

'INTELIGÊNCIA' ARTIFICIAL E 'APRENDIZADO' DE MÁQUINA: PROBLEMAS E SOLUÇÕES

Valdemar W. Setzer

Prof. Titular Sênior, Depto de Ciência da Computação

Instituto de Matemática e Estatística da USP

www.ime.usp.br/~vwsetzer

1. Introdução

Este artigo é uma grande extensão do artigo “Problemas da ‘inteligência’ artificial”, escrito para divulgação do COINS 2019 (Congresso Internacional de Secretariado), onde dei uma palestra sobre o assunto, mas abordando outros aspectos.

Na seção 2 é tratada brevemente a questão das nomenclaturas usadas na computação, seguindo-se na 3 um breve histórico da ‘inteligência artificial’ (IA); na 4 é feita uma curta descrição do assim chamado ‘aprendizado de máquina’ (AM) e na 5 os problemas que ele causa. Na seção 6 é chamada a atenção de que o AM representa um novo paradigma de programação, na 7 é abordado o impacto social da IA e na 8 a questão da eliminação de postos de trabalho. Na 9, são mostradas algumas diferenças entre IA e inteligência humana. Na seção 10 são apresentadas respostas a perguntas feitas em um artigo sobre IA e dadas respostas às mesmas, e na 11 são descritos os grandes perigos causados pela IA e pelo AM para, finalmente, na 12 serem sugeridas algumas soluções para os problemas apresentados.

Para mais detalhes sobre IA e outros aspectos não abordados aqui, veja-se meu antigo artigo a respeito [2].

2. A questão da nomenclatura

Vou utilizar, a contragosto, as denominações usuais, ‘inteligência’ artificial (IA, *artificial intelligence*) e ‘aprendizado’ de máquina (AM, *machine learning*). Isso deve-se ao fato de que elas estão erradas e são enganosas. Não sabemos o que é inteligência, portanto não se pode classificar algum tipo dela como ‘artificial’. Não sabemos como o ser humano aprende (se soubéssemos, o curso de medicina não levaria 6 anos e o de direito não levaria 5 anos), portanto, não se pode dizer que uma máquina aprende. No primeiro caso, uma denominação correta poderia ser ‘simulação de comportamentos humanos’; comportamentos tanto físicos como cognitivos. No segundo caso, poderia ser ‘programas adaptativos’.

Como a tecnologia, nenhuma denominação é neutra: ela acaba influenciando a mentalidade dos seus usuários. As denominações usuais IA e AM produzem uma dupla mistificação: elevam as máquinas, os computadores, ao nível humano, e degradam a imagem que se deveria fazer do ser humano. Por detrás dessas denominações há uma concepção de mundo, a materialista ou fisicalista, que considera que só há substâncias e processos físicos no universo e no ser humano. Essa concepção leva à ideia de que o ser humano é uma máquina e age como uma, o que para mim é uma tragédia pior do que se considerar que o ser humano é um animal, que espalhou-se principalmente devido ao darwinismo. Pode-se ter dó, compaixão dos animais, pois eles sentem dor como nós; mas seria uma aberração psicológica ter dó de uma máquina, por exemplo, ter dó de desligar o computador ou um robô. Existem sociedades protetoras dos animais, mas não há sociedades protetoras das máquinas. O que se está fazendo e se vai fazer com o ser humano, devido a essa mentalidade, será uma aplicação de uma das ‘Leis de Setzer’ (ver a referência [1]): “A miséria causada pelo ser humano ultrapassa o pior pessimismo.” As tragédias provocadas

pelo nazismo e pelo comunismo serão fichinha perto do que está para vir, se não houver mudança na mentalidade.

Tenho muitas evidências para a minha hipótese de trabalho de que o ser humano não é um ser puramente físico, o que abordo em vários artigos disponíveis em meu *site*. Aliás, aplico a mesma hipótese para animais e plantas. Em particular, se o ser humano é um ser puramente físico, não se pode atribuir a ele o livre arbítrio, pois da pura matéria não pode advir liberdade, já que ela está inexoravelmente sujeita às 'leis' e condições físicas. Sem livre arbítrio, não se pode atribuir responsabilidade, e nem mesmo dignidade ao ser humano.

Quando se usam as denominações IA e AM, dever-se-ia estar consciente dos efeitos psicológicos negativos que elas produzem. Especialmente, essa ressalva deveria ser feita quando esses sistemas são abordados no ensino médio ou superior (não cabem de modo algum no ensino fundamental, incluindo a robótica).

É interessante notar que não estou sozinho nessa crítica à nomenclatura "inteligência artificial". Por exemplo, S. Quintarelli escreveu um artigo nesse sentido [13]. Devido à abreviatura (SALAMI) que ele sugere para a IA, pode-se pensar que o artigo é piadista; no entanto, o conteúdo é sério. Em particular, ele formula interessantes questões sobre IA que serão aqui respondidas na seção 10.

Outros exemplos de antropomorfização indevida em nomenclaturas da computação são 'memória' e 'leitura de dados'. Não se sabe como a nossa memória funciona, mas sabe-se como a de um computador funciona. Antigamente ela era chamada muito propriamente de 'unidade de armazenamento', *storage device*). Quanto à leitura, só o ser humano lê; um computador apenas recebe os dados de entrada, e os processa seguindo seus programas. Um aspecto essencial da leitura é a compreensão do texto escrito. No entanto, computadores são máquinas sintáticas, não tem absolutamente nenhuma compreensão, isto é, semântica, no sentido humano.

3. Evolução da IA

A expressão *artificial intelligence* foi inventada pelo grande pioneiro da ciência da computação John McCarthy em 1955, por ocasião de um simpósio que houve no Dartmouth College, nos EUA [3, 4]. Ela suscitou muitas esperanças de que os computadores poderiam equiparar-se ao ser humano em todas as suas funções e ações, especialmente as intelectuais; McCarthy estava crente disso. Durante muito tempo, a IA não cumpriu as visões dos que tinham essa concepção, para decepção dos fisicalistas futuristas. Alguns sucessos foram totalmente mal interpretados, como o do xadrez eletrônico.

Considerava-se que jogar bem xadrez era uma mostra de inteligência. Quando a máquina Deep Blue da IBM venceu em 1997 o segundo torneio da empresa (a primeira máquina, que perdeu por 4 a 2, chamava-se, pouco modestamente, Deep Thought) contra o então campeão mundial Garry Kasparov [5], os visionários da IA cantaram vitória, inclusive afirmando que a máquina tinha ultrapassado o ser humano. Essa última afirmação precisa ser colocada em seu devido contexto: desde sua introdução em 1946 (com a máquina ENIAC) e o primeiro comercial em 1951 (UNIVAC) os computadores já haviam ultrapassado o ser humano em uma área muito restrita: fazer cálculos aritméticos repetitivos – com muito mais rapidez e confiabilidade. O mesmo passou-se com o xadrez eletrônico. Mas o que significou, realmente, o sucesso da Deep Blue? Ocorre que, em primeiro lugar, o desfecho do torneio foi de 2 vitórias da DB, uma do Kasparov e 3 empates. Em segundo lugar, o xadrez é um jogo matemático, isto é, o tabuleiro, as peças e seus movimentos, bem como as regras do jogo podem ser descritos matematicamente. Em terceiro lugar, a DB fazia 200 de milhões de previsões (cálculos) de lances por segundo. Em quarto lugar, Kasparov certamente fazia mentalmente apenas algumas dezenas de lances.

Assim sendo, como foi possível que um ser humano pudesse empatar três partidas e vencer uma contra uma supermáquina matemática, jogando um jogo matemático, quando a máquina tinha capacidade milhões de vezes maior? Essa pergunta não foi formulada na época e constou de meu artigo sobre o xadrez eletrônico [6].

Minha conjectura é que isso foi possível porque Kasparov, ao contrário da DB, *não calculava!* Como todo grande mestre de xadrez, ele provavelmente olhava a posição do tabuleiro e tinha rapidamente a *intuição* do lance correto; aí ele certamente testava sua ideia, simulando mentalmente lances adiante, para ver se a intuição tinha sido correta. Ora, a intuição é uma ideia vinda do nada, portanto, algo inexplicável. De fato, os grandes mestres em xadrez não conseguem explicar como chegam mentalmente ao lance correto, de modo que nunca foi possível colocar as suas estratégias numa máquina.

Qualquer concepção de que o cérebro gera os pensamentos é mera especulação, pois não se sabe como o cérebro funciona. Cientificamente, o máximo que se deveria dizer é que o cérebro participa das atividades mentais. Mais recentemente, o programa AlphaZero usou técnicas de AM para, partindo do zero, apenas com as regras do jogo, jogar contra a própria máquina e calcular parâmetros que o fizeram ganhar de todas as outras máquinas de jogar xadrez [14].

4. 'Aprendizado' de máquina

O maior sucesso recente da IA é o 'aprendizado' de máquina (AM) [7], devendo-se ressaltar que a o AM é apenas uma parte da IA. Trata-se de uma técnica de dar a um programa uma enorme quantidade de conjuntos de dados de entrada e de saída; o programa então, em uma fase denominada de 'treinamento' (novamente uma denominação antropomórfica indevida, pois deveria ser aplicada apenas a seres humanos e a animais), calcula parâmetros para transformar um conjunto de dados de entrada em um conjunto de dados de saída correspondente, também fornecido ao programa. Posteriormente, depois de muitos casos desses, dá-se um novo conjunto de dados de entrada, obtendo-se então dados de saída que devem se aproximar o máximo possível do que seria esperado. Trata-se, portanto, de um problema de otimização matemática: diminuir o erro dos dados de saída em relação ao que seria de esperar. Uma das técnicas é usar uma 'rede neural artificial' [8] – novamente uma denominação errada: não se sabe como um neurônio funciona, e muito menos a rede dos neurônios no cérebro (calcula-se que haja 86 bilhões de neurônios e mais de um trilhão de conexões entre eles, as sinapses). As redes neurais artificiais contêm várias camadas de nós abstratos interligados. A primeira camada de nós recebe os dados de entrada, cada nó recebendo um dado. Na segunda camada de nós, cada nó é ligado a alguns nós da primeira camada; um cálculo combinando os dados de cada nó da primeira camada produz um resultado em cada nó da segunda camada, segundo uma fórmula que usa parâmetros de transformação do dado que sai de cada nó da camada anterior. Um caso simples é usar como entrada em cada nó N a soma de todas as saídas dos nós que são ligados a N e que pertencem ao nível anterior da rede. Uma terceira camada tem cada um de seus nós ligado a alguns da segunda camada, recebendo os resultados desses últimos e combinando esses dados por meio de um cálculo. E assim por diante, até a última camada de nós, a de saída. Cada cálculo que se faz para gerar a entrada em cada nó será aqui determinado como sendo um *parâmetro* da rede. Note-se que a topologia da rede, isto é, as interligações entre os nós, também pode ser considerada como parâmetro da rede.

Um exemplo que se tornou clássico foi dar como entrada a uma rede os dados de muitas fotos digitalizadas de cães e de lobos. Para cada uma, especificava-se se a saída devia indicar um cão ou um lobo. Depois de muito 'treinamento', a rede era capaz de prever com grande chance de acerto se uma nova foto era de um cão ou de um lobo. Voltaremos a esse exemplo mais adiante.

O AM está sendo usado em inúmeras aplicações, por exemplo, no sistema financeiro, para indicar se uma pessoa desejando tomar um empréstimo tem chance de honrá-lo ou de se tornar inadimplente. Ou em entrevistas para contratação de funcionários, dando-se o perfil de um candidato e o sistema concluindo se ele tem boa chance de ser um bom funcionário. Ou dar-se uma foto de um tecido humano e o sistema detectar se a conformação das células indica um tumor.

5. Problemas do ‘aprendizado’ de máquina

Há três problemas fundamentais no AM.

5.1 Uso de programas fechados

Muitas aplicações de AM utilizam programas fechados, isto é, o usuário do sistema não sabe como o programa calcula os parâmetros. Por exemplo, não se conhece quantas camadas a rede tem, quantos nós existem em cada camada e como as saídas de cada nó são calculadas a partir dos dados dos nós da camada anterior a ele ligados. Isto é, o usuário de tal rede não sabe que critérios são usados para transformar os dados de entrada nos de saída.

5.2 Parâmetros desconhecidos

Em geral, usuários empregam um sistema de AM para transformar certos dados de entrada nos de saída, mas não examinam os parâmetros calculados pelos sistemas, isto é, não sabem como o sistema chega aos dados de saída. Em outras palavras, não se sabe que algoritmo o programa usa para chegar aos resultados.

O exemplo da distinção entre cães e lobos citado acima tornou-se famoso, pois os pesquisadores tiveram a curiosidade de examinar o que os parâmetros significavam. Tiveram a surpresa de verificar que as conformações das cabeças dos animais não estavam sendo usadas, como esperado. Em lugar disso, os parâmetros faziam o programa usar as imagens do ambiente: cães em geral estavam em ambientes domésticos, ruas etc., e as fotos dos lobos mostravam savanas, neve, florestas etc.

Somente em casos de pesquisas os parâmetros são examinados depois do ‘treinamento’. Mas há aqui um problema adicional: numa rede muito grande, com milhares de nós, eventualmente é impossível deduzir qual o significado de cada nó, isto é, qual a relação dele com os dados de entrada.

5.3 Combinação dos dois fatores

Muitas aplicações de AM são usadas sem que o usuário saiba como funcionam os algoritmos de cálculos dos parâmetros e, além disso, não conhece esses parâmetros. Isto é, o programa é uma caixa preta.

6. Um novo paradigma de programação

Antes de sistemas que usam o AM, um programador desenvolvia um algoritmo, e depois o expressava por meio de comandos de uma linguagem de programação (por exemplo, ver características da linguagem Python, com exemplos, em [12]). A partir de um programa assim formulado, ele é entregue ao computador, sendo então processado e convertido em “linguagem de máquina”, que é posteriormente interpretada pela máquina.

O programador tinha pensado em cada comando da linguagem que tinha usado, isto é, teoricamente ele compreendia o que cada comando do programa devia fazer. O ‘teoricamente’ deve-se ao fato de que muitos programas contêm erros de programação e em certos processamentos não fazem o que o programador tinha planejado. Por outro lado, quando o programador corrige vários erros ao mesmo tempo, em geral ele não sabe qual correção fez o programa começar a funcionar a contento. É impossível provar formalmente que um programa

de um porte médio funciona; testes comprovam apenas que o programa funciona bem para os dados do teste.

O AM quebrou esse paradigma. Agora um programa de AM modifica seu comportamento por meio dos cálculos dos parâmetros que ele usa. Não se examinando cada um desses parâmetros, não se sabe como o programa funciona.

Em outras palavras, um sistema de AM executa um algoritmo desconhecido e em geral imprevisível.

7. Impacto social da IA

Se um sistema de IA é utilizado para controlar máquinas, ele pode ser muito útil. Por exemplo, a no Brasil 90% dos acidentes de trânsito é devida a falhas humanas [15]. Se um sistema de automação total de veículos diminuir os acidentes, seu emprego é recomendável. A respeito disso, é interessante fazer uma observação: aqui no Brasil, motoristas muitas vezes não respeitam as regras de trânsito, por exemplo, avançando em faixas em semáforos quando há congestionamentos. Um carro autônomo seguirá essas regras, não vai ocupar essas faixas e com isso terá que esperar um tempo extenso até que o trânsito não esteja mais congestionado, o que provocará o desespero dos motoristas atrás dele. Ou todos os motoristas passam a respeitar as regras de trânsito, ou o carro autônomo terá que ser colocado em modo manual para também poder desrespeitá-las.

O grande problema de IA ocorre quando um sistema é usado para tomadas de decisões sociais ou humanas, sem que seres humanos participem da última fase das decisões. É muito importante reconhecer que computadores não tomam decisões, eles fazem escolhas lógicas. Decisões humanas envolvem sentimentos; máquinas não têm sentimentos e em minha conjectura jamais terão [2]. Se uma máquina é usada sozinha para substituir decisões humanas, ela trata as pessoas, às quais as escolhas dela se aplicam, como se fossem máquinas, como seriam os casos citados na seção 3, e mais aplicações desse tipo no sistema judiciário, diagnóstico e procedimentos médicos etc.

Como o ser humano tem sentimentos, toda lei social é estabelecida levando-se-os em conta. Um contraexemplo seria considerar-se que há excesso de população no mundo. Assim sendo, um raciocínio puramente lógico seria o seguinte: vamos passar leis permitindo pessoas matarem umas as outras, até que a população diminua. É por uma questão de sentimentos que uma tal lei parece revoltante – fora o fato de o legislador correr o risco de ser ele próprio morto, o que deve suscitar um sentimento de repulsa nele próprio.

Foi dito que os computadores são máquinas sintáticas. Isso significa que elas não têm semântica (compreensão) e muito menos pragmática, ambas características humanas. Por exemplo, em um julgamento no sistema judiciário poderiam ser levados em conta atenuantes, como por exemplo, uma mãe faminta roubar um pão para seus filhos. Um computador aplicará a lei dentro das regras que lhe foram dadas, e talvez essa atenuante ou outras, inclusive não previstas, não sejam parte do sistema.

Um exemplo é o sistema COMPAS, da Northpointe Inc., usado nos EUA para proferir sentenças judiciais, que emprega um algoritmo ‘proprietário’, portanto secreto; o sistema calcula a probabilidade de uma pessoa reincidir novamente em um crime, levando a uma prisão preventiva. É conhecido o caso de um motorista guiando um carro roubado que não parou em uma perseguição policial. Devido ao seu perfil, o sistema COMPAS recomendou que ele fosse preso preventivamente. O réu solicitou informação sobre quais foram os motivos que levaram à sua prisão. O juiz encaminhou a solicitação ao fabricante do sistema, que se recusou a fornecer essa informação, o que foi acatado pelo juiz. Em apelação à Suprema Corte, esta decidiu que não podia julgar o caso pois não tinha conhecimento técnico suficiente [9].

Um outro exemplo é a geração de timbre de voz de uma pessoa, bem como de um vídeo com a movimentação dos lábios e expressões faciais que acompanham uma certa fala. Com isso é possível gerar vídeos extremamente convincentes de que alguém disse certa coisa que jamais pronunciou. Aqui não se trata de *fake news*, notícias falsas, mas de discurso falso (*fake voice*), o que está sendo denominado de falsidade profunda (*deep fake*). Tornaram-se conhecidos o vídeo falso de um discurso do ex-presidente Obama [10] e da presidente do Congresso americano Nancy Pelosi, em que ela aparece embriagada [11].

8. Eliminação de postos de trabalho

Há muitos anos discute-se se a automação elimina ou não postos de trabalho. Há uma opinião relativamente generalizada de que alguns postos são eliminados, mas outros são criados. Por exemplo, o Bradesco usa um atendente telefônico automático, reconhecendo palavras faladas, por exemplo “saldo”. Muitos atendentes foram eliminados. Outra substituição típica é a de robôs na indústria manufatureira, com a vantagem de que eles não precisam de intervalo para almoço, de tempo de descanso e de sono, de férias, não fazem greve, são mais rápidos etc., sendo assim muito mais baratos.

O grande problema é o que acontecerá com os trabalhadores que forem substituídos. Em geral, os novos postos de trabalho criados são de alta especialização, e muitos trabalhadores substituídos não terão capacidade de adquirir o conhecimento necessário para os novos postos.

Ocorre que o trabalho é psicologicamente essencial para o ser humano, pois aí ele se sente útil. Talvez seja melhor uma pessoa ter um trabalho relativamente degradante (um triste exemplo: ascensorista de elevadores automáticos) do que trabalho nenhum. A esse respeito, não posso me furtar de fazer um comentário sobre educação. É atribuída a Ruy Barbosa uma frase de que o trabalho que mais eleva o ser humano é o ensino. Assim, o uso computadores na educação substituindo parcial ou totalmente um professor deveria levar em conta essa questão. Obviamente, aqui há um grande problema: o que se entende por ‘degradar’ ou ‘elevar’? Para isso é necessário ter uma conceituação profunda do que é um ser humano. Na mentalidade induzida pela IA, de que ele é uma máquina, não há sentido em se empregar essas duas palavras.

9. ‘Inteligência’ artificial vs. inteligência humana

Uma das diferenças fundamentais entre a ‘Inteligência’ artificial e a inteligência humana é que a primeira é sempre especializada. Por exemplo, os programas de jogar xadrez só jogam xadrez, não fazem mais nada. O sistema de AM relatado acima, para distinguir cães de lobos, não distingue mais nada. Tenho a impressão de que um sistema de AM que tivesse várias aplicações distintas iria resultar na maior confusão. Qualquer ser humano, ao contrário, tem uma inteligência extremamente multifacetada, por exemplo, distinguindo uma quantidade muito grande de animais, sabendo fazer inúmeras tarefas diferentes mentais ou com as mãos etc.

Uma outra diferença fundamental é que os computadores são máquinas matemáticas, usando uma matemática muito particular, a discreta e finita. De fato, qualquer dado que esteja dentro de um computador tem que ser expresso numericamente, o que vale também para os programas, que também são dados. É impossível colocar em um computador o conceito de contínuo (por exemplo, dos números reais, ou os pontos de uma reta) e de infinito, apesar de serem conceitos usados na matemática. Não se pode afirmar que a inteligência humana seja discreta e finita. Aliás, a nossa memória é aparentemente ilimitada, pois jamais alguém teve a vivência de querer memorizar algo (p.ex. um número de telefone, o nome de uma pessoa etc.) e não haver mais ‘lugar’ para isso. Isso é uma das indicações de que o ser humano tem elementos que transcendem a matéria, pois se nossa memória fosse simplesmente física, teria necessariamente que ser limitada, finita. Aliás, quando se usa o número 2, nessa representação, na verdade se está usando o conceito puro do dois, que não depende da representação

simbólica (poderia ser II, ii, .., dois, two etc.). Ora o conceito puro não tem representação simbólica, portanto não pode estar armazenado no cérebro.

No meu livro sobre IA [2], caracterizei dois tipos de inteligência: a ‘incorporada’ e a ‘criativa’. Todas as máquinas têm uma inteligência incorporada, devida ao seu projeto, sua construção e funcionalidade, todos devidos a ações humanas, eventualmente com o auxílio de máquinas. Mas também seres humanos têm inteligência incorporada, na infinita sabedoria revelada pelo seu corpo físico. Animais e plantas idem. Mas até a Terra tem uma inteligência incorporada; por exemplo, se ela estivesse um pouco mais próxima ou afastada do Sol, não estaríamos aqui; se houvesse muito mais oxigênio na atmosfera, ‘queimaríamos’. Se o eixo de rotação da Terra não fosse inclinado, não haveria estações do ano, e provavelmente também não estaríamos aqui. E também o mundo físico tem essa inteligência: se houvesse um pequeno desvio nas constantes físicas, até mesmo os átomos não existiriam.

Mas só o ser humano tem o que denominei de ‘inteligência criativa’. É devido a ela que aparecem novas ideias e novos impulsos. É devido a ela que o ser humano está mudando as condições da Terra – infelizmente, em vários aspectos mais para pior do que para melhor.

10. Minhas respostas às questões de Quintarelli

Ver a referência 13; as questões dele estão abaixo em minhas traduções livres.

10.1 IA terá emoções?

Nunca. Computadores são máquinas universais, objetivas. Sentimentos, e também sensações, são absolutamente individuais e subjetivos. Se uma pessoa come um caqui, e tem a sensação do seu gosto, nenhuma outra pessoa poderá ter a sensação que a primeira está sentindo. Sentido aquele gosto, a primeira terá ainda o sentimento de gostar ou não daquela sensação (o caqui pode não estar maduro, cheio de tanino...). Esse sentimento também não pode ser sentido por outra pessoa.

Experimente-se descrever a uma pessoa o gosto de uma fruta desconhecida dela. Isso é impossível.

10.2 IA será capaz de adquirir uma ‘personalidade’ similar a de seres humanos?

Nunca. Seres humanos têm individualidade e atividades interiores absolutamente pessoais e subjetivas, tais como pensar, sentir e querer. A primeira pode ser objetiva, por exemplo no pensar matemático, mas também pode ser subjetiva, por exemplo ao se pensar numa dor de cotovelo que se está sentindo. Por outro lado, todas as máquinas são objetivas, pois seu funcionamento pode ser totalmente compreendido (a menos do nível atômico, pois não se sabe o que é um átomo). Elas são universais, pois o projeto e a fabricação de um certo tipo de máquina são os mesmos para todos os exemplares produzidos. Em particular, no sentido de ‘máquina universal’ de Turing, qualquer computador, dada suficiente capacidade e tempo, pode simular (tecnicamente, ‘emular’) qualquer outro computador: dando-se a uma máquina A um programa P em linguagem de máquina que é processado em uma máquina B, um emulador de B processando em A pode interpretar P e chegar exatamente aos mesmos resultados do programa interpretado por B. (A rigor, computadores não executam uma linguagem de máquina, eles a interpretam.)

10.3 A IA poderá suplantar as limitações humanas e desenvolver uma identidade (*self*) superior à dos humanos?

Depende das limitações. Desde seu invento na década de 1940 computadores suplantaram a capacidade humana de fazer cálculos aritméticos. Mas na minha concepção eles nunca superarão os seres humanos, por exemplo em sentimentos, em criatividade e na intuição (ver a seção 3 acima), talvez em generalidade.

10.4 Uma pessoa pode apaixonar-se por um sistema de IA?

Sim, se a tal pessoa for degenerada.

10.5 Será que poderemos repentinamente ter um sentido de como essas previsões irrealistas parecem um pouco ridículas?

Isso depende da maneira de pensar. Se é uma maneira materialista, fiscalista, então é perfeitamente possível encarar seriamente essas previsões.

A tecnologia está sendo muito mal usada, destruindo a natureza e a humanidade, satisfazendo o egoísmo, esse filho dileto do materialismo. O contrário de egoísmo é o altruísmo, mas uma ação só pode ser realmente altruísta se for feita em total liberdade, o que não faz sentido do ponto de vista fiscalista. Somente uma mudança na maneira de pensar reverterá esse processo destrutivo, que está em contínuo desenvolvimento. Mas para isso será necessário acrescentar outras concepções de mundo à maneira científica materialista de pensar – mas não estou me referindo a religiões dogmáticas! Leiam-se artigos em meu *site* para ver o que pode ser uma maneira diferente, racional e coerente de pensar.

11. Os grandes perigos

O grande sucesso recente da IA, especialmente o AM, representa cinco enormes perigos para a humanidade.

11.1 Possivelmente as pessoas vão passar a confiar cada vez mais nas máquinas, substituindo decisões humanas pelas escolhas lógicas, frias, das últimas. Isso poderá levar a ações desumanas.

11.2 Os dados de entrada que são usados em um sistema de AM podem ser viciados, levando a conclusões também viciadas. Por exemplo, uso de dados racistas levarão a escolhas racistas. Um outro caso é o de uso de dados que, em sua maioria, são apenas de homens, como em certas profissões, e depois os sistemas serem aplicados também a mulheres.

11.3 A tendência é de armazenar em computadores cada vez mais dados pessoais, usados posteriormente em escolhas feitas pelas máquinas. Por exemplo, a empresa Google traça um perfil de cada um de seus usuários, examinando para isso os *e-mails* que ele enviou pelo processador Gmail, quais *sites* visitou usando o navegador Chrome, que textos usou no Google Translate etc. Mesmo se uma pessoa não usa nenhum produto da Google, se ela enviar um *email* para alguém com um endereço do Gmail, a Google já começa a traçar o perfil da primeira. Por exemplo, com esses perfis o sistema de busca do Google elenca os resultados da busca segundo o perfil, sugerindo *sites* comerciais. Afinal, a Google desenvolveu um sistema genial: ela cobra pelo número de acessos que são feitos aos *sites* dos seus anunciantes, que é o que interessa aos últimos. Um dos objetivos da Google é o de se apossar de todos os dados do mundo [16].

11.4 Existe um grande perigo de dados pessoais ficarem armazenados eternamente. Por exemplo, alguma pessoa cometeu uma bobagem quando adolescente. Esse dado poderá ser usado por um sistema de IA muitos anos depois, na seleção de candidatos a algum emprego, sem que a pessoa saiba disso. O ser humano está em permanente mudança; na verdade, ele não deveria ser chamado de 'ser humano', e sim de 'devir humano'. Cada vivência de uma pessoa é incorporada, em geral no inconsciente, podendo mudar sua maneira de pensar e suas atitudes. Em princípio, não se deveria julgar uma pessoa pelo que ela foi; ela pode ter mudado de opinião e de atitude.

11.5 Notícias e falas falsas com som e vídeo podem subverter totalmente a confiança que se tinha em uma pessoa, especialmente na mensagem recebida dela, devido ao seu timbre de voz

e seu aspecto únicos. Já estamos naquilo que denomino 'era da mentira'. Isso vai abalar um dos pontos fundamentais do relacionamento humano, a confiança.

11. Soluções

O que fazer?

12.1 É absolutamente essencial que se saiba como cada programa e sistema funciona, isto é, quais são os critérios de escolha embutidos neles. Deveria haver uma lei obrigando os fornecedores de sistemas de IA e AM a especificarem como cada sistema chega aos seus resultados. Para sistemas muito complexos, isso pode ser até mesmo impossível e, nesse caso, deveriam ser evitados.

12.2 Quaisquer dados pessoais deveriam poder ser examinados pelos cidadãos envolvidos, e deve haver mecanismos judiciais rápidos (rapidez no sistema judiciário é, quem sabe, a maior utopia no Brasil...) para que eles sejam apagados quando inconvenientes, ou corrigidos. Nós ainda presamos a privacidade, mas essa atitude está desaparecendo.

12.3 Não deveria haver sistemas que substituam autonomamente decisões humanas. Deveria haver uma lei obrigando que qualquer escolha feita por uma máquina, envolvendo ações a serem aplicadas a seres humanos, devesse passar pelo crivo de uma ou mais pessoas. Por exemplo, um diagnóstico médico automatizado deveria ser necessariamente examinado e avaliado por um médico antes de ser utilizado; esse profissional iria se responsabilizar pela aplicação do diagnóstico. Isto é, nesse caso no máximo os resultados fornecidos por uma máquina deveriam auxiliar o diagnóstico, mas não determiná-lo. Quem sabe dessa maneira evitar-se-ia um enorme perigo: os profissionais seguirem cada vez mais as escolhas das máquinas, e com isso irem perdendo eles mesmos a capacidade de tomar decisões.

12.4 Para tudo isso, é necessário mudar a mentalidade, a maneira de pensar, conforme foi exposto no item 10.5. Creio ser absolutamente essencial começar-se a encarar o ser humano e a natureza de uma maneira diferente. A maneira usual está destruindo ambos.

13. Referências

- [1] "Leis e aforismos de Setzer". www.ime.usp.br/~vwsetzer/jokes/leis.html (Acesso em 11/8/19).
- [2] Setzer, V.W. *I.A.: Inteligência Artificial ou Imbecilidade Automatizada? As máquinas podem pensar e sentir?* São Paulo: Biblioteca24horas, 2021.
- [3] Artigo bastante completo sobre histórico e detalhes da IA: www.oezratty.net/wordpress/wpcontent/themes/Ezratty5/forcedownload.php?file=/Files/Publications/Usages%20intelligence%20artificielle%202018%20Olivier%20Ezratty%20Compressed.pdf (idem).
- [4] Proposta do simpósio no Dartmouth College: <https://web.archive.org/web/20070826230310/http://wwwformal.stanford.edu/jmc/history/dartmouth/dartmouth.html> (Idem).
- [5] O torneio Kasparov vs. Deep Blue: <https://rafaelleitao.com/o-homem-e-a-mquina-o-match-kasparov-x-deep-blue/> (Idem)
- [6] Setzer, V.W. "Considerações sobre o xadrez eletrônico". www.ime.usp.br/~vwsetzer/xadrez.html (Idem)
- [7] 'Aprendizado' de máquina. https://pt.wikipedia.org/wiki/Aprendizado_de_m%C3%A1quina (Idem)
- [8] Redes 'neurais' artificiais. https://pt.wikipedia.org/wiki/Rede_neural_artificial (Idem)

- [9] Artigo do New York Times sobre o caso do uso do sistema judiciário COMPAS e de outros sistemas. www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html?module=inline (Idem)
- [10] Vídeo falso de Barack Obama. www.youtube.com/watch?v=AmUC4m6w1wo (Idem)
- [11] Vídeo falso de Nancy Pelosi bêbada. www.irishtimes.com/culture/tv-radio-web/the-fakefacebook-video-that-turned-nancy-pelosi-into-an-alcoholic-old-hag-1.3905862 (Idem)
- [12] Setzer, V.W. *Folha de consulta para a linguagem de programação Python*. (Acesso em 15/12/19).
- [13] Quintareli, S. “[Let’s forget the term AI. Let’s call them Systematic Approaches to Learning Algorithms and Machine Inferences \(SALAMI\)](#)”. (Idem)
- [14] Alhazero. <https://en.wikipedia.org/wiki/AlphaZero> (Acesso em 4/10/21)
- [15] Observatório Nacional de Segurança Viária. <https://www.onsv.org.br/90-dos-acidentes-sao-causados-por-falhas-humanas-alerta-observatorio/> (Acesso em 4/10/21)
- [16] Cleland, S. e I. Brodsky. *Busque e destrua. Por que você não pode confiar no Google Inc.* Trad. F.E. de Mello. São Paulo: Matriz, 2012.