

Conhecimento e interação: *fronteiras entre o agir humano e inteligência artificial*

Alex Primo¹

Universidade Federal do Rio Grande do Sul

alex.primo@terra.com.br

Resumo: *Este trabalho pretende aprofundar a crítica às promessas da inteligência artificial em reproduzir tecnologicamente a interação humana. Trata-se, na verdade, de um prolongamento da discussão iniciada no trabalho “Comunicação e inteligência artificial: interagindo com a robô de conversação Cybelle” (Primo e Coelho, 2002). Procurar-se-á aqui expandir a reflexão sobre conhecimento e interação a partir de uma abordagem relacional da comunicação humana e da biologia do conhecimento. Além disso, quer-se diferenciar e criticar duas linhas de pesquisa em inteligência artificial (a simbólica e o connexionismo), e suas respectivas referências ao behaviorismo e cognitivismo.*

Ao fazer citações deste artigo, utilize esta referência bibliográfica:

PRIMO, Alex. Conhecimento e interação: fronteiras entre o agir humano e inteligência artificial. In: LEMOS, A.; P. CUNHA (Eds.). Olhares sobre a cibercultura. Porto Alegre: Sulina, 2003. p. 37-56.

Introdução

Este trabalho pretende aprofundar a crítica às promessas da inteligência artificial em reproduzir tecnologicamente a interação humana. Trata-se, na verdade, de um prolongamento da discussão iniciada no trabalho “Comunicação e inteligência artificial: interagindo com a robô de conversação Cybelle” (Primo e Coelho, 2002)². Procurar-se-á aqui expandir a reflexão sobre conhecimento e interação a partir de uma abordagem relacional da comunicação humana e da biologia do conhecimento. Além disso, quer-se diferenciar e criticar duas linhas de pesquisa em inteligência artificial (a simbólica e o connexionismo), e suas respectivas referências ao behaviorismo e cognitivismo.

¹ Professor de Comunicação (Fabico/PPGCOM//UFRGS), doutor em Informática na Educação (PGIE/UFRGS), mestre em Jornalismo pela Ball State University; coordenador do Laboratório de Interação Mediada por Computador (PPGCOM/UFRGS).

² Trabalho apresentado no GT “Comunicação e Sociedade Tecnológica”, na X reunião da Compós, em Brasília, DF.

Inteligência artificial simbólica

A linha de pesquisa chamada de inteligência artificial simbólica parte do pressuposto que o aparato mental é essencialmente “um dispositivo lógico que pode ser descrito por meio de um conjunto de computações abstratas, onde o que importa são as propriedades formais dos símbolos que são manipulados” (Teixeira, 1998, p. 43). Para essa vertente, a inteligência poderia ser definida como a capacidade de resolver problemas. Para tanto, se faz necessário “um algoritmo que permita a manipulação adequada da atividade simbólica” (p. 44). Logo, pensar “nada mais é do que realizar computações, uma em seguida da outra” (p. 44).

Veja-se por exemplo o pioneiro modelo do “Jogo de von Neumann³”. O “jogador” seria teoricamente capaz de toda computação necessária para resolver qualquer problema apresentado⁴. O “jogador” recebe informações do jogo e age sobre esses dados. Na verdade, não consegue deixar de desempenhar essas funções quando elas são apropriadas. Ao analisar esse sistema, Bateson (2000) conclui que o “jogador” apresenta “aprendizado zero”⁵, não conseguindo aprender por tentativa e erro. Ou seja, é incapaz de modificar e adaptar suas ações futuras em virtude do reconhecimento dos erros anteriores. Sempre que um mesmo problema for reapresentado pelo “jogo”, ele irá seguir “corretamente” suas mesmas computações, ainda que sejam inapropriadas. Isto é, o “jogador” de von Neumann continuará sempre seguindo o que o algoritmo prescreve.

Deve-se atentar que a inteligência artificial simbólica parte de uma teoria representacional da mente. Maturana e Varela (1995), contudo, fazem uma dura crítica ao representacionismo. Segundo eles, essa visão entende que o meio informa um mapa que guiará a rota. Sendo assim, como explicar através dessa abordagem “a extraordinária eficácia operacional do homem e dos animais, nossa enorme capacidade de aprendizagem e manipulação do mundo?” (p. 163).

O entendimento de que a inteligência humana é apenas um aparato para resolver problemas ignora justamente a criatividade e a inventividade humana. Além de resolver problemas, os homens têm a capacidade de criá-los. Problematizando o mundo, o homem modifica o próprio meio em que vive relacionalmente. Logo, o ser humano não está a mercê da imposição do meio, mas estabelece com ele uma relação interativa.

Como defende Freire (2001, p. 96), a realidade não se encontra dicotomizada dos homens, “como se fosse um mundo à parte, misterioso e estranho, que os esmagasse”. Vale lembrar que, segundo o autor, o existir humano dá-se na

³ John von Neuman (1903-1957) foi um brilhante matemático que inspirou a produção de computadores. A chamada “arquitetura von Neumann” permite que programas sejam estocados na memória do computador, e suas instruções sejam executadas sequencialmente (Teixeira, 1998). Entre tantos outros projetos, a Teoria dos Jogos e sua pesquisa sobre autômatos são importantes referências nas raízes da inteligência artificial.

⁴ Segundo Von Neumann e sua definição formal de “jogo”, todos os problemas que o “jogo” apresenta são considerados computáveis.

⁵ Esse tipo de aprendizado, segundo Bateson (2000), caracteriza-se pela especificidade da resposta, que, verdadeira ou falsa, não é sujeita a correção. O aprendizado zero ocorre quando apenas uma pequena mudança (ou nenhuma) ocorre em resposta a um evento repetido.

práxis, na pronúncia do mundo e em sua modificação. Esse mundo problematizado se volta aos sujeitos exigindo deles nova ação-reflexão.

Behaviorismo: estímulo e resposta

Não se pode deixar de identificar claramente na crença que a performance “inteligente” dos robôs depende apenas da ampliação das regras e associações entre informações cadastradas uma concepção behaviorista do comportamento humano

A perspectiva behaviorista (comportamentalista), dedicada ao estudo das relações causais entre estímulo e resposta, presume um mundo ordenado de entidades que se relacionam mecanicamente, cujas regularidades podem ser registradas pela observação repetida. O comportamento humano seria consequência das condições anteriores. Ou seja, dado o antecedente X, Y deve ocorrer – uma relação se/então. O foco investigativo dessa tradição se volta para o comportamento individual.

De forte traço determinístico, o behaviorismo busca prever comportamentos a partir de condições anteriores. Na verdade, o que se passa entre um estímulo e uma resposta não interessa aos behavioristas. O que importa é isolar os inputs e outputs e generalizar a relação entre eles.

A previsibilidade, contudo, nem sempre é possível⁶. Maturana e Varela (1995) mostram que os observadores podem não ter condições de obter os conhecimentos necessários sobre a operação de um sistema. Ora, é justamente a impossibilidade de se fazer uma descrição fechada e completa que inviabiliza a intenção de se dominar o funcionamento do sistema cognitivo humano. O que, claro, impede a previsibilidade do comportamento.

Mesmo assim, suponha-se por um momento que o sistema cognitivo pudesse ser observado em toda sua complexidade. Nesta hipótese, todas as relações mentais poderiam ser visualizadas a olho nu! Mesmo que isso fosse possível, a descrição final já não corresponderia à dinâmica daquele ser. Ora, suas estruturas não são estáticas e imutáveis. Como lembram Maturana e Varela (1995), muitos sistemas se modificam enquanto são observados.

Quanto ao foco na associação estímulo-resposta, Piaget (1996, p. 39) nega explicitamente que o sistema nervoso seja restrito a intervir sob a forma de respostas ou reações ($S \rightarrow R$), pois “está longe de limitar-se a recolher inputs ou informações aferentes, uma vez que reage com movimentos e respostas ativas que modificam o meio”. Para o autor, “os conhecimentos não constituem uma cópia do meio mas um sistema de interações reais, que refletem a organização auto-reguladora da vida” (p. 39). Produto de uma interação, conhecer é agir sobre o real, transformando-o. Nem mesmo o condicionamento⁷ e o desenvolvimento de hábitos se encaixariam na simples relação $S \rightarrow R$, pois “ao adquirir novos condicionamentos e novos hábitos, o

⁶ Deve-se aqui acrescentar que nem diante dos fenômenos físicos a previsibilidade total é possível, como nos lembra a física quântica.

⁷ Piaget (1987, p. 128) avisa que o próprio reflexo condicionado é frágil e instável. O mesmo deixa de ocorrer se não for continuamente reforçado (ou “confirmado”) pelo meio exterior. E, “na medida em que o reflexo condicionado é ‘confirmado’, ele deixa de ser uma simples associação para inserir-se no esquema muito mais complexo das relações entre a necessidade e a satisfação, logo, das relações de assimilação”.

ser vivo assimila os sinais e organiza esquemas de ação que se impõem e ao mesmo tempo se acomodam ao meio” (p. 45).

Conforme mostra Becker (1997), para a concepção piagetiana um processo linear que conduz de S a R não faz sentido. Tal esquema despreza a dimensão temporal, caindo em um círculo vicioso, “explicando ora o estímulo pela resposta, ora a resposta pelo estímulo” (p. 100). Por outro lado, se o tempo for introduzido na análise, perceber-se-á que o círculo transforma-se em espiral. Em outras palavras, o que escapa ao modelo S-R é que a resposta passa a constituir retrospectivamente a classe de respostas.

Ainda sobre o impacto temporal, Maturana (2001, p. 186) vai destacar que uma distinção cabal entre o ser humano (um ente natural, um organismo vivo) e um robô (produto do design humano) é que – além de participarem de sua própria criação (o que não acontece com os robôs) – os homens são seres históricos.

A diferença entre os dois é o modo pelo qual suas respectivas coerências operacionais com suas circunstâncias surgiram em sua história de origem. O robô surge através de um projeto, um design. Um artista ou um engenheiro faz um projeto dispondo um conjunto de elementos e configuração de relações entre eles de uma forma que eles constituam uma totalidade dinâmica em congruência dinâmica com um meio que também, de maneira ad hoc, foi projetado como tal. Assim, o robô, o meio ou circunstâncias nas quais ele funciona e a congruência dinâmica entre os dois são conseqüências de um projeto intencional naquilo que, pode-se dizer, foi um processo aistórico.

Ou seja, os seres humanos, ao contrário das máquinas artificiais, não têm uma configuração inicial que comandará necessariamente suas ações futuras. Para o autor, os homens existem “enquanto seres vivos como entidades sistêmicas num espaço relacional em mudança estrutural contínua” (p. 190). E é no viver com outras pessoas que o homem vem a ser o que é (*Homo sapiens amans* ou *Homo sapiens aggressans*, segundo a crítica irônica de Maturana). Diferentemente do que ocorre com robôs, a vida histórica humana

...se dá na dinâmica relacional na qual a vivemos ao viver em conversações como seres linguajantes. Como uma conseqüência de nossa condição de viver em conversações, nossa história enquanto seres humanos se deu na geração contínua de domínios de coordenações de coordenações de comportamentos que flutuam na conservação de nosso viver como entidades biológicas, num fluir de realidades humanas variáveis, e que é possível porque não importa como nosso viver biológico é conservado, desde que seja conservado (Maturana, 2001, p. 192).

Indeterminação e senso comum

Uma das diferenças fundamentais entre humanos e robôs de inteligência artificial é a capacidade dos primeiros em conviver com a complexidade e com a indeterminação. Segundo Morin (1990), uma das conquistas do estudo do cérebro humano foi mostrar sua superioridade em relação ao computador por conseguir trabalhar com o insuficiente e o vago. É preciso notar que “a complexidade não compreende apenas quantidades de unidades e interações (sic) que desafiam as nossas possibilidades de cálculo; compreende também incertezas, indeterminações, fenômenos (sic) aleatórios” (p. 52). Decorrente disso, o autor entende que é preciso

reconhecer uma certa ambigüidade nas relações sujeito/objeto e que certos fenômenos como a liberdade e a criatividade são inexplicáveis fora do quadro complexo.

Mas como pode o ser humano agir quando um problema não é bem definido, nem as soluções são evidentes? Capra (1996) sugere que nesses casos o comportamento humano inteligente pode recorrer ao senso comum, decorrente das experiências vividas. “No entanto, o senso comum não está disponível aos computadores devido à cegueira destes à abstração e às limitações intrínsecas das operações formais, e, portanto, é impossível programar computadores para serem inteligentes” (Capra, 1996, p. 216-217).

Veja-se o seguinte exemplo de Terry Winograd, a partir do qual Capra (1996, p. 217) explicita aquela limitação da inteligência artificial:

“Tommy tinha acabado de receber um novo conjunto de blocos de montar. Ele estava abrindo a caixa quando viu Jimmy chegando”. Como Winograd explica, um computador não teria uma pista a respeito do que existe dentro da caixa, mas supomos imediatamente que ela contém os novos blocos de Tommy. E supomos isso porque sabemos que os presentes frequentemente vêm em caixas e que abrir a caixa é a coisa adequada a fazer. E o mais importante: nós supomos que as duas sentenças no texto estão ligadas, ao passo que o computador não vê a razão para vincular a caixa com os blocos de armar. Em outras palavras, nossa interpretação desse simples texto baseia-se em várias suposições de senso comum e em várias expectativas que não estão disponíveis ao computador.

Por outro lado, Douglas Lenart afirma que os seus predecessores não haviam se esforçado suficientemente para sistematizar o senso comum (Dreyfus, 1992). Lenart trabalha desde 1984 no programa Cyc que visa cadastrar em um banco de dados justamente informações de senso comum. Encontra-se abaixo uma descrição do projeto encontrada no *site* de sua empresa (<http://www.cyc.com/overview.html>):

The Cyc product family is powered by an immense multi-contextual knowledge base and an efficient inference engine. The knowledge base is built upon a core of over 1,000,000 **hand-entered assertions** (or "rules") designed to capture a large portion of what we normally consider consensus knowledge about the world. For example, Cyc knows that trees are usually outdoors, that once people die they stop buying things, and that glasses of liquid should be carried rightside-up⁸ [grifo meu].

Mas, como se pode ver no trecho salientado em negrito, trata-se do trabalho de previsão e cadastramento de dados por uma equipe externa à máquina. Ou seja, insiste-se nesse projeto – sobre o qual reside grande expectativa do mercado e da comunidade dedicada à inteligência artificial – que o conhecimento pode ser copiado e representado. Bastariam tempo e dedicação para que se conseguisse replicar a inteligência humana. É como se a complexidade pudesse se agarrada em sua totalidade, sistematizada por regras e fechadas em um pacote final. Vale lembrar,

⁸ Tradução do autor: A família de produtos do projeto Cyc é movido por uma imensa base de conhecimento multi-contextual e por um mecanismo de inferência eficiente. A base de conhecimento é construída sobre um núcleo de mais de 1.000.000 de assertivas(ou regras) incluídas manualmente e projetadas para capturar uma grande porção do que nós normalmente consideramos conhecimento consensual sobre o mundo. Por exemplo, Cyc sabe que árvores estão normalmente ao ar livre, que uma vez que as pessoas morram elas param de comprar coisas, e que garrafas de líquido devem ser carregadas como lado certo para cima.

porém, que o chamado senso comum não é algo “ensinado” ou “entregue” ao ser humano, mas produto de interação ativa com seu meio.

Cabe também destacar que a liberdade e criatividade, lembradas por Morin, são totalmente estranhas aos sistemas informáticos determinísticos. Logo, as interações na quais se envolvem aquelas máquinas não são criadas espontaneamente, pois acontecem em virtude de potenciais configurados.

Conexionismo e redes neurais

Enquanto a inteligência artificial simbólica era bombardeada por diversos filósofos, uma linha de pesquisa vizinha reingressa no debate com renovada energia. Ganhando vigor nos anos 80, o conexionismo volta-se contra a ênfase simbólica na transformação seqüencial de símbolos em virtude de regras pré-estabelecidas e de forma independente do hardware.

Segundo Teixeira (1998, p. 166), o conexionismo é uma concepção alternativa para modelagem da mente humana através do computador. Tendo como base o cérebro e suas características biológicas⁹, tenta modelar processos inteligentes, buscando reproduzir artificialmente neurônios e sinapses. Isto é, busca uma simulação do cérebro através do processamento paralelo em “redes neurais”. Teixeira (1998, p. 84) faz a seguir uma diferenciação entre o conexionismo e a inteligência artificial simbólica:

Sistemas conexionistas e simbólicos são sistemas computacionais, mas há uma grande diferença no tipo de computação que eles realizam. Na perspectiva simbólica, a computação é essencialmente a transformação de símbolos de acordo com regras – regras que estão estabelecidas em um programa. A idéia de computação subjacente a um sistema conexionista é diferente: seu princípio é um conjunto de processos causais através do quais as unidades se excitam ou se inibem, sem empregar símbolos ou tampouco regras para manipulá-los.

Logo de pronto deve-se questionar: se Piaget, Maturana e Varela demonstram que não há como pensar a inteligência e o aprendizado dos homens em detrimento de sua biologia, então como simular o conhecimento humano e sua evolução ativa em máquinas não-biológicas? Além disso, Teixeira (1998, 116) diante da complexidade do cérebro e lembrando o chamado “problema da descrição” pergunta: “como simular aquilo que não podemos sequer representar?”. Ou seja, como se poderia copiar aquilo que não se consegue descrever totalmente? E acrescenta: como assegurar que uma suposta descrição do cérebro seria a correta?

Vale agora observar como se entende a questão do aprendizado e tomada de decisões nessa perspectiva de inteligência artificial. O projeto conexionista busca modelar a cognição humana a partir de neurônios artificiais¹⁰. Esses “neurônios”

⁹ Conforme Teixeira (1998, p. 85), “A abordagem conexionista é uma tentativa de construir um modelo de mente mais próximo de sua realidade biológica. Embora estes sistemas não sejam um modelo completo do cérebro e de seu funcionamento, pode-se pelo menos dizer que eles são inspirados na estrutura do cérebro”.

¹⁰ Button et al (1998, p. 151) afirmam que no trabalho original de McCulloch & Pitts “com muita frequência são camufladas as distinções entre um neurônio ‘formal’ de McCulloch-Pitts e um neurônio biológico real”.

teriam pesos que podem ser alterados em virtude da estimulação positiva ou negativa das conexões em que se envolvem. “Cada neurônio tem um valor de ativação, e cada sinapse que chega até ele tem uma força, positiva ou negativa, de conexão” (Teixeira, 1998, p. 84). O “significado” e as decisões emergiriam (segundo o jargão conexionista) do estado global da rede, dos complexos padrões de atividade em jogo.

Mas, apesar da crítica conexionista à inteligência artificial simbólica, por trabalhar com regras pré-definidas na programação, pode-se ainda reconhecer uma orientação associacionista na definição de como a rede “aprende” por repetição. Veja-se: através de recorrentes inputs de um conjunto de características de um certo domínio, a rede pode gerar um protótipo, produzindo um padrão a partir das informações estocadas. Por exemplo, ao fornecer-se à rede os valores f_1 , f_2 , f_3 e f_4 espera-se que ela preencha f_5 , f_6 , f_7 , f_8 , f_9 e f_{10} com os valores apropriados. A partir dos valores iniciais fornecidos, a rede busca um padrão comum entre todos inputs, o que lhe permite trabalhar com entradas incompletas. E, segundo complementa Teixeira (1998, p.100), o padrão de conectividade entre os neurônios artificiais pode ser ajustado através do “treinamento” das redes neurais:

A fase de treinamento é uma fase de aprendizado. A rede recebe um *input* e produz um *output*. Este *output* é, então, comparado com o *output* que seria correto. Calcula-se o erro e a rede então ajusta seus padrões de conectividade para ver se consegue aproximar seu *output* daquilo que se considera o *output* correto. Uma vez tendo feito todo o aprendizado, a rede torna-se capaz não apenas de processar o *input* típico como também suas instâncias mais próximas e a partir dela gerar protótipos.

Deve-se perguntar, porém, quem define o que é erro e em virtude do quê. Certos problemas têm de fato uma resposta única e indiscutível. Mas tantos outros não encontram a definição de uma resposta correta. A própria questão do reconhecimento, tão cara ao conexionismo (tendo em vista que uma significativa parte dos projetos dessa linha se dedicam ao reconhecimento de padrões), não é trivial nem encontra no domínio humano resultados exatos, livres de erro ou ambiguidade.

...reconhecer algo é, frequentemente, mas nem sempre, estar correto sobre o que ele é (de alguma maneira específica relevante para os propósitos de alguém, e de acordo com as convenções de identificação pertinentes). Pode-se também reconhecer algo específico sem sequer tê-lo visto antes, se ele for reconhecido como um sinal de um tipo conhecido de fenômeno (por exemplo, aquele peixe lá é uma “truta”), e pode-se reconhecer algo como tendo sido visto antes, sem se saber exatamente o que seja. (“Aquilo de novo! Que diabos será?”). Além disso, pode-se dizer sinceramente que se reconhece, realmente pensar que se reconheceu algo ou alguém e estar enganado: o objeto ou a pessoa parecia ser assim e assado mas era na realidade tal e tal. (Button et al, 1998, p. 174)

Finalmente, apesar do protesto de muitos conexionistas que negam a influência representacionista, Teixeira (1998, p. 108) sentencia que “a maioria dos sistemas conexionistas ainda funciona com representações que são em grande parte fornecidas pelo programador, e, assim sendo, não podemos afirmar que tais sistemas estão efetivamente em contato com o mundo exterior”. Trata-se de outra gigantesca barreira, pois, segundo a biologia do conhecimento, é na relação entre organismo e o meio que o conhecer humano estrutura-se dinamicamente. Mesmo que as redes neurais reúnam diferentes unidades operando em paralelo, elas não deixam de

constituir, pois, um sistema alopoiético – o que inviabiliza seu aprendizado verdadeiramente autônomo.

O polêmico Dreyfus (1992) vai também denunciar que não existe aprendizado autônomo das redes neurais. Segundo os conexionistas, a rede treinada conseguirá efetuar generalizações, oferecendo o mesmo output para inputs do mesmo tipo que aqueles que participaram de seu treinamento. O autor, no entanto, pergunta: mas o que conta como “mesmo tipo”? Ora, isso deverá ser configurado pelo projetista, pois a rede frequentemente gera resultados descontextualizados.

Veja-se o relato de Dreyfus sobre um pioneiro projeto conexionista de discriminação de padrões desenvolvido junto ao exército. Uma rede neural havia sido treinada através de fotografias para reconhecer tanques parcialmente ocultos numa floresta. O sucesso dos primeiros testes – incluindo generalizações de sucesso com fotos que não fizeram parte do treinamento – deixou a todos motivados. Porém, para confirmar o êxito do aprendizado da rede, novas fotos foram tiradas. Contudo, desta vez o sistema fracassou em reconhecer os tanques atrás das árvores. Mais tarde, percebeu-se que a rede estava fazendo discriminações entre a floresta com sombras (primeiras fotos tiradas em dia ensolarado) e sem (novo lote produzido em dia nublado).

Apesar de os discursos conexionistas insistirem que a rede aprende sozinha, pode-se perceber mais uma vez a dependência dos sistemas informáticos nas determinações externas de seus programadores. No caso descrito, os projetistas precisaram determinar qual a classe de respostas apropriadas ao contexto de interesse. Muito diferente do que ocorre com a inteligência humana, cujo senso comum lhe capacita adaptar-se a diferentes contextos (Dreyfus, 1992).

Neural-network modelers were initially pleased that their nets were blank slate (*tabula rasa*) until trained, so that the designer did not need to identify and provide anything resembling a pretraining intelligence. Recently, however, they have been forced by the problem of producing appropriate, human-like generalizations to the recognition that, unless the class of possible generalizations is restricted in an appropriate a priori manner, nothing resembling human generalizations can be confidently expected. Consequently, after identifying in advance the class of allowable human-like generalizations appropriate to the problem (the hypothesis space), these modelers then attempt to design the architecture of their networks so that they transform inputs into outputs only in ways that are in the hypothesis space. Generalization would then be possible only on the designer's terms. While a few examples will be insufficient to identify uniquely the appropriate member of the hypothesis space, after enough examples only one hypothesis will account for all the examples. The network will then have learned the appropriate generalization principle. That is, all further input will produce what, from the designer's point of view, is the right output¹¹(Dreyfus, 1992, p. xxvii).

¹¹ Tradução do autor: Os modeladores de redes neurais estavam inicialmente satisfeitos que suas redes eram um suporte vazio (*tabula rasa*) até serem treinadas, de modo que o projetista não precisava identificar e fornecer nada que se parecesse com uma inteligência em treinamento prévio. Recentemente, contudo, eles foram forçados pelo problema de produzir generalizações apropriadas e semelhantes às humanas ao reconhecimento de que, a não ser que a classe de generalizações possíveis seja restrita a uma forma prévia apropriada, nada se parecendo com generalizações humanas pode ser esperado com certeza. Consequentemente, depois de identificar previamente a classe permitida de

O pretense aprendido “ativo” das redes neurais poderia ser aproximado, talvez, daquilo que Bateson (2000) chama de aprendizado I¹², ou seja, uma mudança na especificidade da resposta através da correção de erros de escolha dentro de um conjunto de alternativas. Esse tipo de aprendizado seria caracterizado, segundo o autor, pelas atividades de condicionamento e aprendizagem por repetição e reforço realizadas em condições laboratoriais.

Cognitivismo: a mente como programa de computador

É preciso apontar que muitos debates sobre inteligência artificial contaminam-se de metáforas por demais otimistas e imprecisas gerando fantasias que passam a ser vistas como reproduções quase perfeitas do comportamento humano. Questões como inteligência, autonomia, aprendizado, percepção, para citar algumas, são tratadas de forma trivial. Analisa-se as potencialidades tecnológicas com conceitos de outro contexto não-tecnológico. Ou seja, descreve-se o funcionamento informático com explicações psicológicas, ao passo que se joga para baixo do tapete a própria biologia do ser humano. Para nublar ainda mais o contraste entre as diferentes naturezas do ser humano e da máquina informática, postula-se ainda que a mente humana não passa de um programa de computador.

Assim, se a) “A mente é um programa de computador” e b) “Mentes humanas aprendem e demonstram inteligência”, logo c) “Programas de computador aprendem e demonstram inteligência”. Ora, diante dessa distorção intencional, fica fácil legitimar as explicações e resultados da inteligência artificial através de tal silogismo falacioso. Decorrentes dessa interpretação ligeira são as conclusões de que os “agentes inteligentes” e outros programas agem e aprendem autonomamente de forma equivalente à humana.

generalizações de estilo humano apropriadas ao problema (o espaço da hipótese), esses modeladores então procuram projetar a arquitetura de suas redes de forma que elas transformem entradas em saídas apenas em um estilo que esteja dentro do espaço da hipótese. A generalização poderia então ser possível somente nos termos do projetista. Enquanto alguns poucos exemplos serão insuficientes para identificar unicamente o membro apropriado do espaço de hipótese, depois de exemplos suficientes apenas uma hipótese será considerada para todos os exemplos. A rede terá então aprendido o princípio generalizador apropriado. Isto é, toda entrada futura irá produzir o que, do ponto de vista do projetista, é a saída correta.

¹² Harries-Jones (1995) observa que o artigo de Bateson “The logical categories of learning and communication” foi uma dura crítica à tradição behaviorista, que supunha que a aprendizagem dependia da memorização por repetição sob controladas situações de estímulo-resposta. Bateson (2000) argumenta que o estudo da aprendizagem não pode ser conduzido desconsiderando-se os contextos em que ela ocorre (que inclusive oferece limitações ao comportamento). Segundo ele, mesmo o condicionamento animal falha fora das condições laboratoriais. Ou seja, o aprendizado humano não se dá por mera associação entre o estímulo provindo do ambiente e sua correlata resposta. Para o autor, os organismos levam sempre em conta o contexto do aprendizado (característica do que ele chama de aprendizado II e III). Na verdade, esse entendimento acompanha sua recorrente correlação entre conteúdo e contexto (que influenciou significativamente a abordagem relacional da comunicação humana). Enfim, as formas superiores de aprendizagem dependem do que Bateson chama de “aprender a aprender”. Não se trata da acumulação de informações individuais (como supõe o behaviorismo), mas de uma capacidade de problematização, de compreensão contextualizada (“pontuando” a seqüência de eventos). Enfim, diferentemente do que propõe o associacionismo (instrução de alternativas e correção de erros por reforço), Bateson propõe que as formas mais complexas de aprendizado envolvem a própria transformação dos conjuntos de “alternativas”, a partir do qual as decisões são tomadas.

Na verdade, a comparação da mente humana a um software se alicerça e busca legitimidade na Ciência Cognitiva, no chamado cognitivismo.

Enquanto o behaviorismo evitava a mente, vendo-a como uma “caixa preta”, o cognitivismo (entendido por muitos como neo-behaviorismo) quer ocupar-se dos processos e mecanismos internos. Em virtude dessa rotação focal, os cognitivistas acabam por suprimir os problemas da vida cotidiana. E por tratar os sujeitos como “processadores de informações”, também ofuscam os problemas interpessoais (Gergen, 1999).

Diante das aproximações da mente a um software e dos desenvolvimentos tecnológicos que buscam reproduzir características humanas, Maturana (2001, p. 194) responde:

O humano não é uma expressão de algum programa de computador que especifica certos modos de funcionar, é uma **maneira de viver relacional** que implica seu ser fundado numa corporalidade básica. Sim, muitos de nossos órgãos podem ser substituídos por órgãos artificiais, mas haverá substituição apenas se eles substituírem os órgãos originais na realização do viver humano. Sim, é possível eventualmente se fazer robôs que claramente se comportem como nós, mas sua história será presa à sua corporalidade, e à medida que eles existirem como entes compostos em domínios de componentes diferentes dos nossos, os domínios de realidades básicas que eles gerarão serão diferentes de nossos [grifo meu].

O comportamento para a perspectiva cognitivista, segundo conclui Gergen (1999), não se origina a partir do “mundo como ele é” (uma visão ambientalista), mas sim da concepção individual do mundo. Isto é, o mundo é uma projeção ou um subproduto da cognição de um indivíduo. O autor também denuncia que os entes cognitivos são apresentados como máquinas de estruturas estáveis e persistentes. Diante desse salto para uma nova forma de solipsismo, o autor posiciona a seguinte dúvida: como uma categoria cognitiva, uma conjunto de proposições ou uma estrutura representacional produz ação?

Em virtude desses problemas, o movimento cognitivista vem sendo tachado, conforme relata Gergen (1999), de abstrato, impessoal, tecnicista e focado exclusivamente na informação (que será logo substituída por nova informação).

Como se pode detectar até aqui, o expediente de aproximar a cognição humana do operar informático perde-se ora pelo solipsismo, ora pelo representacionismo. Sobre esse perambular errático, Maturana e Varela (1995, p. 195) sentenciam que “A metáfora tão em voga do cérebro como um computador é não só ambígua como francamente equivocada”.

Diante da expectativa de muitos de que a modelização conexionista abriria caminho para uma compreensão das relações entre cérebro e comportamento, Button et al (1998, p. 167) rebatem:

...insistimos na tese de que as análises teóricas da conduta humana não são facilitadas por empréstimos tomados das tecnologias produzidas na computação; de fato, a importação não-crítica para dentro das ciências humanas de modelos e estruturas conceituais inicial e convincentemente elaboradas no desenvolvimento de artefatos computacionais é mais um estorvo do que um auxílio.

Experiência e aprendizagem

Na verdade, as pesquisas em inteligência artificial – tanto na perspectiva simbólica quanto na conexionista – são guiadas por uma epistemologia empirista de aprendizagem. Crítico ferrenho dessa postura epistemológica, Piaget (1987, p. 341) afirma que a essência do empirismo é colocar a “coisa” ou o “dado imediato” como ponto de partida da evolução intelectual – o que implica necessariamente em atitude receptiva do espírito. Isto é, o progresso da inteligência consiste apenas “em construir vias mais reduzidas para as reações ou reações cada vez mais ‘diferidas’, destinadas a contornar ou dispensar o contato direto para só o reencontrar de longe em longe”¹³.

Para o associacionismo empirista, a confirmação da associação provém da experiência, de um reforço que consolide uma determinada conduta. Piaget (1987, p. 339) observa, no entanto, que o empirismo defende uma questionável concepção da experiência e sua ação:

Por uma parte, tende a considerar a experiência como algo que se impõe por si mesmo, sem que o sujeito tenha de organizá-la, isto é, com se ela fosse impressa diretamente no organismo sem que uma atividade do sujeito seja necessária (sic) à sua constituição. Por outra parte, e por consequência, o empirismo encara a experiência como existente em si mesma, quer ela deva o seu valor a um sistema de “coisas” exteriores, totalmente feitas, e de relações dadas entre essas coisas (empirismo metafísico), quer consista num sistema de hábitos e de associações auto-suficientes (fenomenismo).

Por outro lado, Piaget (1987, p. 339) vai entender que mesmo aquilo que se chama de “reflexo condicionado”, ao ser confirmado pela experiência, ingressa “num esquema de conjunto, quer dizer, deixa de estar isolado para converter-se numa parte integrante de uma totalidade real”.

A rigor, Piaget (1987) não deixa de reconhecer o papel fundamental da experiência no desenvolvimento da inteligência. Tampouco nega o papel essencial que o meio exterior tem nesse processo. O autor vai defender ainda que o conhecimento corresponde ao problema das relações entre o organismo e o meio (Piaget, 1996). Entretanto, insiste que “a experiência não é recepção, mas ação e construção progressivas” (1987, p. 342). O conhecimento, pois, não é gerado por simples pressão do meio, pela intervenção de uma causa exterior ao organismo, o que negaria todo o operar endógeno do sujeito. Para Piaget (1987, p. 344), as coisas “nunca poderão ser concebidas independentemente da atividade do sujeito”.

Projetos de inteligência artificial, contudo, dependem de um intenso cadastramento de dados e regras (I.A. simbólica) ou demorado “treinamento” da rede neural¹⁴ e correção de erros. Em contraste à cognição humana que caminha na

¹³ Decorrente desse encaminhamento é aquela questão que o autor apelida de “pergunta americana”: “Que se deve fazer para acelerar este desenvolvimento?”. É justamente nesse intuito que se avolumam teorias e métodos de aprendizagem por associação, por condicionamento ou por reforço externo. Entretanto, segundo o olhar piagetiano, eles produzem pouca mudança no pensamento lógico ou uma extraordinária mudança momentânea, desprovida de uma real compreensão (Becker, 1997).

¹⁴ Button et al (1998, p. 165) diante da comparação entre projetos simbólicos e conexionistas relatam a conclusão de Clarke de que “apresentada a formas inteiramente transpostas de problemas inicialmente postos, uma rede tem de ser maciçamente retreinada, ao passo que uma máquina de Turing pode ser

direção de progressiva adaptação com seu meio (e, portanto, maior flexibilidade diante das perturbações), quanto mais informações um sistema de inteligência artificial contiver, mais demorado e difícil será seu trabalho.

Indeed, AI researchers have long recognized that the more a system knows about a particular state of affairs, the longer it takes to retrieve the relevant information, and this presents a general problem where scaling up is concerned. Conversely, the more a human being knows about a situation or individual, the easier it is to retrieve other relevant information¹⁵ (Dreyfus, 1992, p. *xxi*).

Para que se compreenda bem o processo de aprendizagem, portanto, é preciso entender como o sujeito constrói e inventa o conhecimento, e não somente como ele repete e copia (Becker, 1997). Com esse entendimento, Piaget (1996, p. 45) mostra que o comportamento humano consiste em “um conjunto de escolhas e de ação sobre o meio, que organiza de maneira ótima as trocas”.

Becker (1997, p. 87-88) conclui que:

Se a atividade conceitual é originária de outras formas mais elementares de atividade, construídas mediante tomadas de consciência e abstrações reflexivas (que também são formas de atividade do sujeito), constitui um contra-senso conceber uma aprendizagem à margem desse processo ou contra ele. Assim uma aprendizagem, conforme Piaget ou Freire, será entendida sempre como produtor de uma relação ativa entre o sujeito e o objeto (ou entre sujeitos), entre ação e reflexão, entre teoria e prática, portanto, como uma relação eminentemente transformadora da realidade.

Contextualização

Ainda no que toca à dificuldade de máquinas alopoiéticas¹⁶ (segundo terminologia desenvolvida por Maturana e Varela) aprenderem e interagirem na experiência com seu meio (incluindo aí os outros interagentes), interessa também acompanhar algumas observações de Bateson sobre o ele que chama de aprendizado II. Conforme define, trata-se de uma mudança corretiva no conjunto de alternativas a partir do qual a escolha é feita; ou uma mudança em como a seqüência da experiência é “pontuada”¹⁷. Em sua argumentação, o autor quer chamar atenção para a contextualização, questão sempre cara em sua obra.

Entretanto, diferentemente dos organismos, as máquinas alopoiéticas não conseguem contextualizar suas ações e o que aprendem. Tais máquinas não percebem o que Bateson chama genericamente de “marcadores de contexto” sem que eles sejam informados deliberadamente por um operador. Por exemplo, como distinguir que

ajustada com (relativa) facilidade (por exemplo, mudando seus indicadores de dados ou reconstruindo suas estruturas de dados)”.

¹⁵ De fato, pesquisadores de IA há muito reconheceram que quanto mais um sistema conhece sobre alguma coisa, mais tempo ele leva para recuperar a informação relevante, e isso apresenta um problema geral no que toca o escalonamento. Inversamente, quanto mais um ser humano sabe sobre uma situação ou indivíduo, mais fácil é de recuperar outra informação relevante.

¹⁶ Que não participam de sua própria criação e cujas fronteiras são determinadas por agentes externos.

¹⁷ Para o observador externo a comunicação é uma seqüência ininterrupta de trocas. Por outro lado, Watzlawick et al (1967) afirmam – a partir dos trabalhos de Bateson – que os participantes na interação “pontuam” a seqüência de eventos. Trata-se de uma organização dos eventos comunicativos, a partir da qual emerge a interpretação da relação.

atores representando Romeu e Julieta não estão de fato tentando se matar? Para Bateson, o contexto ajuda o organismo a decidir em qual conjunto de alternativas ele deve basear sua escolha. A tecnologia informática, no entanto, não pode perceber o contexto (no caso do teatro, as cortinas, a platéia, o programa da peça, etc.) sem que alguém digite tais inputs. E, mais uma vez, a rede não compreende o que são teatro, cortinas, platéia, ator. Logo, sem conseguir contextualizar as informações sobre as falas dos atores, provavelmente um sistema informático “inteligente” chamaria imediatamente a ajuda dos bombeiros e da polícia!

As qualidades atribuídas à personalidade de uma pessoa dependem também da contextualização dessas informações. Por exemplo, dizer que Fulano é ansioso, exibicionista, narcisista, competitivo, covarde, fatalista, cuidadoso ou mesmo desleixado é interpretar o que ocorre entre ele e algo ou outra(s) pessoa(s). Em outras palavras, ninguém é competitivo ou fatalista no vácuo. Bateson (2000) afirma que referir-se a alguém dessa forma é descrever a transação entre a pessoa, seus materiais e suas relações humanas.

Em uma amizade duradoura, por exemplo, os amigos tenderiam a um acordo sobre a pontuação do fluxo dos eventos comunicativos e sobre a natureza de seu próprio relacionamento. Por outro lado, segundo observam Watzlawick et al (1967), a discordância na pontuação da seqüência encontra-se na raiz de muitas lutas em torno das relações. Por exemplo, em um casal com problemas, o marido frustrado justifica seu retraimento como defesa contra as implicâncias da esposa. Ela, por sua vez, diz que o critica por causa de sua passividade. Em outras palavras, na seqüência de eventos¹⁸ em que os atos ímpares são a retração do marido e os pares a hostilização da esposa, o primeiro percebe apenas as tríades 2-3-4, 4-5-6, 6-7-8, etc. A pontuação da esposa é diferente, reconhecendo apenas as tríades 1-2-3, 4-5-6, 7-8-9, etc.

É interessante ver Bateson trazendo para a arena de debates sobre aprendizado a questão do contexto e da pontuação dos eventos interativos, visto que, em geral, esses problemas não freqüentam as discussões sobre o tema. Ora, o aprender sobre as coisas do mundo não é atividade desvinculada da historicidade interativa, nem das relações e transações do sujeito em seu meio, com outros sujeitos. Deve-se observar que a impossibilidade atual dos sistemas informáticos em contextualizar os conhecimentos e interpretar a historicidade das relações interindividuais, pontuando-as ativamente, é fator fundamental que lhes impede de participar de interações mútuas¹⁹. Tal limitação também segura a interação entre homem e máquina alopoiética em uma relação reativa, mesmo que o primeiro seja um ser ativo e inventivo. Nas interações em que participam, os interagentes informáticos – mesmo aqueles de inteligência artificial – não conseguem ultrapassar as barreiras impostas por seu operar determinístico e reativo (nem autônomo, nem criativo). E como observam Teixeira (1998) e Dreyfus (1993), nem mesmo os sistemas

¹⁸ Bateson, citado por Watzlawick et al (1967), diz que a classificação de “estímulos”, “reforços” e “respostas” pelo psicólogo behaviorista, cuja atenção se volta sempre para seqüências de permutas muito curtas, é falha, pois todos os atos em uma seqüência interativa extensa poderiam ser rotulados como “estímulos”, “reforços” ou “respostas”. Nesse sentido, a classificação behaviorista é apenas uma pontuação particular do experimentador no intuito de encaixar o fragmento da interação que manipula dentro de sua perspectiva S-R.

¹⁹ Para uma conceituação sobre interação mútua e interação reativa, ver Primo (1998).

conexionistas conseguem interagir efetivamente com seu meio, pois ainda dependem das determinações do programador. Enquanto isso, o homem precisa se adequar às condições da máquina para que a interação entre eles possa se estabelecer.

Considerações finais

Os projetos de inteligência artificial, como se pretendeu mostrar, focam-se basicamente na resolução lógica de problemas. Em sua ânsia de “modelar” a inteligência humana, a reduzem ao processamento de dados, deixando de fora tudo o que é ambíguo e “não-racional” (mas inseparável da subjetividade humana). Capra (1996, p. 216) aponta, porém, que as “decisões humanas nunca são completamente racionais, estando sempre coloridas por emoções, e o pensamento humano está sempre encaixado nas sensações e nos processos corporais que contribuem para o pleno espectro da cognição”.

Enfim, a diferença de natureza entre seres autopoieticos e máquinas alopoiéticas impede uma equiparação das formas como interagem com o meio (incluído aí os outros interagentes). Enquanto o conhecimento dos primeiros dá-se na ação e eles interagem de forma ativa e criadora, as máquinas alopoiéticas reagem segundo as determinações externas inscritas previamente em suas estruturas. Isso enrijesse suas interações e limita inclusive as interações travadas com seres autopoieticos.

Referências Bibliográficas

Bateson, G. Mind and nature: a necessary unity. Nova Iorque: Bantam New Age Books. 1980

_____. Steps to an ecology of mind. Chicago: University of Chicago Press. 2000. 533 p.

Becker, F. Da ação à operação: o caminho da aprendizagem em J. Piaget e P. Freire. Rio de Janeiro: DP&A. 1997. 160 p.

Button, G., J. Coulter, et al. Computadores, mentes e condutas. São Paulo: Unesp. 1998

Capra, F. A teia da vida: uma nova compreensão científica dos sistemas vivos. São Paulo: Cultrix. 1996. 256 p.

Freire, P. Pedagogia do Oprimido. Rio de Janeiro: Paz e Terra, v.21. 2001. 192 p. (O mundo, hoje)

Harries-Jones, P. A recursive vision: ecological understanding and Gregory Bateson. Toronto: University of Toronto Press. 1995. 358 p.

Maturana, H. e F. Varela. A árvore do conhecimento: as bases biológicas do conhecimento. São Paulo: Editorial Psy. 1995. 281 p.

_____. De máquinas e seres vivos: autopoiese - a organização do vivo. Porto Alegre: Artes Médicas. 1997. 138 p.

Maturana, H. Cognição, Ciência e Vida Cotidiana. Belo Horizonte. 2001

Piaget, J. O Nascimento da Inteligência na Criança. Suíça: Editora Guanabara. 1987



<http://www.ufrgs.br/limc>

_____. *Biologia e conhecimento: ensaio sobre as relações entre as regulações orgânicas e os processos cognoscitivos*. Petrópolis: Vozes. 1996. 423 p.

Primo, A. F. T. *Interação Mútua e Interação Reativa: uma proposta de estudo*. Intercom 1998 - XXI Congresso Brasileiro de Ciências da Comunicação. Rio de Janeiro, 1998. p.

Primo, A. F. T. e L. R. Coelho. *Comunicação e inteligência artificial: interagindo com a robô de conversação Cybelle*. In: L. G. M. Motta, M. H. Weber, et al (Ed.). *Estratégias e culturas da comunicação* ed. Brasília. Brasília: Editora Universidade de Brasília, 2002. *Comunicação e inteligência artificial: interagindo com a robô de conversação Cybelle*, p.83-106

Teixeira, J. D. F. *Mentes e máquinas: uma introdução à ciência cognitiva*. Porto Alegre: Artes Médicas. 1998. 179 p.