores em linguística, engajados a décadas na análise filosófica da linguagem.

A linguística moderna e a IA “nasceram” quase ao mesmo tempo, por isso os linguistas não exercem um papel fundamental no desenvolvimento de IA. Em vez disso, os dois cresceram juntos, intersectando-se num campo híbrido chamado **linguística computacional** ou **processamento de linguagem natural**, que concentra-se no problema do uso da linguagem.

## 1.3 A HISTÓRIA DA INTELIGÊNCIA ARTIFICIAL

---

Com o material histórico nos apoiando, estamos prontos agora para esboçar corretamente o desenvolvimento da IA. Poderíamos fazer isso identificando fases sobrepostas e definidas livremente em seu desenvolvimento, ou registrando cronologicamente os vários enlaces conceituais que compõem a área. Nesta sessão, tomaremos a última abordagem, correndo o risco de cometer algum grau de violência contra as reais relações entre suas subáreas. A história de cada subárea é coberta em capítulos individuais mais afrente neste livro.

---

<sup>11</sup> Ela também deu seu nome para Ada, uma linguagem de programação de propósito geral do Departamento de Defesa dos Estados Unidos

## A gestão da Inteligência Artificial (1943-1956)

O primeiro trabalho que agora é geralmente reconhecido como IA foi feito por Warren McCulloch e Walter Pitts (1943). Eles recorreram a 3 fontes: conhecimento da fisiologia básica e função dos neurônios no cérebro; a análise formal da lógica proposicional de Russel e Whitehead e a teoria da computação de Turing. Eles propuseram um modelo de neurônios artificiais em que cada neurônio é caracterizado como estando “ligado” ou “desligado”, com uma chave em “ligado” em resposta ao estímulo dado por um número suficiente de neurônios vizinhos. O estado de um neurônio foi concebido como “factualmente equivalente a uma proposição que propôs seu estímulo adequado”. Eles mostraram, por exemplo, que qualquer função computável poderia ser computada por alguma rede de neurônios conectados, e que todos os conectivos lógicos poderiam ser implementados por estruturas de rede simples. McCulloch e Pitts também sugeriram que as redes definidas apropriadamente poderiam aprender. Donald Hebb (1949) demonstrou uma regra simples de atualização para modificar as forças de conexão entre os neurônios, de forma que a aprendizagem pudesse ocorrer.

O trabalho de McCulloch e Pitts foi provavelmente o precursor tanto da tradição logística em IA como da tradição conexionista. No início dos anos 50, Claude Shannon (1950) e Alan Turing (1953) estavam escrevendo programas de xadrez para os computadores convencionais do estilo von Neuman<sup>12</sup>. Ao mesmo tempo, dois estudantes graduados do departamento de matemática de Princeton, Marvin Minsky e Dean Edmonds, criaram o primeiro computador de rede neural em 1951. O SNARC, como foi chamado, usava 3000 tubos de vácuo e um mecanismo sobressalente de piloto automático de um bombardeio B-24 para simular uma rede de 40 neurônios. O comitê de PhD de Minsky foi céptico se esse tipo de trabalho deveria ser considerado matemática, mas von Neumann estava no comitê e disse enfaticamente, “se não é agora, será algum dia”. Ironicamente, Minsky mais tarde provaria teoremas que contribuiriam para o ocaso de muitas das pesquisas sobre redes neurais durante os anos 70.

Princeton foi o lar de outra figura influente da IA, John McCarthy. Depois da graduação, McCarthy mudou-se para o Dartmouth College, que estava para tornar-se o local de nascimento oficial da área. McCarthy convenceu Minsky, Claude Shannon e Nathaniel Rochester a ajudá-lo a reunir pesquisadores americanos interessados na teoria dos autômatos, redes neurais, e no estudo da inteligência. Eles organizaram um Workshop de dois meses em Dartmouth, durante o verão de 1956. No todo, compareceram dez estudiosos, incluindo Trenchard More, de Princeton, Arthur Samuel, da IBM e Ray Solomonoff e Oliver Selfridge, do MIT.

Dois pesquisadores da Carnegie Tech<sup>13</sup>, Allen Newell e Herbert Simon, quase “roubaram a cena”. Embora os outros tivessem idéias e em alguns casos programas para algumas aplicações particulares para jogos de tabuleiro, Newell e Simon já tinham um programa de raciocínio, o Logist Theorist (LT), sobre o qual Simon aclamava, “nós inventamos um programa de computador capaz de pensar numericamente, e desse modo, resolvemos o venerável problema da mente-corpo”<sup>14</sup>. Logo após o Workshop, o programa foi capaz de provar a maioria dos teoremas do capítulo 2 do livro *Principia Mathematica* de Russel e Whitehead. Russel ficou encantado quando Simon mostrou que o programa tinha sugerido uma prova para um dos teoremas que era menor que o da *Principia*. Os editores do jornal *Journal of Symbolic Logic* ficaram menos impressionados; eles rejeitaram um artigo de co-autoria de Newell, Simon e Logic Theorist.

O Workshop de Dartmouth não levou a nenhum acontecimento, mas introduziu todas as principais figuras umas às outras. Pelos próximos 20 anos, a área seria dominada por estas pessoas, seus alunos e colegas do MIT, CMU, Stanford, e IBM. Talvez a coisa mais durável surgida do workshop foi o acordo para adotar o nome dado por McCarthy para a área: **Inteligência Artificial**.

---

<sup>12</sup> Shannon não tinha na realidade nenhum computador para trabalhar, e Turing teve o seu acesso eventualmente negado aos computadores de sua própria equipe pelo governo britânico, com a justificativa de que a pesquisa em inteligência artificial era certamente frívola.

<sup>13</sup> Hoje, Carnegie Mellon University (CMU).

<sup>14</sup> Newell e Simon também inventaram uma linguagem de processamento de listas, IPL, para escrever o LT. Eles não tinham um compilador, e traduziram o código na máquina manualmente. Para evitar erros, eles trabalhavam em paralelo, chamando números binários de um para outro para escrever as instruções para certificarem que eles estavam de acordo.

## Entusiasmo precoce, grandes expectativas (1952-1969)

Os primeiros anos da IA foram repletos de sucessos – de maneira limitada. Considerando os computadores primitivos e as ferramentas de programação da época, e o fato que apenas poucos anos mais cedo computadores eram vistos como objetos que podiam fazer cálculos aritméticos e nada mais, era espantoso sempre que um computador realizava algo remotamente inteligente. O establishment intelectual preferiu acreditar que “uma máquina nunca poderia fazer X” (veja no capítulo 26 uma longa lista de Xs elencados por Turing). Pesquisadores de IA naturalmente responderam demonstrando um X após o outro. Alguns pesquisadores modernos de IA referem-se a este período a era “Olhe, mamãe, sem as mãos!”

O precoce sucesso de Newell e Simon foi seguido pelo Resolvedor Geral de Problemas (GPS). Diferentemente do Logist Theorist, este programa foi projetado desde o começo para imitar os protocolos de resolução de problemas humanos. Dentre as classes limitadas de quebra-cabeças que ele poderia manipular, ele obteve como resultado a informação que a ordem em que o programa considerava submetas e possíveis ações era similar à maneira como os humanos abordavam o mesmo problema. Daí, o GPS foi provavelmente o primeiro programa a personificar o pensamento humano. A combinação da IA com a ciência cognitiva tem continuado na CMU até os dias atuais.

Na IBM, Nathaniel Rochester e seus colegas produziram alguns dos primeiros programas da IA. Herbert Gelernter (1959) construiu o Geometry Theorem Prover. Assim como o Logic Theorist, ele provou teoremas usando axiomas representados explicitamente. Gelernter descobriria em breve que existiam muitos caminhos de raciocínio possíveis a seguir, sendo que a maioria deles fracassariam. Para ajudar o foco de busca, ele adicionou a capacidade de criar um diagrama de representação numérica - um caso particular do teorema geral a ser provado. Antes do programa tentar provar algo, ele poderia primeiro checar o diagrama para ver se o caso particular era verdadeiro.

A partir de 1952, Arthur Samuel escreveu uma série de programas para jogos de tabuleiros que eventualmente aprenderam a jogar xadrez em nível competitivo. No seu caminho, ele contestou a idéia de que computadores apenas faziam o que lhes eram mandados, enquanto seus programas aprendiam rapidamente a jogar melhor que seu criador. O programa foi exibido na televisão em fevereiro de 1956, criando um forte impressão. Assim como Turing, Samuel teve problemas com o tempo de computação. Trabalhando a noite, ele utilizou máquinas que estavam ainda em testes no setor de manufatura da IBM. O Capítulo 5 cobre jogos e o Capítulo 20 descreve e expande as técnicas de aprendizagem utilizadas por Samuel.

John McCarthy mudou-se de Dartmouth para o MIT e lá produziu três contribuições cruciais em um ano histórico, 1958: no laboratório memorando nº 1 do laboratório de IA do MIT, McCarthy definiu uma linguagem de alto nível denominada **Lisp**, que tornar-se-ia a linguagem de programação dominante para IA. Lisp é a segunda mais velha linguagem em uso corrente<sup>15</sup>. Com Lisp, McCarthy tinha a ferramenta que ele precisava, mas o acesso escasso, caro e os recursos computacionais eram também um sério problema. Assim, ele e outros do MIT, inventaram o tempo compartilhado. Depois de obter um sistema experimental de tempo compartilhado no MIT, McCarthy eventualmente atraiu o interesse do grupo de graduação do MIT, que formou a Digital Equipment Corporation (DEC), que se tornou a segunda maior indústria de computadores do mundo, graças aos minicomputadores de tempo compartilhado. Também em 1958, McCarthy publicou um artigo intitulado *Programs of Common Sense*, em que ele descreveu o Advice Taker, um programa hipotético que pode ser visto como o primeiro programa de IA completo. Como o Logic Theorist e o Geometry Theorem Prover, o programa de McCarthy foi projetado para utilizar conhecimento para buscar soluções para problemas. Mas, diferente dos outros, ele podia incorporar conhecimentos gerais do mundo. Por exemplo, ele mostrou como axiomas simples habilitariam o programa a gerar um plano para dirigir até um aeroporto para tomar um avião. O programa foi também modelado para que pudesse receber novos axiomas em curso normal de operação, permitindo com isso alcançar a competência em novas áreas sem a necessidade de reprogramação. O Advice Taker incorporava então os princípios centrais da representação de conhecimento e de raciocínio: isto é útil para se ter uma representação explícita formal do mundo e da maneira com que as ações de um agente afetam o mundo, e para ser capaz de manipular estas representações com processos dedutivos. É notável como os artigos de 1958 permaneceram relevantes por

---

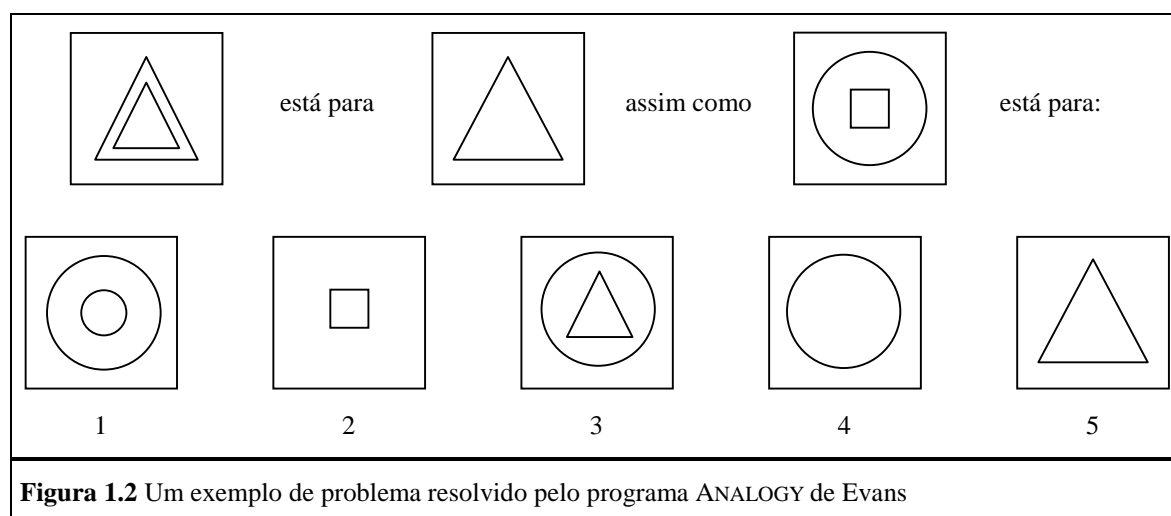
<sup>15</sup> Fortran é 1 ano mais velha que Lisp.

mais de 35 anos.

1958 também marca o ano em que Marvin Minsky mudou-se para o MIT. Por anos ele e McCarthy foram inseparáveis enquanto juntos definiam a área. Mas eles se distanciaram cada vez mais a medida que McCarthy aprofundava-se na representação e raciocínio na lógica formal, enquanto Minsky estava mais interessado em fazer os programas funcionarem e, eventualmente, desenvolveu um sentimento anti-lógico. Em 1963, McCarthy aceitou a oportunidade de ir para Stanford e lá iniciar um laboratório de IA. Sua agenda de pesquisa de utilizar a lógica para construir o Advice Taker mais recente foi alavancada pela descoberta de J. A. Robinson do método da resolução (um algoritmo de prova de teoremas completo por a lógica de primeira ordem; veja Sessão 9.6). O trabalho em Stanford enfatizava métodos de propósitos gerais para raciocínio lógico. Aplicações de lógica incluíam a resposta ao problema de Cordell Green e sistemas de planejamento (Green, 1969b), e o projeto de robótica de Shakey no novo instituto de pesquisa de Stanford (SRI). O último projeto, discutido mais tarde no Capítulo 25, foi o primeiro a demonstrar a completa integração do raciocínio lógico e da atividade física.

Minsky dirigiu uma série de estudantes que escolheram problemas limitados que aparentemente requeriam inteligência para serem resolvidos. Esses domínios limitados tornaram-se conhecidos como **micro-mundos**. O programa SAINT de James Slagle (1963a) foi capaz de resolver problemas de integração definida, típicos dos cursos de cálculo do primeiro ano universitário. O programa ANALOGY de Tom Evans (1968) resolveu problemas de analogia geométrica que apareciam em testes de QI, como o da Figura 1.2. O SIR (Recuperação de Informação Semântica) de Bertram Raphael (1968) era capaz de aceitar declarações de entrada em um conjunto muito restrito do inglês e responder a questões referentes a elas. O programa STUDENT de Daniel Bobrow (1967) resolveu problemas históricos de álgebra como

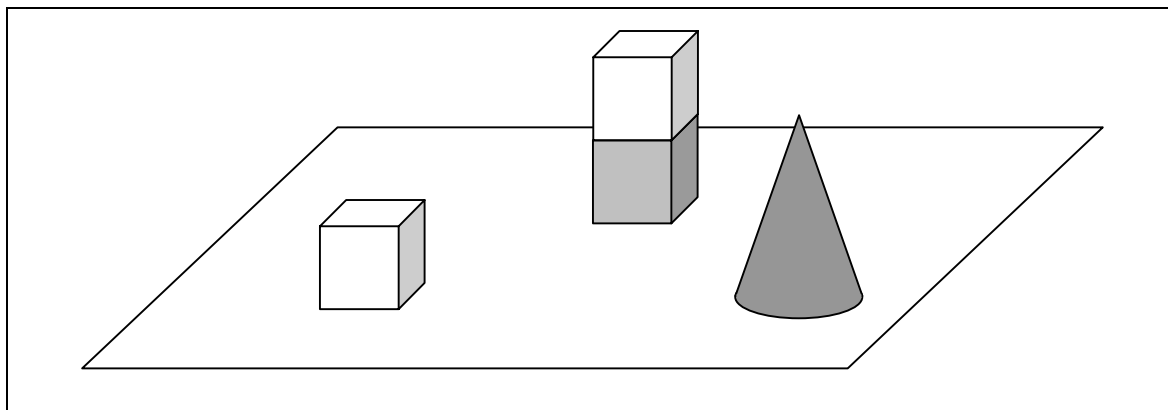
Se o número de clientes que Tom consegue é o dobro do quadrado de 20% do número de anúncios que ele coloca, e o número de anúncios que ele coloca é 45, qual o número de clientes que Tom conseguiu?



O mais famoso micro-mundo foi o mundo de blocos, que consiste de um conjunto de blocos sólidos colocados em um tabuleiro (ou mais frequentemente, uma simulação de um tabuleiro), como mostrada na Figura 1.3. Uma tarefa neste mundo é reorganizar os blocos de certa maneira, usando a mão de um robô que pode pegar um bloco de cada vez. O mundo dos blocos foi o início para o projeto de visão de David Huffman (1971), o trabalho de visão e de propagação de restrições de David Waltz (1975), a teoria de aprendizagem de Patrick Winston (1970), o programa de entendimento da linguagem natural de Terry Winograd (1972) e o planejador de Scott Fahlman (1974).

Os trabalhos sobre as redes neurais de McCulloch e Pitts também prosperaram. O trabalho de Winograd e Cowan (1963) mostrou como um grande número de elementos poderiam representar coletivamente um conceito individual, com um correspondente acréscimo em robustez e paralelismo. Os métodos de aprendizagem de Hebb foram melhorados por Bernie Widrow (Widrow e Hoff, 1960; Widrow, 1962), que chamou suas redes de **adalines**, e por Frank Rosenblatt (1962) com seus **perceptrons**. Rosenblatt provou o famoso teorema da **convergência do perceptron**, mostrando que seu

algoritmo de aprendizagem poderia ajustar a força das conexões de um perceptron para se alinhar com qualquer dado de entrada, desde que isso seja possível. Esses tópicos são cobertos na Seção 19.3.



**Figura 1.3** Uma cena do mundo dos blocos. Uma tarefa para o robô poderia ser "pegar um grande bloco vermelho", expresso em linguagem natural ou em uma notação formal.

### Uma dose de realidade (1966-1974)

Desde o início, pesquisadores de IA não foram modestos em fazer previsões de seus futuros sucessos. O seguinte trecho de Herbert Simon em 1957 é frequentemente citado:

Não é minha intenção surpreender ou chocar vocês — mas a forma mais simples que eu posso resumir é dizer que existem agora no mundo máquinas que pensam, que aprendem e que criam. Mais ainda, suas habilidades para fazer estas coisas estão sendo ampliadas rapidamente até — num futuro visível — o espectro de problemas que elas possam manipular será coextensivo com o espectro em que a mente humana tem sido aplicada.

Embora alguém possa questionar que termos como “futuro visível” possam ser interpretados de várias maneiras, algumas das predições de Simon foram mais concretas. Em 1958, ele previu que em 10 anos um computador poderia ser um campeão de xadrez, e um importante novo teorema matemático poderia ser provado pela máquina. Asserções como essas foram tomadas como desmesuradamente otimistas. A barreira que quase todos os projetos de pesquisa de IA enfrentaram foi que os métodos que eram suficiente para demonstrações em um ou dois exemplos simples falharam miseravelmente quando testados em seluções de problemas mais amplos e em problemas mais difíceis.

O primeiro tipo de dificuldade surgiu por que os primeiros programas frequentemente continham poucos ou nenhum conhecimento do assunto em questão, e tiveram sucesso por meio de simples manipulações sintáticas. O programa ELIZA de Weizenbaum (1965), que podia aparentemente angajar-se em conversações de alto nível em qualquer assunto, na verdade apenas emprestavam e manipulavam as sentenças digitadas nele por um humano. Uma história típica ocorreu durante os primeiros esforços para o desenvolvimento de uma máquina tradutora, que foi generosamente fundada pelo NRC (*National Research Council*) em uma tentativa de acelerar a tradução de artigos científicos russos na época do lançamento do Sputnik, em 1957. Pensava-se inicialmente que transformações sintáticas simples baseadas nas gramáticas russas e inglesas, e a substituição de palavras usando um dicionário eletrônico, seria suficiente para preservar o significado exato das sentenças. De fato, tradução requer conhecimento geral do assunto em questão para resolver ambigüidades e estabelecer o conteúdo de uma sentença. A famosa retradução de “o espírito é forte mas a carne é fraca” como “a vodka é boa mas a carne é estragada” ilustra as dificuldades encontradas. Em 1966, um relatório de um comitê de conselhos descobriu que “não existe máquina de tradução de textos científicos genéricos, e nenhuma está prevista a curto prazo.” Todos os fundos do governo dos Estados Unidos para projetos acadêmicos de tradução foram cancelados.

O segundo tipo de dificuldade foi a intratabilidade de muitos problemas que a Inteligência Artificial tentou resolver. A maioria dos programas iniciais da IA trabalhavam representando os fatos básicos sobre um problema e tentando uma série de passos para resolvê-lo, tentando diferentes combinações de

passos até a solução correta ser encontrada. Os primeiros programas só eram viáveis porque os micro-mundos continham pouquíssimos objetos. Antes da teoria da NP-completude ser desenvolvida, foi amplamente pensado que a escalada para problemas maiores, era simplesmente uma questão de hardware mais rápido e maiores quantidades de memória. O otimismo que acompanhava o desenvolvimento da prova do teorema da resolução, por exemplo, foi imediatamente descartada quando os pesquisadores falharam em provar teoremas envolvendo mais que poucas dúzias de fatos. *O fato de um programa poder encontrar uma solução em princípio não significa que o programa contém qualquer alguns dos mecanismos necessários para encontrá-lo na prática.*

A ilusão do poder computacional ilimitado não se confinou a programas de resolução de problemas. Experimentos iniciais em **evolução de máquina** (agora chamado **algoritmos genéticos**) basearam-se na crença correta e indubitável de que através de séries apropriadas de pequenas mutações para um código de programa de máquina, pode-se gerar um programa com bom desempenho para qualquer tarefa simples particular. A idéia, então, era tentar mutações aleatórias e então aplicar um processo de seleção para preservar mutações que pareciam melhorar comportamentos. Apesar de milhares de horas de tempo da CPU, quase nenhum progresso foi demonstrado.

Os consideráveis fracassos em virtude da explosão combinacional foi uma das principais críticas da IA contida no relatório de Lighthill (Lighthill, 1973), que formou a base para a decisão do governo britânico de suspender o apoio a todas as pesquisas em IA, menos em duas universidades. (a tradição oral pinta um quadro de certa forma diferente e mais colorida, com ambições políticas e animosidades pessoais que não podem ser impressas.)

Uma terceira dificuldade surgiu por causa de algumas limitações fundamentais na estrutura básica usada para gerar conhecimento inteligente. Por exemplo, em 1969, o livro de Minsky e Papert provou que, embora perceptrons pudessem ser mostrados como algo que aprende qualquer coisa que fossem capazes de representar, eles podiam representar muito pouco. Em particular, um perceptron de duas entradas não podia ser treinado para reconhecer quando aquelas duas entradas eram diferentes. Embora seus resultados não se aplicassem a redes multi-camadas mais complexas, os pesquisadores em redes neurais rapidamente reduziram-se a quase nenhum. Ironicamente, os novos algoritmos de aprendizagem com propagação retrógrada (*back-propagation*) para as redes neurais multi-camadas que estavam para causar uma enorme ressurgimento da pesquisa de redes neurais no fim dos anos 80, foram na verdade inicialmente descobertos em 1969 (Bryson e Ho, 1969).

### Sistemas baseados em conhecimento: a chave para o poder?

A visão de resolução de problema que havia surgido durante a primeira década da pesquisa em IA foi a de um mecanismo de busca de propósito geral tentando concatenar passos de raciocínios elementares para encontrar soluções completas. Tais abordagens têm sido denominadas **métodos fracos**, porque usam informações fracas sobre o domínio. Para muitos domínios complexos, resultou-se que sua performance é também fraca. O única maneira de contornar isso é utilizar conhecimento mais voltado para passos de raciocínio mais largos e resolver casos que ocorrem tipicamente em áreas estreitas de conhecimento. Pode-se dizer que para resolver um problema difícil, deve-se quase já saber a resposta.

O programa DENDRAL foi um dos primeiros exemplos desta abordagem. Foi desenvolvido em Stanford, onde Ed Feigenbaum (um antigo estudante de Herbert Simon), Bruce Buchanan (um filósofo voltado para a Ciência da Computação), e Joshua Lederberg (um laureado Nobel geneticista) juntaram-se para resolver o problema de inferência da estrutura molecular a partir da informação fornecida por um espectrômetro de massa. A entrada para o programa consiste da fórmula elementar da molécula (e.g.,  $C_6H_{13}NO_2$ ), e o espectro de massa dando as massas de vários fragmentos da molécula gerada quando é bombardeada por um feixe de elétron. Por exemplo, o espectro de massa pode conter um pico em  $m = 15$ , correspondendo à massa de um fragmento de metil ( $CH_3$ ).

A versão leiga do programa gerou todas as estruturas possíveis consistentes com a fórmula, e então predisse qual espectro de massa seria observado por cada uma delas, comparando com o espectro real. Como se pode esperar, isso rapidamente tornou-se intratável para moléculas de tamanho decrescentes. Os pesquisadores de DENDRAL consultaram analistas químicos e descobriram que eles trabalhavam observando padrões bem conhecidos no espectro que sugeriam subestruturas comuns na molécula. Por



exemplo, a regra seguinte é usada para reconhecer um subgrupo Cetona ( $C=O$ ):

**Se** existem dois picos em  $x_1$  e  $x_2$  tal que

(a)  $x_1 + x_2 = M + 28$  ( $M$  é a massa da molécula inteira),

(b)  $x_1 - 28$  é um pico alto,

(c)  $x_2 - 28$  é um pico alto,

(d) Pelo menos um entre  $x_1$  e  $x_2$  é alto,

**Então** há um subgrupo Cetona.

Tendo reconhecido que a molécula contém uma subestrutura particular, um número de possíveis candidatas é enormemente reduzido. A equipe do DENDRAL concluiu que o novo sistema era poderoso porque:

Todo o conhecimento teórico relevante para resolver esses problemas, tem sido mapeado, a partir de sua forma geral, no componente de predição do espectro (“primeiros princípios”) em formas especiais eficientes (“receitas culinárias”). (Feigenbaum *et al*, 1971).

A importância de DENDRAL foi ter sido considerado o primeiro sistema de conhecimento intensivo a ter sucesso: sua perícia derivou-se de um grande número de regras de propósito especial. Sistemas subsequentes também incorporaram cerne da abordagem do *Advice Taker* de McCarthy - a clara separação entre o conhecimento (na forma de regras) e o componente de raciocínio.

Com esta lição em mente, Feigenbaum e outros em Stanford iniciaram o Projeto de Programação Heurística (HPP), para investigar o quanto a nova metodologia de **sistemas especialistas** poderia ser aplicada a outras áreas da perícia humana. O próximo grande esforço foi na área de diagnóstico médico. Feigenbaum, Buchanan, e Dr. Edward Shortliffe desenvolveram o MYCIN para diagnosticar infecções sanguíneas. Com cerca de 450 regras, MYCIN foi capaz de ter um desempenho tão bom quanto alguns peritos, e consideravelmente melhor que médicos iniciantes. Ele também continha duas grandes diferenças do DENDRAL. Primeiro, diferentemente das regras do DENDRAL, não existia nenhum modelo teórico geral a partir do qual as regras do MYCIN pudessem ser deduzidas. Elas tinham de ser adquiridas a partir de entrevistas extensivas com peritos, que, por sua vez, as adquiriam diretamente da experiência de casos. Segundo, as regras tinham que refletir a incerteza associada ao conhecimento médico. MYCIN incorporou um cálculo de incerteza chamado **fatores de certeza** (ver cap. 14), que aparentavam (naquela época) encaixar-se bem na maneira como os médicos avaliavam o impacto da evidência nos diagnósticos.

Outra abordagem em diagnóstico médico foram também seguidas. Na Universidade de Rutgers, o projeto Computadores em Biomedicina de Saul Amarel iniciou um projeto uma ambiciosa tentativa de diagnosticar doenças baseado em conhecimento explícito dos mecanismos causais dos processos de doenças. Entretanto, grandes grupos no MIT e no New England Medical Center estavam em busca de uma abordagem para diagnóstico e tratamento baseado em teorias de probabilidade e utilidade. O objetivo era a construção de sistemas que forneciam recomendações médicas ótimas prováveis. Na medicina, a abordagem de Stanford usando regras fornecidas por médicos verificou-se mais popular a princípio. Mas um outro sistema de raciocínio probabilístico, o PROSPECTOR (Duda *et al*, 1979), gerou enorme publicidade ao recomendar prospecções exploratórias em um sítio geológico que provou conter um grande depósito de molibdênio.

A importância do domínio de conhecimento era também evidente na área de compreensão de linguagem natural. Embora o sistema SHRDLU de Winograd para a compreensão de linguagem natural houvesse engendrado uma boa dose de estímulo, sua dependência da análise sintática causou alguns dos mesmos problemas que ocorreram nos primeiros trabalhos sobre máquinas de tradução. Ele estava apto a contornar ambigüidades e entender referências pronominais, mas isto devia-se principalmente porque fora projetado especificamente para uma área – o mundo dos blocos. Vários pesquisadores, incluindo Eugene Charniak, um estudante de graduação de Winograd no MIT, sugeriu que a compreensão de linguagens robustas iria requerer conhecimento geral acerca do mundo e um método geral para usar este conhecimento.

Em Yale, o pesquisador de AI voltado para lingüística Roger Schank enfatizou este ponto proclamando “Não existe nada como a sintaxe”, que irritou uma grande quantidade de lingüistas, mas serviu para começar uma útil discussão. Schank e seus estudantes construíram uma série de programas que faziam todas as tarefas de compreensão de linguagem natural. A ênfase, no entanto, era menor sobre

a linguagem em si e maior nos problemas de representação e de raciocínio com o conhecimento requerido para o entendimento da linguagem. Os problemas incluíam a representação de situações estereotipadas (Cullingford, 1981), descrevendo a organização da memória humana (Rieger, 1976; Kolodner, 1983), e compreendendo planos e metas (Wilensky, 1983). William Woods construiu o sistema Lunar, o qual permitiu a geólogos abordarem questões em inglês acerca de modelos de rochas trazidos pela missão lunar Apolo. Lunar foi o primeiro programa em linguagem natural utilizado por pessoas, que não eram autores do sistema, que conseguiram realmente concluir o trabalho. Desde então, muitos programas em linguagem natural vem sendo utilizados em interfaces para databases.

O amplo crescimento de aplicações para problemas do mundo real causou o aumento concomitante na demanda por esquemas viáveis de representação de conhecimento. Um grande número de diferentes linguagens de representação foram desenvolvidas. Algumas eram baseadas na lógica – como por exemplo o PROLOG, que tornou-se popular na Europa, e a família PLANNER nos Estados Unidos. Outros, resultantes da idéia dos “frames” de Minsky, adotaram uma abordagem melhor estruturada, agrupando fatos acerca de objetos e tipos de eventos particulares, e arranjando os tipos em uma grande hierarquia taxonômica análoga a uma taxonomia biológica.

### **IA na indústria (1980-1988)**

O primeiro sucesso comercial de um sistema especialista, R1, iniciou sua operação na DEC - Digital Equipment Corporation (McDermott, 1982). O programa ajudava a configurar novos sistemas de computadores, e por volta de 1986 a companhia estava economizando certa de 40 milhões de dólares por ano. Por volta de 1988, a DEC possuía 40 desses sistemas especialistas desenvolvidos, com mais ainda a caminho. A Du Pont possuía 100 em uso e 500 em desenvolvimento, economizando cerca de US\$ 10 milhões por ano. Quase todas as grandes corporações americanas tinham seu próprio grupo de IA e estavam ou utilizando ou investigando novas tecnologia em sistemas especialistas.

Em 1981, os japoneses anunciaram o projeto “quinta-geração”, um plano de 10 anos para construir computadores inteligentes que rodavam Prolog da mesma maneira que os computadores ordinários executavam códigos de máquina. A idéia era que com a habilidade para fazer milhões de inferências por segundo, os computadores seriam capazes de tirar vantagem de um vasto depósito de fatos. O projeto propunha alcançar o entendimento total em linguagem natural, além de outras metas ambiciosas.

O projeto “quinta-geração” injetou interesse na IA, e tomando vantagem do medo do domínio japonês, pesquisadores e corporações estavam aptos a gerar suporte para um investimento similar nos Estados Unidos. A MCC - Microelectronics and Computer Technology Corporation era formado por um consórcio de pesquisadores para reagir ao projeto japonês. Na Inglaterra, o relatório Alvey reiniciou os fundos públicos para a área que foram cortados em virtude do famigerado relatório Lighthield.<sup>16</sup> Em ambos os casos, IA era parte de um amplo esforço, incluindo projetos de chips e pesquisa em interfaces homem-máquina.

O surto de progresso na indústria de IA também incluiu companhias como Grupo Carnegie, Inference, Intellicorp e Teknowledge, que ofereceram as ferramentas de software para a construção de sistemas especialistas, e companhias de hardware como a Lisp Machines INC, Texas Instruments, Symbolics, e Xerox, que estavam construindo estações de trabalho otimizadas para o desenvolvimento de programas LISP. Mais de uma centena de companhias construíram sistemas industriais de visão robóticos. Por toda parte, a indústria saiu de alguns poucos milhões em vendas em 1980 a US\$ 2 bilhões em 1988.

### **O retorno das redes neurais (1980 - dias presentes)**

Ainda que a Ciência da Computação tivesse negligenciado o campo das redes neurais depois do livro “Perceptrons” de Minsky e Papert, os trabalhos haviam continuado em outros campos, particularmente na física. Grandes conjuntos de neurônios simples podiam ser compreendidos da mesma

---

<sup>16</sup> Para salvar do embaraço, um novo campo chamado KBS (Intelligent Knowledge-Based Systems) era definido porque a Inteligência Artificial tinha sido oficialmente cancelada.

maneira que grandes conjuntos de átomos nos sólidos. Físicos como Hopfield (1982) utilizaram técnicas de estatística mecânica para analisar as propriedades de armazenamento e de otimização das redes, trazendo um importante cruzamento de idéias. Psicólogos incluindo David Rumelhart e Geoff Hinton continuaram o estudo de modelos de memória em redes neurais. Como discutimos no capítulo 19, o impacto real veio em meados de 1980, quando pelo menos quatro grupos diferentes reinventaram o algoritmo de aprendizagem back-propagation, primeiramente encontrado em 1969 por Bryson e Ho. O algoritmo foi aplicado em muitos problemas de aprendizagem na Ciência da Computação e na psicologia, e os resultados foram divulgados na coleção *Parallel Distributed Processing* (Rumelhart, McClelland, 1986), despertando grande interesse.

Na mesma época, algumas desilusões ocorriam no tocante à aplicabilidade de sistemas especialistas derivados do sistema Mycin. Muitas corporações e grupos de pesquisas achavam que a construção de um sistema especialista de sucesso envolveria muito mais que simplesmente comprar um sistema de raciocínio e enchê-lo com REGRAS. Alguns prediziam um "congelamento na IA", onde a aplicação de capital seria rigorosamente apertada. Foi talvez este medo, e o fator histórico pelo lado das redes neurais, que levaram a um período em que as redes neurais e a IA tradicional eram vistas como campos rivais, ao invés de abordagens de suporte mútuo para o mesmo problema.

### Eventos recentes (1987 - dias presentes)

Os anos recentes têm presenciado um mar de mudanças no conteúdo e na metodologia da pesquisa em inteligência artificial<sup>177</sup>. Agora é mais comum construir sobre teorias existentes que propor teorias novas em folha, basear alegações em teoremas rigorosos ou evidências experimentais difíceis ao invés da intuição, e mostrar relevância às aplicações do mundo real ao invés de exemplos "brinquedos".

O campo do reconhecimento de voz ilustra esse padrão. Na década de 1970, uma grande variedade de arquiteturas e abordagens foram tentadas. Várias dessas eram "ad hoc" e frágeis, e foram demonstradas com poucos exemplos especialmente selecionados. Nos anos recentes, abordagens baseadas nos **modelos ocultos de Markov** (MEMs) têm dominado a área. Dois aspectos dos MEM são relevantes para a discussão presente. Primeiro, eles são baseados em uma rigorosa teoria matemática. Isto tem permitido a pesquisadores de voz construir, por várias décadas, resultados matemáticos desenvolvidos em outras áreas. Segundo, eles são gerados por um processo de treinamento em um grande corpo de dados de voz real. Isso certifica que a desempenho é robusto, e em vários testes cegos rigorosos os MEMs têm melhorado constantemente esses resultados. A tecnologia de voz e o campo relacionado do reconhecimento de caracteres manuscritos já estão fazendo a transição para aplicações industriais e domésticas bastante difundidas.

Outra área que aparentemente tem sido beneficiada com a formalização é o planejamento. O trabalho primitivo de Austin Tate (1977), seguido por David Chapman (1987), tem resultado em uma elegante síntese dos programas de planejamento existentes em uma estrutura simples. Tem havido um grande número de avanços que se constroem baseados um no outro ao invés de se começar do zero a cada vez. O resultado é que sistemas de planejamento que eram bons apenas para micromundos nos anos 70 são agora utilizados para programar trabalhos de fábrica e missões espaciais, entre outras coisas. Veja os capítulos 11 e 12 para maiores detalhes.

O *Raciocínio Probabilístico de Sistemas Inteligentes* de Judea Pearl (1988) marcou uma nova aceitação da probabilidade e da teoria da decisão em IA, seguindo um ressurgimento de interesse a partir do artigo "Em Defesa da Probabilidade" de Peter CheeseMan (1985). O formalismo da **rede de crenças** foi inventado para permitir raciocínio eficiente sobre a combinação de evidência incerta. Essa abordagem superou largamente os problemas com sistemas de raciocínio probabilístico das décadas de 60 e 70, e tem vindo a dominar a pesquisa em IA sobre raciocínio incerto e sistemas especialistas. O trabalho de Judea

---

<sup>177</sup> Alguns têm caracterizado essa mudança como uma vitória dos arrumados - aqueles que pensam que as teorias da IA deveriam ser fundamentadas no rigor matemático- sobre os <scruffies> - aqueles que preferem experimentar várias idéias, escrevem alguns programas e depois avaliam o que parece estar trabalhando. Ambas as abordagens são importantes. Uma mudança na direção à arrumação aumentada implica que o campo alcançou um nível de estabilidade e maturidade. (Se esta estabilidade será quebrada por uma nova idéia <scruffy> é outra questão).

Pearl (1982a) e Eric Horvitz e David Heckerman (Horvitz e Heckerman, 1986; Horvitz *et al.*, 1986) promoveu a idéia de sistemas especialistas normativos: aqueles que agem racionalmente de acordo com as leis da teoria da decisão e não tentam imitar peritos humanos. Os capítulos 14 e 16 cobrem esta área.

Notáveis revoluções similares têm ocorrido em robótica, visão do computador, aprendizagem mecânica (incluindo redes neurais) e representação do conhecimento. Um melhor entendimento dos problemas e de suas propriedades de complexidade, combinadas com um refinamento matemático, tem levado a agendas de pesquisa de interesse e métodos robustos. Talvez encorajado pelo progresso na resolução de subproblemas da IA, pesquisadores têm também começado novamente a considerar o problema do “agente como um todo”. O trabalho de Allen Newell, John Laird e Paul Rosenbloom em SOAR (Newell, 1990; Laird *et al.*, 1987) é o melhor exemplo conhecido de uma arquitetura completa de agentes em IA. O assim chamado movimento “situado” tem a finalidade de entender os trabalhos de agentes inseridos em ambientes reais com entradas sensoriais contínuas. Muitos resultados interessantes estão surgindo de tal trabalho, incluindo a percepção que as subáreas anteriormente isoladas da IA podem necessitar ser reorganizadas de maneira que seus resultados devam ser fundidos em um único desenho de agente.

## 1.4 O ESTADO DA ARTE

---

O grande mestre internacional Arnold Denker estuda as peças no tabuleiro em frente a ele. Ele compreende que não há esperança; ele deve renunciar ao jogo. Seu oponente, HITECH, torna-se o primeiro programa de computador a derrotar um grande mestre em um jogo de xadrez (Berliner, 1989).

“Eu quero ir de Boston a São Francisco”, fala a viajante no microfone. “Em que data você deseja viajar?” é a resposta. A viajante explica que ela quer ir em 20 de outubro, sem paradas, com a tarifa mais barata disponível, retornando no domingo. Um programa que entende voz chamado PEGASUS manipula toda a transação, o que resulta numa reserva confirmada que economiza para a viajante \$894 do preço normal da passagem de ônibus. Embora o reconhecedor de voz cometa um erro em dez tentativas<sup>18</sup>, ele também está apto a recuperar-se desses erros pelo seu entendimento de como os diálogos são organizados (Zue *et al.*, 1994).

Um analista na sala de Operação de Missões do Laboratório de Jatopropulsão de repente começa a prestar atenção. Uma mensagem vermelha apareceu na tela indicando uma anomalia com a espaçonave Voyager, que está em algum lugar na vizinhança de Netuno. Felizmente, o analista está apto a corrigir o problema a partir da Terra. O pessoal de operações acredita que o problema poderia não ter sido detectado se não fosse o MARVEL, um sistema especialista de tempo real que monitora a massiva quantidade de dados transmitida pela espaçonave, manuseando tarefas de rotina e alertando os analistas para problemas mais sérios (Schwuttke, 1992).

Cruzando a autoestrada nas imediações de Pittsburgh em confortáveis 55 mph, o homem no assento do motorista parece relaxado. Ele deve estar - ao longo das últimas 90 milhas, ele não tocou no volante, freio ou acelerador. O motorista real é um sistema robótico que coleta dados de entrada de câmeras de vídeo, sonares e captadores de distância a laser fixados ao carro. Ele Combina essas entradas com a experiência aprendida no treinamento de corridas anteriores e computa com sucesso como guiar o veículo (Pomerleau, 1993).

Um perito em patologia do nodo linfático descreve um caso extremamente difícil para um sistema inteligente e examina o diagnóstico do sistema. Ele zomba da resposta do sistema. Apenas levemente preocupados, os criadores do sistema sugerem que ele peça ao computador uma explicação do diagnóstico. A máquina aponta os maiores fatores que influenciaram sua decisão e explica a sutil interação de vários sintomas desse caso. Finalmente, o perito admite seu erro (Heckerman, 1991).

De uma câmera situada em um semáforo de um cruzamento, o monitor de tráfego observa a cena. Se qualquer pessoa estivesse acordada para ler a tela principal, iria ver um "Citröen 2CV vindo da Place de la Concorde para Champs Elysées", "Grande caminhão de marca desconhecida parado na Place de La Concorde", e assim por diante noite adentro. E ocasionalmente, "Grave acidente na Place de La

---

<sup>18</sup> Alguns outros sistemas existentes erram apenas metade do usual nessa tarefa.

Concorde, veículo em alta velocidade colidiu com motociclista", e uma chamada automática acionaria os serviços de emergência (King *et al.*, 1993; Koller *et al.*, 1994).

Estes são apenas alguns poucos exemplos dos sistemas de inteligência artificial que existem hoje. Não é magia ou ficção científica, mas ciência, engenharia, e matemática, para os quais este livro fornece uma introdução.

## 1.5 RESUMO

---

Este capítulo define IA e estabelece o cenário cultural no qual ela tem se desenvolvido. Alguns dos pontos importantes são os seguintes:

- Pessoas diferentes pensam diferente sobre IA. Duas questões importantes são: você está preocupado com o pensamento ou com o comportamento? Você quer modelar humanos, ou trabalhar a partir de um padrão ideal?
- Neste livro, adotamos o ponto de vista de que a inteligência diz respeito principalmente à **ação racional**. Idealmente, um **agente inteligente** executa a melhor ação possível em uma situação. Estudaremos o problema de construir agentes que são inteligentes neste sentido.
- Filósofos (voltando ao ano 400 a.C.) tornaram a IA concebível por considerar as idéias de que a mente é, de alguma maneira, como uma máquina, que opera sobre o conhecimento codificado em alguma linguagem interna, e que o pensamento pode ser utilizado para definir as ações corretas a serem tomadas.
- Matemáticos forneceram as ferramentas para manipular declarações de lógica certa bem como de lógica incerta, que são as declarações probabilísticas. Eles também estabeleceram as bases do raciocínio sobre algoritmos.
- Psicólogos fortaleceram a idéia de que humanos e outros animais podem ser considerados máquinas de processamento de informação. Linguistas mostraram que o uso da linguagem se adequa a esse modelo.
- A Engenharia da Computação forneceu o artefato que torna possíveis aplicações de IA. Programas de IA tendem a ser grandes, e eles não poderiam trabalhar sem os grandes avanços em velocidade e memória que a indústria da computação tem fornecido.
- A história da IA tem tido ciclos de sucesso, otimismo inadequados, e conseqüentemente cortes ou reduções no entusiasmo e recursos financeiros. Também tem havido ciclos de introdução de novas abordagens criativas e sistematicamente o refinamento das melhores.
- Progresso recente na compreensão da base teórica para a inteligência tem passado de mão em mão com aperfeiçoamentos nas capacidades dos sistemas reais.

## NOTAS BIBLIOGRÁFICAS E HISTÓRICAS

---

*Inteligência Artificial* de Daniel Crevier (1993) dá uma história completa deste campo; e *A Idade de Máquinas Inteligentes* de Raymond Kurzweil (1990) situa a IA no contexto mais amplo de Ciência da Computação e na história de intelectualidade em geral. Dianne Martin (1993) documentou o grau com o qual os primeiros computadores foram considerados pela mídia como portadores de poderes místicos de inteligência.

O status metodológico de inteligência artificial é discutido em *As Ciências do Artificial*, de Simon (1981), que discute pesquisa na área que com relação a artefatos complexos. Ele explica como a IA pode ser vista tanto como Ciência como Matemática.

*Inteligência Artificial: A Grande Idéia*, de John Haugeland (1985), nos dá uma noção legível dos problemas filosóficos e práticos da IA. A Ciência Cognitiva é bem descrita em *O Computador e a Mente: Uma Introdução à Ciência Cognitiva*, por Johnson Laird. Baker (1989) cobre a parte sintática da lingüística moderna, e Chierchia e McConnell-Genet (1990) cobre a semântica. Allen (1995) cobre a lingüística do ponto de vista da IA.

Os trabalhos primordiais sobre IA são cobertos por Feigenbaum e Feldman em *Os Computadores e o Pensamento*, por *Processo de Informação Semântica*, de Minsky, e pela série *Inteligência de Máquinas*, editada por Donald Michie. Um grande número de documentos influentes é colecionado em *Leituras em Inteligência Artificial* (Webber e Nilsson, 1981). Trabalhos primordiais em redes neurais estão colecionados em *Neurocomputação* (Anderson e Rosenfeld, 1988). A *Enciclopédia de IA* (Shapiro, 1992) contém artigos de pesquisa em quase todos os tópicos da IA. Esses artigos usualmente provêm um bom ponto de partida na literatura de pesquisa sobre cada tópico. Os quatro volumes do *Manual de Inteligência Artificial* (Barr e Feigenbaum, 1981) contém descrições de quase todos os principais sistemas de IA publicados antes das 1981.

O mais recente trabalho aparece nos anais das principais conferências de IA: a Conferência Conjunta Internacional Bienal em IA (IJCAI), e a Conferência Nacional Anual em IA, mais freqüentemente conhecida como AAAI, depois de sua organização patrocinada. Os principais jornais para IA: *Artificial Intelligence*, *Computational Intelligence*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, e o *Journal of Artificial Intelligence Research*. Existem ainda muitos jornais dedicados a áreas específicas que nós cobriremos nos capítulos apropriados. Produtos comerciais são cobertos na revista *AI Expert* e *PC IA*. As principais sociedades profissionais para IA são a American Association for Artificial Intelligence (AAAI), o Grupo de Interesse Especial em Inteligência Artificial da ACM (SIGART), e a Sociedade para Inteligência Artificial e Simulação de Behaviour (AISB). A Revista de IA de AAAI e o Boletim de SIGART contém muitos tópicos e artigos de tutorais como também anúncios de conferências e seminários.