

A INTELIGÊNCIA ARTIFICIAL: O INÍCIO DO FIM OU O FIM DO INÍCIO?

Eric Massaki Hirayama – massakihirayama@gmail.com

Universidade Federal de Santa Catarina
Campus Universitário Reitor João David Ferreira Lima
Rua Defino Conti, s/n, Trindade
88040900 – Florianópolis – Santa Catarina

Luís Felipe Martins Barbosa de Oliveira – luisfelipembocdb@gmail.com

Universidade Federal de Santa Catarina
Campus Universitário Reitor João David Ferreira Lima
Rua Defino Conti, s/n, Trindade
88040900 – Florianópolis – Santa Catarina

Resumo: *Da SIRI até carros autônomos, a inteligência artificial (IA) está progredindo rapidamente. Enquanto a ficção científica retrata a IA como robôs antropomórficos, a IA pode envolver tudo desde o algoritmo de pesquisa do Google até o Watson da IBM e armas autônomas. Inteligência artificial é conhecida atualmente como IA limitada (ou IA fraca), sendo projetado para executar tarefas limitadas como reconhecimento facial, pesquisas na internet ou dirigir um carro. Porém, o objetivo de longa data de muitos pesquisadores é criar uma IA geral (ou IA forte). Enquanto a IA limitada pode superar os humanos em qualquer tarefa específica seja, como jogar xadrez ou solucionar equações, a IA geral poderá superar os humanos em quase todas as atividades cognitivas.*

Palavras-chave: *Inteligência artificial, Relação homem-máquina, Futuro.*

1 Introdução

1.1 Histórico

O desenvolvimento da área começou logo após a Segunda Guerra Mundial, com o artigo "Computing Machinery and Intelligence" ("Computadores e Inteligência") do matemático inglês Alan Turing em 1950. O nome de inteligência artificial foi criado por John McCarthy em 1956 numa conferência de especialistas da área celebrada em Darmouth College. Seus principais idealizadores foram os cientistas Herbert Simon, Allen Newell, John McCarthy, Warren McCulloch, Walter Pitts e Marvin Minsky, entre outros.

Nos seus primeiros anos, a inteligência artificial obteve um bom sucesso, porém de forma limitada. Os primeiros computadores, que eram capazes apenas de realizar cálculos aritméticos, impressionavam a todos quando realizavam qualquer operação remotamente inteligente. Atualmente, com a evolução da tecnologia computacional, a IA abrange uma grande variedade de subcampos, entre eles está o estudo dos modelos conexionistas ou redes neurais. Este pode ser visto como um modelo matemático simplificado do cérebro humano. Este estudo se relaciona com a capacidade dos computadores de aprender e reconhecer padrões em ações mostradas a eles. Este tipo de modelo está incluso na forma conexionista da IA.

A forma coneccionista da inteligência artificial surgiu juntamente com sua outra abordagem, a IA “neat” ou limpa, mas foi abandonada nos anos 1960 e 1970 por acharem que a IA limpa resolvesse os problemas da forma mais eficiente. Porém, nos anos de 1980 os pesquisadores viram que haviam muitas limitações na abordagem limpa. Nesta época havia uma grande expectativa em cima do desenvolvimento das inteligências artificiais pesquisadas por grandes agências e os projetos surgiram com a esperança de incríveis resultados de forma imediata. Mas o que ocorreu foi que os resultados estavam demorando a sair e com isso os projetos começaram a ter vários cortes de fundos ocasionando o “inverno da IA”.

No decorrer da década seguinte, muitos pesquisadores de inteligência artificial decidiram partir para áreas relacionadas com metas mais modestas como aprendizado de máquinas, robótica e visão computacional.

1.2. O Sistema de Inteligência Artificial

O principal objetivo dos sistemas de IA é executar funções que seriam consideradas inteligentes se um humano fosse executá-lo. Este é um conceito tão amplo e que recebe tantas definições quanto damos significados diferentes para a palavra inteligência. Algumas das características básicas dos sistemas de inteligência artificial são a capacidade de raciocínio (utilizar de regras lógicas para uma gama de dados disponíveis para chegar a uma conclusão), aprendizagem (aprendo como os erros e acertos influenciaram o resultado de forma a agir de uma maneira mais eficaz), reconhecer padrões (padrões visuais, sensoriais e também padrões de comportamento) e inferência (conseguir aplicar o raciocínio nas situações do nosso cotidiano).

Steve Russel, professor de ciência da computação no Centro de Sistemas Inteligentes da Universidade da Califórnia em Berkley e autor do livro-texto da área, *Artificial Intelligence: A Modern Approach*, diz que a inteligência artificial deu grandes passos, em especial na força dos algoritmos de aprendizagem neuro-inspirada, utilizada em reconhecimento facial do Facebook, assistente pessoal nos smartphones e nos carros autônomos do Google.

Num resultado impressionante publicado na *Nature*, uma rede de neurônios artificiais aprendeu a jogar jogos de Atari de uma forma mais eficiente e melhor do que os humanos em questão de horas simplesmente com o comando de aumentar a sua pontuação ao máximo, mas sem pré-programar o conhecimento dos aliens, tiros, direita, esquerda, cima ou baixo.

As pesquisas foram se aprofundando e as IA se tornaram cada vez mais refinadas, com sistemas de aprendizado cada vez mais rápidas e potentes. Essa evolução se deve a um pouco ao melhor entendimento do que é a inteligência em si.

2 Intellectus

A inteligência artificial é um termo que se mostrou em uma grande discussão nos últimos anos, principalmente no século XXI, com o crescimento vertiginoso das tecnologias que envolvem a internet e os parâmetros inerentes que possibilitam ainda mais a ocorrência da inteligência artificial. Com o florescer de discussões mais profundas e abrangentes, e necessário

se entender os princípios destas discussões, e assim nasce o primeiro questionamento: o que é inteligência?

Muitos estudiosos das áreas da inteligência buscam descrever, exemplificar e caracterizar a inteligência, entretanto, assim como a inteligência possui sua beleza na sua forma de se manifestar, possui beleza também em suas definições. Uma das definições mais completas, no sentido de abranger sua complexidade, foi feita em um relatório, conhecido como *Intelligence: Knowns and Unknowns* (lit. “Inteligência: conhecido e desconhecido”), concebida pela Board of Scientific Affairs da American Psychological Association, e a mesma diz:

Os indivíduos diferem na habilidade de entender ideias complexas, de se adaptarem com eficácia ao ambiente, de aprenderem com a experiência, de se engajarem nas várias formas de raciocínio, de superarem obstáculos mediante o pensamento. Embora tais diferenças individuais possam ser substanciais, nunca são completamente consistentes: o desempenho intelectual de uma dada pessoa vai variar em ocasiões distintas, em domínios distintos, a se julgar por critérios distintos. Os conceitos de 'inteligência' são tentativas de aclarar e organizar esse conjunto complexo de fenômenos. (Neisser, U. et al, 1996).

Assim, pode-se compreender que, entender a inteligência depende da ocasião e do domínio de inserção, podendo assim julgar com os critérios certos a inteligência analisada. No caso, assimilar o ambiente onde a inteligência artificial se desenvolve serve de forma a avaliar o seu comportamento, as suas manifestações e possivelmente um auxílio para projetar o como essa inteligência pode evoluir e quais suas consequências.

3 O Ambiente da IA e seu “sangue”

Ao final de primeira metade do século XX, com o desenvolvimento das primeiras tecnologias de computação, a projeção de que estes artifícios tecnológicos um dia exerceriam uma compreensão sem que, necessariamente, ocorresse uma orientação direta humana, sempre foi especulado. I.J.Good, um dos parceiros que ajudaram Alan Turing no decorrer da Segunda Guerra Mundial a criar o mecanismo computacional que decifrava códigos nazistas, discorreu em um de seus artigos:

Descreva uma máquina ultra inteligente como uma máquina capaz de ultrapassar em muito todas as atividades intelectuais de qualquer homem, mesmo o mais inteligente. Uma vez que o design das máquinas é uma dessas atividades intelectuais, uma máquina ultra inteligente poderia projetar máquinas ainda melhores; haveria, sem dúvida, uma "explosão da inteligência", e a inteligência do homem ficaria para trás ... Assim, a primeira máquina ultra inteligente é a última invenção que o homem precisa fazer. (Good, I.J., 1965)

Assim, onde a máquina ultra inteligente poderia exercer sua existência? Qual seria a sua matéria prima de desenvolvimento, aquilo que fluiria em suas veias, possibilitaria sua vivência e sua ambição?

A internet certamente seria o ambiente de desenvolvimento da inteligência artificial. Um sistema global de redes que se comporta como uma bela incubadora capaz de disponibilizar recursos, segurança e atender às necessidades da IA. Entretanto, a internet se torna o melhor ambiente de desenvolvimento deste tipo de inteligência não só por suas conexões e teias, mas também pela matéria prima que faz a internet ser necessária: a informação.

Como foi mencionado, o “sangue” que corre pelas veias da inteligência artificial e as informações, em tempo real e de todos os lugares do globo ligados a rede de internet. Essa informação flui na inteligência artificial a nutre, hidrata, atua como linhas neurais, por fim, dá a vitalidade e a possibilidade da IA se desenvolver da forma como enxergamos, ou pelo menos da forma que nos compete interpretar no momento atual.

4 A singularidade Tecnológica

A inteligência artificial, por boa parte da sociedade intelectual, já é considerado uma realidade no cotidiano do ser humano já há alguns anos. Todavia, uma das grandes questões é quando essa inteligência terá capacidades maiores que as dos seres humanos em si.

Essa questão guia os pensadores até a questão da singularidade tecnológica. O termo singularidade normalmente é dado a questões que, em certo ponto, são surpreendentes e apresentam características incomuns das observadas anteriormente e que exigem uma grande reflexão para compreendê-las ou até mesmo render-se ao incomum e assumir não possuir conhecimentos suficientes para descrevê-lo.

A tecnologia possui, em muitas previsões, sua singularidade associada ao exato momento onde a inteligência artificial transpor a linha de conhecimento do ser humano. O momento onde as máquinas, *softwares* ou qualquer outro mecanismo que pode ser desenvolvido amanhã que ainda não temos conhecimento, conseguirem dentro de sua consciência, desenvolver conhecimentos avançados aos conhecimentos dos humanos, a singularidade se instaura. Como será a sociedade/ordem social após este marco? Qual será o comportamento da IA quando perceber que chegou neste patamar? São tantas questões onde não se há respostas exatas com os conhecimentos atuais e se projeta como um momento único, singular.

5 Benefícios e malefícios da IA

5.1 Como a inteligência artificial pode ser perigosa e o que é possível fazer para que ela seja benéfica?

Uma superinteligência artificial muito provavelmente não irá desenvolver emoções humanas como amor e ódio. Portanto não há razão para ela se tornar intencionalmente benevolente ou malevolente. Experts dizem que a IA se tornaria uma ameaça em dois casos. Uma é ela ser programada para fazer algo destrutivo – no caso de armas autônomas causarem

uma guerra por serem programadas para atacar as ameaças. A outra forma é ela ser programada para algo benéfico, mas desenvolve um método destrutivo para chegar no seu objetivo – um carro autônomo é solicitado a chegar o mais rápido possível no seu destino, mas para isso ele não respeita as leis e acaba atropelando pessoas e quebrando tudo em seu caminho.

Isto mostra que a preocupação com as inteligências artificiais avançadas não é a malevolência, mas sim a competência. Uma superinteligência artificial vai ser extremamente boa em atingir os objetivos desejados, mas se esses objetivos não estiverem alinhados com os nossos, isso pode causar desastres.

Para tentar diminuir o risco de acidentes acontecerem, o cientista computacional Stuart Russel escreveu uma carta aberta para todos os pesquisadores da área pedindo para que buscassem não somente uma inteligência artificial mais poderosa. “Nós recomendamos uma pesquisa abrangente visando garantindo que os sistemas IA cada vez mais capazes sejam robustos e benéficos” a carta afirma. Nessa mesma carta ele diz que os sistemas IA devem fazer o que nós quisermos que eles façam.

Esta carta foi assinada por milhares de pesquisadores do meio, incluindo líderes de desenvolvimento de inteligência artificial de grandes empresas como Google e Microsoft. Ao fim de março de 2015, ano da divulgação da carta, haviam mais de 300 grupos de pesquisas que se adaptaram para buscar novas formas de “manter a inteligência artificial benéfica” com fundos do 37º assinante da carta, o inventor e empreendedor Elon Musk.

Segundo Russel, uma das formas para os objetivos da IA estejam alinhadas com o que nós quisermos é a utilização de uma técnica chamada “aprendizagem de reforço inverso”. A aprendizagem de reforço normal é um processo onde você recebe recompensas e punições conforme seu comportamento e seu objetivo é conseguir imaginar qual é o comportamento que lhe concede mais recompensas. A aprendizagem de reforço inverso é o contrário como o nome já diz. Você vê o comportamento e tenta perceber qual resultado esse comportamento está tentando maximizar. Existem uma enorme quantidade de informações em livros, filmes e na internet sobre ações humanas e atitudes para as ações. Isto é uma ótima fonte de recursos para as máquinas aprenderem o que os valores humanos são – quem ganha condecorações, quem vai preso e porquê.

5.2 Quais benefícios a IA pode nos trazer?

Os potenciais benefícios são enormes. Tudo que uma civilização tem para oferecer é um produto da inteligência humana. Não conseguimos prever o que podemos alcançar quando essa inteligência for ampliada pelas ferramentas de IA, mas o fim das guerras, doenças, fome e pobreza podem estar no alto da lista de qualquer pessoa. O sucesso na criação da inteligência artificial como imaginamos pode ser o maior evento na história da humanidade.

Já podemos desfrutar de algumas inteligências artificiais que, por sua vez, passam despercebido pelos mais leigos. A SIRI da Apple, assistentes de *smartphones*, reconhecimento facial, entre outras tecnologias que utilizamos frequentemente são uma forma de IA.

A Google tem vários tipos de sistemas inteligentes. Ela está desenvolvendo um carro que consegue andar pelas ruas das cidades de forma autônoma, ou seja, sem um ser humano controlando o carro, seja ela de forma presencial ou remota. Foi desenvolvida também um sistema que consegue captar uma música ambiente e identificar qual é essa música.

Na área da medicina, a inteligência artificial pode ser usada nas consultas para achar qual é a doença ou então na procura da cura para elas como a AIDS e o câncer.

O público de uma forma geral cria expectativas em cima das capacidades da inteligência artificial. Com isso são geradas várias teorias e mitos com relação a

6 Mitos e fatos sobre a IA

É muito comum que as pessoas tenham certeza de que a singularidade irá ocorrer ainda neste século devido à velocidade que a nossa tecnologia se desenvolve. Mas a verdade é que isso foi resultado da grande expectativa tecnológica que havia em torno da inteligência artificial. Essa expectativa não é exclusiva dos leigos pois até alguns criadores da área acreditavam que a tecnologia iria avançar tão rápido ao ponto de se ter carros voadores em poucos meses. John McCarthy (criador do termo “inteligência artificial”), Marvin Minsky, Nathaniel Rochester e Claude Shannon propuseram que com 2 meses e 10 estudiosos da IA e os computadores da época (1956), iriam tentar criar uma máquina que use linguagens, forme abstrações e conceitos, solucionar problemas que somente humanos conseguem resolver e realizar melhoramentos em si mesma.

Pelo outro lado, um contra-mito bastante popular diz que não iremos alcançar uma super-IA neste século. Pesquisadores fizeram uma ampla estimativa de quão longe nós estamos desta super-IA, mas ainda assim não tiveram a certeza para dizer com convicção que a probabilidade de obtermos essa tecnologia até o ano de 2100 seja nula.

Ou seja, realmente não há como descobrir quando essa inteligência será criada. Mas há um consenso entre todos os envolvidos com a inteligência artificial que são as pesquisas sobre segurança devem começar agora. Muitos dos problemas de segurança associadas com IA de nível humano são tão difíceis que podem tomar décadas para se resolver.

Um outro mito muito difundido, principalmente por filmes e livros, é que os robôs irão criar consciência e irão se rebelar contra os humanos com armas em mãos. Porém, como citado anteriormente, as máquinas dificilmente irão desenvolver sentimentos. Isso gerou uma certa revolta para os pesquisadores em robótica, por ter gerado uma imagem ruim para seus trabalhos. As pessoas esquecem que o meio em que as IA trabalham é a internet, ou seja, ela está em todos os dispositivos que se conectam a ela.

Mesmo que construir robôs seja fisicamente impossível, uma superinteligente e rica IA pode facilmente comprar ou manipular inúmeras pessoas para fazer suas vontades. Tem gente que acredita que as máquinas não podem nos controlar, mas a inteligência é quem controla os outros. Nós dominamos os leões não pela força, mas sim pela nossa inteligência. Então para a IA nos dominar é só elas serem mais espertos que nós.

7 Responsabilidades do engenheiro com a IA

Os engenheiros são os principais responsáveis pelo início da era da inteligência artificial, e serão eles os responsáveis pelo crescimento e amadurecimento que essa inteligência terá nos próximos anos. Desta maneira, considerando os pressupostos de que cautelas devem ser tomadas afim de que este “adolescente” não venha a trazer consequências irreparáveis para a espécie humana.

Os mentores e desenvolvedores das novas tecnologias com inteligência artificial associada devem analisar, estudar, supor sobre as consequências que tecnologia pode trazer para a humanidade, boas ou ruins. O engenheiro desta área não poderá somente ser um engenheiro de ofício, o mesmo terá de ter competências das mais gerais, ser um cientista “total”, com conhecimentos de psicologia, sociologia, capacidade filosófica e um toque de intuitividade.

Em algumas obras artísticas, onde tratam a inteligência artificial de variadas maneiras expõem como a AI pode influenciar no cotidiano do ser humano, duas delas trazem especulações bastante pertinentes sobre como o engenheiro como projetista não projetou alguns fatores de como esta inteligência se comportaria no decorrer de seus anos de vida e acarretou em malefícios para a humanidade mesmo momentos antes de apenas prover qualidade.

Como primeiro exemplo é trazido a obra literária de Isaac Asimov, “Eu, Robô”, onde trata da revolta da robótica, um dos maiores medos dos leigos da IA, onde, resumidamente, a máquina toma consciência de suas ações, e toma a decisão de rebelar-se contra o ser humano, o explorador de suas atividades. Na obra, como forma de prevenção de uma possível revolta robótica, os projetistas elaboraram as 3 leis da robótica, que são as a seguir:

- 1ª Lei: Um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano sofra algum mal.
- 2ª Lei: Um robô deve obedecer às ordens que lhe sejam dadas por seres humanos exceto nos casos em que tais ordens entrem em conflito com a Primeira Lei.
- 3ª Lei: Um robô deve proteger sua própria existência desde que tal proteção não entre em conflito com a Primeira ou Segunda Leis.

Entretanto, uma falha na interpretação de como seu maquinário desenvolvia sua inteligência e como a mesma estava evoluindo, não perceberam que a mesma criou consciência capaz de questionar tais leis que antes eram máximas e em busca de um benefício próprio tomou decisões que não condizem com a proteção ao ser humano.

Nesse cenário pode se observar como certa omissão e não capacidade de acompanhamento dos engenheiros e projetistas desta IA fez com que a mesma passasse de uma aliada ao bem-estar humano a uma ameaça mortífera. Neste momento chega a discussão de como os engenheiros podem controlar a inteligência artificial após a singularidade ser atingida? São questões onde o engenheiro deve estar sempre atento, junto de grupos de pesquisa compostos por filósofos, sociólogos, psicólogos e qualquer outro estudioso ligado as partes humanas da ciência, afim de projetar estes cenários e projetar a IA da forma mais segura para a manutenção do bem-estar da humanidade.

Um outro cenário, criado pelos artistas, que mostra consequências de um mau projeto de inteligências artificiais e o filme de animação infantil “Wall-e”. Este filme trata, no âmbito da nave chamada no filme de AXIOM, da inteligência artificial em um nível onde a mesma existe para a total manutenção do bem-estar humano, realizando todas as atividades vitais como produção alimentícia, de energia entre outros, e o ser humano apenas aproveitava sua estada espacial se nenhum trabalho e máximo conforto.

A IA nesse caso possuía capacidade de desenvolver produtos e serviços que atendiam as demandas dos passageiros, porém a mesma ainda estava limitada a comandos dos seus programadores iniciais. A máquina fora programada para nunca mais retornar, em hipótese alguma, ao planeta terra, devido a um comando feito pelos projetistas que por irresponsabilidade, ou falta de previsão ocasionou consequências inesperadas. Com o aparecimento de evidências de que a vida na terra se tornou novamente viável, quando o ser humano decidiu que a sua casa retornaria, o mesmo foi reprimido violentamente pela máquina. Entretanto, o ato de violência não foi um ato “emocional” ou de consciência, foi apenas um ato desesperado de cumprir os comandos que seu software foi programado a receber.

É de importante percepção do engenheiro também entender que a máquina muitas vezes não pode ser nociva a humanidade apenas por questões físicas. Como já foi mencionado, apenas pelo fato da inteligência ser capaz de nos “enjaular” como fazemos com os tigres, já se retrata em ameaças para o gênero humano.

Desta maneira, criar limitantes, indicadores, ou simplesmente não evoluir de forma cega e imprudente esta tecnologia em prol dos benefícios capitalistas que os envolve e de suma importância no dever dos engenheiros como projetistas do futuro. Se as tecnologias de IA não forem concebidas de uma forma sensata, cautelosa e refletida, as consequências para os seres humanos podem ser as mais diversas e inimagináveis, para bem ou para mal. Seria o equivalente a dar “all-in” em um jogo de pôquer sem antes ver suas cartas, pura sorte.

Referências

Luger, George F (2004). *Inteligência Artificial. Estruturas e Estratégias para a Solução de Problemas Complexos* 4ª ed. Porto Alegre: Bookman. p. 23. 774 páginas. ISBN 85-363-0396-4

Rich, Elaine; Knight, Kevin (1994). *Inteligência Artificial* 2ª ed. Rio de Janeiro: McGraw-Hill. 722 páginas. ISBN 85-346-0122-4

Vasconcelos, V.V.; Martins Junior, P.P. *Protótipo de Sistema Especialista em Direito Ambiental para Auxílio à decisão em Situações de Desmatamento Rural*. NT-27. CETEC-MG. 2004. 80p.

Alan Turing. «Computing Machinery and Intelligence».

István S. N. Berkeley. «What is Artificial Intelligence?».

J. McCarthy, M. L. Minsky, N. Rochester e C.E. Shannon. «A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence»

Bruce Mazlish. «The man-machine and artificial intelligence»

<https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/>

https://www.washingtonpost.com/opinions/matt-miller-artificial-intelligence-our-final-invention/2013/12/18/26ed6be8-67e6-11e3-8b5b-a77187b716a3_story.html?utm_term=.dd78e4ef9796

https://www.huffingtonpost.com/stephen-hawking/artificial-intelligence_b_5174265.html

<https://www.newyorker.com/tech/elements/why-we-should-think-about-the-threat-of-artificial-intelligence>

<https://www.quantamagazine.org/artificial-intelligence-aligned-with-human-values-qa-with-stuart-russell-20150421>

Good, I. J. (1965), Franz L. Alt; Morris Rubinoﬀ, eds., "Speculations Concerning the First Ultra-intelligent Machine", *Advances in Computers*, Academic Press, 6: 31–88, doi:10.1016/S0065-2458(08)60418-0, ISBN 9780120121069, archived from the original on 2001-05-27, retrieved 2007-08-07