

Inteligência Artificial: uma utopia, uma distopia

Fabio Gagliardi Cozman¹

Resumo: A busca por inteligências artificiais parece ter atingido um ponto de inflexão no qual várias tecnologias há muito prometidas têm se tornado realidade. Em particular, o uso intensivo de grandes bases de dados tem levado ao desenvolvimento de sistemas de reconhecimento de imagens, compreensão de linguagem natural e tomada de decisão, cujo desempenho chega a igualar, e em alguns casos superar, o desempenho humano. Esses avanços tecnológicos têm gerado reações de otimismo e pessimismo. Algumas opiniões são entusiasticamente positivas, entendendo que o futuro trará riquezas imensuráveis a todos; outras vislumbram o fim da espécie humana. Entre esses extremos pode-se encontrar um leque matizado de posições. Este artigo procura capturar essas posições por meio de um conjunto de “distopias” e “utopias”. Ao final, chega-se a uma “utopia realista”, que apresenta um objetivo plausível para a sociedade atual, em contraponto a uma “distopia realista” que representa um modelo verossímil mas que deve ser evitado.

Palavras-chave: Inteligência Artificial. Aprendizado de máquina. Mercado de trabalho.

Abstract: The search for artificial intelligences seems to have reached a point where many long-promised technologies have become reality. In particular, the heavy use of large databases has led to the development of systems for image recognition, natural language understanding, and decision making, whose performance often reaches, and sometimes surpasses, human performance. This technological advance has produced optimistic and pessimistic reactions. Some opinions are enthusiastically positive in believing that the future will bring enormous riches to everyone, while others foresee the end of humanity. Between these extremes, one can find a varied set of positions. This paper tries to capture these positions using a set of “dystopias” and “utopias”. In the end we reach a “realistic utopia” that serves as a plausible goal for society, in contrast to a “realistic dystopia” that represents a credible model that must be avoided.

Keywords: Artificial Intelligence. Machine Learning. Job market.

Inteligência artificial: um verão escaldante, mas com possíveis trovoadas

Embora a busca por uma Inteligência Artificial seduza a mente humana há séculos (NILSSON, 2009), foi só em 1950 que Alan Turing apresentou a primeira formalização do conceito de uma “máquina que pensa” (TURING, 1950). Poucos anos

¹ Professor Titular da Poli - USP/Departamento de Engenharia Mecatrônica, PhD Carnegie Mellon University, Livre-docência pela USP. Atualmente é coordenador da comissão de Inteligência Artificial da SBC, Associate Editor do Int. Journal on Approximate Reasoning, membro do Editorial Board do Artificial Intelligence Journal, e Associate Editor do Journal of Artificial Intelligence Research. E-mail: fgcozman@usp.br.

depois ocorreu o primeiro encontro de pesquisadores interessados no tema em que foi cunhado o termo “Inteligência Artificial” (MCCARTHY, 1955).

Desde então, a ideia de uma “Inteligência Artificial” tem fascinado a sociedade. Esse interesse se reflete em referências culturais, por exemplo, na literatura e no cinema e também em discussões científicas e econômicas.

Grandes promessas fazem parte da história da Inteligência Artificial. Durante os seus anos iniciais, a área foi cercada de previsões otimistas e em alguns casos mirabolantes. O entusiasmo sofreu um duro golpe em 1974, quando um relatório encomendado pelo parlamento britânico indicou que a área falhava na prática. Durante vários anos subsequentes o financiamento da área foi reduzido em todo o mundo, um período que hoje é conhecido como “inverno da IA”. Mas o pessimismo não durou para sempre: durante os anos oitenta houve um novo momento de euforia, marcado pelo desenvolvimento de “sistemas especialistas” que teriam condições de reproduzir as regras usadas por especialistas na solução de problemas específicos. E, de novo, a área enfrentou um longo inverno durante o final dos anos 80 e começo dos anos 90. Olhando para esse período, pode-se ver que muitas técnicas fundamentais da área estavam em gestação, mas ainda não haviam maturado ao ponto de resolver problemas práticos significativos. O inverno foi longo, e bastante frio.

Hoje a área vive um verão escaldante. Algumas das maiores empresas do mundo, como Google e Facebook, baseiam-se em métodos de Inteligência Artificial, como entendimento de textos e reconhecimento de faces. Grande parte das empresas anuncia estratégias que incluem técnicas de Inteligência Artificial para processar montanhas de dados acumuladas por toda parte. A sociedade é bombardeada com notícias sobre avanços da tecnologia e de seus impactos.

Como ocorreu essa mudança climática? Em primeiro lugar, houve exponencial aumento na capacidade de processamento de computadores, em particular computadores com processamento paralelo. E, além do aumento de capacidade, computadores e equipamentos como câmeras de vídeo tiveram enormes reduções de preço. Em segundo lugar, a área se beneficiou do acúmulo exponencial de dados na sociedade e indústria em função da redução de custos de sensores e da melhoria das redes de comunicação.

Mas, além desses dois pontos práticos importantes, houve também a maturação de técnicas computacionais ao longo de duas décadas de esforços. Houve grande avanço em algoritmos de aprendizado, que extraem padrões de grandes bases de dados, bem como algoritmos de otimização aplicados a problemas de planejamento. Além disso, houve significativa maturação de técnicas de modelagem baseadas em lógica e probabilidades. Isso levou a um grande progresso na geração automática de diagnósticos e de planos, e na compreensão de imagens e de linguagem natural. A área conseguiu finalmente concretizar algumas das promessas feitas no seu início, logrando realizar computacionalmente atividades que de fato parecem inteligentes.

Esse progresso naturalmente gera reações, algumas otimistas, outras pessimistas. Os otimistas creem, em resumo, que o aumento de produtividade causado por máquinas inteligentes será extraordinário, e o conseqüente aumento de riqueza mais que compensará alguns desconfortos causados por erros eventuais de serviços automáticos. Os pessimistas creem que o aumento de produtividade virá, mas acompanhado de tantos problemas que o balanço geral será negativo: alguns julgam que os problemas passam pela própria destruição da humanidade, enquanto outros preveem uma inaceitável perda de privacidade e de controle, além de um aumento da desigualdade entre seres humanos. Previsões sobre o futuro variam e são de difícil classificação.

Para tentar compreender melhor como se dá a discussão sobre Inteligência Artificial, este artigo procura dividir as várias opiniões em “modelos de futuro”, alguns utópicos, outros distópicos. Um futuro utópico, mesmo que improvável, oferece um objetivo a perseguir. Um futuro distópico, mesmo que pouco verossímil, descreve circunstâncias a evitar, suscitando debate sobre como a sociedade deve se organizar hoje para desviar-se de um desastre amanhã.

O artigo passa por várias utopias e distopias de baixa probabilidade; ao final, o artigo procura descrever uma “utopia realista” e uma “distopia realista”. Ou seja, o artigo procura encontrar um modelo de futuro otimista e que pode de fato ser atingido se houver esforço para tal, e um modelo de futuro pessimista que pode de fato ocorrer se não houver preocupação nem prudência.

Distopias Catastróficas

Existem muitas distopias totalmente pessimistas que se baseiam na evolução da Inteligência Artificial. Esse é um tema que tem recebido grande atenção na literatura e no cinema, e que está profundamente aninhado no imaginário popular. Em linhas gerais, pode-se identificar duas diferentes distopias dessa natureza.

A primeira distopia se baseia na noção de que computadores podem ter uma ação fundamentalmente maléfica devido a falhas de operação ou devido a uma maldade intrínseca (o “desejo de dominar o mundo”). O cinema tem sido pródigo em explorar esse tema há muito tempo; provavelmente, um dos exemplos mais populares está no filme *Terminator* e suas sequências. Por anos, essa foi a distopia mais comum relacionada à Inteligência Artificial.

A segunda distopia se baseia na noção hoje designada por superinteligência artificial (BOSTRON, 2014). Nessa visão, os computadores terão uma melhoria exponencial em sua capacidade cognitiva, em um certo momento superando em muito o ser humano. Nesse ponto o ser humano terá se tornado, para a superinteligência artificial, um ser inútil e inconveniente – como um enxame de mosquitos que apenas perturbam. A superinteligência poderá então decidir destruir os seres humanos, baseado em argumentos estritamente racionais; por exemplo, para evitar divisão de recursos energéticos. A distopia envolve a destruição da humanidade ou sua redução a uma situação de servidão completa (talvez uma servidão na qual os seres humanos remanescentes estejam subjugados e inconscientes da sua real situação).

Embora tais distopias catastróficas mereçam atenção, esse artigo não as discutirá em profundidade. Essas distopias estão hoje muito mais próximas de experimentos mentais sobre a sociedade humana do que de reais possibilidades de futuro. A Inteligência Artificial hoje em desenvolvimento não tem condições de planejamento em escala global nem de negociação sofisticada; tampouco tem condições de assumir aparatos físicos que possam realmente oferecer dano a populações de grande porte.

Mas não é pela sua improbabilidade que essas distopias são aqui deixadas em segundo plano. O ponto aqui é que, para que distopias catastróficas possam acontecer,

distopias mais sutis muito provavelmente terão sido antes atingidas. Vale a pena entender o que seriam essas distopias mais sutis.

Distopias: um guia para o pessimista

Considere agora que uma distopia catastrófica não ocorre, portanto a humanidade continua sobrevivendo e interagindo com suas máquinas. Quais ameaças são contempladas por uma visão pessimista da Inteligência Artificial?

As ameaças parecem se dividir em dois grupos.

Em primeiro lugar, há ameaças que podem ser resumidas como “perda do controle”. A mais simples é relacionada à perda de privacidade: máquinas que usam dados sigilosos sem autorização. Outra ameaça é que decisões discriminatórias sejam tomadas por algoritmos sem que haja controle social: já existem preocupações sobre decisões judiciais tomadas com base em dados enviesados (OSOBA; WELSER IV, 2017). Uma outra ameaça é a perda de entendimento, por parte de seres humanos, das razões pelas quais decisões automáticas são tomadas. O ser humano se transformaria em uma minúscula engrenagem em uma sociedade kafkiana em que decisões são tomadas com base em algoritmos que ninguém consegue entender. Aqui há uma ameaça ao próprio debate democrático na sociedade: como é possível debater assuntos de interesse geral se ninguém sabe realmente como o sistema funciona? Uma sociedade em que todas as ações são mediadas por dispositivos artificiais pode se tornar uma sociedade em que relacionamentos profundos são raros ou impossíveis. Finalmente, máquinas podem falhar, falhas podem ter efeitos tão mais nefastos quanto mais responsáveis forem as máquinas. Uma falha pode disparar mísseis erradamente ou pode causar acidentes aéreos de grande porte; uma sequência de falhas pode causar uma guerra ou uma epidemia. E uma sociedade controlada por máquinas pode ser muito mais vulnerável a ataques perpetrados por humanos mal-intencionados que podem danificar elementos centrais do processo decisório.

Em certa medida, todas as ameaças citadas no parágrafo anterior já podem ser observadas em alguma medida na sociedade atual: já somos submetidos a “sistemas” de controle bancário com pouca transparência, já perdemos alguma privacidade, já temos que lidar com decisões automáticas que não são explicadas e já nos

defrontamos com “falhas não humanas”. O pessimista pode se sentir plenamente justificado ao raciocinar que a perda de controle já em curso só pode se acentuar, em certo momento tornando a vida insuportável.

Em segundo lugar, há ameaças relacionadas ao mercado de trabalho. Existe muita discussão sobre quais profissões serão afetadas pelo aumento de capacidade cognitiva de computadores (FREY; OSBORNE, 2013). Algumas profissões parecem ter pouco futuro: funcionários de empresas de propaganda por telefone terão menos relevância no momento em que essas propagandas forem realizadas de forma eletrônica. Mas ninguém sabe como será o mercado de trabalho no futuro, e essa insegurança gera angústia: cada trabalhador se pergunta se sua profissão está em risco.

Mas quais são exatamente as ameaças relacionadas ao mercado de trabalho?

A substituição de postos de trabalho por máquinas pode gerar dramas pessoais momentâneos, mas se a sociedade tiver condições de reeducar pessoas, em princípio será possível atingir um patamar mais produtivo em que novos empregos florescerão. A simples necessidade de modificar a força de trabalho atual, levando-a a um patamar mais refinado, talvez a um custo alto, não parece ser elemento suficiente para uma distopia. Porém, outras consequências da automação de empregos parecem estar na raiz de várias distopias. Uma possível consequência é que a turbulência no mercado de trabalho nunca acabe: que o ser humano esteja sempre submetido a um processo humilhante de obsolescência, passando de profissão em profissão, perseguido por máquinas cada vez mais capazes. Outra possível consequência é que ocorra uma concentração cada vez maior da renda, com máquinas controladas por uma minoria da população – enquanto a maior parte da humanidade estaria competindo arduamente com as máquinas, um grupo pequeno poderia auferir lucros extraordinários. Ou poderia haver uma crescente desigualdade no mercado de trabalho: seres humanos com certas habilidades de difícil automação se distanciariam cada vez mais de trabalhadores com habilidades facilmente reproduzidas artificialmente (BRYNJOLFSSON; McAFEE, 2014). Uma terceira possível consequência é a captura dos empregos com maior capacidade cognitiva por agentes artificiais, relegando o ser humano a tarefas repetitivas e de baixa remuneração (FREEMAN, 2016). Nessa última visão, o problema

não é necessariamente que faltarão empregos: novos tipos de trabalho poderão ser criados, mas a vantagem competitiva dos seres humanos sobre as máquinas desaparecerá e elas controlarão as atividades mais nobres.

O pessimista parece ter amplo espectro de opções. O que diz o otimista?

A grande utopia

A Grande Utopia da Inteligência Artificial (GUIA) é, em essência, simples. Máquinas realizam todo o trabalho repetitivo com extraordinária precisão e produtividade, enquanto seres humanos gastam algum tempo controlando o processo produtivo e democraticamente debatendo como regular a sociedade, trabalhando de forma prazerosa e dedicando tempo considerável à saúde e bem-estar.

Na GUIA a infraestrutura artificial opera de forma clara e ética, seguindo diretrizes produzidas pela sociedade humana em um processo democrático amplo. Nessa utopia, máquinas oferecem diagnósticos médicos a todos os cidadãos, e médicos humanos especializados se ocupam apenas de novos casos e também da interação mais pessoal com pacientes; máquinas oferecem serviços legais simples, enquanto advogados humanos se ocupam de negociações mais complexas e da concepção de novos instrumentos jurídicos; e assim por diante. Essa sociedade utópica também dará grande assistência ao envelhecimento da população: assistentes artificiais estarão disponíveis de forma ininterrupta, até para conversar e confortar. Um melhor planejamento urbano, acoplado a um melhor planejamento individual sobre opções de transporte, pode reduzir níveis de poluição e evitar escassez de recursos; similares ganhos podem ser obtidos em todos os aspectos da vida humana.

Na GUIA o mercado de trabalho é completamente subvertido pela Inteligência Artificial, mas o resultado é bem-vindo: atividades não desejadas são ocupadas por máquinas que operam com grande eficiência, e seres humanos se ocupam de atividades de seu interesse.

Há possíveis variantes dessa utopia no que diz respeito à distribuição de riquezas e de empregos. Uma possível GUIA seria obtida com uma população humana altamente qualificada operando em um mercado de trabalho competitivo (mas talvez com regras sobre competição não-humana). Uma outra GUIA contemplaria mecanismos

de distribuição de renda, por exemplo, oferecendo programas de renda mínima para garantir que toda a população tenha benefícios relativamente equânimes.

Uma utopia “realista”

Embora a Grande Utopia da Inteligência Artificial sirva como um objetivo meritório, alguns de seus aspectos parecem excessivamente idealizados. Mesmo em um mundo com máquinas superinteligentes e dóceis poderá haver problemas com proteção de privacidade, falta de explicação, falhas – talvez causadas por seres humanos incompetentes ou mal-intencionados. Também poderá haver desigualdade, até mesmo disputa sobre os meios de produção. Parece razoável buscar um objetivo mais realista, e em particular uma utopia que sirva para um país como o Brasil.

Como seria essa utopia?

Em primeiro lugar, o uso intensivo de máquinas elevará enormemente a produtividade. Portanto a riqueza total gerada pela sociedade crescerá. Os ganhos de produtividade melhorarão a vida da sociedade em geral, aprimorando os serviços e produtos do setor privado, oferecendo mais recursos para o setor público. Produtos serão mais úteis e mais eficientes, reduzindo por exemplo a poluição. Os recursos públicos, administrados e fiscalizados com suporte de agentes artificiais, serão utilizados com maior eficiência e sob maior controle social.

Em segundo lugar, as ameaças relativas à perda de controle (violação de privacidade, decisões automáticas discriminatórias ou sem explicação) serão controladas por meio de legislação apropriada. Aqui se trata de encontrar legislação que proteja a sociedade sem impedir a inovação; proibições genéricas baseadas em medos abstratos só aumentarão a burocracia e reduzirão a produtividade. Uma boa legislação deve incentivar o progresso e evitar as ameaças.

Em terceiro lugar, e com uma boa dose de otimismo utópico, a sociedade encontrará diretrizes éticas que serão codificadas em máquinas. Em particular, máquinas receberão objetivos claros e éticos. Parece difícil que todos os dilemas éticos possam ser assim equacionados, mas será importante que os casos mais simples sejam codificados, deixando os casos complexos para debate entre os seres humanos. Da mesma forma, questões legais sobre atribuição de responsabilidade sobre

comportamento artificial serão em grande parte codificadas, deixando os casos complexos para análise por agentes humanos.

Em quarto lugar, o número de “falhas não humanas” será reduzido a um ponto tão baixo que não haverá questionamento quanto ao uso de máquinas nas mais variadas atividades. Afinal, o próprio ser humano comete falhas; o objetivo realista deve ser produzir dispositivos que apresentem muito menos falhas que os operadores humanos, e não produzir dispositivos que nunca falham.

Em resumo, a infraestrutura artificial da sociedade será onipresente mas transparente, justa e ética, praticamente sem falhas, suprimindo os seres humanos de suas necessidades.

Em uma sociedade com essas características, com menor pressão do trabalho e melhor distribuição de riquezas, relacionamentos humanos de qualidade poderão ocorrer naturalmente. Mesmo o uso de “amigos artificiais” poderá ser útil para lidar com ocasionais episódios de solidão ou de doença. Seres humanos poderão dedicar mais tempo a esportes, artes, lazer.

Para completar essa utopia, é necessário imaginar como ocorreria de fato a distribuição de trabalho e riquezas. Essa é uma questão que tem suscitado as mais diversas propostas. Parece razoável supor que a maior produtividade na sociedade poderá levar a programas universais de educação e saúde. Além disso, parece razoável supor que o sistema de educação evolua de tal forma que os seres humanos estejam continuamente se aprimorando e se colocando à frente das máquinas em atividades fundamentais. Mas e se houver um contingente de pessoas sem colocação? Seria o caso de oferecer programas de renda mínima a todas essas pessoas? Ou isso será um incentivo para que os seres humanos deixem de trabalhar e se coloquem em uma posição subalterna às máquinas? Esse é um debate necessário para os próximos anos, à medida em que a tecnologia avança. Na versão utópica, a sociedade conseguirá encontrar mecanismos de distribuição equânime de oportunidades que permitam a todos manter sua relevância e dignidade.

E uma distopia “realista”

Se a utopia realista da Inteligência Artificial serve como um objetivo a almejar, uma distopia realista pode servir como indicação sobre o que evitar. Na verdade, a simples continuidade de várias mazelas da sociedade já parece ser suficiente para gerar uma distopia realista bastante preocupante. Suponha que, ao tentar atingir a utopia realista, façamos alguns erros e, ao mesmo tempo, não consigamos nos libertar de nossos medos e preconceitos. O que pode acontecer, em particular pensando no Brasil?

Pode ocorrer uma resistência corporativa tremenda à Inteligência Artificial; leis podem ser promulgadas para impedir o avanço dessa tecnologia ou para criar tanta burocracia que o avanço se torne impossível. Nesse caso o país pode perder a oportunidade de empreender na área, no futuro tornando-se importador da tecnologia. Por exemplo, imagine que associações de médicos proibam a análise automática de radiografias: isso impedirá a existência de empresas no Brasil com essa tecnologia, e em pouco tempo a população (grande parte da qual hoje não tem acesso a médicos) poderá enviar suas radiografias a outros países para análise. Outro exemplo: imagine uma legislação que, no afã de evitar comportamento não-ético, torne impossível qualquer teste de um dispositivo artificial que se comunique com seres humanos. Ou ainda se pode tentar controlar o mercado de trabalho, impedindo que a tecnologia floresça e que novas formas de ocupação sejam criadas – nesse caso também podemos nos ver em pouco tempo importando máquinas, bens e serviços de outros países, enquanto nossa força de trabalho míngua devido a sua falta de preparo.

É importante entender que a tecnologia de Inteligência Artificial é relativamente barata e depende apenas de boas ideias e boa formação, permitindo que empreendedores descubram novas maneiras de aumentar a qualidade de vida da população. O país não pode perder essa oportunidade por medo de enfrentar suas consequências.

De uma forma ou de outra, a produtividade geral deve aumentar com a implantação de máquinas inteligentes. Pode então ocorrer um crescimento da desigualdade social. Alguns poucos controlam a produção; alguns outros se dedicam a tarefas criativas; muitos se dedicam a tarefas menores. Alguns trabalham longas horas;

outros não têm o que fazer. Para evitar um colapso social, pode ser então instituída uma alta taxa redistributiva, gerenciada por um governo ineficiente e pouco transparente, em que decisões humanas e artificiais são igualmente incompreensíveis para a maioria da população. Nesse cenário, uns poucos se beneficiam muito do avanço tecnológico, enquanto a maioria vive uma situação ambígua: todos podem obter análises médicas de forma automática, e todos podem obter muitos outros benefícios da tecnologia onipresente, mas praticamente todos vivem com meios limitados, cercados por máquinas de alta produtividade. Todo esse cenário é realmente distópico e significa um aumento de problemas que já nos acompanham. Para evitá-lo, é preciso dar suporte à tecnologia e regrá-la adequadamente, além de conduzir um debate sério sobre a forma mais efetiva de distribuir riquezas geradas por inteligências artificiais.

Em resumo, é preciso evitar essa distopia realista a todo custo, mantendo todo o foco na utopia realista que é de fato possível.

Enviado: 1 março 2018

Aprovado: 31 abril 2018

Referências

BOSTRON, Nic. *Superintelligence: paths, dangers, strategies*. Oxford: Oxford University Press, 2014.

BRYNJOLFSSON, Erik; McAfee, Andrew. *The second machine age*. New York, NY: Norton, 2014.

FREEMAN, Richard B. Who owns the robots rules the world, *Harvard Magazine*, vol. 118, nº. 5, May/June, p. 37-39, 2016.

FREY, Carl Benedikt ; OSBORNE, Michael A. The future of employment: How susceptible are jobs to computerisation? *Technological forecasting and social change*, vol. 114, issue C, p. 254-280, 2017.

NILSSON, Nils. *The quest for artificial intelligence: a history of ideas and achievements*. Cambridge: Cambridge University Press, 2009.

MCCARTHY, J. MINSKY, M. L. ROCHESTER, N. SHANNON, C. E. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence. August 31, 1955. *AI Magazine*, vol. 27, nº 4. Disponível em: <<https://aaai.org/ojs/index.php/aimagazine/article/view/1904/1802>>. Acesso em: 3 jun, 2018.

OSOBA, Osonde. WELSER IV, William. *An intelligence in our image: the risks of bias and errors in artificial intelligence*. Santa Monica, CA: Rand, 2017.

TURING, Alan. Computing machinery and intelligence, *Mind*, vol. 59, issue 236, October 1950, p. 433-460.