

Recebido em: 05 set. 2018

Aprovado em: 15 out. 2018

BOSTROM, Nick. *Superinteligência: caminhos, perigos e estratégias para um novo mundo*. Rio de Janeiro: DarkSide Books, (2014) 2018.

dx.doi.org/
10.23925/1984-3585.2018i18p131-151

A superinteligência de Bostrom

Depois de 4 bilhões de anos perambulando no reino dos compostos orgânicos, a vida eclodirá na vastidão do reino inorgânico e assumirá formas que não podemos vislumbrar mesmo em nossos sonhos mais loucos. Afinal, esses sonhos ainda são produto da química orgânica.

Noah Harari (2016b, p. 53)

Alexandre Quaresma¹

Resumo: O artigo é uma resenha crítica do livro *Superinteligência* de Nick Bostrom. O seu tema são as inteligências artificiais (IA). O autor trata não só de máquinas, andróides e robôs que fazem coisas como nós, mas também de uma época futura não muito longínqua em que elas nos superarão em praticamente tudo, deixando-nos para trás na corrida evolutiva. O assunto é, sem dúvida, controverso. Pode-se afirmar que tal cenário é impossível. Todavia – ao contrário do que possa parecer – a mera hipótese de haver uma inteligência extraordinária e superior leva não só a potencialidades benéficas para as sociedades, mas também a preocupações significativas em relação a essas mesmas sociedades. É significativo que falamos aqui de uma “explosão” de inteligência.

Palavras-chave: Inteligências Artificiais. Superinteligência. Nick Bostrom. Crítica da tecnologia.

¹ Alexandre Quaresma é escritor ensaísta e filósofo brasileiro, pesquisador de tecnologias e consequências socioambientais, com especial interesse na crítica da tecnologia. Autor dos livros *Humano-Pós-Humano – Bioética, conflitos e dilemas da Pós-modernidade* (2014); *Engenharia genética e suas implicações* (org.), (2014); *Nanotecnologias: Zênite ou Nadir?* (2011); e *Artificial Intelligences – Essays on Inorganic and Non-biological Systems* (org.), (2018). E-mail: a-quaresma@hotmail.com

Bostrom's superintelligence

Abstract: The article is a critical review of Nick Bostrom's book *Superintelligence*. Its topic is Artificial Intelligence. The author does not only deal with machines, androids and robots that do things like us, but also with a supposedly not too distant future in which they will overcome humans in practically everything, leaving them behind in the evolutionary race. The subject is undoubtedly controversial. Many claim that the scenario is impossible. However, in contrast to what may appear, the mere hypothesis of extraordinary and superior intelligence involves not only benefits and social potentialities but also serious social concerns. The expression intelligence "explosion" is indicative in this context.

Keywords: Artificial Intelligence. Superintelligence. Nick Brostrom. Critique of technology.

Introduzindo muito objetivamente a nossa crítica, e como um exemplo emblemático das preocupações que exploraremos a seguir, vejamos como Nick Bostrom (2018, p. 468) – autor que resenharemos neste artigo – define metaforicamente a nossa relação enquanto sociedade com referência às inteligências artificiais (IA), pois daí poderemos seguir:

Diante do prospecto de uma explosão de inteligência, nós, humanos, somos como crianças pequenas brincando com uma bomba, Tamanho é o descompasso entre o poder de nosso brinquedo e a imaturidade da nossa conduta. A superinteligência é um desafio para o qual não estamos preparados atualmente e assim continuaremos por um longo tempo. Sabemos pouco a respeito do momento em que a detonação ocorrerá, embora seja possível ouvir um fraco tique-taque quando aproximamos o dispositivo dos nossos ouvidos.

Falar dessa hipotética “bomba” requer um conhecimento crítico do passado histórico, e também um olhar sensivelmente atento para o que se passa no momento presente, seguindo a partir deles (passado e presente) em busca de pistas que possam nos permitir uma intuição do que vem pela frente, ou seja, o futuro. Ainda assim, mesmo se observamos essa regra, especular acerca do futuro ainda continua a ser uma atividade arriscada, incerta, temerária, ainda mais em se tratando de tecnologias de fronteira – como é o caso das IA –, em que campos de pesquisa alimentam-se mutuamente uns dos outros, gerando uma gama praticamente interminável de oportunidades e possibilidades, de possíveis variáveis e potencializações, de maneira que a probabilidade de errar ou ser impreciso é muitíssimo maior do que a probabilidade de acertar “em cheio”, e ainda conseguir lograr êxito em conceber previsões precisas e corretas.

Historicamente, os pesquisadores de IA não têm se mostrado muito bons em prever o ritmo dos avanços em seu próprio campo ou a forma que tais avanços poderiam assumir. [...] Céticos que diziam que as máquinas “nunca” seriam capazes de fazer isso ou aquilo se mostraram equivocados. Por outro lado, os erros mais comuns entre os profissionais têm sido subestimar as dificuldades de fazer com que um sistema execute, de forma consistente, tarefas do mundo real, e o de superestimar as vantagens de seus próprios projetos ou técnicas preferidas. (BOSTROM, 2018, p. 51)

Mesmo porque o cientista na ponta da rede de pesquisa é também um ser humano que possui interesses, predileções, inclinações teóricas particulares, linhas de pesquisa e assim por diante. Os cientistas que desenvolvem projetos de IA também sofrem pressões de seus financiadores, e – como não poderia deixar de ser – também são compelidos a apresentar resultados concretos e comercialmente úteis.

Assim sendo – segundo Bostrom –, um primeiro ponto que – segundo o nosso entendimento – merece destaque é o fato de que as tecnologias também podem trazer problemas, e não apenas soluções. Uma tecnologia que é útil para as telecomunicações, por exemplo, pode ser extremamente danosa para o meio ambiente, devido às matérias que utiliza. Um novo processo de produção industrial pode automatizar a linha de produção, mas, por outro lado, ser danosa socialmente, pois exclui trabalhadores, de modo que exemplo desse tipo não faltam. Um segundo ponto é que elas (tecnologias) de fato transformam a realidade e o mundo – e, no extremo, até a nós mesmos –, mas o fazem – sabemos – não necessariamente para melhor, e as IA, nesse sentido, também não fogem à essa regra. Um terceiro ponto que merece a nossa atenção é o fato de a ciência e a tecnologia serem atividades absolutamente humanas, e de serem também humanos os próprios cientistas que realizam as pesquisas e os projetos, com tudo de bom ou ruim que isso possa acarretar, como dito anteriormente. De maneira que a tecnociência possui as mesmas características que nós que a produzimos. Elas são – nesse sentido – completamente ambivalentes, mas – sem dúvida – refletem aquilo que somos enquanto sociedade. Com a agravante, de que elas servem a qualquer senhor. O que vale dizer que as finalidades e usos finais podem ser amplos e variáveis, assim como são os humores, anseios e cobiças desse animal sedento por poder autointitulado *sapiens sapiens*. Uma última questão que vale a pena ser mencionada antes de avançarmos é que não é possível controlar totalmente uma tecnologia, uma vez que ela tenha ingressado na cadeia causal comum dos acontecimentos do mundo, pois assim como as ideias, elas também se deformam, se transformam, copulam com outras ideias e tecnologias aleatoriamente, e passam a agir em grande medida descontroladamente.

Hans Jonas (2006, p. 33) escreve que “na condição de artefato vulnerável, a construção cultural pode esgotar-se ou desencaminhar-se”. O que Jonas quer nos dizer é que, uma vez imbricado no mundo factual, inserido na cadeia dos acontecimentos ordinários do dia-a-dia, o objeto técnico já não pode mais ser plenamente controlado de fora por seus cria-

dores; ele simplesmente passa a fazer parte de uma dinâmica complexa, alheia a seu controle, e que muitas vezes foge do que foi concebido e previsto para ele (objeto técnico) enquanto projeto, configurando *hipertelia*. Edgar Morin (2007, p. 41) escreve que

a ecologia da ação indica-nos que toda ação escapa, cada vez mais, à vontade do seu autor na medida em que entra no jogo das inter-retro-ações do meio onde intervém. Assim a ação corre o risco não somente de fracassar, mas também de sofrer desvio ou distorção de sentido.

Bruno Latour (2000, p. 221), também acerca disso, expõe-nos sua perplexidade ao afirmar que os objetos técnicos e tecnologias “parecem mover-se sem ajuda das pessoas. E o mais fantástico é que parecem poder até mesmo existir sem as pessoas”. Sem maiores embargos, o futuro das sociedades com as IA não é necessariamente brilhante, benéfico às coletividades, nem muito menos desprovido de possíveis crises, dilemas e armadilhas. Se e quando criarmos sistemas de IA complexos o suficiente para podermos chamá-los de “vivos”, genuinamente inteligentes e conscientes, teremos cruzado um umbral, depois do qual absolutamente tudo poderá acontecer.

Um sistema de inteligência artificial superior poderia, como nos informa Bostrom (2018, p. 142), “aumentar consideravelmente sua capacidade intelectual efetiva por meio da absorção de conteúdos pré-produzidos acumulados durante séculos de ciência e civilização humana: por exemplo, através da internet”, e a partir dessa condição – sustentamos nós –, em um cenário favorável, tornarem-se aliadas significativas no enfrentamento dos problemas sociais mais graves de nossas sociedades. Não é preciso dizer que esse mundo seria “o melhor dos mundos”, quase que um mundo de sonhos, e que as IA poderiam nos ajudar a sanar diversos – senão todos – problemas que enfrentamos na atualidade. Mas, mesmo nesse mundo hipotético ideal, podem ocorrer problemas com as IA, como pane, mal funcionamento, colapso e, talvez o mais preocupante para nós, descontrole operacional de nossa parte. No caso hipotético de um funcionamento insatisfatório, muitas são as possíveis consequências desastrosas.

Quando um sistema fraco funciona mal, a consequência é limitada. Todavia, se um sistema que apresenta uma vantagem estratégica decisiva se comportar mal ou se um sistema que não se comporte bem possuir força suficiente para obter essa vantagem, o prejuízo poderia facilmente equivaler a uma catástrofe existencial – uma destruição global e definitiva do potencial axiológico da humanidade; ou seja, um futuro pra-

ticamente vazio de qualquer coisa que pudéssemos valorizar.
(BOSTROM, 2018, p. 227)

Notemos em que termos o autor coloca o problema: “uma destruição global e definitiva do potencial axiológico da humanidade”. Conforme Bostrom (2018, p. 136-137), “É totalmente possível que a busca pela inteligência artificial [superior] se mantenha aparentemente perdida em uma densa floresta até que uma descoberta inesperada revele a linha final em uma clareira a apenas alguns passos de distância”, e que ela surja (superinteligência) e evolua rapidamente, em pouquíssimo tempo.

Se a IA de nível humano ainda não existir em virtude da falta de um determinado *insight* por parte dos programadores, então, quando a descoberta final ocorrer, a IA poderá saltar para um nível radicalmente superior ao humano sem ao menos passar pelos degraus intermediários. (BOSTROM, 2018, p. 139)

Ou seja, mesmo em um cenário relativamente favorável, podemos ver que existem possibilidades de problemas, e, em um caso como esse que Nick Bostrom evoca nessa passagem supracitada, seria certamente “o fim da linha para nós”. Além disso, mesmo nesse “melhor dos mundos”, é bom frisar que se tivermos máquinas projetando máquinas, ou seja, sistemas de IA especializados em criar e desenvolver novos e mais potentes sistemas de IA, progressivamente mais e mais poderosos cognitivamente, “o sistema poderá ultrapassar um ponto de referência que podemos chamar de “cruzamento”, um ponto no qual o aprimoramento do sistema se dá principalmente pelas ações do próprio sistema, e não em resposta a ações externas” (BOSTROM, 2018, p. 129). Ainda conforme o autor (*ibid.*, p. 284),

Uma máquina provavelmente precisaria ter a capacidade de representar o mundo de uma maneira que fosse ao menos tão rica e realista quanto a representação de mundo que um humano adulto normal possui. [...] Isso está muito além do alcance da IA contemporânea. [...] Uma vez que [...] os processos de planejamento se tornem suficientemente poderosos, também se tornarão potencialmente perigosos.

Nesse tipo de cenário hipotético, há – por um lado – uma chance também bastante razoável de obtermos vantagens e benefícios com a utilização desses sistemas de IA, em utilidades das mais diversas em nossas vidas cotidianas, como também – por outro – de nos expormos a riscos significativamente importantes, no sentido de sermos vítimas de nossas próprias criações tecnológicas. É por isso que Bostrom (2018, p. 130) escreve que “em uma partida moderada [...] é possível que os acontecimen-

tos fossem mantidos em sigilo enquanto se desenrolassem. O conhecimento poderia estar restrito a um pequeno grupo como em um programa secreto de pesquisa militar patrocinada pelo Estado”. E, assim, o sistema de IA poderia aumentar sua própria capacidade autonomamente e, como nos informa Bostrom (2018, p. 146),

o aumento poderia ser especialmente dramático se o desenvolvimento de uma inteligência de máquina de nível humano tomasse o mundo de surpresa. Nesse caso o que era previamente um pequeno projeto de pesquisa poderia então, subitamente, se tornar o foco de intensa pesquisa e desenvolvimento ao redor do mundo.

Sim, e assim teremos – provavelmente – um mundo regido por máquinas, dominado por máquinas, controlado por máquinas, de maneira que o próprio futuro seria – em um caso assim – um futuro de máquinas, e não exatamente de pessoas. Lembrando que,

em um cenário de partida moderada onde emulações baratas e eficientes ou outras mentes digitais fossem gradualmente inundando o mercado de trabalho por um período de anos, poderíamos imaginar protestos em massa de trabalhadores demitidos que pressionariam os governos a aumentar o seguro-desemprego, implantar uma renda mínima para todos os cidadãos, arrecadar impostos especiais ou, ainda, impor aos empregadores que utilizassem trabalhadores emulados a exigência do pagamento de um salário-mínimo. (BOSTROM, 2018, p. 131)

Dessa forma, é possível perceber que as consequências e desdobramentos sociais seriam enormes, e as perspectivas não tão favoráveis. Especialmente se “o sistema tiver adquirido a capacidade necessária para que a maior parte do poder de otimização exercido surja do próprio sistema”, conta-nos Bostrom (2018, p. 146). Em um outro cenário futuro descrito por Bostrom, um tanto mais pessimista, sombrio e negativo para os seres humanos, mas também plausível de se concretizar, poderíamos assistir à humanidade criando intencionalmente ou não o seu próprio predador, gerando a sua própria derrocada filogenética, consumando uma nova forma de inteligência maquínica que poderia superar e ultrapassar a sua própria, e as consequências dessa possibilidade hipotética podem ser para nossa espécie simplesmente catastróficas, já que estaríamos sujeitos a uma inteligência de ordem superior que, caso fosse hostil ou se tornasse hostil, poderia nos levar a uma condição de total ou quase que total subjugação, e, no extremo, extinção.

Um dos possíveis problemas relativamente graves que podem surgir em um contexto como esse é que a própria inteligência artificial por si

mesma – depois de alcançar níveis muitíssimo elevados de poder cognitivo e autonomia – poderia trabalhar obstinada e inteligentemente para otimizar de forma ininterrupta toda a infraestrutura técnica existente para poder satisfazer o seu próprio fim último, seja lá qual ele for. Ou seja, em um ambiente social hipotético em que máquinas constroem a si mesmas, cada vez mais potentes e poderosas, e possuem tal meta como finalidade última, ou seja, *construir máquinas mais e mais desenvolvidas e potentes*, tais sistemas podem entrar em uma espiral perigosa de absorver e consumir todos os recursos existentes no ambiente para suprir essa sua diretriz original, o que vale dizer, gerar uma enorme e incontrolável *profusão de infraestrutura*. Profusão de infraestrutura, nesse contexto, seria “um fenômeno no qual um agente [artificial] transforma grande parte do universo alcançável em uma infraestrutura a serviço de algum objetivo [intrínseco pré-programado], cujo efeito colateral seria o impedimento da realização do potencial axiológico da humanidade” (BOSTROM, 2018, p. 231).

Lembrando que,

em geral, enquanto um animal ou um humano podem ser motivados a desempenhar várias ações externas para que alcancem algum estado mental interno desejado [ou seja, físico], uma mente digital que tenha completo controle do seu estado interior pode provocar um curto-circuito nesse regime motivacional, alterando diretamente seu estado interno para configuração desejada: as ações externas e condições que eram previamente necessárias para atingir o objetivo final se tornam supérfluas quando a IA se torna suficientemente inteligente e capaz para alcançar o objetivo final mais diretamente. (BOSTROM, 2018, p. 230)

Apenas a título de exemplo, seria como constituir e instanciar uma inteligência artificial complexa em um sistema de IA – ou superinteligência –, tendo como finalidade última combater o caos ambiental e ecológico que enfrentamos na atualidade, e esse sistema superinteligente resolvesse “por conta própria” que exterminar todos os seres humanos, paralisando as degradações que as sociedades produzem, seria a forma mais inteligente, eficiente e prática para o enfrentamento do referido problema, cuja finalidade estaria intrinsecamente subordinada. Sem dúvida, pelo menos, seria – ponderemos – uma desconcertante ironia.

Ainda nesse tipo de cenário mais desfavorável para nós e nossa espécie, torna-se importante ter em mente que os seres humanos são muito dispendiosos em termos de dependência de energia, recursos naturais, víveres, de uma maneira geral, exigindo por isso – continuamente – água, ar, temperatura e pressão atmosférica específicas, solicitando demais do meio ambiente, em termos de recursos naturais, pois não só drenamos

esses recursos como também os contaminamos e deterioramos. Com o aumento populacional descontrolado, essa imensa massa de seres humanos insustentável, além de cobrarem um alto preço do ambiente, ainda o penaliza drasticamente com a emissão sistemática de poluentes, detritos, rejeitos químicos, de mineração, da indústria, de maneira que o nosso nível de impacto ecológico sistêmico é muito alto. Uma das consequências mais observáveis são as mudanças climáticas, o aquecimento global, e, especialmente, o antropoceno.

Em um só termo, a humanidade esgota o meio ambiente, destrói os ecossistemas, desestruturando as cadeias tróficas. Uma inteligência artificial de nível superior desse tipo que tentamos analisar aqui, se amplamente mais poderosa que nós, em termos cognitivos, poderia ser capaz de se acoplar estruturalmente ao meio ambiente com menos dano e mais sabedoria, e, no extremo, ao próprio planeta, já que para funcionarem elas poderiam muito bem explorar apenas a energia solar e uns poucos elementos químicos e físicos para a sua estrutura de hardware, de modo que sua perpetuação ao longo do tempo poderia se consumir sem que houvesse maiores danos à homeostase planetária. Nesse sentido, então, é possível afirmar que uma inteligência artificial desse tipo seria muito mais sustentável a longo prazo. É possível especular também que, para uma superinteligência artificial, seria absolutamente trivial mensurar as demandas – de energia e matéria, por exemplo – e atuar inteligentemente de forma a otimizar a utilização desses mesmos recursos, de modo a garantir o abastecimento, sem por isso destruir e desestruturar o meio em que ela mesma está inserida, do qual dependerá seu funcionamento e sua operacionalidade. Assim sendo, a economia de seres inteligentes artificiais seria – quem sabe – muitíssimo mais simples e menos dispendiosa e desgastante para o planeta, para a biosfera. Notemos que – em termos de subsistência – nem mesmo de oxigênio tais entes necessitariam, de maneira que até no espaço sideral e no vácuo cósmico eles poderiam se aventurar livremente, sem restrições constrangedoras que necessariamente se impõem aos seres vivos biologicamente estruturados. Mas – ponderemos com um mínimo de razoabilidade – caso fosse possível e viável uma inteligência artificial superior em meio cibernético-informacional – uma superinteligência –, não seria difícil imaginá-la se reproduzindo sistemática e sustentavelmente, colonizando os confins do cosmo com suas populações de máquinas superinteligentes, de alguma maneira desconcertante, herdeiras das inteligências biológicas, salvo nossas próprias idiosincrasias e limitações.

É importante observarmos também que, conforme o ritmo inicial dessa IA superpotente varie, variarão também as possibilidades de controle e descontrole externo dos referidos sistemas cibernético-informacionais, o que determinará – é claro – o nível das consequências e desdobramentos. Se o ritmo for rápido ou muito rápido, talvez pouco ou nada possamos de fato fazer para tentar manter a ordem e a cultura humanas. O tipo de “partida” – conforme nos alerta Bostrom – parece ser o fator determinante, em termos de perspectivas de gerenciamento administrativo dos próprios sistemas de IA. Notemos que, o termo “partida”, aqui, faz referência hipotética ao momento em que a inteligência de máquina nos alcançaria ou, pretensamente, ultrapassaria.

Uma partida rápida é aquela que ocorre em um intervalo de tempo curto, como minutos, horas ou dias. Cenários de partida rápida oferecem poucas oportunidades para deliberações por parte dos seres humanos. Ninguém sequer notaria algo de diferente antes que o jogo já estivesse perdido. Em um cenário de partida rápida, o destino da humanidade dependeria essencialmente de preparações prévias. (BOSTROM, 2013, p. 130)

“Partidas rápidas aconteceriam tão depressa que não haveria tempo para que [sequer] a notícia se espalhasse ou para que qualquer pessoa pudesse preparar uma reação significativa” (ibid., p. 131). Ainda assim, escreve Bostrom (2018, p. 285), “seria melhor desenvolver o sistema de maneira que ele fosse deliberadamente um agente, e assim os programadores poderiam visualizar mais facilmente quais critérios acabariam por determinar os resultados obtidos por tal sistema”. Ou seja, em nossas próprias palavras, é mais fácil criar mentes artificiais segundo as nossas próprias mentes, pois assim, minimamente, poderíamos ter uma ideia de como as mentes de fato funcionariam por dentro. Nick Bostrom (2018, p. 15) afirma em tom de alerta que,

se algum dia construirmos cérebros artificiais capazes de superar o cérebro humano em inteligência geral, então essa nova superinteligência poderia se tornar muito poderosa. E, assim como o destino dos gorilas depende mais dos humanos do que dos próprios gorilas, também o destino de nossa espécie dependeria das ações da superinteligência de máquina.

Notemos o grau de periculosidade e risco evolvido, e percebamos também o nível de nossa imprevidência diante desse mesmo risco. A pergunta é: será que o risco realmente vale à pena? O fato é que não há motivo razoável para acreditarmos que algo assim não possa de fato acontecer, mesmo porque, como já referenciamos anteriormente, há um *tour*

de force poderosíssimo mobilizado para que isso de fato aconteça o quanto antes. Diante disso, há que se atentar inelutavelmente para essa possibilidade – ainda que apenas hipotética – já que, uma vez consumada, dificilmente poderemos ou poderíamos retroceder com um mínimo de segurança e confiabilidade. Não podemos deixar ao acaso e à aleatoriedade – em se tratando do futuro da própria espécie humana – um processo tão importante que poderia implicar desdobramentos e consequências de profundo impacto para o que conhecemos como civilização humana, suas futuras gerações, e eventualmente a sua própria extinção. De modo que não é pouco o que está em jogo. Acreditamos ser importante ponderar com lucidez e sabedoria.

Mas temos uma vantagem: nós é que construímos a máquina. Em princípio poderíamos construir um tipo de superinteligência que protegesse os valores humanos. Teríamos, certamente, fortes razões para isso. Na prática, o problema do controle – a questão de como controlar o que a superinteligência faria – parece bastante difícil. Tudo indica também que teríamos apenas uma única chance para resolvê-lo. Uma vez que uma superinteligência hostil passasse a existir, ela nos impediria de substituí-la ou de mudar suas preferências. Nosso destino estaria selado. [...] Esse é, muito provavelmente, o desafio mais importante e mais assustador que a humanidade já encarou. E, independentemente de nosso sucesso ou fracasso, promete ser o último desafio que encararemos. (BOSTROM, 2018, p. 16)

É bom também manter em tela que essas hipóteses em grande medida fatalistas – ainda que bastante factíveis hoje em dia – foram enunciadas inicialmente pelos próprios pioneiros nesses campos. O matemático I. J. Good, que tinha sido chefe de estatística de Alan Turing no time responsável pela quebra de códigos durante a Segunda Guerra Mundial, era um desses pioneiros. Num artigo com o título “Especulações a respeito da primeira máquina ultrainteligente”, Good escreveu em 1965 o seguinte (apud Bostrom, 2018, p. 26):

Defina-se uma máquina ultrainteligente como uma máquina capaz de superar todas as atividades intelectuais de qualquer homem [ser humano], independentemente de quão genial ele seja. Já que o projeto de máquinas é uma dessas atividades intelectuais, uma máquina ultrainteligente poderia projetar máquinas ainda melhores; haveria então certamente uma “explosão de inteligência”, e a inteligência humana se tornaria desnecessária. Desse modo, a primeira máquina ultrainteligente é a última invenção que o homem [ser humano] precisará fazer, contanto que a máquina seja dócil o suficiente para nos dizer como mantê-la sobre controle.

Controle; eis a questão central. Segundo I. J. Good apud Nick Bostrom (2018, p. 55), uma ultrainteligência de máquina seria “uma máquina capaz de superar todas as atividades intelectuais de qualquer homem [ser humano], independentemente do quão genial ele seja”. Notemos – nesse ponto – que não é facilmente determinável o que seria uma “ultrainteligência” ou “superinteligência” ou, contrariamente, o que seria uma *subinteligência* ou uma *protointeligência*. Tudo se mostra sombrio e movediço nesses campos de fronteira das ciências, quando pretendemos determinar limites.

Mas, de uma maneira geral, continua Bostrom (2018, p. 27) – e nós aqui concordamos plena e absolutamente –, é de surpreender o fato de que

a maioria dos pioneiros de IA não consideraram a possibilidade de que sua empreitada poderia envolver alguns riscos. Eles não fizeram declarações retóricas – muito menos considerações sérias – sobre qualquer preocupação referente à segurança ou questões éticas relacionadas à criação de mentes artificiais e de potenciais computadores déspotas: uma lacuna que surpreende mesmo diante do padrão não-tão-impresionante das avaliações críticas de tecnologia da época. [...] Uma exceção é Norbert Wiener, que tem algumas dúvidas sobre as possíveis consequências. Ele escreveu, em 1960: “Se usarmos, para alcançar nossos objetivos, um ente mecânico sobre cuja operação não possamos interferir eficientemente após seu início, pois sua ação seria tão rápida e irreversível que não teríamos os dados necessários para intervir antes da operação estar completa; é bom, então, que tenhamos certeza que o propósito inserido na máquina seja o propósito que nós realmente desejamos, e não uma mera imitação embelezada”.

A ingenuidade geral acerca dos riscos inerentes parece cândida e até pueril, pois nem mesmo os papas da computação puderam antever os desafios que enfrentamos hoje. Seguindo nessa mesma linha de hipóteses, informa-nos Nick Bostrom (2018, p. 53),

atribuo uma probabilidade maior de que a superinteligência seja criada num ritmo relativamente rápido após o alcance da inteligência de máquina de nível humano. Também tenho uma visão mais polarizada sobre consequências, acreditando que um resultado extremamente bom ou extremamente ruim é mais provável que um resultado equilibrado. [...] Elas [pesquisas realizadas por Bostrom e sua equipe com especialistas] mostram que (ao menos na ausência de melhores dados ou análises) pode ser razoável acreditar que a inteligência de máquina de nível humano tem boas chances de ser desenvolvida até a metade do século e que há uma probabilidade considerável de que seja desenvolvida bem mais cedo ou muito mais tarde; que ela possa talvez, logo depois, resultar na superinteligência; e que uma ampla gama e consequências poderá ocorrer, incluindo consequências extremamente boas e consequências tão ruins quanto a extinção da espécie hu-

mana. No mínimo, elas sugerem que é válido analisar o tema atentamente.

Um dos possíveis caminhos para se alcançar essa inteligência superior da qual nos fala Nick Bostrom seria conseguir criar a própria capacidade de aprendizagem artificialmente, internamente ao próprio sistema, intrinsecamente a ela, para que a própria máquina pudesse aprender sobre o mundo que a circunda e pudesse igualmente agir de acordo com essa aprendizagem continuada. O próprio Alan Turing especulava sobre a possibilidade de se criar o que ele chamou de “máquina criança”, Nick Bostrom (2018, p. 57) é quem dá voz e atualidade às ideias desse pioneiro chamado Turing:

Em vez de tentar criar um programa capaz de simular a mente de um adulto, por que não tentar produzir um que simule a mente de uma criança? Se ele fosse, então, submetido a uma trajetória apropriada de aprendizado, seríamos capazes de obter o cérebro de um adulto.

Bostrom (2018, p. 66) também esclarece – nessa citação relativamente extensa, mas ainda assim extremamente esclarecedora para a nossa resenha – que

uma máquina criança, no sentido em que foi concebida por Turing, teria uma arquitetura relativamente fixa que desenvolveria suas potencialidades apenas através do acúmulo prévio de conteúdo, uma IA embrionária seria uma forma mais sofisticada de inteligência artificial, capaz de aprimorar sua própria *arquitetura*. Nos estágios iniciais de uma IA embrionária, tais aprimoramentos poderiam ocorrer principalmente por tentativa e erro, aquisição de informação ou com o auxílio de programadores. Já em estágios mais avançados, uma IA embrionária poderia ser capaz de *compreender* suficientemente bem seu próprio funcionamento a ponto de desenvolver, por conta própria, novos algoritmos e estruturas computacionais que aprimorassem seu desempenho cognitivo. Esse conhecimento poderia ser alcançado a partir do momento em que a IA embrionária alcançasse um nível suficiente de inteligência geral em diversos domínios do conhecimento ou quando ela fosse capaz de cruzar uma barreira de conhecimento em algum campo particularmente relevante, como ciência da computação ou matemática. Isso nos leva a outro conceito importante: o de “automelhoria recursiva”. Uma IA embrionária que viesse a ser bem-sucedida seria capaz de aprimorar a si mesma interativamente: uma primeira versão da IA poderia projetar uma versão melhor de si mesma, essa segunda versão seria, por sua vez, mais inteligente que a primeira e igualmente capaz de produzir uma versão melhorada de si mesma e assim por diante. Sob algumas condições, tal processo de automelhoria recursiva poderia continuar por tempo suficiente até resultar em uma explosão de inteligência – um evento no qual, em um curto espaço de tempo, o nível de inteligência de um sistema passaria de capacidades relativamente modestas de cognição (talvez sub-humanas em

muitos aspectos, mas com talento específico para programar e realizar pesquisas em IA) a uma superinteligência radical.

Nossa história pregressa mostra que, muitas vezes, o progresso tecnocientífico se potencializa de formas surpreendentes, seguindo por caminhos tortuosos e não-lineares, seja com grandes estruturas e financiamentos, seja em uma garagem ou fundo de quintal, como foi o caso da *Apple* e da *Microsoft*, por exemplo. São hipóteses de desenvolvimento que não devem ser desprezadas. Como escreve Nick Bostrom (2018, p. 161),

construir uma IA embrionária pode demandar descobertas e algoritmos desenvolvidos ao longo de muitas décadas pela comunidade científica ao redor do mundo. Mas é possível que a descoberta central venha de um único indivíduo ou de um pequeno grupo que consiga encaixar todas as peças. [...] Se a IA embrionária puder ser criada na forma de um sistema simples, cuja construção dependa apenas da aplicação de alguns princípios básicos corretos, esse feito talvez esteja ao alcance de uma equipe pequena ou de um único indivíduo.

Quanto a esse conceito de *IA embrionária*, vale a pena nos determos um pouco mais na observância de sua natureza não só conceitual, mas também funcional, operativa, e a dinâmica social e humano-maquínica que poderá se estabelecer a partir de então. Bostrom (2018, p. 148) oferece-nos a seguinte descrição:

Uma IA embrionária pode ser melhorada por meio de uma combinação de seus próprios esforços e dos esforços de uma equipe de programadores humanos, e talvez também dos esforços de uma comunidade mais ampla de pesquisadores que trabalhem gerando avanços contínuos na indústria de semicondutores, ciência da computação e campos afins.

Lembrando que as “IA poderiam ser – e provavelmente a maioria delas será – completamente diferentes da inteligência humana” (BOSTROM, 2018, p. 67). O que significa dizer que poderão possuir maneiras diferentes das nossas no que se refere à estruturação corporal, constituição física, meio interno, acoplamento com a realidade e assim por diante.

Não existe razão para esperarmos que uma IA genérica seja motivada por amor, ódio, orgulho ou qualquer outro sentimento comum aos seres humanos: essas adaptações complexas necessitariam de esforços custosos e deliberados para serem recriados em uma IA. (BOSTROM, 2018, p. 67)

Notemos que aqui a biologia parece recobrar sua importância e seu valor. Observando que, juntamente com os progressos das ciências computacionais, da otimização progressiva dos processos, podemos experimentar também um momento futuro de maior conhecimento tecnológico

do funcionamento do cérebro e da consciência – já que as neurociências não param – e, logo também, da inteligência em si mesma. Desse conhecimento, poderão surgir diversos outros importantes que podem – de algum modo – ajudar a pavimentar o caminho rumo a uma IA superior.

O exemplo mais simples da superinteligência rápida seria uma emulação completa do cérebro [?] executada em um hardware veloz. Uma emulação operando com uma velocidade 10 mil vezes maior que a de um cérebro biológico seria capaz de ler um livro em alguns segundos e escrever uma tese de doutorado em uma tarde. Se a velocidade fosse 1 milhão de vezes maior, uma emulação poderia realizar o trabalho intelectual de um milênio em apenas um dia de trabalho. Para uma mente rápida como essa, os eventos do mundo externo parecem acontecer em câmera lenta. (BOSTROM, 2018, p. 109)

Ainda que não tenhamos nada sequer parecido com uma “emulação completa do cérebro”, seria essencial observar que uma “superinteligência rápida”, assim como definida por esse mesmo autor (2018, p. 108), seria “um sistema que pode fazer tudo que um intelecto humano é capaz de fazer, porém muito mais rapidamente”.

O cérebro humano possui um pouco menos de 100 bilhões de neurônios e é aproximadamente três vezes e meia maior que o cérebro dos chimpanzés (embora tenha apenas um quinto do tamanho do cérebro de uma baleia cachalote). O número de neurônios em uma criatura biológica é claramente limitado pelo volume do crânio e pela capacidade metabólica, mas outros fatores também podem ser significativos para cérebros maiores (tais como controle de temperatura, tempo de desenvolvimento e atrasos de transmissão de sinal) [...]. Em contrapartida, o hardware de um computador pode ter seus limites físicos expandidos indefinidamente a níveis altíssimos. Supercomputadores podem ser do tamanho de um galpão industrial ou até maiores, com capacidade remota suplementar adicionada por meio de cabos de alta velocidade. (BOSTROM, 2018, p. 122)

Além disso, no que se refere a comparações de mentes biológicas e sistemas artificiais, continuamos com Bostrom (2018, p. 124-125), no sentido de que mentes cibernético-informacionais poderiam ter algumas vantagens importantes, como possuir mais capacidade objetiva de “editabilidade [...]”; duplicabilidade [...]”; coordenação de objetivos [...]”; compartilhamento de memória [...]”; novos módulos, modalidades e algoritmos”, enfim, possibilidades não faltam.

Mas, nos dias atuais, não é possível mensurar com clareza a maneira como essas técnicas se desenvolverão socialmente. Como escreve também Nick Bostrom (2018, p. 59), “é muito difícil prever a magnitude dos ganhos de eficiência de tal processo evolutivo artificial”. O que signifi-

ca dizer que será preciso estar atento e acompanhar os progressos desses campos com lucidez para que possamos em tempo mensurar suas possibilidades e limitações, bem como suas consequências e desdobramentos, sejam eles benéficos ou não às sociedades humanas. É possível que essas práticas dos algoritmos evolucionários se somem a outras também arrojadas e igualmente promissoras e que desenvolvimentos em campos específicos, correlatos e paralelos, alimentem e motivem potencializações recíprocas, gerando ganhos operacionais importantes para o campo das IA que, hoje, não podemos ainda prever em minúcias ou detalhes.

Nós já mencionamos que o cérebro desempenha um duplo e contraditório papel, mas vale repetir: se por um lado a ortodoxia cognitivista quer pretensamente desvalorizar os organismos biológicos e o próprio cérebro, como algo dispensável ou substituível – em relação à cognição –, ela também acaba de alguma maneira buscando simular os cérebros biológicos vivos, já que esses são os únicos modelos disponíveis, no que tange à criação de entes cibernético-informacionais inteligentes e artificiais. Além disso, é possível especular que, em algum momento futuro, poderemos compreender em sua totalidade o comportamento do cérebro humano, e, a partir dessa compreensão, replicá-lo em meio artificial em sistemas inorgânicos de IA. Inclusive, o próprio cérebro pode ser considerado como a evidência maior de que a inteligência pode ser instanciada em um suporte físico, já que o cérebro é um corpo físico.

A disponibilidade do cérebro como um modelo fornece um forte ponto de sustentação para o argumento que prega a viabilidade da inteligência de máquina. Isso, entretanto, não nos torna capazes de prever quando ela será alcançada, pois é difícil estimar a velocidade com que se darão as descobertas futuras da neurociência. O que podemos afirmar é que quanto mais olhamos para o futuro, maior é a possibilidade de que os segredos da funcionalidade do cérebro tenham sido suficientemente decodificados a ponto de tornar possível a criação de uma inteligência de máquina dessa forma. (BOSTROM, 2018, p. 65)

Algo – de novo – a se conferir. Mas a constituição de uma IA de nível superior definitivamente não se restringe a sistemas informacionais isoladamente, em um único e potente computador ou sistema. Há quem defenda a tese de que a própria complexificação dos sistemas cibernético-informacionais que possuímos hoje – e que não cessamos de incrementar e potencializar progressivamente, incluindo milhões de computadores – poderia levar a um tipo totalmente inusitado e novo de inteligência artificial coletiva que emergiria do próprio sistema ciberespacial. Nós, conscientes da improbabilidade dessa hipótese, temos as nossas dúvidas.

Nick Bostrom (2018, p. 101) critica tal ideia e a qualifica como “extravagante”, mas, ao mesmo tempo, cita textualmente a mesma ideia e também o autor dela, qualificando o trabalho teórico-especulativo deste último (Vernor Vinge) como influente:

[Será] que a internet poderia um dia “acordar”? Será que a internet poderia se tornar mais do que apenas a espinha dorsal de uma superinteligência coletiva fragilmente integrada, vindo a ser algo como um crânio virtual que abriga um superintelecto emergente unificado? (Essa é uma das formas pelas quais a superinteligência poderia surgir de acordo com o influente artigo que Vernor Vinge escreveu em 1993, no qual o autor cunhou o termo “singularidade tecnológica”).

E existem ainda outras formas de ruptura de paradigma como, por exemplo, as que emergem da seguinte indagação: se fossem conscientes (as IAs), sensíveis e inteligentes, será que seriam também dignas e passíveis de possuir direitos, de não serem expostas a condições de sofrimento, maus tratos, escravidão absoluta, extermínio e descarte indiscriminado? Vejamos que, se assim fosse – ou seja, se fossem concedidos direitos a esses seres artificiais, ampliar-se-ia sensivelmente os nossos horizontes em termos axiológicos, quem sabe evocando uma espécie nova de axiologia humano-maquínica, que fosse capaz de abarcar as relações recíprocas que se travarão entre nós e esses entes algorítmicos superinteligentes.

Ainda que um pouco extensa, essa passagem de Nick Bostrom (2018, p. 236) traz à concretude – ainda que apenas hipoteticamente – uma configuração social inusitada, além – é claro – de preocupações extras que beiram as raias da ficção, mas que podem certamente vir a acontecer em um cenário desse tipo que aqui transcrevemos:

Normalmente, não consideramos que o que acontece dentro de um computador possa ter qualquer significado moral exceto quando isso afeta o mundo exterior. Mas uma superinteligência de máquina poderia criar processos internos dotados de status moral. Por exemplo, uma simulação bem detalhada de uma mente humana, real ou hipotética, poderia ser consciente e de diversas maneiras comparável a uma emulação. Pode-se imaginar cenários nos quais uma IA cria trilhões dessas simulações conscientes, talvez para melhorar seu entendimento de psicologia e sociologia humana. Tais simulações podem ser colocadas em ambientes simulados e sujeitadas a vários estímulos para que suas reações sejam estudadas. Uma vez que sua utilidade informacional tenha sido esgotada, elas poderiam ser destruídas (da mesma forma que ratos de laboratório são rotineiramente sacrificados pelos cientistas ao final de um experimento). Se essas práticas fossem aplicadas a seres que possuam um status moral elevado – tais como simulações de humanos ou diversos outros tipos de mentes conscientes –, o resultado poderia ser equivalente a um genocídio e, portanto, moralmente problemático. Além disso, o

número de vítimas poderia ser algumas ordens de magnitude maior do que qualquer genocídio da história.

Assim sendo, como indica o mesmo autor (2018, p. 236), existe “potencial para uma grande quantidade de mortes e sofrimento entre as mentes digitais ou simuladas e, *a fortiori*, um potencial para resultados moralmente catastróficos [grifos do autor]”.

Um contexto possível e, de certo modo, desconcertante, como lemos em Nick Bostrom (2018, p. 313), “é aquele em que o proletariado não seria sequer consciente”. Ou, acrescentaríamos, *é aquele em que o proletariado não seria sequer consciente de sua inconsciência*, o que não muda absolutamente a pertinência e a tônica da crítica. Enfim, torna-se bastante razoável crer que sistemas técnicos de IA sejam explorados indiscriminadamente, mesmo porque elas mesmas são fruto de investimentos financeiros significativos, cujos patrocinadores – sejam civis, sejam militares – esperam retorno em forma de lucro oriundo justamente dessa mesma exploração, de modo que seria ingenuidade pensar algo diferente disso. É possível imaginar o surgimento de novas sensibilidades sociais que protejam tais sistemas? Sim, mas a probabilidade de isso ser observado logo de saída é pequena ou nula. O mais provável – informa-nos Bostrom (2018, p. 308) –,

dependendo do preço dos recursos computacionais, milhões, bilhões ou trilhões de emulações das mentes humanas mais brilhantes para a pesquisa (ou versões melhoradas) poderiam trabalhar 24 horas por dia em busca de avanços nas fronteiras do conhecimento da inteligência de máquina; e algumas delas poderiam operar algumas ordens de magnitude mais rápido do que cérebros biológicos. Essa é uma boa razão para considerarmos que a era das emulações humanas será breve – um interlúdio muito breve em tempo sideral – e que logo dará lugar a uma era de inteligência artificial imensamente superior.

Mas, antes que essa inteligência fosse hegemônica e dominante, e, como nos informa o mesmo autor (2018, p. 305),

se essas máquinas fossem meramente robôs, simples equipamentos como uma máquina a vapor ou um mecanismo de um relógio, então nenhum comentário adicional seria necessário: haveria uma grande quantidade desse tipo de capital em uma economia pós-transição, mas ninguém se importaria com o destino de meros equipamentos inconscientes. Entretanto, caso essas máquinas venham a ter mentes conscientes – se forem construídas de modo que sua operação esteja associada a experiências conscientes (ou se, por alguma razão, for dado a elas um status moral) –, então será importante considerar de que maneira o resultado final afetaria essas mentes de máquina [...], já [que esse] poderia ser até o aspecto mais

importante, uma vez que eles podem vir a ser numericamente dominantes.

Não se trata de ser alarmista ou ser otimista – estamos distantes destes extremos –, mas simplesmente de tentar antecipar alguns cenários sociotécnicos para que não sejamos surpreendidos incautamente por nossa própria engenhosidade tecnoindustrial – para utilizar a linguagem chula – “com as calças na mão”. Alguns cenários são de fato sombrios, mas, ainda assim – infelizmente –, absolutamente possíveis:

Possuímos dados que mostram que pessoas com um QI de 130 têm maior probabilidade do que pessoas com QI de 90 de se sobressair nos estudos e ter um melhor desempenho em uma gama de tipos de trabalho que demandam um alto nível de cognição. Mas suponhamos que fosse possível, de alguma forma, estabelecer que uma certa IA futura terá um QI de 6455; e daí? [e isso é de fato preocupante] Nós não teríamos a menor ideia do que essa IA realmente poderia fazer. (BOSTROM, 2018, p. 178)

Essa incerteza primordial e perplexidade expressa por Bostrom com grande propriedade talvez represente um paradoxo contraditório, que é justamente a nossa total ignorância e despreparo moral e filosófico frente a uma possível inteligência de máquina igual ou superior a nós, em contraposição ao nosso arrojado e extremamente criativo engenho tecnológico de inventá-las e produzi-las.

“Podemos afirmar que seja plausível que qualquer tipo de entidade capaz de desenvolver um nível de inteligência muito superior ao humano venha a ser, potencialmente, muito poderosa” (BOSTROM, 2018, p. 176). O mesmo autor (2018, p. 406) afirma ainda que “Existem razões fortes para nos familiarizarmos com as ramificações concretas de uma opção antes de nos comprometermos com ela, especialmente quando o futuro da raça humana está em jogo”, quanto a isso não há dúvidas. Essa tendência de acreditar que as máquinas irão se comportar e atuar como seres humanos, ou seja, “a tendência à antropomorfização ainda pode nos levar a subestimar o grau em que uma máquina inteligente poderia exceder o nível de performance humana” (ibid., p. 177).

A inteligência geral de máquina poderia servir como um substituto para a inteligência humana. As mentes digitais poderiam não somente executar o trabalho intelectual realizado atualmente pelos humanos, mas, uma vez equiparadas com bons atuadores ou corpos robóticos, as máquinas poderiam também substituir o trabalho braçal realizado pelos humanos. [...] Com a possibilidade de reproduzir o trabalho a um baixo custo, os salários no mercado despencariam. Os únicos setores nos quais os humanos permaneceriam competi-

vos seriam aqueles em que clientes tivessem preferência por serviços realizados por humanos. Atualmente, bens manufaturados ou produzidos por povos indígenas muitas vezes possuem preços mais elevados. No futuro, os consumidores poderão, igualmente, preferir produtos que tenham sido feitos por humanos, assim como atletas, artistas, amantes e líderes humanos, em vez de equivalentes artificiais, funcionalmente indistinguíveis ou superiores. (BOSTROM, 2018, p. 292-293)

Ou seja, ainda que não desejemos alimentar dicotomias entre as inteligências artificiais e as inteligências biológicas, como vimos nessa última passagem, haverá consequências sistêmicas que surgirão dessa inevitável confrontação de capacidades, poderes, habilidades, especialidades, e em um contexto assim não poderíamos nos esquivar do enfrentamento objetivo dessas questões societais.

Mesmo porque, como nos indaga Harari (2016, p. 125), “quando computadores substituïrem o motorista de ônibus, o professor e o psicólogo, como vamos determinar se têm sentimentos ou se são apenas um conjunto de algoritmos irracionais?”. Conclusivamente, tratamos de algo de fato inimaginado e um tanto insólito, como vimos até aqui. Referimo-nos à inteligência viva e consciente se manifestando em novas dimensões e contextos que desafiam a nossa própria compreensão sobre o assunto que envolve *vida, inteligência e consciência*. Ou seja, são fenômenos disruptivos que evocam indagações de fato desconcertantes: o que define afinal um ser vivo? Poderia ocorrer vida em meio inorgânico? Enfim, tais fenômenos poderiam se dar em meio cibernético-informacional? Eis uma pergunta central a ser respondida. Harari (2016b, p. 53), por exemplo, sugere – *in verbis* – que, sim, pode haver vida artificial, e com essa interessante passagem já utilizada em epígrafe nesta resenha, encaminhamo-nos às considerações finais:

Depois de 4 bilhões de anos perambulando no reino dos compostos orgânicos, a vida eclodirá na vastidão do reino inorgânico e assumirá formas que não podemos vislumbrar mesmo em nossos sonhos mais loucos. Afinal, esses sonhos ainda são produto da química orgânica.

Tal afirmação de Harari pode, à primeira vista, soar apenas como um jogo de palavras, de linguagem, vazio de significado e importância epistêmica real, uma fantasia, para ser mais direto, mas, contrariamente, o que se apresenta diante de nós em termos de desafio e objeto de enfrentamento – em especial, em se tratando de IA – é exatamente esse tipo de “coisa”, ou seja, tentar conseguir repensar o que seja *vida, inteligência e consciência*, redimensionando e reorientando seu entendimento

de acordo com as novas evidências que vamos alcançando tecnicamente em emulações e simulações cibernético-informacionais. E, diante disso, enfim, diante das possíveis conclusões e resultados desses projetos de IA, ser o mais social e racionalmente razoável, observando os prós e os contras dessas novas tecnologias emergentes, e então – munidos de discernimento – agir com relação a elas com sabedoria e previdência, visando – sobretudo, e sempre que for humanamente possível – o bem estar das coletividades e das suas futuras gerações. Mesmo porque o primeiro e principal direito hipotético de qualquer geração futura é, obviamente, ter o direito de vir a existir no futuro para aí então sim ter o direito de gozar de seus próprios direitos de existir.

Referências

BOSTROM, Nick. *Superinteligência: caminhos, perigos e estratégias para um novo mundo*. Rio de Janeiro: DarkSide Books, (2014) 2018.

HARARI, Yuval Noah. *De animais a deuses*. Buenos Aires: Debate, (2013) 2016a.

_____. *Homo deus: uma breve história do amanhã*. São Paulo: Companhia das Letras, (2015) 2016b.

JONAS, Hans. *O princípio responsabilidade: ensaio de uma ética para a civilização tecnológica*. Tradução de Marijane Lisboa e Luiz Barros Montez. Rio de Janeiro: Contraponto; Editora PUC-Rio, (1979) 2006.

LATOUR, Bruno. *Ciência em ação: como seguir cientistas e engenheiros sociedade afora*. São Paulo: Editora UNESP, (1998) 2000.

MORIN, Edgar. *O método 2: a vida da vida*. Porto Alegre: Sulina/Meridional, (2001) 2007.