

A ÚLTIMA GRANDE INVENÇÃO OU O FIM DO HUMANO

Paulo Alexandre e Castro

paecastro@gmail.com

Universidade do Algarve - Centro de Investigação em Artes e Comunicação

Desde que o homem construiu a primeira máquina, certamente rudimentar e sem um nome adequado para lhe dar, que alterou o seu mundo. Não sabemos nem podemos saber que máquina seria, nem tão pouco a sua designação, mas a sua construção modificou desde esse momento a própria forma de pensar do homem. Esse momento inaugural deu-lhe acesso não apenas a uma visão diferente mas a um modo de estar e fazer diferentes. A máquina inaugural configurava um cenário de possibilidades; era possível ver de outro modo, fazer de outro modo, viver de outro modo com essa ou outras máquinas que se pudessem construir. Tal foi o fascínio primordial por essa criação que nunca mais desapareceu. Das simples arquiteturas como a criação do movimento de rodas às mais elaboradas com circuitos integrados, o deslumbramento está presente e assenta em algo profundamente humano: a *vanitas*. Uma tal vaidade não é necessariamente negativa, pois expressa muito daquilo que é o desejo de concretização e sucesso do homem, e nesse sentido, ela espelha também a capacidade inventiva do homem.

Uma tal visão das máquinas trouxe igualmente a imagem surpreendente dos entes que se encontravam no mundo como máquinas. O homem, como máquina, fosse pela forma estrutural do seu corpo, fosse pela forma como raciocinava, ajudou a acentuar o fascínio originário. Desde os finais do século XVI, considerado o início da visão científica do mundo, que a comparação da operatividade da mente com uma “calculadora” se dava a conhecer. Por exemplo, para Hobbes pensar mais não seria que calcular, e para La Mettrie na sua famosa obra *L’Homme Machine* a perspetiva mecânica/orgânica do homem seria uma realidade incontornável. Esta forma de pensar o homem levaria a que Pascal e Leibniz apresentassem as primeiras máquinas de calcular (ainda que a um nível muito elementar) e que viessem a assumir o estatuto de aparelhos que

“raciocinavam” (tal como os homens). Com a visão mecanicista deu-se um grande desenvolvimento das maquinarias de que o maior exemplo são os relógios, não apenas os de pêndulo mas sobretudo os das igrejas. O apuramento de relógios de igrejas eram frequentemente acompanhados de autômatos (figuras humanas, animais ou santos que executam movimentos de saudação ou movimentos de acerto de horas entre outros). Recorde-se neste contexto que o próprio mundo poderia ser visto como uma gigantesca máquina posto em andamento por um grande relojoeiro universal (de resto esta ideia está hoje ainda presente na sociedade sob o nome de teoria do “*designer* inteligente”, que infelizmente se vê a preencher os manuais escolares em detrimento da teoria evolucionista).

Retome-se ainda os autômatos. Estes manequins cinéticos geralmente procuravam assemelhar-se aos seres humanos e era quase inevitável estabelecer-se uma comparação no modo de funcionamento (há inúmeros exemplos de autômatos desta época em que eles articulam movimentos de braços, boca e olhos). Note-se que a ideia de um autômato (latinização da palavra grega $\mu\tau\omicron\mu\alpha\tau\omicron$, onde está implícita a noção de “vontade própria”) não é algo que apareça com a visão científica do mundo. Há relatos de autômatos na antiguidade (e do mesmo período na China), na época medieval e até ao Renascimento. A existência de tais autômatos fomentaram inúmeras e fantasiosas narrativas na literatura de que a mais famosa talvez seja o jogador de xadrez conhecido por “o Turco”, de Wolfgang von Keplem, (mas que não passava de uma máquina operada por um ser humano) e que Edgar Allan Poe poria em conto. O mecanicismo e o fascínio do humano pelas máquinas (no caso, os autômatos), sobretudo se essa máquina revelasse algum tipo de inteligência ou raciocínio (como jogar xadrez ou servir chá como os autômatos japoneses do século XVII e XVIII do período Edo) permitiu alimentar uma ideia que atravessaria séculos e ao mesmo tempo estimular o debate em torno dessa criação.

Permitindo-me produzir aqui um salto cronológico, avanço para as pesquisas de Alan Turing que patenteiam um debate começado no início do século XX sobre a possibilidade de se construir uma máquina com a qual se pudesse demonstrar a veracidade ou falsidade de afirmações matemáticas, sem qualquer intervenção humana. Com o desafio matemático provocado pelo “teorema da incompletude” de Gödel e os estudos de David Hilbert, a investigação de Turing começaria a ficar traçada: incitado pelo debate é elaborada a “Tese Church-Turing” e idealizada a máquina de Turing. A questão que desafiou Turing e os seus pares, a saber, se “as máquinas podem pensar?”, conduziria ao debate que ainda hoje se

mantém. É preciso compreender que a cibernética aparece no mesmo período, e que surpreendentemente Warren S. McCulloch e Walter Pitts acabassem por inventar o neurónio artificial. É a partir deste período que as analogias biológicas se acentuam e se tornam inevitáveis e permitem o aparecimento da biónica e os diferentes movimentos de fundação (e compreensão) de linguagens de comunicação e informação (Pereira, 2016). Acrescente-se ainda o rápido desenvolvimento de máquinas computacionais e o facto de através delas se associar o pensamento humano com a capacidade de elas estarem a pensar para produzirem um resultado. Não é assim de estranhar que no debate em torno da mente, Hilary Putnam tenha usado a metáfora do computador para afirmar que os estados mentais são propriedades funcionais, o que equivaleria a dizer que a mente agiria como um programa de computador. Embora a conhecida ilustração proposta por Searle do “quarto chinês” coloque em causa tal funcionalismo – uma vez que para Searle tratar-se-ia de mera manipulação de símbolos –, não deixa de ser curioso observar o facto de a pessoa estar precisamente a manipular bem os símbolos (não sabendo falar chinês) tal como se de uma máquina se tratasse (Searle, 1980, p. 419). Isto não significa que desconhecamos o seu sentido e a importância do ensaio que visa refutar essa visão reducionista de Putnam, mas tão-somente, do ponto de vista funcional, estar a agir como uma máquina.

Naturalmente neste cruzamento entre autómatos e inteligência artificial não pode nem deve ser esquecido o grande pioneiro da robótica, Hiroshi Ishiguro, que tem dedicado a sua vida à robótica e à compreensão daquilo que significa ser humano. No seu laboratório procura não só criar andróides semelhantes a humanos – a que dá o nome de “Geminoids” – mas criá-los com capacidade de resposta emocional e portanto, com resposta adequadas à simulação de consciência (se se atentar na literatura elaborado pelos investigadores do Hiroshi *Ishiguro Laboratories*, perceber-se-á claramente a direção da pesquisa). Os modelos tele-operados *Geminoid HI-2* (modelado à semelhança do próprio Ishiguro) e o *Geminoid F*, um *android* (corretamente deveria ser uma *ginoid*) com aparência de uma mulher, são usados para investigar a capacidade de influenciar as pessoas e para se saber o que as pessoas experienciam nessa conversação. Note-se que já existe o *Geminoid HI-4* (também cópia de Hiroshi Ishiguro) em que se afirma que este dispõe de dezasseis graus de liberdade, que lhe permite, aparentemente, comportar-se como um humano. Na verdade, a investigação faz aparecer o lado mais humano nos robôs para que estes sejam

aceites. (dir-se-ia uma outra espécie de teste de Turing)¹. A recente demonstração feita (no início de 2016) pela *Boston Dynamics* (detida pela Google), em que um humanoide corre no meio de uma floresta, conseguindo gerir e ultrapassar os diferentes obstáculos, é prova do avanço na área.

As investigações em computação e robótica (em concreto em inteligência artificial) tem aberto inúmeras possibilidades e permitem já sustentar a inevitabilidade do aparecimento de andróides surgirem com um grau de inteligência artificial muito superior ao já existente, e eventualmente até com consciência (artificial). Alguns ensaístas têm procurado expor e gerir as condições de tais aparições futuras. Veja-se para o efeito a posição de Alva Noë e as suas premissas: as máquinas inteligentes tem a mesma possibilidade de afirmação que os homens, considerando (isto é admitindo) que o cérebro humano não seja um elemento fundacional (ou não é apenas o elemento fundacional segundo Noë) da consciência. Colocando a questão de outro modo, a consciência não sendo algo que esteja na “cabeça” (Noë, 2009, p. 24) e, que a consciência se baseia fundamentalmente na construção de hábitos desenvolvidos na interação com o mundo (“a minha hipótese: só um ser com hábitos pode ter uma mente como a nossa” (Noë, 2009, pp. 97-98). Então é legítimo conceber-se a aparição de uma “consciência artificial”, proveniente de uma “vida artificial” (Noë, 2009, p. 45). Também e neste mesmo sentido, Luís Moniz Pereira sustenta que tal será uma realidade, nomeadamente, através do que se concebeu como *Cognoma*, referindo assim que

é essencial começarmos por introduzir uma noção com raízes na Psicologia Evolucionária. Referimo-nos à noção de *Cognoma*, por justaposição e analogia com a de Genoma. *Cognoma* significará a cognição no homem, na máquina, e no respetivo interface. (...) Inevitavelmente, vamos integrar-nos na teoria da evolução, tal como, desde os anos 70, a criatividade genética induziu John H. Holland (1929-2015) ao desenvolvimento de toda a computação genética, por via do conceito de algoritmos genéticos. (Pereira, 2016, p. 11).

Embora nos pareça uma ideia demasiado rebuscada, quer dizer, uma ideia demasiado ficcional, os avanços na área parecem ir nesse sentido, na senda de Daniel Dennett, ao sugerir-se que há uma possibilidade legítima de reflexão que

1 O “Uncanny Valley” diz que quando réplicas humanas se comportam de forma muito parecida — mas não idêntica — a seres humanos reais, provocam repulsa entre observadores humanos.

considera a possibilidade de haver conteúdos de consciência que possam vir a ser re-inventados. É ao mesmo tempo uma possibilidade perturbadora e fascinante que abre ainda para outra hipótese extraordinária do ponto de vista da filosofia da mente: *qualia* não sentidos por humanos mas *qualia* apenas sentidos por robôs portadores de consciência artificial. Na verdade, não estamos a sugerir nada de particularmente original, pois parece que a investigação em robótica se dirige muito para reinventar a consciência, para reinventar os *qualia*, com as consequências (dramáticas) que a sétima arte nos tem dado a ver e que ilustram o pior dos mundos possíveis.

Hoje, a velocidade de processamento e de armazenamento de informação, de *data*, de um qualquer computador doméstico permite avaliar como em meras duas décadas se avançou vertiginosamente. A inteligência, apanágio do ser humano durante séculos, passou a ser atribuída com muita facilidade a qualquer instrumento. Calum Chace, um teórico da área, tem alertado para essa vertigem da computação. Refere o ensaísta que um *Smartphone* com o jogo de xadrez como o “Pocket Friz 4” ultrapassa largamente o computador “Deep Blue” da IBM (Chace, 2015, p. 23), que terá vencido o extraordinário jogador de xadrez Gary Kasparov.

Se aqui ainda se pode pressupor que tudo o que o “Deep Blue” fez foi o resultado de tudo aquilo que lá colocaram e pôde executar (estariam programadas todas as combinações possíveis), nos casos seguintes a realidade é outra. Tome-se o “Watson” (outro sistema da IBM) que conseguiu pela primeira vez em 2011 bater concorrentes humanos no famoso jogo *Jeopardy*, em que são dadas as respostas e os concorrentes tem de adivinhar as perguntas. Dito assim, o leitor é tentado a ver tudo isto como um grande jogo em que os computadores acabam por sair vencedores pois sabem de antemão todas as jogadas possíveis. Para que seja claro darei mais um exemplo concreto. O supercomputador “KnIT” surgido em 2013 (IBM e Baylor College of Medicine) leu cerca de 190.000 publicações (artigos científicos) em duas horas, gerindo a informação sobre a proteína p53, e dando a conhecer hipóteses de trabalho ainda não estabelecidas pelos cientistas. Mas é com o “DeepMind” que a revolução se está a operar: o sistema ensina-se a si mesmo (e resolve problemas) sem ter sido especificamente programado para esse fim. Semelhante capacidade deixa antever a batalha que se começa a perder e que poderá ser imparável (Chace, 2015, p. 24). Note-se ainda a este propósito aquilo que Nick Bostrom refere a propósito da quantidade de jogos em que os humanos já são completamente ultrapassados por computadores/máquinas (Bostrom, 2014, pp. 15-17).

Nesse sentido, as indicações de Vernor Vinge, Bill Joy ou Raymond Kurzweil sobre o avanço tecnológico da inteligência artificial devem servir como ponto de partida para uma reflexão ética. Alguns teóricos têm manifestado essa preocupação.

Raymond Kurzweil sugere com a noção de “Singularidade” para referir esse ponto em que a inteligência artificial terá superado a inteligência humana, e Nick Bostrom postula que a possibilidade de uma superinteligência pode permitir fugir ao controle humano. Não se pense que se trata de ficção científica. A conhecida tese da “emulação do cérebro” coloca a hipótese da construção de uma mente (copiar ou replicar ao mais pequeno pormenor as estruturas do cérebro humano de tal modo que as “respostas” possam ser iguais ao original). Esta possibilidade surge, é de crer, sustentada por projetos paralelos que já decorrem há alguns anos como o “Blue Brain Project” ou o “Human Connectome Project”, que procuram mapear o cérebro humano (neurónios e respetivas ligações). Na revista *Science*, um dos cientistas responsáveis, Henry Markram declarou que: “deve ser possível dentro de dez anos modelar o cérebro humano em silício”. Calum Chace refere a este propósito: “olhem para as três maneiras de construir uma mente - um sistema artificial que pode realizar todas as atividades intelectuais que qualquer adulto humano pode. Elas são: 1. A emulação inteira do cérebro; 2. Baseando-se em inteligência estreitamente artificial; 3. Uma teoria da mente abrangente (Chace, 2015, p. 67).

Projetos de reconhecimento da linguagem através de impulsos neuronais são já hoje uma realidade (não de reconhecimento de voz como os que se utilizam nos *smartphones* ou *GPS*, mas de reconhecimento de sinais neuronais como pensamento). Calum Chace refere que é logicamente possível que uma “mente” dessas não só possa ter vontade própria como possivelmente será bem mais eficiente na resolução de problemas (Chace, 2015). Repare-se que um dos argumentos mais referidos para contestar a capacidade de um computador simular uma mente humana baseia-se na habilidade que ele teria de poder manter uma conversa normal (como se fosse uma pessoa humana). Stuart Armstrong refere-se assim:

Considere o que aconteceria se a IA atingisse a capacidade de funcionar socialmente – manter uma conversa com uma razoável *fac-simile* fluência humana. (...) Com a habilidade de conversar vem a habilidade de convencer e de manipular. Com um bom quadro estatístico, teorias sociais válidas, e a habilidade de ler as reações de uma audiência em tempo real, las poderiam aprender a fazer o mais convincente e emocionante

discurso. Em pouco tempo, o nosso cenário político poderia ser dominado por IAs ou por humanos capacitados com IA (...) Se uma IA possuir alguma destas capacidades – habilidade social, desenvolvimento tecnológico, capacidade económica – a um nível sobre-humano, é muito provável que rapidamente pudesse dominar o mundo de uma forma ou de outra (Armstrong, 2014, p. 37).

O risco (existencial) surge para o tempo vindouro. Já não se trata pura e simplesmente de simulação mas de verdadeira inteligência artificial (possivelmente a designação artificial acabará por cair em desuso) ou de superinteligência que “processe” informação em tempo real.

A preocupação de alguns cientistas sobre os riscos do desenvolvimento da inteligência artificial ou como também é chamada robotização inteligente, que se espera venha a ser alargada à sociedade em geral, teve uma das suas primeiras manifestações (a notícia original surgiu no *The Independent* de 28 de julho de 2015), com a assinatura de uma carta aberta em que participaram Stephen Hawking e Noam Chomsky (entre mais de mil cientistas e filósofos, em que se inclui o cofundador da Apple, Steve Wozniak) a pedir a proibição dos “robôs assassinos”². Esta iniciativa coordenada pelo *Future of Life Institute*, procura mitigar os riscos existenciais com que a humanidade se depara³. A que se deve este receio? Na verdade, as questões a colocar são de ordem ética, pois nada neste cenário parece claramente identificável. Assim, onde entra a ética ou a bioética neste processo imparável? Que valores assumiram essas superinteligências? Que quadro de valores morais e jurídicos terão?

Um breve parêntesis: embora se encontrem na vasta literatura de ficção científica diversas considerações éticas relevantes sobre o futuro da humanidade, recupere-se uma feita por Searle. O filósofo proferiu uma conferência no *Office of Technology*, em 9 de novembro de 1984, intitulada “Computers, The Mind and Responsibility”, em que referiu que as máquinas desenvolvidas com inteligência artificial não poderiam ser chamadas de “responsáveis” (leia-se

-
- 2 A preocupação advém do facto de estar a ocorrer uma corrida ao armamento militar/policial inteligente. Tem-se desenvolvido de forma paralela à investigação científica o armamento de armas autónomas de ataque, ou, como também são apelidados “robôs assassinos” (as armas robóticas são aviões não tripulados e os *drones*).
 - 3 Recorde-se que, já em 2012, um grupo de organizações não-governamentais se tinha associado para criar uma campanha para “parar os robôs assassinos”, e cujo objetivo seria eliminar as armas robóticas da investigação e do uso. Crê-se que esta matéria esteja a ser abordada pela Convenção da ONU das Armas Convencionais de Genebra, no sentido de estipular a proibição deste tipo de armamento, tal como foram proibidas as minas antipessoais e as armas de raios laser.

responsabilizadas). Permito-me finalizar com a seguinte pergunta de Paul Virilio: estaremos na era da ciência póstuma, da era “pós-científica” em que a ciência e o homem, não sobrevive a si mesmo? Se se pensar num cenário futurístico catastrófico (que pode e deve ser equacionado), nada impedirá que um tal ser não se possa desenvolver de acordo com as suas capacidades e perpetrar uma série de neurocrimes aos seres inferiores, os humanos, ou que os possa submeter como animais (afinal, seres menos inteligentes) ao trabalho escravo. É sobre cenários desta natureza que a Ética e o Direito tem a obrigação de começar a pensar.

A construção de uma Inteligência Artificial amigável (Bostrom e Muehlauser, 2014), tal como sugerem, Chace, Armstrong e Bostrom entre muitos outros, é desejável para que não sejamos como crianças a brincar com explosivos, sem qualquer preparação (Bostrom, 2014, p. 319).

Tal é o fascínio que a última grande invenção produz.

Referências bibliográficas

- ARMSTRONG, S. (2014). *Smarter Than Us. The rise of machine intelligence*. Berkeley: MIRI.
- BOSTROM, N. (2014). *Superintelligence. Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- BOSTROM, N. & MUEHLAUSER, L. (2014). Why We Need Friendly AI. *Think*, 13(36), 41-47.
- CHACE, C. (2015). *Surviving AI. The promise and peril of artificial intelligence*. Londres: Three Cs Publishing.
- HERFF, C.; HEGER, D.; DE PESTERS, A.; TELAAR, D.; BRUNNER, P.; SCHALK, G. & SCHULTZ, T. (2015). Brain-to-text: decoding spoken phrases from phone representations in the brain. *Front. Neurosci*, 9(217). doi: 10.3389/fnins.2015.00217.
- JOY, B. (2000). Why the Future Doesn't Need Us. *Wired*, 8(4), 1-18.
- KURZWEIL, R. (2006). *Singularity is Near*. Londres: Gerald Duckworth.
- MARTIN, S.; BRUNNER, P.; HOLDGRAF, C.; HEINZE, H.-J.; CRONE, N. E.; RIEGER, J.; SHALCK, G.; KNIGHT, R. & PASLEY, B. (2014). Decoding spectrotemporal features of overt and covert speech from the human cortex. *Front. Neuroeng*. 7(14). doi: 10.3389/fneng.2014.00014.
- MILLER, Greg (2011), Blue Brain Founder Responds to Critics, Clarifies His Goals. *Science*, 334(6057), 748-749. doi: 10.1126/science.334.6057.748.
- MUGLER, E. M.; PATTON, J. L.; FLINT, R. D.; WRIGHT, Z. A.; SCHUELE, S. U.; ROSENOW, J.; SHIH, J.; KRUSIENSKI, D.; SLUTZKY, M. (2014). Direct classification of all american english

- phonemes using signals from functional speech motor cortex. *Journal of Neural Engineering*, 11(3). doi: 10.1088/1741-2560/11/3/035015.
- PEREIRA, L. (2016). *A Máquina Iluminada*. Porto: Fronteira do Caos.
- PUTNAM, H. (1960). Minds and Machines. In S. Hook (Eds), *Dimensions of Mind* (pp. 148-180). Nova Iorque: New York University Press.
- PUTNAM, H. (1999). Brains in a Vat. In K. DeRose & T. Warfield (Eds), *Skepticism: A Contemporary Reader* (pp. 27-42). Oxford: Oxford University Press.
- RUSSEL, S. & NORVIG, P. (2003). *Artificial Intelligence: A Modern Approach*. Upper Saddle River: Prentice Hall.
- SEARLE, J. (1980). Minds, Brains and Programs. *Behavioral and Brain Sciences*, 3(3), 417-457.
- TURING, A. (1950). Computing Machinery and Intelligence. *Mind*, 236, 433-460.
- VIRILIO, P. & LOTRINGER, S. (2002). *Crepuscular Dawn*. Los Angeles: Semiotext.