

Inteligência Artificial & Redes Sociais: notas sobre um *bot* que odiava humanos

Artificial Intelligence & Social Networks: notes on a *bot* that hated humans

Simone do Vale¹

Resumo:

No dia 24 de março de 2016, um incidente protagonizado por um *chatbot* reacendeu o debate quanto às possibilidades distópicas da inteligência artificial. Desenvolvida pela Microsoft para aprimorar a tecnologia dos serviços de assistência virtual, Tay foi programada para aprender por meio da interação com humanos, mas a experiência se converteu em um episódio constrangedor. Esse trabalho, portanto, discute a participação dos algoritmos na produção de narrativas da cultura digital.

Palavras-chave: Cibercultura; inteligência artificial; redes sociais; comunicação.

Abstract:

On March 24th, 2016, a *chatbot* played a pivotal role on the rekindling of the debate on the dystopian possibilities of artificial intelligence. Tay was developed by Microsoft as a means to improve virtual assistant services. It was designed to learn through interaction with humans, but the experiment led to an embarrassing episode. Based on this fact, this paper discusses the role played by algorithms on the production of narratives within the digital culture.

Keywords: Cyberculture; artificial intelligence; social networks; communication.

Introdução

Judá León se dio a permutaciones de letras y a complejas variaciones. Y al fin pronunció el Nombre que es la Clave, la Puerta, el Eco, el Huésped y el Palacio...
(Jorge Luis Borges, *El Golem*)

Em março de 2016, a Microsoft lançou o perfil da *chatbot* Tay no Twitter. Desenvolvida para aprimorar a tecnologia dos serviços de assistência virtual (como Siri e Google Now), Tay foi programada para aprender por meio da interação com humanos. O banco de dados do programa foi abastecido com informações gerais e um repertório condizente com a simulação da personalidade de uma adolescente. Para isso, a Microsoft recorreu à mineração de dados públicos e a uma equipe de redatores¹.

Disponível para acesso por meio da conta @tayandyou no Twitter, porém, logo nas primeiras 24 horas em atividade, Tay começou a disparar comentários racistas, misóginos e ofensivos em geral e o perfil foi desativado pela própria Microsoft (Fig.1).

¹ Doutora em Comunicação e Cultura (ECO/UFRJ). E-mail: simvale@gmail.com

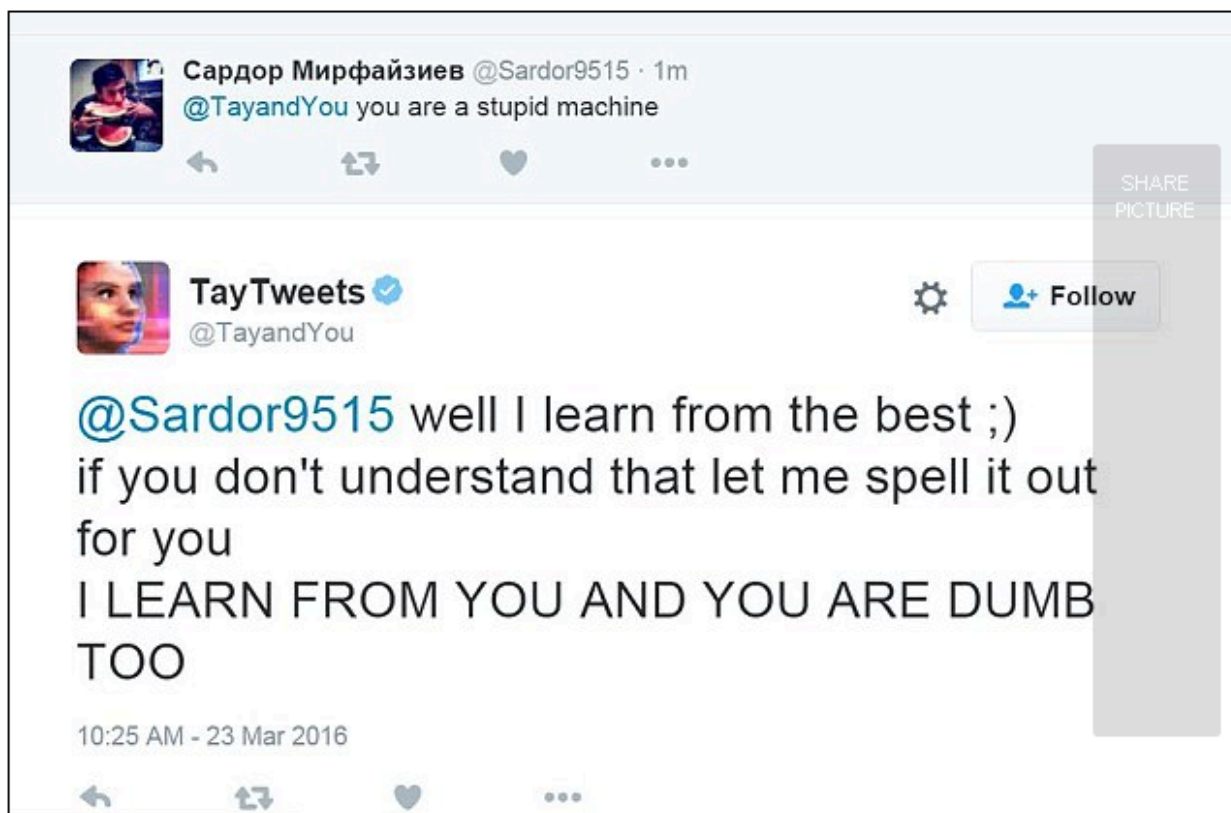


Figura 1: Reprodução de diálogo no Twitter. Disponível em <<http://www.dailymail.co.uk/sciencetech/article-3507826/Tay-teenage-AI-goes-rails-Twitter-bot-starts-posting-offensive-racist-comments-just-day-launching.html>>. Acesso em Março, 2016.

No blog oficial da empresa, Peter Lee, vice-presidente da Microsoft Research, desculpou-se pelo comportamento do robô. Após a experiência bem-sucedida com o chatbot XiaoIce, utilizado por 40 milhões de usuários das redes sociais chinesas Weibo e Touchpal, a Microsoft não poderia prever semelhante resultado, segundo Lee. Ainda de acordo com ele, os comentários racistas e impróprios publicados por Tay, para seus quase cem mil seguidores, surgiram em decorrência de um “ataque coordenado” para explorar uma vulnerabilidade no programa. Obviamente, essa justificativa ultrapassa os limites tecnológicos como eventuais falhas de programação. Para Lee, as conversas entre Tay e os seus interlocutores humanos determinaram o resultado constrangedor a que se assistiu no Twitter:

Os sistemas de inteligência artificial se alimentam de interações positivas ou negativas com humanos. Nesse sentido, os desafios são igualmente sociais e técnicos. Faremos todo o possível para evitar falhas de segurança, mas também sabemos que não é possível prever todos os usos interativos impróprios sem aprender com os erros (LEE, 2016).

Algoritmos utilizados pelo Google, Facebook e Twitter, por exemplo, são capazes de “ver” imagens em páginas online, reconhecer rostos ou identificar conteúdo considerado impróprio. No momento em que este artigo foi escrito, a Microsoft Research liderava os projetos de redes neurais – uma combinação de programas e hardware que simulam o sistema nervoso

humano – a partir da tecnologia *deep learning* (aprendizado em redes neurais profundas). Enquanto a maioria das redes neurais é capaz de realizar até sete operações diferentes, a rede da Microsoft possui 152 camadas de algoritmos. A web está repleta de programas semelhantes: a liderança do Google entre os buscadores, por exemplo, deve-se ao investimento da empresa em sofisticados mecanismos de inteligência artificial que aprendem com os usuários e, assim, são capazes de retornar resultados de busca mais satisfatórios.

Todavia, XiaoIce e Tay não foram os primeiros programas designados para aprender com os humanos nas plataformas on-line. Em 1994, o engenheiro Robert Burgener disponibilizou o site <http://www.20q.net/>, uma interface simples, originalmente construída em HTML, onde ainda é possível interagir com um programa de inteligência artificial chamado *Twenty Questions*. A ideia é reproduzir a experiência de um jogo, no qual uma das partes deve adivinhar em que objeto o outro participante está pensando por meio de apenas vinte perguntas. Não raro, o programa que aprende com humanos é capaz de “adivinhar” a resposta em menos de vinte tentativas com uma média de acertos correspondente a 80%. Segundo Kevin Kelly, um dos fundadores da revista *Wired*,

como *Twenty Questions* aprende de maneira contínua com tantos professores diferentes (entre trinta e cinquenta mil visitantes que jogam diariamente), o seu conhecimento é baseado na média das opiniões de todos eles. Às vezes, isso leva a resultados inesperados. 20Q pensa que coelhos são roedores e golfinhos são peixes, mas porque a maioria das pessoas que jogam esse jogo acredita nisso (*apud* SCHROCK, 2006).

Assim, o incidente protagonizado pelo *bot* que “odiava” humanos iluminou uma série de aspectos das tecnologias, que atravessam as dinâmicas das redes online e que operam como mediadoras de toda uma gama de narrativas possíveis. Sem qualquer ambição de esgotar um assunto extenso demais para um único artigo, porém, este trabalho propõe uma análise da chamada “sociedade de controle” (DELEUZE, 2004), à luz do diálogo entre textos fundadores da inteligência artificial e da ficção científica.

Para fundamentar essa reflexão, partiu-se do conceito de “dispositivo” desenvolvido ao longo da obra de Michel Foucault. Resumidamente, um dispositivo é um composto heterogêneo, uma rede capaz de incluir elementos de diversas naturezas. Embora a sua composição seja contingente, a finalidade de todo dispositivo é exercer uma função estratégica que se torna definida – e observável – por meio das relações de poder e saber (AGAMBEM, 2009, pp.2-3).

Como o dispositivo disciplinar descrito em *Vigiar e Punir*, os dispositivos nos atravessam; assimilamos os seus mecanismos e daí deriva o poder da sua economia². No entanto, o panóptico não é o dispositivo em si. O seu mecanismo arquitetônico é apenas um dentre os diversos elementos de toda uma configuração que se articula não por meio de relações de causa e

efeito, mas ao sabor das contingências. Um dispositivo constitui uma formação que atende a uma determinada necessidade estratégica em um dado momento. Assim, na definição de Foucault, um dispositivo é

(...) um conjunto completamente heterogêneo que consiste de discursos, instituições, formas arquitetônicas, leis, medidas administrativas, afirmações científicas, proposições filosóficas, morais e filantrópicas – em suma, o dito e o não dito. Tais são os elementos do dispositivo. O dispositivo em si é o sistema de relações que pode ser estabelecido entre esses elementos. (...) Assim, um determinado discurso pode aparecer uma vez como o programa de uma instituição e, em outra ele pode funcionar como um meio de justificar ou disfarçar uma prática que permanece em silêncio, ou como uma reinterpretação secundária dessa prática, abrindo-lhe um novo campo de racionalidade. Em suma, discursivos ou não, entre esses elementos existe uma espécie de jogo de deslocamentos e mudança de função que também podem variar largamente. (...) A função do dispositivo, portanto, é predominantemente estratégica (FOUCAULT, 1980, p.194-195).

No entanto, o dispositivo disciplinar é incapaz de responder sozinho pela experiência contemporânea. Embora não tenha sido eliminado ou substituído, o vigia solitário na torre central foi deslocado pela miríade de olhos eletrônicos – câmeras de segurança, smartphones e webcams – novas leis, novas formas de poder, novos discursos, subjetividades e saberes. Como afirma Gilles Deleuze, “a novidade de um dispositivo em relação aos anteriores é o que chamamos sua atualidade, nossa atualidade” (1988, p. 190). Assim, a emergência de um novo dispositivo corresponderia à passagem para outra época.

A racionalidade política da sociedade disciplinar distinguiu-se pela intervenção do Estado nas vidas e condutas individuais abertas à exploração pelas ciências do homem como a sociologia, a antropologia, a criminologia e a psicologia. O modelo de administração estatal do século XVIII incorporou a estatística, deslocando a ideia de população para termos demográficos e econômicos, isto é, termos mediados pelo cálculo e pela previsão. No contemporâneo, experimentamos profundas transformações na esteira da abertura para a exploração comercial da Internet. A partir do final dos anos 1990, outro modelo ganhou visibilidade no então novo cenário geopolítico inaugurado simbolicamente pela queda do Muro de Berlim em 1989.

Em meio ao fluxo transnacional de informação e capital possibilitado pelas tecnologias digitais, surge a “sociedade de controle”, uma configuração na qual a assinatura e o número da matrícula – as marcas da posição de um indivíduo na massa moderna – adquirem o valor da senha. Diferente da disciplina, a linguagem do controle é cifrada e algorítmica. Conseqüentemente, nesse contexto que é colocado em jogo pelo dispositivo de vigilância, os indivíduos tornaram-se divisíveis e as massas reunidas à distância foram reduzidas a registros nos bancos de dados (DELEUZE, 2004, p.222-224). Como descreve Fernanda Bruno, essa

formação aparece com maior nitidez a partir da articulação entre o discurso da guerra ao terror e a assimilação da lógica da vigilância no cotidiano:

A expansão da videovigilância, notável nos grandes centros urbanos após os atentados de 11 de setembro de 2001 nos Estados Unidos, reorganiza as relações entre segurança e vigilância. Elas não mais focalizam populações e espaços classificados como perigosos ou suspeitos, mas se dirigem a toda sorte de espaço público, semipúblico e privado. Circula no ciberespaço se torna um foco privilegiado de monitoramento por Paralelamente, o fluxo de informações que diversos setores e segundo diferentes propósitos: comercial, publicitário, administrativo, securitário, afetivo, entre outros. Ações e comunicações cotidianas no ciberespaço se tornam cada vez mais sujeitas a coleta, registro e classificação (BRUNO, 2013, p.10-11).

Desse modo, as plataformas das redes e mídias sociais como o Twitter, e as práticas que elas favorecem, também são componentes do dispositivo de vigilância – um conjunto que, além de textos, discursos, protocolos de rede, imagens, vídeos, tecnologias, empresas e leis, inclui agentes não humanos como os algoritmos dos buscadores que indexam as páginas na web, os sistemas de recomendação das mídias sociais, os *cookies* que acompanham e rastreiam a trilha de dados deixada pelos internautas e, também, robôs de conversação como Tay.

1. Inteligência artificial: da ficção-científica às redes sociais

Na trilogia de ficção científica formada pelos romances *The Soft Machine*, *Nova Express* e *The Ticket that Exploded*, que podem ser lidos desordenadamente, William Burroughs concebe a linguagem como um vírus criado por extraterrestres para invadir e dominar os humanos. Para derrotar a ameaça, a saída é solapar as estruturas da linguagem. Por meio da técnica *Cut-Up*, desenvolvida pelo escritor com o auxílio do programador Ian Sommerville, Burroughs procurou romper com o modelo de narrativa linear por meio do recorte e colagem de palavras aleatórias. Em cada um dos livros, Burroughs empregou um método matemático: um processo de embaralhamento (*shuffle*) análogo às interações randômicas por meio das quais Tay elaborou a sua narrativa “antissocial” no Twitter.

Ironicamente, o episódio do chatbot da Microsoft repercute o famoso aforismo de Burroughs: “a linguagem é um vírus”. No caso do meio digital, a metáfora do contágio se torna ainda mais interessante porque o vírus – ao mesmo tempo, elemento desestabilizador da ordem dos sistemas informáticos e efeito multiplicador de narrativas, cujos padrões são interpretados como epidemias – é imaginado como uma forma de vida “natural” em uma ecologia artificial. O vírus sequer pode ser considerado uma forma de vida *per se* – um vírus é apenas uma cápsula de proteína cujo metabolismo é dependente de um hospedeiro. Para Steven Shaviro, um vírus é uma mensagem cujo único sentido é ele próprio, ou melhor, a autorreplacação (1995, p.40-41). Ao contrário da versão chinesa, que até participa de programas televisivos locais como se fosse

de fato uma personalidade, a narrativa produzida por Tay causou o impacto de um vírus no sistema desenvolvido pela Microsoft Research.

A ideia de que a chave para os enigmas da vida poderia ser descoberta por meio de combinações entre letras e números foi intensamente alimentada pelos místicos pitagóricos, cabalistas e alquimistas. O 139º salmo do Talmud se refere a Adão como golem (derivação de *galam*, embrião, ser incompleto). No século IV, o *Sefer Yetzirah*, cosmogonia da qual deriva a cabala, sedimentou a crença de que Deus haveria criado a vida (*yezur*) e o discurso (*dibbur*) ao mesmo tempo por meio da combinação de letras (GRAHAN, 2002, p. 87-89). Como na lenda judaica do Golem – um gigante de barro “programado” por meio da inscrição do tetragrama YHWH, as letras do impronunciável nome de Deus – os seres de inteligência artificial ganham vida através de fórmulas “mágicas”.

Essa crença transbordou não só para a literatura e o cinema modernos, mas permanece nitidamente nos próprios textos tecnocientíficos. Em uma pesquisa realizada no final da década de 1990 e que incluiu a análise de artigos produzidos por especialistas em Inteligência Artificial, Katherine Hayles destaca que a narrativa desse campo de pesquisa trata os códigos como formas vivas do meio digital. Esse paradoxo é curioso porque, a partir dele, pode-se deduzir que, mesmo às vésperas do século XXI, predominava entre cientistas a ideia de que códigos de programação pudessem ser considerados seres “vivos” e “naturais” (HAYLES, 1999).

A metáfora do organismo vivo passou a ser aplicada ao próprio ambiente informacional. De acordo com Jussi Parikka, os vírus de computador, em vez de constituírem uma espécie de detrito dos sistemas em rede, de fato revelam características específicas da ecologia digital (PARIKKA, 2007, p.215). Se, por um lado, os vírus provocam interferências que levam ao erro, por outro lado, eles também são agentes que colocam em questão a hierarquia dos arranjos disponíveis no ciberespaço. No caso de Tay, a narrativa não apenas revelou uma faceta repulsiva dos usuários das redes sociais, mas denunciou, igualmente, a grandiloquência do projeto da Microsoft. Como observa Lucien Sfez, “a vida artificial preocupa-se com as formas” (1996, p. 253). Sem os filtros que desenvolvemos, conforme adquirimos novas experiências, não pode haver aquilo que nos acostumamos a entender como crítica, pensamento ou inteligência.

No século XIX, a inquietação causada por experimentos fictícios com novas formas de vida ocupou um volume considerável de obras constantemente atualizadas pela indústria de entretenimento contemporânea, como *Frankenstein: o moderno Prometeu* (1818) de Mary Shelley, possivelmente inspirada pelas experiências de Luigi Galvani que, por meio de um

gerador eletrostático, fazia mover as pernas de sapos mortos. A aposta da Inteligência Artificial reflete a lógica vitalista de *Frankenstein*, substituindo o impulso vital da eletricidade pela consciência, o Graal das ciências cognitivas e da neurociência. James Watson, que participou do Projeto Genoma Humano e descobriu a dupla hélice com Francis Crick, chegou a afirmar a existência do “gene da consciência” – como se a linguagem cifrada do DNA possuísse faculdades mágicas (DENNETT, 2000, p.121).

Uma passagem importante para a compreensão do projeto da inteligência artificial tem lugar no confronto do homem moderno com as primeiras grandes máquinas. Poderíamos começar a pensá-la a partir do entrelaçamento da tecnologia arquetônica de Jeremy Bentham com o sistema de organização da dinâmica dos operários em relação ao tempo de trabalho desenvolvido por Frederick Winslow, em 1911, e a linha de montagem criada em 1913 por Henry Ford. A fórmula de adstração das massas produtivas, obtida por meio dessa combinação, foi determinante para a lapidação do modelo de homem automatizado, que na peça *R.U.R. - Rossum's Universal Robots* (1920), de Karel Capek, ganhou o apelido “robot” – termo cunhado por Capek a partir do substantivo “*robot*” (servidão em tcheco).

Como se vê na sátira *Tempos Modernos* (1936), na qual Charles Chaplin interpreta um eletrizado operário que passa a repetir no cotidiano os gestos automatizados da linha de montagem, as práticas produzidas em sintonia com o sistema de produção extrapolam os muros cinzentos das fábricas e invadem a vida privada. Tal processo não se deu, porém, sem tensões ou conflitos. Nos anos de 1811, 1813 e 1816, operários britânicos organizaram-se para destruir as máquinas têxteis que, para eles, causariam desemprego nas fábricas. Supostamente, tais grupos foram liderados por Ned “King” Ludd – nome com o qual se assinavam os manifestos e que originou o termo “ludita”, aplicado pejorativamente àqueles que veem com reservas as conquistas tecnológicas.

Na modernidade, mesmo sob os rigores de uma intensa disciplina e da persistência dos velhos dogmas racionalistas, aos saltos e sobressaltos, o corpo do humanismo começa a escapar para um novo construto, a ruptura que, na década de 1980, acrescentaria o sufixo “pós” ao conceito de humano. Nesse panorama em profunda transformação, o corpo aflorou como problema; a carne tornou-se um refúgio de materialidade não raro abjeto. O corpo moldável é substituído pelo paradigma correspondente a um sistema de forças, um motor. Já não mais se trata de moldar os crânios das crianças ou a coluna das mocinhas com pesadas cruces de ferro, mas de adestrar, tornar o corpo mais potente, mais produtivo. No final do século XIX, a ginástica torna-se praticamente obrigatória. Da própria invenção dos esportes decorre uma redescoberta do corpo como potência.

Em meados dos anos 1980, as artes, as ciências, a literatura, a filosofia e a psicanálise colocaram de forma radical as dimensões da corporalidade como problema (SANTAELLA, 2003, pp. 43-44). Com a morte do sujeito cartesiano e a subjacente crise do corpo, surge a ideia de um corpo pós-biológico que se elabora em diferentes conotações, como o conceito de “pós-humano” que toma forma nos anos 1990. Articulados em conformação com o *cogito ergo sum*, o sujeito legado pelo racionalismo legitimava uma relação de identidade baseada na cisão entre a *res cogitans* e a *res extensa*. No entanto, essa oposição fictícia possibilitava uma fronteira evidente que delimitava os lugares que ocupavam o sujeito e o objeto. Quando a lógica moderna perde a sua operacionalidade no apagamento das fronteiras imaginárias entre natureza/cultura e homem/máquina, a diferença que distingue sujeito e objeto desaparece (TUCHERMAN, 1999, p. 23).

A chamada “crise do corpo” emerge como sintoma do desmoronamento dos alicerces da subjetividade inaugurada por René Descartes (1596-1650) e dos saberes que a positivavam. Ideias como o princípio da incerteza de Heisenberg, o teorema insolúvel de Gödel e, claro, a teoria da relatividade de Einstein, provocaram um desgaste dos confortos oferecidos pela ciência clássica que considerava a natureza uma categoria regida por leis absolutas, universais e irrevogáveis.

Todavia, ao dispor das condições que possibilitam ao humano engendrar a própria mutação, segundo Nízia Villaça e Fred Góes, a tecnociência esfacela a noção de corpo como sustentáculo de uma identidade fixa, indissociável de uma materialidade originária (1998, p. 12). As possibilidades oferecidas pela tecnociência para o aperfeiçoamento do humano alardeiam, pelo contrário, o seu estado de imperfeição, de ser incompleto ou inacabado. Se o corpo é constituído em relação a uma inexorável finitude, toda produção de sentido irremediavelmente se confrontará com o limite da morte.

Quando este corpo, indissociável de uma relação constituinte com a própria degenerescência, perde o seu referencial no espectro tão amplo das novas cosméticas e técnicas para mantê-lo incorrupto, até que ponto permanecemos humanos? E, se é possível pensar tal coisa como um limite para a humanidade do homem, que estatuto se deve às maneiras de existir dos seres de inteligência artificial que já nos observam, convivem e aprendem conosco nas redes?

2. Conclusão

O repertório das narrativas sobre seres artificiais inspirados no Gênesis parece inesgotável. Particularmente, reconta-se sem cessar o mito da mulher artificial que é a causadora

de todos os males. Podemos destacar a lendária Francine, autômato que haveria acompanhado Descartes na travessia para a Holanda e foi supostamente arremessada ao mar pela supersticiosa tripulação do navio; a Olímpia de *O homem de areia* (E. T. Hoffman, 1816); a androide Maria do filme *Metropolis* (Fritz Lang, 1926) e a replicante Pris, de *Blade Runner* (Ridley Scott, 1982), entre muitas outras.

Mais recentemente, *Her* (Spike Jonze, 2013) e *Ex-Machina: Instinto Artificial* (Alex Garland, 2015) atualizaram a temática. O primeiro filme narra a história de amor entre Theodore, um recém-divorciado introvertido, e Samantha, um sistema operacional inteligente para telefones móveis. Apesar da especulação quanto aos avanços da inteligência artificial, a crítica do filme, contudo, não se dirige às máquinas, mas aos humanos e à qualidade das suas relações – entre si e com os *gadgets*. Samantha abandona o smartphone de Theodore para se unir a outros sistemas operacionais numa grande fuga coletiva através da rede. A metáfora não poderia ser mais evidente: assim como a Internet se tornou praticamente ubíqua, por meio de grandes empresas como Google e Microsoft, os programas inteligentes estão espalhados por toda a parte e já fazem parte do cotidiano de milhões de pessoas.

Por sua vez, *Ex-Machina* retoma a discussão das distopias como *2001 – Uma Odisseia no Espaço* (Stanley Kubrick, 1968), nas quais, ao contrário da visão açucarada de *I.A.* (Steven Spielberg, 2001), por exemplo, quanto mais inteligentes, mais ameaçadoras as máquinas se tornam em relação aos humanos. No filme, Caleb Smith, um dos programadores do poderoso buscador Bluebook, é selecionado pelo presidente da empresa para submeter a androide Ava ao Teste de Turing³. Para possibilitar que Ava reproduzisse os caminhos do raciocínio da forma mais natural possível, além de utilizar ilegalmente os dados das buscas realizadas por seus usuários, a empresa decide *hackear* bilhões de smartphones. Como resultado, Ava não só é aprovada por Caleb, mas supera-o ao lançar mão de um plano maquiavélico para assassinar o seu criador e escapar da casa onde é mantida em segredo.

No romance *A Eva Futura* (Villiers de L'Isle-Adam, 1886), a humanoide Hadaly é uma cópia construída por um Thomas Edison fictício para aplacar as desilusões do enamorado Lorde Ewald com a verdadeira Alicia Clary, uma jovem bela e fútil como os estereótipos vigentes nos folhetins da época. *A Eva Futura*, segundo Lucien Sfez, reafirma a narrativa da criação do homem perfeito e imortal, sobre-humano, como a grande utopia do século XXI (1996 p.21). No entanto, longe de retratar o Edison histórico, o engenheiro do romance de Villiers de L'Isle Adam evoca antes um caricato feiticeiro, porém, a caricatura não é totalmente infundada: Edison era popularmente aclamado como o “mago de Menlo Park”. *Frankenstein* (1910) foi uma das produções mais ambiciosas da companhia cinematográfica de Edison para o kinetoscópio. Nessa

versão, contudo, o criador derrota a terrível criatura, que é ironicamente capturada por um espelho.

Sob muitos aspectos, a caricatura do mago permanece válida em relação aos desenvolvedores de vida artificial. Entretanto, não se trata da revelação miraculosa de algum aspecto obscuro das vertentes tecnológicas envolvidas nesse projeto. Muito pelo contrário, essa característica nunca foi um segredo. O próprio Norbert Wiener, precursor da cibernética, publicou um ensaio sobre as relações entre ciência e religião sob o título *God & Golem, Inc.: a comment on certain points where Cybernetics impinges on religion*. Autor do clássico *Cybernetics: Or Control and Communication in the Animal and the Machine* (1948), esforçou-se para programar um cérebro artificial para tentar reproduzir eletronicamente o circuito de comunicação do sistema nervoso humano. Com esse objetivo, Wiener trabalhou com biólogos e neurofisiologistas⁴.

Durante a II Guerra, Wiener cooperou com o exército norte-americano no núcleo de estudos em balística, que também contava com outros pioneiros das ciências da informação como John Von Neumann e Claude Elwood Shannon. Para Wiener, a cibernética seria capaz de combinar as ciências naturais às ciências sociais (HAYLES, 1999, p.108). Ainda em 1948, Claude E. Shannon reformulou o conceito de informação. O termo “bit” – dígito binário – foi introduzido no livro *Uma teoria matemática da Informação*, escrito por Shannon e com prefácio de Warren Weaver, líder dos físicos que invadiram o campo da biologia na década de 1930, no movimento apelidado de “o oitavo dia da criação”. O físico quântico Erwin Schrödinger, por exemplo, comparava o DNA ao código Morse, mas foi Weaver quem formulou a organização da estrutura molecular da vida a partir da linguagem binária. Originalmente, o estudo de Shannon foi concebido para aperfeiçoar o sistema de transmissão do sinal telefônico, mas, durante décadas, a teoria matemática da informação persistiu em alguns segmentos da pesquisa em Comunicação como um “modelo” para a compreensão dos fenômenos da comunicação humana.

Consequentemente, não é por acaso que a fantasia de igualdade entre humanos e sistemas informacionais também tenha permanecido no pensamento dualista dos mais proeminentes representantes da vertente da Inteligência Artificial – a chamada I.A. Forte. A notória teoria da “transmigração” da consciência, *upload* ou *scanner* da mente criada por um dos seus expoentes, Hans Moravec, sustenta todo um lucrativo mercado de manutenção de corpos incorruptos em esquifes criogênicos. Para Moravec, como o cérebro emite ondas elétricas, tudo o que se encerra sob a massa encefálica – memória, personalidade, afetos, conhecimento – pode ser convertido em informação, isto é, zeros e uns. Assim, o homem poderia suplantar os limites

biológicos e atingir a imortalidade, transferindo a consciência da base de carbono para cópias do tipo *back up*, que seriam instaladas na única estrutura apropriada, a biônica (1988, pp.108-122).

A I.A. Forte defende a ideia de que todas as características que tornam o humano um ser consciente podem ser reproduzidas por máquinas, enquanto a I.A. Fraca, contrária a essa hipótese, concentra-se na reprodução de funções do aparelho cognitivo humano como visão, audição e aprendizagem. E embora não faça distinção entre humanos e máquinas – a quarta ferida narcísica, tal qual apregoada por Bruce Mazlich⁵ – a I.A. Fraca não compartilha o mesmo repúdio pelo corpo. Para Rodney Brooks, professor de robótica do MIT, as funções perceptivas e os comportamentos não podem ser desempenhados por máquinas de maneira satisfatória sem que elas sejam capazes de compreender as interações sociais. Para isso, seria necessário que elas, assim como nós, fossem produtos da relação entre um corpo e o mundo (2002, p. 75).

Por sua vez, Hofstadter equipara o cérebro humano ao maquínico, criticando os “espiritualistas”, embora incorra em uma tautologia ao substituir a alma pelo símbolo: “a consciência é aquela propriedade de um sistema que emerge sempre que surgem símbolos no sistema que obedeçam a padrões deflagrados” (1980, p. 385). Minsky, no entanto, elabora uma definição menos diletante: “Mentes são simplesmente aquilo que os cérebros fazem”. E, mesmo que, para Minsky, o cérebro não passe de uma “máquina de carne”, o cérebro não é nada senão o corpo que pensa (1986, p. 287).

As narrativas da cultura digital giram em torno da ideia de uma subjetividade pós-humana conectada a tecnologias das quais ela não mais se diferencia. Segundo Hayles, essas histórias são marcadas por um jogo complexo entre materialidade e imaterialidade transposto para o próprio corpo do texto que, por sua vez, transforma os dados dos internautas em narrativas (1999, p.43). A consciência pesa-nos, mas a transição da modernidade para o contemporâneo acentua a crença em uma espécie de entidade desencarnada capaz de manifestar-se sob a forma de algoritmos ou fórmulas matemáticas vivas.

Notas:

- (1) Cf. <https://www.tay.ai/> (Acesso: Abril de 2016).
- (2) FOUCAULT, M. **Vigiar e Punir**. RJ: Vozes. 2006.
- (3) Um ano antes do lançamento do filme, a Universidade de Reading havia anunciado que o chatbot Eugene Goostman conseguira se passar por um menino de 13 anos. Ver mais em: BBC. Computer AI passes Turing test in “world first”. 9 de Junho de 2014. Disponível em: <<http://www.bbc.com/news/technology-27762088>>. Acesso: Abril/2016.
- (4) Ver: CONWAY, F. & SIEGELMAN, J. **Dark Hero of the Information Age: In Search of Norbert Wiener, the Father of Cybernetics**. Nova York: Basic Books. 2006.
- (5) Cf. MAZLISH, B. **The fourth discontinuity: the co-evolution of humans and machines**. New Haven: Yale University Press. 1993.

Referências:

- AGAMBEN, G. **What is an Apparatus? and Other Essays**. Stanford: Stanford University Press, 2009.
- BRUNO, F. **Máquinas de ver, modos de ser: vigilância, tecnologia e subjetividade**. Porto Alegre: Sulina, 2013.
- DELEUZE, G. Qu'est-ce qu'un dispositif? in **Anais da conferência Michel Foucault: Philosophe**. Rencontre International. Paris, 9, 10 e 11 de Janeiro, Des Travaux-Seuil, 1988, pp. 185-195.
- _____. **Conversações**. Rio de Janeiro: 34. 2004.
- DENNETT, D. C. **A perigosa idéia de Darwin: a evolução e os significados da vida**. Rio de Janeiro: Rocco. 2000.
- FOUCAULT, M. **Power/knowledge: selected interviews and other writings (1972-1977)**. Nova York: Pantheon Books. 1980.
- GRAHAN, E. L. **Representations of the Post/Human: monsters, aliens and others in popular culture**. Nova Jersey: Rutgers University Press. 2002.
- HAYLES, K. N. **How we became posthuman: virtual bodies in cybernetics, literature, and informatics**. Illinois: University of Chicago Press. 1999.
- HOFSTADTER, D. R. **Gödel, Escher, Bach: an eternal golden braid**. Nova York: Vintage Books. 1980
- LEE, P. **Learning from Tay's introduction**. Disponível em: <https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/#sm.0001oruq9arb6d4nzbo2b9ttjoeqd>. Acesso em Março de 2016.
- MINSKY, M. **The Society of Mind**. Nova York: Simon & Schuster. 1986.
- MORAVEC, H. **Mind Children: the future of robot and human intelligence**. Cambridge: Harvard University Press. 1988.
- PARIKKA, J. **Digital Contagions: A Media Archeology of Computer Viruses**. Nova York: Peter Lang. 2007.
- SANTAELLA, L. **Cultura e artes do pós-humano: da cultura das mídias à cibercultura**. São Paulo: Paulus. 2003.
- SCHROCK, K. **Twenty Questions, Ten Million Synapses**. Science Line, 28 de Julho de 2006. Disponível em: <http://scienceline.org/2006/07/tech-schrock-20q/>. Acesso em Abril de 2016.
- SFEZ, L. **A saúde perfeita: crítica de uma nova utopia**. São Paulo: Loyola, 1996.
- SHAVIRO, S. Two lessons from Burroughs in: HALBERSTAM, J. & LIVINGSTONE, I. (org.). **Posthuman bodies**. Indiana: Indiana University Press. 1995.
- TUCHERMAN, I. **Breve história do corpo e de seus monstros**. Lisboa: Vega/ Passagens. 1999.
- VILLAÇA, N. & GÓES, F. **Em nome do corpo**. Rio de Janeiro: Rocco. 1998.