

UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE PSICOLOGIA

**Um estudo introdutório dos aspectos psicossociais da  
artificialização das inteligências**

Luiz Joaquim Dias de Lima Nunes

São Paulo

2021



# **Um estudo introdutório dos aspectos psicossociais da artificialização das inteligências**

**Versão Corrigida**

Luiz Joaquim Dias de Lima Nunes

Dissertação apresentada ao Instituto de  
Psicologia da Universidade de São Paulo para  
obtenção do título de Mestre em Ciências.

Área de Concentração: Psicologia Social

Orientadora: Profa. Dra. Sandra Maria Patrício  
Ribeiro

São Paulo

2021

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE  
TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO,  
PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Catálogo na publicação  
Biblioteca Dante Moreira Leite  
Instituto de Psicologia da Universidade de São Paulo  
Dados fornecidos pelo(a) autor(a)

Nunes, Luiz J. D. L.

Um estudo introdutório dos aspectos psicossociais da artificialização das  
inteligências / Luiz J. D. L. Nunes; orientadora Sandra Maria Patrício Ribeiro. --  
São Paulo, 2021.

73 f.

Dissertação (Mestrado - Programa de Pós-Graduação em Psicologia Social) --  
Instituto de Psicologia, Universidade de São Paulo, 2021.

1. INTELIGÊNCIA ARTIFICIAL. 2. PSICOLOGIA SOCIAL. 3. HISTÓRIA DA  
CIÊNCIA. 4. FUTUROLOGIA. 5. INFLUÊNCIAS SOCIAIS. I. Patrício Ribeiro,  
Sandra Maria, orient. II. Título.

Ao meu marido, à minha irmã, a todos que tenham  
acreditado em mim, e àqueles que ainda virão.

## Resumo

NUNES, L. J. D. L. **Um estudo introdutório dos aspectos psicossociais da artificialização das inteligências**. Dissertação (Mestrado) - Instituto de Psicologia, Universidade de São Paulo, São Paulo, 2021.

A velocidade acelerada com que novas aplicações de inteligência artificial são hoje realizadas é fruto do potencial já observado nessas tecnologias por diferentes interesses ao longo do tempo, seja na ficção literária, nas Revoluções Industriais ou Guerras Mundiais. Essa corrida por aplicações é acompanhada, porém, de uma visão menos crítica quanto ao que as tecnologias são realmente capazes de fazer, e a inteligência artificial existente hoje é apenas um grupo de possibilidades dentre outras que, por motivos diversos, não foram exploradas. Os estudos de Piaget sobre inteligência combinados às perspectivas de Russell e Norvig sobre IA traçam um campo inicial de análise, que nos permite assumir um rigor interdisciplinar sobre do que se trata a inteligência, e a partir daí podemos ponderar, dentro da psicologia social, sobre o passado e o futuro da IA. Quando nos deparamos com o funcionamento dessas tecnologias e seu desenvolvimento ao longo do tempo, podemos identificar quais forças direcionaram sua trajetividade, reconhecer suas limitações e ter uma classificação mais nítida sobre os cenários em que podem ser aplicadas e quais pontos de atenção são necessários em cada um destes cenários.

Palavras-chave: Inteligência artificial. Psicologia social. Trajetividade. História da ciência. Futuro do trabalho.

## **Abstract**

NUNES, L. J. D. L. **An introductory study of the psychosocial aspects of the artificialization of intelligences.** Dissertação (Mestrado) - Instituto de Psicologia, Universidade de São Paulo, São Paulo, 2021.

The accelerated speed at which new applications of artificial intelligence are carried out today is the result of the potential already observed in these technologies by different interests over time, whether in literary fiction, Industrial Revolutions or World Wars. This race for applications is accompanied, however, by a less critical view of what technologies are really capable of doing, and the artificial intelligence that exists today is just a group of possibilities among others that, for various reasons, have not been explored. Piaget's studies on intelligence combined with Russell and Norvig's perspectives on AI trace an initial field of analysis, which allows us to assume an interdisciplinary rigor about what intelligence is all about, and from there we can ponder, within social psychology, about the past and the future of AI. When faced with the functioning of these technologies and their development over time, we can identify which forces have driven their trajectory, recognize their limitations and have a clearer classification of the scenarios in which they can be applied and which points of attention are needed in each one of these scenarios.

Keywords: Artificial intelligence. Social Psychology. Trajectory. History of science. Future of work.

## Sumário

Introdução	1
1 Definindo Inteligência	2
2 Imaginário e ambições que moldaram ou moldam a artificialização das inteligências	9
3 A trajetória realizada pela IA	16
4 Analisando a linha do tempo da IA	25
5 Imaginário da IA no futuro presente	33
6 O futuro da IA, suas inteligências e renúncias	45
Conclusão	65
Bibliografia	67



## Introdução

Apesar da especial atenção que recebe hoje, a ideia de inteligência artificial se estende por um longo período na história: desde o mito judaico-cabalístico dos *golens* - criaturas também vindas do barro porém sem livre arbítrio e com menor capacidade intelectual - até a visão cristã sobre a criação da vida a partir do barro, passando também por estórias infantis como a do *Pinóquio*, apresentam-se em culturas diversas estas construções humanas com aparente capacidade inteligente, mesmo anteriores à introdução dos computadores como conhecemos ao nosso imaginário. Como tecnologia emergente no cenário social, o termo “inteligência artificial” pode nem sempre ser utilizado de maneira correta, e assim sua definição se torna ainda mais difusa e incerta, levantando problemáticas relevantes para a comunicação científica e a psicologia social.

Com isso, mostra-se conveniente uma verificação na psicologia - desde sempre uma inspiração ao desenvolvimento da inteligência artificial - a respeito dos diferentes conceitos existentes para a inteligência em si, assim como avaliar os conceitos utilizados pela inteligência artificial com o passar das décadas. Quando falamos sobre *aprendizado de máquina*, por exemplo, estamos nos referindo a uma reprodução artificial do aprendizado humano ou a alguma outra capacidade semelhante a um aprendizado, mais conveniente à computação?

Ao se construir tal comparativo entre as definições e idéias aplicadas, podemos ter base para atingir ao menos dois objetivos com este projeto:

- comparar as formas de inteligência que desejaram-se reproduzir artificialmente através do tempo e identificar quais (e como) foram alcançadas, além de quais outras ainda não nos mostramos capazes de reproduzir;

- compreender que forças direcionaram o desenvolvimento destas tecnologias e quais são os seus interesses hoje evidenciados, assim montando uma perspectiva sobre seus futuros possíveis e as consequências destes.

E então, tendo essa perspectiva completa sobre o desenvolvimento da inteligência artificial - ou, como muito apelidada, IA - pode ser possível visualizar como o uso institucional desta tecnologia ou outros fatores sociais possam ter afetado ou direcionado este desenvolvimento, traçando assim para a psicologia social um caminho mais evidente quanto às suas possíveis contribuições a este tema.

## 1 Definindo Inteligência

Para que se possa investigar a inteligência artificial de modo aprofundado em cada uma de suas possíveis formas, fazem-se necessárias definições adequadas tanto para a *inteligência* quanto para a *inteligência artificial*. Mas levando em conta o tema em questão, essa necessidade pode ser vista como um desafio à parte.

Jorge Luis Borges inicia seu poema “El Golem” (1974, pg. 885) fazendo referência à significância dos nomes dados às coisas.

*Si (como afirma el griego en el Cratilo)  
el nombre es arquetipo de la cosa,  
en las letras de rosa está la rosa  
y todo el Nilo en la palabra Nilo.*

### **Crerios epistemológicos para boas definições**

Em um capítulo de seu livro “Introduction to Logic” (1999, pg. 151), Patrick Suppes trata da teoria da definição dentro da filosofia. Segundo seu pensamento, uma boa definição do ponto de vista epistemológico deveria se enquadrar nas seguintes regras:

- I. deve mostrar a essência daquilo a ser definido;
- II. não deve ser circular;
- III. não deve ser montada na forma negativa caso possa ser montada na forma positiva;
- IV. não deve ser expressa em linguagem figurativa ou obscura.

Mas outra noção importante trazida por Suppes é a de que, na prática, dentro da forma como a comunicação humana acontece, as definições ocorrem através da transcrição do termo a ser definido (*definiendum*) em função de outros termos já definidos (*definiens*). Especialmente no tema aqui proposto esta noção deve ser considerada, pois com o uso coloquial dos termos *inteligência* e *inteligência artificial* surge o risco de que cada indivíduo instantaneamente assuma alguma definição pessoal para tais termos - o que infringe a regra I do próprio Suppes, afinal a múltipla interpretação individual indica que não há uma essência em específico sendo mostrada na definição.

Ou seja, para que a regra I seja cumprida corretamente, é preciso fugir de termos com maior chance de múltipla interpretação individual, considerando ao mesmo tempo que toda definição inexoravelmente assume algum risco de que isto ocorra, visto que, de acordo com o propósito e o público-alvo de uma definição, tais chances podem ser maiores ou menores. É possível conjecturar que, em qualquer estudo de psicologia social que vise explorar a percepção de uma população sobre certo fenômeno ou objeto, há de haver uma atenção suficiente às definições selecionadas, de modo a reduzir o risco assumido.

Outro critério a ser aqui proposto é o critério instrumental, ou seja, com que propósitos tais definições serão utilizadas no contexto do presente trabalho. Para se estudar o desenvolvimento histórico da inteligência artificial e ao mesmo tempo identificar que possível aspecto de inteligência se está simulando artificialmente em cada momento, surgem alguns critérios hipoteticamente valiosos para o andamento deste trabalho, tal qual propostos:

I. deve ser possível conectar a definição de inteligência à definição de inteligência artificial, de modo que sua relação fique clara e que toda afirmação sobre inteligência seja suficientemente aplicável - ou seja, dentro das proporções adequadas - também à inteligência artificial; este critério é importante para que não seja assumida aqui uma definição de inteligência que englobe somente inteligência humana ou biológica, o que impossibilitaria uma leitura clara da inteligência artificial dentro dos estudos sobre inteligência na psicologia, conseqüentemente impossibilitando o método aqui seguido;

II. deve ser possível entender a existência de diferentes formas de inteligência dentro da definição de inteligência, e um processo análogo deve ser possível para a inteligência artificial também; este critério é importante para que se evitem definições de inteligência que se refiram apenas a uma forma de inteligência, como aquelas que se referem apenas ao raciocínio lógico ou ao aprendizado como inteligência; havendo uma definição suficientemente particionável, pode ser possível reconhecer diferentes formas de inteligência sendo estudadas ou simuladas pela inteligência artificial em cada momento de sua história.

Assumindo-se então as regras de Suppes junto aos critérios aqui propostos como premissas, podem-se utilizar estes todos como os fatores de validação quanto a uma definição de inteligência ser apropriada ou não para o trabalho aqui presente; as regras de Suppes buscam a qualidade epistemológica da definição enquanto os dois critérios instrumentais acima estabelecidos buscam a consistência entre a definição de inteligência e seu paralelo na inteligência artificial.

## **Definindo inteligência**

Em “Psychology of Intelligence” (1960, pg. 9), ao buscar uma definição concisa para a inteligência, Jean Piaget cita algumas definições anteriormente dadas. Para Edouard Claparède e William Stern, "inteligência é uma adaptação mental a nova circunstâncias", o que para Piaget indica uma oposição da inteligência ao instinto e ao hábito, que seriam “adaptações adquiridas hereditariamente ou através de circunstâncias recorrentes”; essa definição dá um destaque especial ao aprendizado enquanto forma de inteligência. Já para Karl Bühler e Wolfgang Köhler, a inteligência estaria ligada a insight - ideias -, enquanto tentativa e erro seriam formas de treinamento não-enquadráveis como inteligência; essa definição afasta da inteligência o processo de aprendizado através de comportamentos e contingências.

Piaget comenta diante dessas referências o seguinte: "É possível definir a inteligência pela direção de seu desenvolvimento, sem nos apegarmos à definição de fronteiras ou limitações, o que se torna apenas uma questão de estágios ou de sucessivas formas de equilíbrio. Podemos então abordar esse tema a partir do ponto de vista tanto da situação funcional quanto do mecanismo estrutural." Esse comentário condiz com o critério II para a seleção das definições, como citado anteriormente.

Dentre mais algumas reflexões, Piaget chega à declaração de que “a inteligência constitui o estado de equilíbrio ao qual tendem todas as adaptações sucessivas de natureza sensório-motora e cognitiva, assim como todas as interações assimilatórias e acomodatórias entre o organismo e o ambiente”. Para melhor compreensão dessa declaração, é válido acessar os conceitos de assimilação e acomodação por Piaget: a primeira diz respeito a compreender algo a partir do conhecimento já existente, enquanto a segunda se refere à modificação de conhecimentos existentes para possibilitar tal compreensão.

Essa declaração de Piaget é também trazida no artigo “As estruturas da inteligência, segundo Piaget: ritmos, regulações e operações” (1980) do professor Lino de Macedo. Nesse artigo a obra de Piaget é explorada de forma mais ampla, sintetizando as ideias construídas e amadurecidas ao longo de sua jornada, apresentando uma consolidação interessante e bem aplicável ao presente trabalho. Segundo ele, a inteligência toma a forma de uma estruturação

que gera padrões entre sujeito(s) e objeto(s), e esses padrões podem ser separados dentre três principais tipos - ou três *formas ou estruturas da inteligência*: ritmos, regulações e operações.

Instintos são extensões funcionais das estruturas dos órgãos, algo como a lógica dos órgãos, como a fome, a sede e os reflexos. Os instintos geram então *ritmos*, que seriam a primeira forma da inteligência. De modo geral, eles garantem as interações do sujeito com o meio desde o nascimento e mantém o sujeito existindo e funcionando como já funciona.

Após os ritmos há as *regulações*, processos que, através da experiência, modificam e evoluem os ritmos, a fim de adequá-los a contextos diversos e a levar resultados passados (feedbacks) em consideração. Esta forma de inteligência é especialmente estudada também pela análise do comportamento.

Por último há as *operações*, processos mais diversos que surgem através da experimentação (individual ou observadora). Em “A Psicologia da Criança” (1974), Piaget e Bärbel Inhelder listam as principais características das operações:

- São ações escolhidas entre as mais gerais, isto é, trata-se de coordenações de ações, como as que permite a criança reunir duas classes numa terceira, ordenar objetos, classificá-los, etc;
- São ações interiorizáveis, isto é, que podem ser realizadas tanto física quanto mentalmente;
- São ações reversíveis, isto é, que podem ser anuladas (via inversão) ou compensadas por uma outra ação (via reciprocidade);
- Nunca estão isoladas, mas sempre coordenadas em sistemas de conjuntos;
- São comuns a todos os indivíduos do mesmo nível mental;
- Intervêm tanto nos raciocínios privados quanto nas trocas cognitivas com outros membros do grupo.

Há também a noção de que as operações se distinguem em dois níveis: desde a ordenação e a classificação (primeiro nível, das operações sobre objetos concretos) até a implicação e a negação (segundo nível, do pensamento hipotético-dedutivo).

Esta distinção das três formas de inteligência traz uma estrutura bem clara àquilo a que Piaget havia se referido como a “direção do movimento” da inteligência, que foi a forma como concluiu sua definição para a inteligência em “Psychology of Intelligence”. Cabe então validar se essa definição é apropriada a este trabalho.

## Definindo inteligência artificial

Stuart J. Russell e Peter Norvig compilam no início do livro “Artificial Intelligence: A Modern Approach” definições existentes para a inteligência artificial, formando quatro definições gerais a partir destas, a fim de representar os diferentes métodos e objetivos de cada abordagem. Seriam elas como a seguir:

- Sistemas que agem como humanos (teste de Turing): abordagem na qual uma inteligência artificial existe quando uma máquina consegue se passar por um ser humano, simulando seu comportamento; baseia-se no teste proposto pelo matemático Alan Turing em 1950, em que uma máquina seria classificada como inteligente caso consiga, dentro de uma certa configuração, demonstrar seu comportamento para humanos de modo que estes afirmassem estar lidando com outro humano e não com uma máquina;
- Sistemas que pensam como humanos (modelagem cognitiva): abordagem na qual se considera que a inteligência artificial deve ser construída a partir da modelagem sistemática das habilidades cognitivas humanas, ou seja, a partir dos entendimentos da psicologia e da neurociência quanto à inteligência humana e da sua reconstituição aproximada dentro de um sistema; esta é a abordagem que mais utiliza conceitos da psicologia para o desenvolvimento da IA;
- Sistemas que pensam racionalmente (leis do pensamento): abordagem na qual se procura encontrar métodos lógicos para solução de problemas e reconstituir estes métodos dentro de sistemas, a fim de solucionar tais problemas não a partir da forma como humanos o fariam, mas a partir do método lógico correto; o termo *racionalidade* é aqui empregado neste sentido de “forma lógica correta”, partindo do princípio de que nem sempre humanos resolvem problemas da forma mais logicamente correta, por não pensarem de forma unicamente lógica; as maiores limitações para esta abordagem estão no fato de que nem toda informação é expressa de forma compreensível totalmente pela lógica clássica - como ocorre com as linguagens naturais humanas, por exemplo - e no outro fato de que a capacidade de processamento das máquinas se torna também um obstáculo para solução de problemas mais complexos.

- Sistemas que agem racionalmente (agentes racionais): esta abordagem parte do princípio de uma utilização prática da mesma racionalidade que a abordagem anterior; devido a este fator prático sobre a inteligência, assume-se que uma ação não precisa se prender a uma racionalidade perfeita, mas a uma racionalidade suficiente combinada a uma capacidade de tomada de ação.

Dentro desta segmentação, os autores apontam que estas quatro definições são determinadas por dois eixos: por um destes, duas delas tratam dos *processos de pensamento* e do *raciocínio* e as outras tratam de *comportamento*. Enquanto pelo outro eixo, duas delas tratam da inteligência *humana* e outras duas tratam de uma inteligência *ideal*, independente da performance humana.

### **Validação epistemológica**

Combinando-se então as definições aqui trazidas por Piaget, Macedo, Russell e Norvig, podemos aplicar sobre elas os fatores de validação decididos anteriormente.

A verificação quanto às regras de Suppes é direta, validando as três formas de inteligência (que juntas podem compor uma definição para a inteligência em si) como também as quatro abordagens de inteligência artificial (que funcionam como quatro definições distintas para IA, mas que devido à sua importância histórica, como apontado pelos autores, não devem ser selecionadas em detrimento uma da outra, e sim tratadas como complementares dos pontos de vista do seu desenvolvimento e de sua instrumentação).

Quanto ao critério instrumental proposto I, é possível observar um paralelo entre as três formas de inteligência e qualquer uma das abordagens da IA:

- Enquanto automação de funcionamento e “lógica dos órgãos”, os ritmos são análogos aos algoritmos principais de qualquer sistema, que agem de fato como instruções e operações de seu funcionamento programado;
- As regulações, enquanto modificadoras e causadoras da evolução dos ritmos, são análogas ao chamado *aprendizado de máquina*, que utiliza feedbacks de testes ou simulações a fim de descobrir o melhor funcionamento dentro de situações específicas, agindo como otimizações do algoritmo principal;

- Já as operações, enquanto combinações de ações mais simples ou de outras operações, são análogas a aplicações computacionais para solução de problemas específicos, como ordenação, classificação e inferência.

Além disso, existe uma relação hierárquica entre estes algoritmos principais, aprendizados de máquina e soluções específicas: cada uma se aplica sobre a anterior, melhorando a eficiência da máquina ou dando a ela novas capacidades, de fato constituindo um desenvolvimento da tecnologia. Esta noção não muda pela decisão quanto a uma abordagem da inteligência artificial ou outra, pois provém do funcionamento da tecnologia em si. Esse tema será aprofundado no capítulo 4, onde o funcionamento da inteligência artificial deverá estar mais claro.

Portanto é possível aplicar esta analogia para observar que a definição de inteligência aqui proposta é aderente também à inteligência artificial, ou seja, podemos observar assim a inteligência artificial como um tipo de inteligência de fato, como necessário pelo critério I.

Quanto ao critério instrumental proposto II, seu resultado fica exposto pelo mesmo raciocínio anterior. É possível observar diferentes formas de inteligência dentro da definição de Piaget, assim como observar diferentes formas de inteligência artificial independentemente de uma abordagem selecionada dentre as mostradas por Russell e Norvig.

Com isso temos que as definições aqui tratadas se mostram apropriadas para o desenvolvimento deste trabalho. Nos próximos dois capítulos, contudo, é possível notar que a conceituação da inteligência artificial nem sempre foi tão homogênea e que impasses têm ocorrido há décadas no que tange conceitos em torno dessa tecnologia e de suas implicações.



## **2 Imaginário e ambições que moldaram ou moldam a artificialização das inteligências**

Compreendermos as possíveis origens de qualquer estudo científico ainda em níveis filosóficos ou ideativos, anteriores a qualquer experimentação ou resultado, pode nos dar uma visão interessante sobre a inserção desse estudo no contexto social.

Em “Chronos, o Tempo. Da Mitologia Grega ao Relógio” (2015), Eda e Marcello Tassara explicam essa ideia:

Interpretar o mundo significa narrar a origem, mas a narração manifesta-se em percursos necessariamente diacrônicos, implicando, por sua vez, a anterioridade dos acontecimentos que compõem a sua história.

Cada acontecimento dentro de uma história é antecedido por outros que o contextualizam, constroem e motivam, sendo que sua ligação com o tema objetivado pode não ser tão clara à primeira vista.

Ou seja, recorrermos à história para a compreensão da ciência pode acabar nos levando a elementos mais distantes da ciência em si, na forma como hoje a reconhecemos e interpretamos. Considerando a ciência e a tecnologia como frutos da vivência humana e a intenção deste trabalho de compreender a inteligência artificial sob um olhar da psicologia social, mais importante que mapear as produções históricas em IA é encontrar os trajetos percorridos em busca dessas produções. Nesse trajeto não se encontra apenas a factibilidade tecnológica, mas as intenções, discussões e movimentações em torno da IA e que podem de fato ser ferramentas para a compreensão de suas relações com a imaginação simbólica e a vivência humanas.

Esse conceito é melhor aprofundado por Ribeiro (Ribeiro & Araújo, 2015) no que abrange as perspectivas de Durand, Berque e Piaget sobre a chamada *trajetividade*. Cabe, em busca de uma visão mais completa sobre a história da IA, explorar tanto indícios subjetivos quanto objetivos ao longo do tempo, para que assim se possa encontrar não somente a trajetória realizada pela IA quanto a trajetividade na qual ela se deu.

### **Idéias sobre seres artificiais ao longo da história**

Múltiplas mitologias apresentam histórias de seres dotados de características - inclusive a inteligência - comparáveis de alguma forma à humana, por exemplo. Cabe citar alguns exemplos para discussão.

Na mitologia nórdica, segundo Faur (2011), os deuses Odin, Vili e Ve moldaram Midgard - a terra humana - a partir do corpo de um gigante que derrotaram, e com base em troncos soltos criaram os primeiros humanos. Criaram também os elfos e anões a partir de outras partes mais ou menos nobres do mesmo gigante, dando a cada um destes povos um diferente lugar a habitar.

Nas religiões é também comum a existência de histórias a respeito da criação do ser humano. Nas religiões abraâmicas e no candomblé, por exemplo, o ser humano é moldado a partir do barro por uma divindade segundo sua própria imagem e então abençoado com a vida.

Há também aqui o já citado *golem* que, segundo Idel (1990) é um ser definido na cabala, repartição mística do judaísmo, como um ser humanóide, construído a partir de matéria inanimada manuseável - comumente o barro - por um sacerdote que, ao fim da construção, dá a ele vida e propósito. A técnica seria possível apenas para os praticantes mais poderosos do sistema, capazes de reproduzir o processo divino de criação do próprio ser humano. Carl Jung (1968) também cita o golem em “Alchemical Studies”, quando discute a figura do *homúnculo*, um ser idêntico em forma a um ser humano, porém pequenino, cuja criação foi um dos grandes objetivos da alquimia.

*Pinóquio* é uma figura bem conhecida das histórias infantis, surgida no século XIX, tratando-se de um boneco de madeira construído aos moldes de uma criança pequena por um carpinteiro que viria a ser seu pai quando, de modo mágico, o boneco recebesse vida. A história mostra o boneco aprendendo a levar a vida de uma criança, mas ao mesmo tempo compreendendo suas diferenças para com as crianças à sua volta.

Outra figura semelhante, porém vinda de um diferente estilo literário, é o *Frankenstein*, criatura retratada na obra de mesmo nome pela escritora Mary Shelley, também no século XIX. Na história, Victor Frankenstein cria um ser de forma humana a partir dos restos mortais de outras pessoas, dando a ela a vida e também o seu próprio nome.

Para se discutir criticamente o papel dos mitos na construção simbólica de algo, Gilbert Durand (1996) tem observações bem pertinentes a respeito:

O mito seria, de algum modo, o modelo matricial de toda a narrativa, estruturado pelos esquemas e arquétipos fundamentais da psique do *sapiens sapiens*, a nossa. É, portanto, necessário procurar qual (ou quais) o mito mais ou menos explícito (ou latente) que anima a expressão de uma linguagem segunda, não mítica. Por quê? Porque uma obra, um autor, uma época - ou, pelo menos, um momento de uma época - está obcecada de forma explícita ou implícita por um um (ou mais de um) mito que dá conta de modo paradigmático das suas operações, dos seus desejos, dos seus receios e dos seus terrores.

É possível afirmar que temos, então, a hipótese de uma origem mitológica para a inteligência artificial, e que a partir da compreensão dos mitos possamos traçar possíveis origens simbólicas de tais ideias.

Podemos notar, seja no contexto mitológico, religioso ou artístico, que algumas estruturas se repetem no que tange a histórias sobre a criação dos seres ou, em outros termos, às origens de suas “vidas”. Dentre tais estruturas podemos citar, por exemplo:

- A formação da matéria viva a partir da matéria não-viva;
- A preexistência de um ser criador em relação ao ser criatura;
- A forma do ser criador servindo como inspiração para a forma do ser criatura;
- E a imperfeição/inferioridade da forma do ser criatura quando comparada à forma do ser criador.

Embora o termo “inteligência” em geral não seja diretamente citado, o criador reconhece a criatura como viva a partir do momento em que suas formas e ações pareçam similares, ou mesmo imitações às suas próprias. Não apenas a forma física é replicada na geração da criatura, como também linguagem e comportamentos, sendo estes dois elementos bastante relacionados à teoria do desenvolvimento por Piaget (1960).

De diferentes maneiras é retratado um percurso similar de desenvolvimento: a matéria não-viva recebe forma análoga à do criador; depois, por vezes realiza algumas ações próprias, colocadas como mais primitivas ou inferiores; por fim, por alguma virada na história, tomar ações mais similares às de seu criador e ser encarada com maior apreço ou dignidade de respeito. Esse momento do reconhecimento da semelhança criador-criatura é encarado como um momento de louvor nas histórias, e é onde é atribuído o predicado de *viva* ou de *humana* a tal criatura. Aliás, é possível visualizar esse momento na famosa pintura “A Criação de Adão”, de Michelangelo.

A partir dessa noção é possível nos questionarmos sobre o quanto pensamos em inteligência artificial também como uma alegoria à própria inteligência humana, e sobre o porquê disto acontecer. Isto é colocado por Russell e Norvig (1995) inclusive como um dos dois aspectos para diferenciar as abordagens da IA, como já discutido no capítulo anterior. Também é possível nos questionarmos se o ser humano seria capaz de pensar na IA sem recorrer a uma tal auto-referencialidade, ou mesmo sem utilizar referencial algum. Seríamos capazes de reconhecer inteligência em algo sem que antes a pudéssemos reconhecer em nós mesmos? Responder tais questões exigiria uma discussão que foge ao presente trabalho, mas a partir do simples questionamento podem ser levantadas algumas hipóteses quanto à origem simbólica da IA.

O mais importante neste momento é que fique evidente esta possível relação da origem da IA para com as mitologias e religiões construídas ao longo da história da humanidade. Pode ser bastante difícil, se sequer possível, identificar as origens exatas destas ideias no passado longínquo dos povos humanos, mas podemos considerar que a relação é possível e tomar os devidos cuidados para a discussão científica a partir daí.

Ou seja, quando qualquer pessoa no contexto científico levanta uma hipótese quanto à construção de uma inteligência artificial, não podemos negar que tal ideia pode ter raízes simbólicas em algum aspecto cultural - aspecto este que pode inclusive não ser a ciência, mas a arte ou a religião. Ao dizermos isso, não fazemos nenhuma inferência que invalide ou desmereça a hipótese em si; na verdade, estamos tentando abrir caminhos para a geração de mais hipóteses para investigação.

### **As ambições históricas em torno da IA**

Embora seja milenar essa ideia de “seres artificiais dotados de vida e inteligência, construídos à imagem dos seres humanos”, a ideia de “seres que realizam o trabalho até então humano” é comparativamente recente, sendo evidenciada em especial na Revolução Industrial, a partir de 1750. Para o historiador Daniel Headrick (2009), mais ainda, isso constitui uma das características essenciais do industrialismo. Na sua estrutura, há quatro destas características, a seguir: a divisão do trabalho em tarefas menores e mais simples, a serem distribuídas e especializadas; a substituição da mão-de-obra humana por máquinas; o

aumento na produção acompanhado da redução de custos (consequência das duas primeiras); e a geração de energia a partir de combustíveis fósseis.

A ideia de se utilizar inteligência artificial para execução de trabalho humano talvez possa ser vista como, por fim, uma combinação da figura mitológica dos seres artificiais à figura industrial da máquina, almejada enquanto realizadora de trabalho. Não é necessário, tanto quanto é provavelmente impossível (Durand, 1996), determinar a origem exata de um símbolo, mas dada a frequência com que a indústria apareceria depois na história da IA, é válido destacarmos esse momento para posterior análise. Por agora, cumpre resgatar as imaginativas figurações que, historicamente, preconizaram a construção efetiva de máquinas inteligentes.

Charles Babbage (Turing, 1950), na *University of Cambridge*, propôs em 1837 uma máquina chamada Máquina Analítica (*Analytical Engine*, originalmente) que teria capacidade potencialmente infinita de processamento de cálculos, dependendo do tamanho com o qual pudesse ser construída, embora não houvesse tecnologia disponível à época para construção de tal máquina. Ada Lovelace reconheceu em sua proposta, porém, um potencial não apenas em cálculos, mas no processamento de informações representáveis através de números.

Das notas que adicionou ao trabalho se reconhecem o primeiro algoritmo computacional na história, assim como o primeiro computador digital, mas também uma observação se referindo à inteligência artificial, em tradução livre a seguir:

A Máquina Analítica não tem qualquer pretensão de *originar* algo. Pode fazer *o que quer que consigamos ordená-la* a fazer. Pode seguir análises, mas não tem poder de antecipar quaisquer relações analíticas ou verdades.

A partir desta nota podemos entender que, na percepção de Lovelace, as possibilidades de ação de uma máquina estão limitadas superiormente às possibilidades direcionadas pelas pessoas que a controlam. Ao mesmo tempo, se for possível transcrever uma ação utilizando não mais do que uma série de operações matemáticas constituintes, então deve ser possível direcionar uma máquina a realizar tal ação.

Além disso, a partir do momento em que Lovelace aponta a factibilidade de uma máquina realizar ações até então de responsabilidade humana, aponta também para a existência de limites a essa possibilidade. A citação de Lovelace a respeito da falta de originalidade no funcionamento de máquinas demonstra isso, e é possível nela enxergar o

quanto a ideia de inteligência artificial enquanto tecnologia vem junto à ideia da própria computação desde os seus primórdios.

Em maio de 1941, a revista *Astounding Science Fiction* publica um conto de Isaac Asimov chamado “Liar!”, onde Asimov postula as suas chamadas Leis da Robótica, como a seguir:

- 1) um robô não pode ferir um humano ou permitir que um humano sofra algum mal;
- 2) os robôs devem obedecer às ordens dos humanos, exceto nos casos em que tais ordens entrem em conflito com a primeira lei;
- 3) um robô deve proteger sua própria existência, desde que isto não entre em conflito com as leis anteriores.

Tais leis são criadas, no contexto das histórias do autor, para possibilitar a coexistência entre humano e máquina de modo seguro para os humanos. Note-se que a terceira lei institui algo semelhante a uma “pulsão de sobrevivência” nos robôs, mas apenas desde que, em todo caso, seja garantida a sobrevivência humana. O robô deve, então, defender sua existência como algo menos importante do que a existência humana.

Embora essas leis tenham surgido num contexto de ficção científica, Asimov era também pesquisador sobre temáticas como a da IA, e apresentava preocupações compreensíveis em suas obras, ainda mais quando considerados o rápido desenvolvimento tecnológico e as tensões políticas existentes na época. Suas ideias repercutiram bastante não apenas entre os adoradores do seu estilo literário, como também no meio científico. Suas leis da robótica, por exemplo, podem ser encaradas como uma versão artificial da ética, a fim de suprir a impossibilidade - ainda hoje existente - de máquinas construírem uma por conta própria. Isso se dá, porém, pela inserção na IA de leis construídas de forma estritamente lógica, sem aderência de qualquer fator afetivo, o que torna essa “ética artificial” fundamentalmente distinta da ética como compreendida no contexto humano. Ainda assim, para fins de direcionar uma máquina a se adequar à ética e aos interesses humanos, ela pode ser vista como uma simulação simplisticamente suficiente.

Em junho de 1949, o jornal britânico *The Times* publica uma coluna falando sobre os avanços tidos na *Manchester University* na criação de um chamado “cérebro mecânico” pelos professores Max H. A. Newman e Alan Turing. Através de avanços tecnológicos no

armazenamento e na velocidade de acesso a informações armazenadas, conseguiram construir uma máquina com maior capacidade computacional, utilizada inicialmente para a resolução de um problema matemático até então em aberto. A descoberta, porém, trouxe à tona a discussão sobre máquinas conseguirem efetuar algo que até humanos não haviam conseguido. Sobre isso, Turing afirma:

Esta é apenas uma amostra do que está por vir, e apenas a sombra do que será. Temos que ter alguma experiência com a máquina antes de realmente conhecer suas capacidades. Podem se levar anos até nos acostumarmos às novas possibilidades, mas não vejo por que ela não deve entrar em qualquer dos campos que normalmente abordam o intelecto humano e, eventualmente, competir em igualdade de condições.

No ano seguinte (Turing, 1950, e Russell & Norvig, 1995), a partir da questão “máquinas podem pensar?”, Turing disserta então sobre o que seria necessário para se reconhecer capacidade de pensamento em uma máquina. Sua argumentação o leva a propor a hipótese de como isto seria possível - um experimento onde uma máquina e uma pessoa são colocadas para responder a questões de observadores, num contexto onde:

- As respostas fossem “computáveis” tanto por uma pessoa quanto por uma máquina (sem questões envolvendo música, por exemplo, àquela época infactível de se processar computacionalmente);
- O contexto da comunicação não se desse de modo que identificasse uma das partes para os observadores (o que implica em tanto a máquina quanto a pessoa não ficarem visíveis, assim como a comunicação não poder ser realizada através da fala, por exemplo).

Turing traz algumas possíveis perspectivas contrárias para agregar à discussão, evoluindo-a sem tentar dar a ela uma resposta final. O texto deixa claro o quanto não haveria uma resposta simples ou consensual tão cedo. De fato, diversas outras publicações foram feitas posteriormente, dando continuidade e expansão a esse debate.

Estes feitos de Lovelace e Turing abriram caminho para diversas hipóteses de estudo no campo da inteligência artificial, tornando claras diversas possibilidades iniciais. Após a publicação desse último artigo de Turing, experimentos práticos começaram a ser publicados e receber destaque, sendo reconhecidos como inteligências artificiais de fato funcionais. Estes serão descritos no capítulo seguinte.

### **3 A trajetória realizada pela IA**

Após compreendermos a história das ideias e conceituações sobre a inteligência artificial, é então adequado que vejamos como a área de estudo de fato se desenvolveu ao longo do tempo.

É importante ressaltar que, para fins de análise das origens das diferentes formas de inteligência artificial, recebem maior destaque aqui apenas trabalhos que tenham destacadamente contribuído para com tais origens. A partir destes, é inumerável a quantidade de trabalhos existentes e que contribuíram para o aprofundamento dos conceitos e evolução das tecnologias por fim; tais trabalhos têm suas próprias importâncias destacáveis para o estudo da IA mas não caberiam no método adotado para o presente trabalho.

Além disso, a história da inteligência artificial se mostra estreitamente conectada à história da robótica durante todo o seu percurso (Buchanan, 2015, Russell & Norvig, 1995, e Headrick, 2009), visto que tantos estudos ao longo do tempo abordavam simultaneamente ambas as temáticas. Porém, quando estudamos robôs apenas de uma perspectiva de hardware, sem a temática da artificialização da inteligência envolvida, há também um afastamento quanto à proposta do presente trabalho. Dado isso, tais trabalhos não serão diretamente pontuados aqui, o que não significa, de forma alguma, que se queira negligenciar sua relevância para a história da IA.

#### **Século XIX**

Segundo Headrick (2009), a primeira nação a relacionar a pesquisa científica à indústria foi a Alemanha, na metade do século XIX, havendo laboratórios nas universidades ligados a indústrias específicas.

Em 1865 foi criado nos Estados Unidos o *Massachusetts Institute of Technology* (MIT), instituição de ensino construída com a mesma proposta. Em 1876, o famoso inventor Thomas Edison criou um laboratório em Menlo Park, também nos Estados Unidos, lá tendo construído alguns protótipos de grande visibilidade até hoje, como a lâmpada elétrica. Headrick atribui à criação do laboratório de Edison a expansão que viria a seguir, quando ao longo dos quarenta anos posteriores diversas indústrias viriam a criar seus próprios laboratórios de pesquisa.



## Década de 1940

Em 1943, McCulloch e Pitts (1990), no MIT, a partir dos seus estudos em neurofisiologia, apresentaram uma modelagem matemática para a sinapse, com base no conhecimento sobre o funcionamento dos neurônios à época. O modelo consistia na geração de uma reação binária (*sim* ou *não*) por parte de um neurônio, proporcional à intensidade do estímulo recebido. Depois, conectando-se os modelos de neurônios individuais, uma rede é construída, onde cada neurônio recebe o estímulo dos anteriores de modo sucessivo até uma resposta final ser atingida - o que pode ocorrer pelo número de interações realizadas ou pela convergência a um resultado.

Esta composição poderia ser utilizada para tomada de decisão - ou, extensivamente, previsão estatística - de variáveis numéricas, sendo chamada de *rede neural artificial*, uma tecnologia ainda hoje muito utilizada para simulações e previsões (Russell & Norvig, 1995).

Diversas tecnologias foram impulsionadas pelas Primeira e Segunda Guerra Mundiais, inclusive a computação (Headrick, 2009). O diferencial trazido, a começar pela velocidade de processamento de cálculos no caso da computação, mostrou-se como uma vantagem num contexto em que todo o mundo era afligido por tensões militares, acelerando investimentos em tecnologias mesmo em nações até então menos envolvidas no tema e enfim aproximando a pesquisa em ciência e tecnologia dos interesses governamentais também. A maioria dessas tecnologias foi originada nos Estados Unidos, na Alemanha, no Reino Unido e na União Soviética.

Durante a Segunda Guerra, Alan Turing foi aplicado na inteligência britânica e seus estudos lá contribuíram para a geração de uma tecnologia capaz de descriptografar mensagens das nações inimigas durante a guerra. Tal criação foi um marco histórico na área da criptografia, pois tinha uma performance muito superior a máquinas existentes na época.

A partir de sua atuação durante a Segunda Guerra Mundial, Norbert Wiener lança em 1948 seu livro “Cybernetics”, que trata sobre sistemas de controle e inteligência artificial. Uma das principais contribuições do livro é pensar a inserção de sistemas baseados em inteligência artificial num contexto de interação humano-máquina.

Também após a Segunda Guerra, governos por todo o mundo aumentaram consideravelmente seus investimentos em tecnologia (Headrick, 2009), em especial àquelas

que se mostraram tão impactantes durante as guerras. Esse aumento nos investimentos é, não à toa, um dos principais aspectos da Guerra Fria a partir de então.

Em 1945, John von Neumann (Headrick, 2009, e Neumann, 1945), na *University of Pennsylvania*, publica um artigo onde introduz o conceito de *programa armazenado* e mostra seu funcionamento, que permitiria a um computador armazenar diferentes algoritmos e alternar entre eles sem a necessidade de modificar algo no hardware em si. Com essa capacidade, computadores passam a receber de todos os interesses uma atenção ainda maior, com sua produção aumentando no começo dos anos 50, como também sua aplicação em governos, universidades e grandes empresas privadas.

Com o decorrer das décadas seguintes, os avanços tecnológicos permitiriam a produção de computadores com menores custos e tamanhos, aos poucos se tornando acessíveis para empresas menores e, por fim, para pessoas. A empresa *IBM* teve grande participação - e cresceu bastante - nesse processo.

## **Década de 1950**

Ocorre em 1956 (McCorduck, 2004) na *Dartmouth College* uma conferência que reúne pesquisadores de matemática, psicologia e ciência da computação para discutir temas acerca da resolução de problemas através das máquinas e suas consequências. É dito que neste evento é que realmente se utilizou o termo “inteligência artificial” pela primeira vez para se discutir essa temática.

Organizaram o evento John McCarthy, professor de matemática na mesma universidade, Marvin Minsky, pesquisador de *Harvard University*, Nathaniel Rochester, pesquisador da *IBM*, e Claude Shannon, pesquisador na empresa *Bell Telephone Laboratories* - um dos laboratórios de pesquisa mais importantes nos Estados Unidos à época (Headrick, 2009), com foco em eletrônica. Participaram do evento pesquisadores de cenários semelhantes - tanto de universidades quanto de empresas de tecnologia.

Frank Rosenblatt (1957), no *Cornell Aeronautical Laboratory, Inc.*, parte dos seus estudos a respeito da percepção de estímulos pelo cérebro e cria um modelo matemático para representar tal funcionamento, com o intuito de explorar a possível reprodução dessa

capacidade do cérebro em uma máquina. Em resumo, seu modelo é constituído por três subsistemas, um que capte os sinais em questão - através de luminosidade, eletricidade, som, etc. -, um intermediário que interprete os sinais e os converta em possíveis reações e um que acione as reações selecionadas pelo intermediário, ou seja, Rosenblatt representou a percepção através de uma sequência captação-interpretação-reação. Então transcreveu seu modelo eletronicamente e criou o chamado *perceptron*, uma estrutura computacional funcional, reconhecida como uma inteligência artificial.

Vale observar que, assim como Lovelace havia antes comentado, a máquina proposta por Rosenblatt faz aquilo que é direcionada a fazer: todas as regras de interpretação e reação são especificamente determinadas e desenhadas pelo cientista.

No ano seguinte (1958), ele apresenta novas ideias a partir do *perceptron*, mostrando como a mesma estrutura poderia ser utilizada para cálculo de probabilidades. Para tanto, ao invés de determinar as interpretações e reações minuciosamente como antes, poder-se-ia utilizar como reação a variável que se quer prever - também chamada de variável “alvo” -, como estímulos às variáveis que se queiram utilizar para previsão, e substituir a interpretação por um cálculo de distribuição de probabilidade que, ao ser encontrada, permitiria a previsão da variável alvo. Ou seja, ao invés de criar as regras para interpretação e reação, nesse cenário, as reações a estímulos no passado são utilizadas para inferir que interpretação é necessária para gerar tais reações a partir de tais estímulos.

Essa alteração e expansão do conceito cria, então, a possibilidade de uma máquina inferir interpretações a partir de históricos para então “imitá-las” para os novos estímulos recebidos. Esse processo é chamado até hoje, no contexto da inteligência artificial, de *aprendizado de máquina* (Russell & Norvig, 1995). Embora esse nome não seja utilizado em tal trabalho, é reconhecível a contribuição deste para com a construção do conceito.

Ainda em 1958 (Minsky, 1968, e Buchanan, 2005), John McCarthy publica “Programs with Common Sense”, onde apresenta uma proposta factível para representação de conhecimento no contexto artificial, incluindo uma linguagem de programação chamada LISP. Através de estruturas provenientes da lógica predicativa, da linguística e da epistemologia, a tecnologia permite armazenamento tanto de informações de estrutura mais complexa quanto de inferências recursivas sobre tais informações, compondo uma

representação artificial para o conhecimento, abrindo caminho para aplicação da IA em contextos mais diversos, como o das linguagens naturais.

### **Décadas de 1960, 1970 e 1980**

Em 1961 (Nof, 1999), um robô chamado Unimate foi o primeiro a ser aplicado na produção industrial, na empresa General Motors. Ele foi produzido a partir de uma patente de 1954 pelo inventor George Devol, que fundou uma empresa dedicada à produção de robôs como esse, a Unimation.

Feigenbaum e Feldman (1963) publicaram o primeiro livro de programas reconhecidos como inteligência artificial (Buchanan, 2015), contendo descrições de soluções funcionais.

Dentre elas, uma que pode ser citada em destaque é um programa de Arthur Samuel capaz de jogar xadrez, não só pela complexidade do jogo em questão - que acabava exigindo uma maior capacidade de processamento da máquina - mas também pela aplicação de aprendizado de máquina em sua composição.

Outra aplicação que vale comentar é um experimento de Feigenbaum que tentava simular o funcionamento da memória de seres humanos num contexto artificial, através de memória associativa.

Além destas, há também soluções no livro dedicadas à resolução de problemas científicos utilizando a reprodução de métodos já conhecidos, provenientes de áreas como a física. Em geral, porém, estas soluções se baseiam na reprodução de regras lógicas transcritas por especialistas nas áreas envolvidas.

Os anos 60 foram um período crucial para a consolidação de certas organizações como grandes investidoras no desenvolvimento da IA (Buchanan, 2005), dentre elas se destacando:

- Massachusetts Institute of Technology (MIT);
- Carnegie Mellon University (CMU);
- Stanford University;
- University of Edinburgh;

Essa visão de Buchanan é exemplificada pelo grande volume de conteúdo produzido por estas instituições durante - no mínimo - esse período.

Também nos anos 60 (Headrick, 2009), o Departamento de Defesa dos Estados Unidos atentou à vulnerabilidade de seus sistemas a ataques nucleares. Estudos militares levaram à construção de algumas formas de comunicação sistêmica entre múltiplos computadores, com protótipos funcionais a partir de 1969, que mais tarde viriam a originar a internet em 1983.

Weizenbaum (1965, 1976), no MIT, Winograd (1973), na *Stanford University*, e Schank & Abelson (1977), na *Yale University*, realizam diferentes experimentos de compreensão de linguagem natural por máquinas, nos quais aplicam conceitos em torno linguagem humana escrita e tentam reproduzi-los no contexto artificial.

As abordagens variam de acordo com o experimento proposto, mas possuem um questionamento em comum: como tornar uma máquina capaz de compreender a linguagem humana? Mais ainda: no que consiste a compreensão da linguagem natural?

Algumas respostas hipotéticas no trabalho giram em torno de:

- A máquina responder perguntas a respeito das mensagens que recebe;
- A máquina gera uma representação computacional das mensagens que recebe;
- A máquina age de acordo com um comando interpretado a partir das mensagens que recebe.

Os resultados alcançados em cada experimento podem ser favoráveis dependendo de como definimos “compreensão de linguagem” em dado contexto e repartimos tal operação em operações menores, logicamente representáveis.

## **Dimensões de desenvolvimento da IA**

Ao longo da história da inteligência artificial - e da ciência da computação como um todo - é possível notar diversos momentos em que conceitos foram criados, soluções foram demonstradas teoricamente e protótipos foram arquitetados sem que ainda houvesse conhecimento ou recursos disponíveis à época para construção de tecnologias com base em tais produções. Esse processo deve ser inevitável nas ciências, ainda mais quando tomamos o ponto de vista de que os próprios avanços teóricos podem orientar as pesquisas em tecnologia.

Como dito anteriormente, há muita produção na história da IA que não caberia incluir neste trabalho devido ao seu foco, mas que tem também suas devidas importâncias em cada

momento. Ocorreram esforços em diferentes linhas de pesquisa em tecnologia para viabilizar a produção de protótipos funcionais.

Mais ainda, é possível tentar traçar alguns dos aspectos essenciais para esta viabilização, não apenas no último século e nos primórdios da IA como ainda hoje, visto que tantas pesquisas em torno desses temas continuam existindo. Desse modo se pode visualizar as distintas *dimensões* nas quais a tecnologia da IA se desenvolveu.

Esses aspectos podem ser exemplificados nos trabalhos anteriormente citados e também ser consolidados como os seguintes:

- Captação de informações: para que máquinas “percebam” o ambiente ao seu redor, são necessárias interfaces suas para com este. Essas interfaces vão desde os periféricos aos quais nos habituamos no dia-a-dia, como teclados e mouses, até implementações mais complexas como câmeras e sensores. Alguns dos primeiros protótipos de IA envolviam a necessidade de um humano inserir texto como entrada de informação, por exemplo. Com o desenvolvimento de formas de captação de luzes, imagens, sons e, finalmente, vídeos, abriram-se as possibilidades de protótipos. Ainda hoje nem todo tipo de informação é comumente captável por máquinas, como aquelas captadas pelo paladar humano, por exemplo - embora alguns experimentos a respeito do tema sejam já realizados (Fleming, 2008).
- Representação de informações: após captadas, informações precisam ser representadas em um formato adequado ao seu armazenamento nas máquinas. O processo envolve a construção de estruturas computacionais adequadas às representações de tais informações. Além de possibilitar o armazenamento e a organização, as representações utilizadas afetam também a velocidade de acesso e as possibilidades de edição. Uma frase em português pode, por exemplo, ser representada em uma máquina como uma sequência de palavras ou como uma hierarquia de palavras - montada de acordo com a relação dessas entre si. Estudos em representação de informações podem então ter foco em criar estruturas para novos tipos de informação ou propor novas e melhores estruturas para as mesmas informações. Há também o conceito de representação de conhecimento, que utiliza estruturas ainda mais complexas a fim de incluir não apenas informações, mas também sistemas a respeito destas. Vale notar que a

epistemologia tem contribuições valiosas a essa dimensão, tanto já exploradas quanto ainda em aberto.

- Armazenamento de informações: um dos maiores desafios citados nas primeiras idealizações da IA é a questão da capacidade de armazenar informações para utilização posterior. Isso porque, embora algoritmos tenham começado a ser escritos e demonstrados desde os tempos de Lovelace, esses por vezes precisariam armazenar um volume de informações superior à capacidade na época. A física, a ciência dos materiais e a engenharia foram fundamentais, dado isso, para viabilizar essa dimensão (Headrick, 2009).
- Velocidade de processamento: de modo semelhante à anterior, esta dimensão foi uma limitação em especial nos primórdios da tecnologia da IA. Dado o envolvimento de pessoas da área da matemática com seu desenvolvimento, muito se tentou utilizar a IA para, por exemplo, demonstrar teoremas e testar conjecturas. Uma abordagem para se fazer isso, por vezes aplicável, é a chamada *força bruta*, que consiste em testar cenários possíveis em busca de contra-exemplos, por exemplo. Essa abordagem nada otimizada, baseada na repetição potencialmente infinita de geração de cenários e posterior verificação, torna-se menos viável quando a máquina envolvida não tem uma boa velocidade, por exemplo.
- Modelagem de soluções: quando Lovelace afirma que a máquina analítica “pode fazer o que quer que consigamos ordená-la a fazer”, é esta dimensão que entra em questão. Toda solução de um problema utilizando inteligência artificial depende diretamente de uma forma de transcrevê-la em comandos lógicos e então em um código de dígitos binários. Esta transcrição depende tanto da representação das informações quanto da estruturação de seu funcionamento nas devidas etapas lógicas, organizadas sistemicamente. Quando se pensa, por exemplo, em um algoritmo que identifique se um número digitado é par, basta que efetue uma operação simples de álgebra e retorne o resultado - mas se o problema em questão é identificar as expressões emocionais de uma pessoa a partir de uma foto sua, a complexidade é outra. A evolução do nosso conhecimento a respeito de um problema, portanto, pode expandir as possibilidades de abordagens utilizáveis em uma inteligência artificial. Isso é

possível, como já comentado, somente se tal abordagem puder ser transcrita em operações lógicas.

- Capacidade de ação mecânica: essa última dimensão diz respeito à capacidade de ação de uma máquina sobre o ambiente em que está inserida, de acordo com qualquer motivação baseada em inteligência artificial ou não. Quando uma máquina consegue identificar qual ação deve ser tomada para resolver um problema, mas não é capaz - por limitações físicas suas - de agir de fato, na prática a máquina não está sendo eficaz diante de tal problema. Esse aspecto pode ainda hoje ser um impedimento para utilização de inteligência artificial em contextos específicos que requeiram precisão das ações - como no caso de máquinas utilizadas hoje para realização de cirurgias, que embora já sejam uma realidade ainda possuem avanços a serem realizados.

Essas seis dimensões apresentadas sintetizam os avanços necessários da inteligência artificial dentro das suas respectivas naturezas e podem nos dar uma visão de mundo quanto aos diferentes fatores que encurtam ou expandem as possibilidades da tecnologia. Isso não significa, porém, que basta a tecnologia continuar avançando nessas seis dimensões para que seja possível solucionar qualquer tipo de problema utilizando-a. No próximo capítulo será possível analisar essa questão com melhor foco.



## 4 Analisando a linha do tempo da IA

Tendo então uma perspectiva sobre os eventos que direcionaram o desenvolvimento da inteligência artificial ao longo de sua história, é possível observar que ela passou por diferentes ambientes, dotados de seus próprios objetivos, dentro dos quais se redefiniu, reorientou e expandiu.

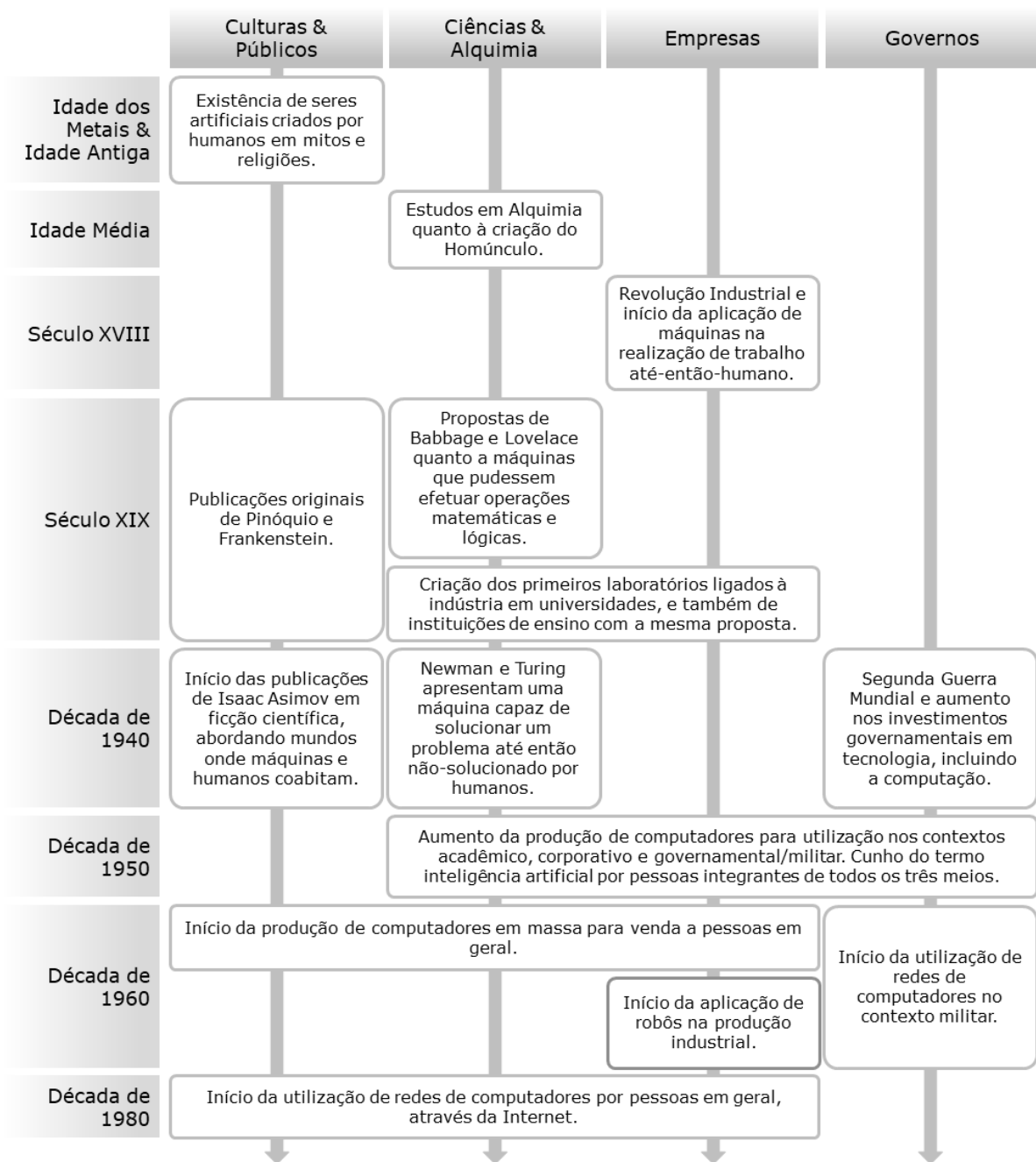
Dentre esses meios é possível delinear quatro que parecem ser adequados para destacar tais modificações nessa história:

- Culturas & Públicos: compreende aquilo que é de conhecimento comum, senso comum ou acesso público, englobando desde fatores religiosos e mitológicos até aqueles artísticos ou de acesso à tecnologia, sendo o meio mais abrangente dentre os quatro, tanto em temas incluídos e relacionados quanto na quantidade de pessoas atingidas direta ou indiretamente; é importante para notarmos os momentos nos quais a própria humanidade orientou o desenvolvimento da inteligência artificial ou passou a ter um maior ou menor acesso a ela;
- Ciências & Alquimia: compreende os estudos formais, filosóficos e metódicos a respeito do tema, tanto nas ciências quanto na alquimia; abrange, na prática, todo avanço conceitual, científico ou tecnológico realizado em busca de conhecimento, sem foco específico em uma aplicação ou problema existente; note-se que, nesse meio, assim como descrito no capítulo 2, não deve ser possível afirmar certamente que qualquer avanço seja isento de influências anteriores provenientes de outro dos meios;
- Empresas: compreende os momentos em que empresas privadas intervieram no desenvolvimento da inteligência artificial, geralmente com objetivos financeiros, não necessariamente de curto prazo;
- Governos: compreende os momentos em que governos de nações intervieram, fosse para fins de desenvolvimento tecnológico interno ou fins geopolíticos.

A partir da classificação desses quatro meios é possível construir uma linha do tempo como no diagrama 1, em que podemos destacar os meios envolvidos em cada momento da história. Note-se que há momentos em que, por haver uma interação entre dois ou mais meios, os momentos aparecem classificados em todos os envolvidos ao mesmo tempo. Note-se também que alguns dos momentos históricos citados nos capítulos anteriores não são

mostrados neste diagrama por serem interpretáveis como continuidades ou consequências dos eventos anteriores, sem que houvesse uma real mudança de direção ou aceleração no desenvolvimento da inteligência artificial ou no envolvimento ou não de algum dos quatro meios considerados.

**Diagrama 1: Linha do tempo da história da Inteligência Artificial**



A partir do diagrama é possível analisar a inserção da inteligência artificial dentro da sociedade, seus caminhos e também os momentos nos quais a intervenção de algum dos meios possa ter causado uma alteração drástica, não necessariamente positiva ou negativa, ao desenvolvimento desta área das ciências da computação.

Como mostrado no capítulo 2, a ideia de seres artificiais teve origem há alguns milênios, nos mitos e religiões, mas passou a ser experimentada de fato apenas na Idade Média, através da alquimia, que buscava a criação de Homúnculos dentre outros objetivos. Mesmo que a alquimia não seja considerada como uma ciência, ela possui seu envolvimento com a história da ciência como um todo, e pode ser vista nessa situação como uma ponte entre as culturas humanas e a experimentação científica que viria a seguir. E quando pensamos na academia no contexto social, temos a forte possibilidade de que tais ideias a serem experimentadas fossem sempre influenciadas, de diferentes maneiras, pelas culturas nas quais os pesquisadores estavam inseridos.

A partir da Revolução Industrial, começando no século XVIII, surgiu o crescente e lucrativo interesse no setor privado de substituir a mão-de-obra humana por máquinas, reduzindo custos e aumentando *rentabilidade*, até que no século XIX começassem a ser criadas instituições de ensino e laboratórios dedicados à produção científica e tecnológica com foco na aplicação pela indústria. Nessa época já existiam trabalhos em computação apontando para a possibilidade de se efetuarem operações ainda mais complexas do que aquelas puramente mecânicas ou eletrônicas que as máquinas eram capazes de realizar durante as primeiras décadas da Revolução Industrial, o que naturalmente colocava a computação como, potencialmente, uma grande fonte de rentabilidade para a indústria naquele momento. A partir da mesma época podemos ver uma mudança na produção científica a respeito da ideia de vida e inteligência artificiais, com os trabalhos tendo um caráter filosófico menos frequente e intenso, dando espaço a um caráter prático com foco em aceleração de operações e consequente aumento na rentabilidade.

### **Abordagens para a IA ao longo do tempo**

Através das quatro abordagens para a IA citadas por Russell & Norvig (1995) é possível realizar um comparativo quanto aos momentos em que elas tiveram mais ou menos espaço nos estudos científicos. Retomando-as, seriam as quatro:

- I. sistemas que agem como humanos;
- II. sistemas que pensam como humanos;
- III. sistemas que pensam racionalmente;
- IV. sistemas que agem racionalmente;

Para fins de rentabilidade no contexto produtivo industrial, uma máquina pensar ou agir como um ser humano não é prioridade, visto que àquela época isso não impactaria na produtividade da máquina em si - a rentabilidade se dava por maior precisão, agilidade e eficiência na atividade, o que independe de a ação ser similar à humana ou não. No contexto das relações entre indústrias a partir da Revolução Industrial, então, onde a corrida por tecnologia gerava vantagens competitivas, investir tempo na construção de algo que parecesse humano poderia ser inclusive ruim, por exigir mais tempo - quando sequer possível - e não ter necessariamente um impacto sobre a velocidade ou a qualidade das produções.

Até então, sem essa competição envolvida, nos trabalhos nos quais se pode notar uma ponderação sobre algo *semelhante* a uma inteligência artificial - ainda sem receber esse nome -, mostravam-se questões a respeito do que constitui a vida ou o pensamento humano e como replicá-los. Nas próprias histórias de origem mitológica ou artística, o momento em que a vida artificial é reconhecida é em geral aquele onde ela age de forma similar à vida original, a do ser criador. A artificialidade parece estar associada nesse contexto a uma ideia de imitação de algo original e anterior.

Ou seja, antes da Revolução Industrial as abordagens I e II citadas por Russell e Norvig tinham maior espaço no questionamento científico. Quando o interesse em rentabilidade passou a intervir mais sobre a pesquisa em IA, é possível dizer que ocorreu uma transição de destaque - e investimento - para as duas últimas abordagens, III e IV.

Isso não quer dizer que as abordagens I e II tenham desaparecido, porém. Podemos observar em inúmeros trabalhos de inteligência artificial que a questão sobre “como funciona e age o ser humano” inspirou e possibilitou o avanço científico, com a criação de soluções para problemas mais complexos do que aqueles de natureza puramente lógica, numérica ou mecânica. É o caso, por exemplo, dos protótipos dedicados a interagir de algum modo com a linguagem humana, onde a linguística e a epistemologia mostraram ter tanta influência, assim como daqueles envolvendo a reação a estímulos e a tomada de decisões, onde a psicologia e a neurociência foram essenciais. Então, embora para a indústria num primeiro momento as

abordagens III e IV fossem mais prioritárias, a pesquisa científica se dissolveu entre todas as quatro, tendo elas portanto papéis complementares dentro da sociedade e da história.

Outro tópico a respeito disso, tanto citado por Headrick (2009) quanto visível nas pesquisas e publicações em computação na época, é que esse contexto industrial intervindo sobre as ciências tornou mais comum a replicação de trabalhos entre diferentes pesquisadores e empresas, devido à corrida tecnológica existente. No contexto da pesquisa em inteligência artificial isso tanto aumentou a quantidade de pessoas envolvidas no tema quanto incentivou uma descentralização dos estudos em novas soluções de IA. Uma consequência direta disso é um aumento perceptível na dificuldade de se traçarem origens exatas para certas formas de IA a partir desse envolvimento da indústria nas pesquisas.

A partir da Segunda Guerra Mundial os governos passaram a ter maior envolvimento com a ciência e a tecnologia como um todo, devido ao impacto que tais temas tiveram sobre os resultados da guerra. No caso da computação em específico, mostram-se interesses a princípio quanto à execução mais rápida de tarefas antes realizadas por humanos, o que é semelhante em forma ao interesse industrial em rentabilidade, porém as finalidades são mais abrangentes, envolvendo segurança, logística e comunicação, por exemplo, e não exatamente produção. O aumento nos investimentos em ciência e tecnologia iria impulsionar, então, a construção de soluções aplicáveis a contextos que envolvam essas finalidades que, em comum, têm por objetivo a geração de algum bem à sociedade ligada a tal governo.

Embora seja possível enxergar que essa maior abrangência nas finalidades crie mais caminhos para que o desenvolvimento da inteligência artificial beneficie os seres humanos, é importante pontuar que havia ainda uma corrida tecnológica acontecendo entre os governos globais e os recursos disponíveis para investimento em ciências estavam bastante concentrados em nações específicas (Headrick, 2009). Então, ao mesmo tempo em que a tecnologia poderia se aproximar mais das pessoas dessas nações, viria a crescer a desigualdade do acesso à tecnologia entre as nações ao redor do mundo.

As pesquisas científicas em IA continuaram distribuídas entre as quatro classes de Russell e Norvig a partir daí, sendo favorecidas pelo maior investimento vindo do governo e da maior diversidade de finalidades envolvidas.

Mais tarde, com a maior produção de computadores e a criação das redes de computadores - ambas num primeiro momento impulsionadas para a utilização em empresas e órgãos governamentais -, aos poucos empresas menores e pessoas comuns iriam conseguindo

acesso a máquinas de uso pessoal e a computação passaria a fazer uma parte maior de seus cotidianos. O acesso a linguagens de programação viria a permitir que as próprias pessoas passassem por fim a poder construir suas próprias soluções computacionais caseiras e, conseqüentemente, inteligências artificiais também.

### **Dimensões de desenvolvimento da IA ao longo do tempo**

É possível perceber, ainda, que as seis dimensões definidas no capítulo 3 conseguem sintetizar as formas de avanço tecnológico que continuaram a acontecer nas últimas décadas:

- Pranchas de desenho, microfones e câmeras passaram a possibilitar a captação de mais tipos de informações;
- A quantidade de formatos de arquivos cresceu imensuravelmente, hoje havendo múltiplas opções para múltiplos tipos de informação a serem guardados e reproduzidos; as diferenças entre tais formatos de arquivos, quando se trata do mesmo tipo de informação, dizem respeito em geral a diferentes representações utilizadas para esses tipos;
- Maiores capacidades de armazenamento e processamento tiveram, só nas últimas duas décadas, um aumento incomparável também, sendo que um disquete era capaz de armazenar não mais do que uma foto tirada com a qualidade das câmeras de celulares populares hoje;
- As linguagens de programação para uso público, assim como programas para construção automatizada de algoritmos, possibilitam a criação de soluções lógicas por qualquer pessoa que saiba manusear tais ferramentas;
- E recentemente começaram a se popularizar impressoras 3D para uso doméstico, capazes de construir objetos físicos moldados dentro de computadores.

Uma observação a ser feita quanto a essas seis dimensões, por fim, é o fato de que seus avanços continuam sendo impulsionados hoje por diferentes interesses, inclusive pelas necessidades de empresas e governos. É comum que tais órgãos tenham, por exemplo, maiores quantidades de dados a processar e mais problemas novos a serem resolvidos do que indivíduos comuns então, dados os impactos que tais meios tiveram sobre a história da ciência e da tecnologia, hoje é inevitável que as seis dimensões continuem a ser impulsionadas por demandas provenientes destes meios.

## **Estruturas da inteligência ao longo do tempo**

Por fim, cabe também analisar o quanto as três estruturas da inteligência definidas por Piaget (Macedo, 1980) conseguem representar, igualmente bem, o desenvolvimento que a inteligência artificial teve em sua história. No capítulo 1 isso foi comentado superficialmente, o que cabe agora validar tendo a linha do tempo construída.

Desde antes do século XVIII temos máquinas sendo aplicadas no auxílio a atividades humanas. As primeiras possuem funcionamento regido por alguma engenharia orientada a um objetivo específico, aplicando os conhecimentos em física e química à época para que se conseguisse um funcionamento naturalmente explicável e interpretável, sem que ocorresse nada análogo ao raciocínio, ao pensamento ou a decisões por parte de alguém além do humano operando a máquina e das leis do mundo natural. Essas máquinas efetuam aquilo que são levadas físico-quimicamente a realizar, de forma consideravelmente simples quando comparada aos fenômenos da inteligência humana. Esses funcionamentos podem ser traçados numa analogia, então, aos ritmos, como definidos por Piaget e Macedo.

O *perceptron*, proposto por Frank Rosenblatt (1957 e 1958), é uma das primeiras aplicações de inteligência artificial nas quais ocorre o chamado *aprendizado de máquina*, onde, a partir da interação e experimentação da máquina ao tomar uma ação sobre o ambiente em que está inserida, a máquina capta o resultado ou a reação do ambiente frente à sua ação e utiliza tal informação captada para reajustar sua tomada de ações. Podemos traçar uma semelhança do aprendizado de máquina para com as regulações de Piaget, então, sendo essas as causadoras da modificação e melhor adaptação dos ritmos ao ambiente.

Já as operações, a última estrutura, funcionam como algo semelhante ao princípio de Lovelace quanto a máquinas funcionarem de acordo com uma série de operações lógicas ou numéricas menores e mais simples - princípio esse que continuou a ser utilizado no século XX quando estudos buscaram soluções para problemas mais complexos, como a interpretação e a tradução de linguagens naturais. A fim de se construírem soluções para tais problemas, estes foram comumente particionados em problemas menores, de forma sucessiva, até o ponto em que talvez pudessem ser solucionados com base apenas em operações lógicas. Esse reparticionamento de operações menores para construção de operações mais complexas que

solucionem problemas naturalmente mais complexos cria uma semelhança desse processo computacional para com as operações como vistas em Piaget.

Ao se fazer isso na computação, por vezes os problemas são de fato solucionados dentro das devidas condições, mas também surgem polêmicas quanto à real natureza dos problemas em questão. Por exemplo, para traduzirmos uma frase do inglês ao português, basta que façamos a tradução de cada palavra separadamente? Pode funcionar, dependendo da frase, mas sabemos que nem sempre será o caso. Num exemplo mais complexo, como poderíamos tornar uma máquina capaz de reconhecer emoções em seres humanos? Seria possível particionar esse reconhecimento em etapas puramente lógicas?

Estudos em inteligência artificial continuam existindo hoje mais do que nunca e alguns buscam responder a esse tipo de pergunta. Para que possamos discutir isso, porém, é preciso investigar mais a fundo as atuais ambições e imagens que direcionam a IA, assim como mapear suas limitações, presentes ou eternas. Essa investigação depende, então, do próximo capítulo.



## 5 Imaginário da IA no futuro presente

No atual momento deste trabalho e da história da inteligência artificial, temos já algumas ferramentas preparadas: um esquema para identificar e categorizar os interesses direcionadores do tema (capítulo 4) e um para as seis dimensões nas quais a IA se desenvolveu ao longo do tempo (capítulo 3), além de outros dois esquemas para classificação - sendo um deles para as três estruturas da inteligência, por Piaget e Macedo (1980), e o outro para as quatro formas de inteligência artificial, por Russell & Norvig (1995). A partir dessas ferramentas podemos agora observar criticamente o que está sendo considerado, figurado, debatido ou produzido para as tecnologias que se podem acomodar sobre esse guarda-chuva.

### As tendências para o futuro da IA nos negócios

Há anos a revista *Forbes*, voltada principalmente a negócios e temas correlatos - incluindo tecnologia -, vem apresentando artigos sobre tendências para a IA no mundo corporativo. É comum que tais artigos sigam um formato anual, anunciando possíveis tendências para o ano seguinte ou que se inicia de acordo com quais temas tiverem discussões emergentes no momento (Schoette, 2018; Spangler, 2018; Marr, 2020; Newman, 2020; Kiran, 2021). Para os últimos quatro anos são citadas, por exemplo, as seguintes tendências (note-se que a quantidade anual é variável), apresentas aqui resumidamente:

- 2018.1: preparação de dados acelerada através de automações possíveis, a fim de reduzir o tempo de trabalho necessário para desenvolvimento ou aplicação de tecnologias de inteligência artificial;
- 2018.2: adoção de ferramentas com maior capacidade de armazenamento de dados, através de tecnologias de *nuvem* (serviço terceirizado de armazenamento de dados), possibilitando mais aplicações e em novas áreas de negócio;
- 2018.3: desenvolvimento de infraestruturas que facilitam a gestão e o acesso a grandes quantidades de dados, viabilizando seu uso;
- 2018.4: adequação do compliance das empresas às mais recentes legislações quanto a dados pessoais, como a GDPR;

- 2018.5: adequação das culturas corporativas ao maior armazenamento de dados ocorrendo na rotina de trabalho;
- 2019.1: transição das pessoas trabalhadoras para novas funções devido à automação de alguns trabalhos e à criação de outros relacionados a tecnologia;
- 2019.2: aumento de eficiência na área de logística através da automação de atividades;
- 2019.3: aplicações de IA na gestão da comunicação com clientes a fim de impulsionar vendas e intensificar o acompanhamento de performance dos operadores;
- 2019.4: aumento da demanda no mercado pelas pessoas chamadas de *cientistas de dados*, que apliquem tecnologias relacionadas a estatística, computação e inteligência artificial;
- 2019.5: aplicação de IA em estudos da medicina, acelerando a produção de medicamentos e seus testes;
- 2020.1: monitoramento e aprimoração dos processos do negócio;
- 2020.2: personalização das estratégias de relacionamento com clientes a partir dos dados históricos;
- 2020.3: desenvolvimento de simulações baseadas em IA para testar veículos e equipamentos em ambientes virtuais antes do físico, reduzindo-se tempo e investimento necessários no processo de prototipação;
- 2020.4: integração de inteligências artificiais a tecnologias que já fazem parte do dia-a-dia de muitos humanos, como celulares e relógios;
- 2020.5: aumento da integração entre humanos e tecnologias e aumento da capacitação das pessoas trabalhadoras a fim de se adequarem ao mercado de trabalho;
- 2020.6: aumento na frequência com que dispositivos são capazes de interferir nas atividades humanas, tanto na vida pessoal quanto no trabalho, devido ao aumento da capacidade de armazenamento de dados e da velocidade de processamento;
- 2020.7: aplicações de IA no desenvolvimento de mídias como música, vídeo e jogos, tanto como ferramentas que facilitem etapas do processo quanto como automações de produções em certos contextos;

- 2020.8: aplicações de IA em segurança digital, em resposta à crescente capacidade de ataques cibernéticos devidos à própria IA;
- 2020.9: convivência de humanos com inteligências artificiais e robôs se tornando mais natural com o passar do tempo, desde a infância;
- 2020.10: aplicações de IA ao reconhecimento de imagem, inclusive na identificação de pessoas;
- 2021.1: continuidade das aplicações de IA na automação de atividades operacionais;
- 2021.2: maior adoção de ferramentas de segurança digital e gestão de operações baseada em IA;
- 2021.3: integração de aplicações de IA a aplicações de sensores conectados à internet, intensificando a produção e utilização de “objetos inteligentes”, que captam informações e reajam às mesmas de acordo com as inteligências artificiais implementadas em si;
- 2021.4: mais aplicações de IA à gestão de relacionamento com clientes através do marketing (similar a 2019.3);
- 2021.5: especialização das empresas fornecedoras de tecnologia em áreas de negócios mais restritas, devido à multitudine de aplicações possíveis em tantas áreas diferentes;
- 2021.6: integração de IA a sistemas de vigilância e registro;
- 2021.7: digitalização de serviços e atendimento através de IA, com necessidade potencializada pela pandemia de Covid-19.

As seis dimensões nas quais a IA se desenvolveu, como definidas no capítulo 3, continuam aparecendo em meio às tendências apontadas para os últimos anos. Na prática, cada uma destas pode ser enquadrada em pelo menos uma das dimensões, como mostrado na Tabela 1. Essa classificação, cabe destacar, leva em conta não apenas os resumos aqui dados, mas os materiais originais na revista.

Note-se que, embora a quantidade de tendências categorizadas em cada dimensão varie tanto, todas têm diversos exemplos para si, mostrando que são elementos recorrentes nas discussões quanto ao futuro da IA e dos próximos avanços possíveis. Ao mesmo tempo, muitas das tendências são um tanto similares àquelas integrantes das mesmas dimensões, como se fossem na prática reflexões daquela dimensão sobre contextos similares, levemente

distintos entre si. É possível que tais dimensões componham, portanto, uma ferramenta a partir da qual possamos vislumbrar mais tendências a partir de cada momento da história, como uma bússola de possibilidades.

**Tabela 1:** Comparação das tendências apontadas na revista Forbes às dimensões de desenvolvimento da IA

<b>Captação de informações</b>	<b>Representação de informações</b>	<b>Armazenamento de informações</b>
2020.1, 2020.4, 2020.10, 2021.3, 2021.6	2018.3, 2019.5, 2020.3, 2020.7, 2020.10	2018.2, 2018.5, 2020.6, 2021.6
<b>Velocidade de processamento</b>	<b>Modelagem de soluções</b>	<b>Capacidade de ação mecânica</b>
2019.2, 2020.4, 2020.6	2018.1, 2018.2, 2018.3, 2019.2, 2019.3, 2019.5, 2020.2, 2020.3, 2020.4, 2020.7, 2020.10, 2021.4	2019.2, 2020.4, 2020.6

Pode se observar nesta lista de tendências que as intenções dos negócios quanto à inteligência artificial não sofreram nenhuma alteração muito grande: automação de trabalhos, redução de custos, aumentos de efetividade e de qualidade continuam sendo interesses nítidos ao longo de todos os anos. Mas algumas observações podem ser feitas ainda assim quanto à forma e ao contexto de tais tendências e interesses no momento atual:

- Há uma maior integrabilidade de inteligências artificiais a outras tecnologias, de modo a se complementarem e viabilizarem em novos cenários e aplicações;
- Há também maiores utilizações de IA dentro de mais áreas de negócio e de pesquisa, o que potencializa o desenvolvimento de novas tecnologias e novos estudos; no caso da medicina e do combate à pandemia de Covid-19, por exemplo, o desenvolvimento acelerado de vacinas tem impacto direto sobre a sociedade e não apenas sobre negócios;
- Observa-se hoje que as culturas organizacionais são impactadas pela inserção de novas tecnologias, inclusive a IA, dentro dos processos e modelos dos negócios; a modificação cultural não é trazida nas tendências, porém, como um objetivo

por si só, mas como uma consequência que pode ser administrada e otimizada a fim de viabilizar uma maior integração das tecnologias à empresa; ou seja, ao menos quando falamos de tendências para o futuro da IA, tais mudanças culturais são um meio e uma consequência, mas não realmente uma finalidade;

- A questão do impacto da IA sobre trabalhos realizados por humanos é trazida na Forbes (2019.1, 2019.4, 2020.5, 2020.9, 2021.1, 2021.5 e 2021.7) em tom otimista, encarado como uma oportunidade para pessoas e empresas, mais do que como uma preocupação; essa questão carece de uma discussão mais aprofundada, a ocorrer no próximo capítulo;
- A cibersegurança aparece mais frequentemente (2018.4, 2020.8 e 2021.2) como tema relevante nesse contexto contemporâneo onde uma parcela muito maior da humanidade tem acesso a computadores pessoais e à internet; a temática da segurança pública já era relevante nas aplicações de IA que governos vinham fazendo há décadas, mas agora a temática da privacidade atinge mais diretamente as vidas de indivíduos que convivem com tecnologia.

### **A questão dos dados pessoais**

Em 1983, o Tribunal Constitucional Federal da Alemanha tomou o que pode ter sido a primeira decisão a reconhecer a proteção de dados como direito fundamental do indivíduo, proibindo a coleta de dados pessoais que não estivessem mascarados (sem possibilidade de identificação da pessoa) e não tivessem finalidade de coleta definida e declarada ao indivíduo (DLA Piper, 2021). Dessa forma as pessoas passavam a ter alguma autoridade sobre as informações referentes a si e para que estas seriam utilizadas dentro ou fora dos contextos tecnológicos, o que afetaria a possibilidade de aplicações de inteligência artificial sobre tais dados.

Entrou em vigor na Europa em 2018 a chamada *General Data Protection Regulation*, ou GDPR, regulamento sobre ações e cuidados necessários no tratamento de dados sobre pessoas com alguma nacionalidade europeia em todo o mundo. E em 2020 entrou em vigor no Brasil uma lei similar chamada *Lei Geral de Proteção de Dados Pessoais*, ou LGPD, que se aplica a pessoas de nacionalidade brasileira em território nacional. Com base em um relatório

do escritório DLA Piper (2021), mais de 70% do território global está hoje sujeito a regulamentos quanto à utilização de dados sobre pessoas.

Vale notar que as discussões sobre a GDPR e a LGPD foram iniciadas anos antes de entrarem em vigor, e a preocupação quanto à privacidade das informações individuais pode ser observada no mínimo a partir dos anos 80. Para mais, é importante considerar essas informações ao observar as tendências apontadas nos últimos anos tanto na revista Forbes quanto em outras não citadas aqui.

A alta capacidade de captura de dados através da internet e o grande volume de pessoas conectadas através de seus computadores são fatores facilmente visíveis como relevantes para entendermos por que a privacidade se tornou uma prioridade tão grande quando falamos em tecnologia na atualidade. Mas é também válido considerar que, décadas antes, como visto no capítulo anterior, governos já vinham utilizando tecnologias, inclusive a inteligência artificial, para fins de segurança e privacidade, e que já vinham sendo publicadas famosas ficções científicas sobre a inserção de robôs na sociedade humana. Não podemos aqui desconsiderar que estes fatores possam ter tido alguma relevância sobre a preocupação dos indivíduos com a utilização dos dados que dizem respeito a si próprio.

### **Algumas interações humano-máquina observadas**

Mas as preocupações quanto à inteligência artificial não se limitam ao uso de dados sobre pessoas, estendendo-se também às decisões tomadas por máquinas quanto à vida humana.

Nos últimos anos algumas unidades policiais no Brasil têm feito uso de tecnologias de reconhecimento facial sobre a população a fim de identificar pessoas suspeitas de crimes ou já incriminadas. Tais ferramentas operam em geral através da comparação de imagens de câmeras de vigilância a bancos de imagens provenientes de documentos, de modo a associar as imagens das câmeras às pessoas mais similares.

A proposta é que tais ferramentas sejam um apoio à localização de pessoas procuradas por crimes, mas há duas questões especialmente sensíveis a respeito disso. A primeira é que tecnologias de reconhecimento de imagem não têm resultados perfeitos, resultando ocasionalmente em a ferramenta apontar, de modo errado, que uma pessoa esteja entre as procuradas. A segunda questão é que tais tecnologias se enquadram nos tipos “sistemas que

pensam racionalmente” ou “sistemas que pensam como humanos” dentre as inteligências artificiais, ou seja, sua função é tentar identificar pessoas e dar uma resposta quanto a isso, mas as ações tomadas a respeito ficam por conta da polícia, e nessa etapa também podem ocorrer certos problemas.

A Secretaria de Segurança Pública da Bahia (SSP-BA) implementou em 2018 um destes sistemas (Palma & Pacheco, 2020) e, em setembro de 2019, ocorreu uma situação onde um homem de 25 anos foi reconhecido erroneamente pelo sistema e abordado pela polícia de forma violenta, com uma arma de fogo já apontada para si antes de qualquer questionamento. Entretanto, a mãe do homem estava presente e o defendeu, impedindo que a situação se agravasse ainda mais.

Na situação descrita, o ocorrido poderia claramente ter sido evitado ao se seguir um procedimento diferente diante do resultado gerado pela tecnologia, ou seja, não havendo por via de regra uma abordagem inicial violenta. Não se sabe no momento o quão precisos e sensíveis podem vir a ser os resultados de reconhecimentos faciais no futuro; mas, mesmo num cenário onde tais resultados fossem perfeitos, a interação entre a tecnologia e seus usuários estaria sujeita a falhas e, ainda mais num contexto policial como esse, a transferência de responsabilidade do ser humano para a máquina, havendo consciência sobre a possibilidade de uma nova falha no futuro, poderia ser vista por si só como irresponsável.

Falhas como essa não se restringem, contudo, às inteligências artificiais que se enquadram como “sistemas que pensam racionalmente”. Ocorreu em 2018 o primeiro acidente com um veículo autônomo envolvendo a morte de um pedestre (Roberto & Lopes, 2018). Havia uma motorista reserva num carro autônomo da empresa Uber sendo testado na ocasião, sendo função da motorista intervir caso o carro tomasse alguma ação indevida, mas, a princípio, este operaria por conta própria.

Inesperadamente uma pedestre atravessou uma rua escura e o carro a atropelou, sem ter sido divulgado o que exatamente fez com que o carro não freasse. É possível que as câmeras do carro não tenham reconhecido que uma pessoa atravessava a rua, seja pela rua estar escura ou pela imagem não ter sido identificada como a de uma pessoa; é também possível que, apesar de uma identificação efetiva, o carro não tenha *decidido* parar, a partir das lógicas construídas em seu algoritmo, ou não tenha decidido tomar essa ação a tempo.

Esse caso reacendeu uma discussão à época quanto à responsabilidade legal pela morte da pedestre, havendo algumas possibilidades: a empresa, as pessoas que implementaram a

tecnologia no carro, a motorista e a própria pedestre. Esta discussão não se restringia, porém, àquele caso específico, tornando-o um infeliz precedente para a futura responsabilização em casos similares no futuro.

Foi identificado mais tarde no mesmo ano (Somerville, 2018) que a motorista estava assistindo a um programa de televisão dentro do carro, sem prestar a devida atenção ao trajeto, o que atribuiu a ela homicídio culposo pela morte da pedestre. Se não fosse por tal elemento, porém, não se sabe como o caso teria evoluído e se alguém teria sido incriminado. Na prática não houve uma conclusão clara sobre como casos do tipo viriam a ser tratados.

Pode-se pontuar que um elemento que torna esta situação mais complexa de se avaliar é o tipo de inteligência artificial que estamos avaliando agora: ao invés de um “sistema que pensa racionalmente”, como na situação do reconhecimento facial para a SSP-BA, trata-se de um “sistema que age racionalmente”, tanto gerando respostas algorítmicas para problemas quanto tomando ações a partir de suas próprias conclusões.

Quando a ação a partir da tecnologia é tomada por um ser humano adulto e consciente sobre seus atos, pode ser mais intuitivo e simples responsabilizá-lo por tal ação, pois a tal pessoa cabe a decisão final sobre de fato realizar ou não a ação. Já quando a tecnologia toma uma ação por conta própria e temos tantas pessoas participando do contexto que permite à tecnologia tomar tal ação, não há apenas uma responsabilidade clara.

Vale questionarmos, mais ainda, o quanto uma pessoa que utiliza tecnologia e realiza ações baseadas nesta está realmente consciente sobre seus atos, afinal não necessariamente a pessoa usuária entende como a tecnologia opera e a lógica que a leva às suas conclusões. E se o desconhecimento sobre tal lógica leva a pessoa a seguir as sugestões da tecnologia sem levar em conta seus critérios pessoais, há uma transferência de responsabilidade sendo feita, mas que na prática é irreal: conscientemente ou não, a decisão quanto a “obedecer” a tecnologia foi tomada pela pessoa.

Então independentemente de qual abordagem para IA esteja sendo aplicada, dentre as quatro definidas por Russell & Norvig, há um complexo cenário que precisa ser avaliado, com múltiplos papéis e responsabilidades envolvidas. Na prática, tal cenário poderia ser avaliado inclusive antes de se decidir qual abordagem para IA utilizar, o que é um tema a ser tratado mais a fundo no próximo capítulo.

O’Neil (2016) trata ainda de outros casos que podemos discutir quanto à utilização de dados e à inteligência artificial, e podemos trazer alguns destes para agregar ao entendimento.



Há múltiplas ferramentas disponíveis no mercado hoje, inclusive no Brasil, que se propõem a apoiar empresas na seleção de pessoas para suas vagas através de inteligência artificial. As versões mais simples aplicam filtros pré-construídos às informações das pessoas candidatas de modo a entregar à área de recursos humanos uma lista menor de pessoas selecionadas para continuar o processo. Versões mais complexas, porém, se propõem a tentar selecionar pessoas que irão ficar por mais tempo na empresa ou que possuem, aparentemente, traços culturais mais próximos à cultura da organização.

Nestas últimas há algoritmos de maior complexidade que, em resumo, visam maximizar um certo indicador na hora da contratação - o tempo esperado de continuidade da pessoa na empresa ou um índice de “similaridade cultural”, por exemplo. Essa metodologia consiste em realizar uma previsão estatística para o indicador desejado de acordo com as características coletadas sobre a pessoa - ou seja, dentre as pessoas candidatas, as selecionadas são aquelas cujas características mais se assemelham às pessoas da empresa com os valores mais altos para o indicador em questão.

De todo modo, em ambos os casos há alguns riscos pouco comentados, a seguir:

- Quando se pensa em contratar pessoas que tenham um maior tempo esperado de continuidade, esse pensamento provém de uma lógica de redução de custos, pois o tempo de treinamento de uma nova pessoa para uma posição na empresa e o tempo de reposição da vaga após a saída de uma pessoa geram gastos e impactos sobre a estabilidade da produção. Mas há sempre a possibilidade de que a continuidade de pessoas em empresas não se dê pelo motivo esperado pela pessoa que desenvolveu o algoritmo: numa empresa que não se atente a comportamentos discriminatórios, por exemplo, o ambiente pode ser especialmente menos agradável e saudável para mulheres, o que poderia fazer com que o tempo esperado de continuidade de mulheres se tornasse menor que o de homens. Se isso não é levado em consideração pela tecnologia em sua análise, temos o risco de que as conclusões indiquem apenas homens para serem contratados numa empresa como essa. Isso de fato aumentaria o tempo médio de continuidade entre as pessoas contratadas, mas através de uma estratégia socialmente questionável.
- Já quando se pensa em contratar pessoas que apresentem traços culturais similares aos da empresa, a questão é ao mesmo tempo mais complexa, de um

ponto de vista da psicologia organizacional, e mais simples, de um ponto de vista tecnológico. Empresas podem enxergar sua cultura organizacional como um certo conjunto de valores e comportamentos julgados preferíveis ou aceitáveis. Por definição há duas sensibilidades no tema: como criar tal definição e como ela afeta as pessoas que trabalham ou não na empresa - ou seja, qual é o papel da cultura organizacional dentro daquela empresa especificamente. Já quanto à questão técnica, temos uma falha de modelagem na forma de se tentar mensurar a cultura individual: em geral o método utilizado é um cálculo comparando as respostas de funcionários da empresa para um questionário com as respostas dadas pelas pessoas candidatas ao mesmo questionário. Os problemas metodológicos envolvidos vão desde a seleção de perguntas feitas (e o quanto estas de fato têm a ver com a cultura organizacional definida) até questões mais complexas - e provavelmente incontornáveis - como a questão da linguagem utilizada nas respostas (pessoas que escrevem de modo parecido podem ter suas respostas vistas como parecidas, mesmo que demonstrem posturas opostas, dependendo de como o algoritmo operar) e a questão do quão pouco significativa é a comunicação escrita quando se querem observar valores e comportamentos de uma pessoa. Na prática o que a ferramenta pode estar fazendo nesse cenário é justamente selecionar pessoas que escrevem de forma parecida com as pessoas que já trabalham na empresa ou que conseguem imaginar quais respostas são esperadas pela empresa.

Em ambos os casos, para uma pessoa usuária da ferramenta e desavisada pode parecer que a inteligência artificial de fato selecione as pessoas adequadas, pois as limitações da tecnologia e seu funcionamento podem não ficar tão claros a princípio e, dado o quanto se fala hoje em inteligência artificial nas empresas, a ideia de utilizar uma pode ser ilusoriamente atrativa. Assim como no exemplo anterior sobre recrutamento de pessoas utilizando IA, a falta de compreensão sobre a tecnologia pode levar a uma inadequada transferência de responsabilidade para uma máquina cuja lógica é desconhecida.

Outro exemplo de grande repercussão nos últimos anos foi a utilização de dados de redes sociais para direcionamento da opinião pública, como no caso da *Cambridge Analytica*, empresa investigada por acusação de ter utilizado recomendações de conteúdos nas redes sociais com base em contas pessoais na rede social Facebook. Esses conteúdos teriam por

objetivo direcionar a opinião pública sobre temas como Donald Trump e Brexit, cumprindo certas agendas políticas através da transmissão especificamente selecionada de conteúdos dentro da rede social.

O'Neil aponta que tal tipo de manipulação continua sendo uma possibilidade nas redes sociais, que operam em geral através de conteúdos recomendados para seus usuários e têm a possibilidade de utilizar os dados pessoais a fim de mostrar conteúdos específicos que incentivem as pessoas a participarem e se engajarem mais na rede social em si, criando um ciclo perigoso de estímulos que incrementam o lucro da ferramenta e não necessariamente o bem-estar da pessoa usuária e de sua visão sobre o mundo.

### **Singularidade tecnológica**

Uma outra hipótese bem impactante para o futuro da IA está na chamada *singularidade tecnológica*. Introduzida por Neumann (Shanahan, 2015) e discutida mais a fundo na atualidade por Walsh (2018), o termo se refere à possibilidade de que em certo momento uma inteligência artificial com capacidade de aprendizado entre num ciclo de aprendizado desenfreado e consiga influência mundial, o que de diferentes formas poderia afetar a vida humana.

Essa preocupação surge a partir da observação de que as capacidades de consumo de informações, armazenamento, processamento e resolução de problemas complexos - ou seja, algumas das dimensões de desenvolvimento da IA definidas no capítulo 3 - têm crescido exponencialmente nas últimas décadas e não há um limite claro para estas no futuro. À medida que máquinas são construídas de modo a ter autonomia dentro de certos contextos, temos uma aparente perda de autonomia enquanto seres humanos.

As formas como a singularidade poderia se dar são múltiplas e hipotéticas, dependendo sempre da autonomia à qual as tecnologias tiverem acesso e da capacidade de integração entre múltiplas tecnologias. Tais hipóteses, contudo, costumam depender da forma como compreendemos as tecnologias hoje, o que faz com que seja importante repensá-las à medida do desenvolvimento tecnológico.

Também quanto aos impactos possíveis desta singularidade há múltiplos cenários possíveis, desde um mal funcionamento generalizado de tecnologias integradas a redes ao redor do mundo até uma tomada de ações mais graves por parte da tecnologia sem

consideração pela vida humana, o que, num cenário extremo, poderia de fato ser uma ameaça à espécie como um todo. Isso faz com que a singularidade tecnológica seja talvez o maior risco imaginado para o futuro da IA.

Os exemplos apresentados aqui são apenas recortes muito pequenos de todo o universo de aplicações de inteligência artificial, problemas envolvidos ou pontos de preocupação para o futuro. Esses recortes têm por objetivo atualizar a visão trazida no capítulo 4 e viabilizar a discussão sobre o futuro da IA no próximo capítulo.

## 6 O futuro da IA, suas inteligências e renúncias

Como notado no capítulo 5, a inteligência artificial abrange na atualidade muito mais aplicações, áreas de negócio e de pesquisa, mas, também, preocupações quanto ao seu futuro. Mas algumas das ferramentas definidas nos capítulos anteriores podem, comparativamente, ajudar na percepção sobre o quanto o cenário atual é diferente do observado desde o surgimento da tecnologia, como a seguir.

- Quando consideramos os quatro diferentes meios pelos quais permeou a história da IA, definidas no capítulo 4 - culturas & públicos, ciência & alquimia, empresas e governos -, temos uma distinção ainda válida quanto às suas posições quanto à tecnologia: a ciência tem feito maior uso da tecnologia em estudos, assim como empresas expandiram suas aplicações, não apenas em novas áreas de negócio, mas em áreas efetivamente criadas pela introdução da tecnologia; por parte do público em geral foi surgindo uma preocupação maior quanto ao uso de dados pessoais e ao possível desempenho causado pela automação dos trabalhos, e em resposta à primeira preocupação vêm surgindo legislações quanto a dados pessoais a fim de proteger a população, mas a questão da automação dos trabalhos permanece sem uma solução nítida;
- As seis dimensões do desenvolvimento da IA, propostas no capítulo 3, continuam sendo observadas na atualidade, com avanços notórios em todas elas; na prática, a preocupante hipótese da singularidade tecnológica (Shanahan, 2015, e Walsh, 2018) surge justamente da acelerada evolução da IA nessas dimensões;
- Já sobre as três estruturas da inteligência de Piaget (Macedo, 1980) temos que, como discutido no capítulo 4, desde a segunda metade do século XX existem tecnologias análogas às três e que surgiram na mesma sequência proposta por Piaget; se refizermos essa comparação na atualidade, o cenário não é muito diferente; podemos pontuar que há formas mais avançadas de aprendizado de máquina, análogas às regulações, e uma vastidão de operações diferentes, partindo das múltiplas aplicações que fazem uso de IA;
- Por fim, quanto às quatro formas de inteligência artificial de Russell & Norvig (1995), já era observável no histórico que cada uma tinha seu próprio espaço e

aplicabilidade, e o cenário permanece o mesmo; cabe ainda assim comentar, a partir de exemplos mostrados no capítulo anterior, que mesmo com uma inserção maior da IA em múltiplas atividades da vida humana hoje, tornando essa tecnologia mais presente no dia-a-dia, tende a não ser geralmente claro para as pessoas usuárias qual das quatro formas de IA está ocorrendo em cada aplicação da tecnologia, exceto para aquelas que tenham algum conhecimento técnico sobre ela; como no caso do recrutamento, por exemplo, a interpretação difusa sobre como a tecnologia opera e sobre qual papel ela de fato cumpre pode levar a uma transferência perigosa de responsabilidades do ser humano para a tecnologia; e a partir dessa difusão podemos rediscutir o papel da tecnologia neste capítulo final.

## **Inteligência e afetividade**

Uma questão relevante para Piaget nos seus estudos sobre a cognição e a inteligência é o papel que a afetividade cumpre em seu desenvolvimento e sua ocorrência (2005). Para ele, a afetividade pode, sim, ser observada em certos contextos, com múltiplos efeitos possíveis, podendo tornar uma atividade mais rápida, mais difícil ou mesmo mais complexa.

A afetividade humana pode permitir que uma pessoa identifique emoções em outras através da forma como elas a fazem sentir, por exemplo, assim como permitir que reconheça uma dor que alguém possa estar sentindo. Quando uma pessoa se depara com algo que seus sentidos possam captar, a interpretação daquela coisa não ocorre apenas num nível racional e descritivo, mas também num nível afetivo, ou seja, através de como aquela coisa faz com que a pessoa se sinta. E a afetividade não cumpre um papel apenas temporário durante a compreensão da coisa, pois mesmo que a racionalidade consiga descrever e interpretar logicamente aquilo apontado pelos sentidos, aquelas sensações continuarão fazendo parte da percepção do indivíduo sobre a coisa.

Quando construímos uma inteligência artificial com o objetivo de identificar emoções em pessoas, geralmente isso ocorre dentro de um certo processo: a máquina recebe acesso a um banco de imagens de pessoas, onde cada imagem tem uma emoção já atribuída a si, e ao receber uma nova imagem a máquina tenta mensurar o quão parecida aquela nova imagem é com as imagens já existentes no seu banco, dando como resposta a mesma emoção atribuída

àquelas imagens consideradas similares. A definição de uma imagem “similar” a outra pode ser construída de inúmeras diferentes formas, levando em conta desde as cores da imagem até as formas geométricas identificadas nas feições das pessoas, e tais lógicas para mensurar “similaridade” devem ser pré-construídas como parte da IA. Ou seja, até é possível que uma máquina consiga identificar emoções em pessoas através de suas imagens, mas isso se dá estritamente através de um cálculo matemático sobre a imagem, sem que nenhum elemento próximo à afetividade esteja envolvido.

Essa questão da afetividade pode, então, apontar a toda uma gama de operações que inteligências artificiais, na forma como existem hoje, seriam incapazes de realizar, assim como outras em que pode haver uma forma alternativa e estritamente racional de realizar. Esses diferentes cenários são bastante relevantes para que se discuta que papéis a IA pode não assumir para a sociedade humana.

### **Inteligência e moral**

Segundo Piaget (1974), ocorre na infância o processo de construção cognitiva do real para a criança. Durante esse processo são desenvolvidas as noções de objeto, espaço, tempo e causalidade. Ele percebe que, quando a criança demonstra ter construído a noção de objeto, demonstra também a construção da noção de si mesmo: no começo da vida a criança não demonstra - pelo menos não claramente - compreender que objetos continuam existindo de forma independente à sua percepção individual, mas a partir de algum momento essa compreensão se torna patente. E é neste ponto que a noção de si mesmo é perceptível do ponto de vista cognitivo, ao se posicionar no cosmos como um objeto entre outros objetos.

Essa noção também permite que o indivíduo reconheça outros indivíduos de algum modo análogos a si, o que depois permite as construções de si mesmo nos pontos de vista cultural e moral. O estudo de Piaget sobre a moral, continuado por Kohlberg (1994), Turiel (1983) e La Taille (2002), define inicialmente a moral como um sistema de regras construídas através da interação social, mas essa perspectiva é criticada como sendo demasiado racional. A partir daí surgem estudos mais preocupados com o aspecto afetivo da moralidade, e o próprio Piaget demonstra em seus trabalhos ao longo do tempo atribuir uma grande importância à dimensão afetiva para a compreensão da moral.

A partir do si mesmo cognitivo e da afetividade um indivíduo seria capaz de coordenar suas necessidades, interesses e pontos de vista junto de outros indivíduos, tendo assim a noção de si mesmo moral construída. Ou seja, a noção cognitiva de si mesmo e a afetividade são necessárias para o si mesmo moral e, portanto, para a moral em si.

Como comentado na seção anterior, máquinas como as conhecemos hoje são incapazes de processar afetividade, pois é algo alheio às suas naturezas puramente racionais. Mas, além disso, máquinas possuem uma separação cognitiva bem distinta de seres vivos: software e hardware, dois elementos que atuam em conjunto, mas não são exclusivamente dependentes. Se um software é colocado em outro hardware absolutamente idêntico, a inteligência artificial não está sendo afetada, pois todos os seus elementos, recursos e capacidades permanecem, na prática, os mesmos. Se simetricamente o software atrelado a um hardware é substituído por outro idêntico, o mesmo se aplica. Isso faz com que seja um tanto mais desafiador definir onde começa e termina o “si mesmo” da inteligência artificial, ao menos enquanto tentamos utilizar a mesma definição aplicada a seres humanos.

Mesmo uma IA cujo software identifique fisicamente onde está - e o que a compõe - o hardware atrelado a si não tem razões lógicas nem capacidade cognitiva para temer pela destruição do mesmo, a menos que seja construída e instruída para tanto, como no caso das Leis da Robótica propostas por Asimov (1941).

Ou seja, por si só uma máquina não tem capacidade para adquirir uma noção cognitiva de si mesma e, portanto, uma moral. Uma alternativa possível é que seja implementado um sistema de regras sobre as conclusões e ações da máquina, algo similar às próprias Leis da Robótica e àquela noção inicial de Piaget sobre a moralidade. Embora pudesse funcionar de um modo próprio, distinto do modo humano, isso poderia cumprir um papel de condição de contorno dentro do contexto das interações humano-máquina.

Condições de contorno são um elemento utilizado em problemas de otimização de sistemas dentro da matemática. São um conjunto de axiomas dentro dos quais a solução encontrada para o problema precisa se encaixar, sendo inválida caso contrário. Na prática, a resolução de uma otimização dentro de condições de contorno consiste em encontrar a solução possível que simultaneamente otimize o sistema e se enquadre nas condições definidas. Esse tipo de elemento já é utilizado em certos tipos de inteligência artificial, mas em geral numa versão numérica mais tradicional, e talvez seu uso possa ser expandido para outros tipos de condições, como as morais.



Há ainda um certo problema um tanto utilizado para se discutir a tomada de decisões onde há um impasse moral envolvido, talvez o problema mais famoso desta categoria: o dilema do bonde. Introduzido e aprofundado por Foot (1967) e Thompson (1976), respectivamente, o dilema trata de situações onde se aplica o chamado *princípio do duplo efeito*: em resumo, uma decisão é necessária dentro de um domínio de decisões possíveis, onde cada uma tem simultaneamente consequências consideradas positivas e outras consideradas negativas - no caso do bonde, a situação é um bonde já em movimento que deverá passar por uma dentre duas vias possíveis, onde estão dois grupos diferentes de pessoas, ferindo-as mortalmente.

Mais do que um dilema quanto a qual decisão deva ser tomada, o dilema trata também da situação em que a decisão se dá e do quanto ela pode interferir sobre a percepção humana a respeito das consequências e, portanto, sobre a tomada de decisão. Dessa forma, a tomada de decisões por pessoas pode ser inconsistente à medida de suas consequências.

Pode valer a pena levar esse princípio em consideração se pensarmos em colocar uma inteligência artificial para tomar uma decisão com conotações morais: se é possível que entre humanos não haja consenso quanto à melhor alternativa a ser seguida, o que seria da convivência humana com máquinas que tomassem tais decisões por nós? Mais ainda, devido à incapacidade de compreensão moral, seria mais difícil que uma máquina identificasse quando há consequências negativas e positivas a serem levadas em consideração. Talvez, como discutido na seção anterior, a máquina até possa identificar certas consequências negativas - como, num cenário hipotético, a possibilidade de uma pessoa ser ferida devido a uma ação da máquina -, só que isso não aconteceria com base em um julgamento moral, mas apenas lógico ou algébrico, e que deve ser antecipadamente instruído à máquina.

### **Inteligência artificial, cognição e criatividade**

Quando se trata de encarar como a inteligência artificial se dá na prática, trazendo à luz suas restrições e cuidados necessários ao invés de apenas levantar possibilidades encantadoras para o futuro da tecnologia, a obra “What Computers Can’t Do: The Limits of Artificial Intelligence” é uma das mais famosas, sendo Dreyfus (1972, 1991, 2007) um grande crítico da inteligência artificial no último século.

Como mostrado por Gomes (2018), Dreyfus retrata sua convivência com o departamento de engenharia do MIT durante o estudo sobre inteligência artificial que acontecia ali, partindo do trabalho de Heidegger (1927) para apontar que a abordagem utilizada pelo departamento tinha alguns pontos questionáveis.

A principal questão, que muito tem a ver com o presente trabalho, é que a inteligência artificial desde seus primórdios operava a partir de representações e operações determinadas anteriormente ao seu funcionamento, estando limitada às estruturas que sua ontologia embutida é capaz de compreender. Isso gera uma ruptura entre a máquina e o mundo em que ela opera, pois sua capacidade de processar qualquer informação recebida fica sujeita àquelas estruturas construídas previamente, ao contrário do que ocorre com seres humanos: antes de compreender o mundo à nossa volta, compreendemo-nos como parte dele, de maneira integrada e, a princípio, indistinta, para então identificarmos as distinções entre os outros e si mesmo (Piaget, 1974).

O mundo pré-existente à cognição artificial construída em uma máquina permanece como uma premissa àquela cognição e, portanto, essa não se constrói a partir do mundo, mas para ele e apenas para ele - num mundo diferente tal cognição não se sustentaria, e assim a inteligência artificial de nada serviria. A máquina construída enquanto ferramenta para o ser humano, mais ainda, não opera sobre o mundo a partir de sua existência nele, mas a partir da existência dos humanos a partir dos quais foi criada. A captação de novas informações do mundo pode levar uma inteligência artificial a alterar a medida que efetua sobre algo, mas não quais medidas ela é capaz de realizar: há paradigmas hoje indiscutíveis e provisoriamente irreversíveis, inerentes ao funcionamento da máquina, quando levamos em conta a forma como a tecnologia se dá hoje e como conseguimos avaliá-la. Esse aspecto da inteligência artificial - frequentemente presente em projetos de IA -, em resumo, é o que leva Dreyfus a chamá-la, criticamente, de *inteligência artificial heideggeriana*.

Embora o desenvolvimento do chamado aprendizado de máquina tenha permitido que inteligências artificiais captem informações do mundo e modifiquem suas conclusões e ações a partir daí, tais informações continuam a ser processadas através das mesmas estruturas existentes, como números diferentes sendo aplicados a funções matemáticas já desenhadas. E se quisermos que uma máquina produza funções nas quais aplicar os números, isso também é possível: nesse caso, implementamos na máquina alguma forma de *função geratriz* que, a partir dos inputs determinados, dê como resultado uma função.

De forma infinitamente escalável, mais e mais estruturas implementadas na máquina podem permitir operações algébricas cada vez mais complexas, e é este fenômeno que Dreyfus se refere em sua tese de que a solução para a IA heideggeriana é que ela se torne cada vez mais heideggeriana.

Ou seja, essa forma de inteligência artificial como a conhecemos parte da existência anterior de uma cognição ao invés de construir uma a partir da sua inserção no mundo (como fazem os seres humanos); mais ainda, a existência prévia dessa cognição através da ontologia interna pode viabilizar algum nível de interpretação sobre o mundo, mas sem viabilizar a expansão das próprias capacidades cognitivas. Assim, a compreensão substancial do mundo e do outro permanecem vazias, não obstante qualquer aprendizado, pois na prática a máquina só é capaz de compreender algo em nível representativo, definindo algo a partir das informações que possui e não de uma experiência própria quanto àquilo.

Para os fins deste trabalho, cabe entrelaçar essa crítica de Dreyfus à psicologia de Piaget, tão relacionada nos últimos capítulos à inteligência artificial. Apesar de estruturalmente o resultado das formas da inteligência ter similaridades para com as estruturas por trás da inteligência artificial, Piaget (1960) e Macedo (1980) abordam as estruturas também como etapas no desenvolvimento da criança. No capítulo 4 foi mostrado que é possível traçar uma analogia dessa sequência com a história da inteligência artificial, mas aqui podemos concluir que isso não se aplica às unidades concretas de inteligência artificial, como uma data máquina ou software, ou seja, uma tecnologia não é dotada de ritmos e vai desenvolvendo regulações e operações à medida do tempo. A analogia da inteligência piagetiana à inteligência artificial é possível unicamente pelo fato da IA, em si mesma, ter sido inteiramente desenvolvida como uma analogia.

Considerando isso, temos que as estruturas de Piaget nos ajudam a distinguir certos tipos de máquinas e inteligências artificiais, mas toda a questão do desenvolvimento infantil, sobretudo em relação ao estágio sensório-motor e à função simbólica, não apenas não é aplicável às máquinas como também representa uma ordem inversa à da construção de uma inteligência artificial, em que uma cognição inicial é implementada antes de qualquer interação com o mundo físico.

Essa diferença é relevante para se pensar na possibilidade de máquinas solucionarem problemas ainda não solucionados por seres humanos. Dado o seu funcionamento, toda inteligência artificial requer algum nível de instrução originado em um ser humano, e as

estruturas artificiais que vem a receber invariavelmente traçam um limite para suas capacidades. Ou seja, as estruturas cognitivas que podem existir em uma inteligência artificial são necessariamente um subconjunto daquelas que seres humanos já tenham sido capazes de conceber, traçando um limite à artificialização que perdurará pelo futuro, ao menos enquanto computadores continuarem funcionando como hoje funcionam. Qualquer definição de criatividade que não seja como uma simples recombinação de estruturas matemáticas fica portanto fora do alcance de uma IA.

### **Inteligência artificial e ética**

Após falarmos sobre as questões da afetividade, da moral e da cognição, há mais um conjunto de temas bem relevantes para entendermos as limitações e impactos da IA: a ética. Um termo definido de modos muito distintos, a depender do ambiente em que o estamos utilizando, mas que possivelmente em todos os seus casos possui alguma conexão com este trabalho.

Ribeiro (2018; 2019) considera que a substância, a essência, de tudo quanto é subsumido à palavra “ética”, de tudo quanto subjaz a qualquer situação em que se possa reconhecer um “dilema ético”, corresponderia aos sentidos que, etimológica e filologicamente, foi atribuído ao termo “ethos”. Nesta perspectiva, a autora apresenta uma definição para o *ethos humano* como “*uma relação natural (como que) de moldagem contínua e recíproca entre um lugar e a vida de seus habitantes*”. Nesta formulação, cumpre esclarecer que a autora considera que a vida se manifesta empiricamente como *comportamento* e como *subjetividade*.

Para compreender essa definição, podemos partir dos seus elementos constituintes:

- O *comportamento* se refere às ações humanas sujeitas a contingências que possam suprimi-las ou reforçá-las, como estudadas, p.e., pela área da análise do comportamento;
- A *subjetividade* se refere a tudo o que ocorre abaixo da *pele percebida* pelos indivíduos humanos, como estudado, p.e., pela psicanálise e pela psicobiologia;
- O *lugar* se refere ao espaço, tanto material quanto relacional e simbólico, em meio ao qual o comportamento e a subjetividade se dão, como estudado, p.e., pela geografia e pela arquitetura;

- A ideia de *moldagem*, nesse contexto, se refere ao fato dos três elementos se influenciarem mutuamente, de modo a se modificarem, tanto do ponto de vista material quanto do ponto de vista simbólico (ou seja, tanto do ponto de vista formal quanto de suas representações), em conjunto e ao longo do tempo; algo estruturalmente similar a uma série temporal na matemática, a um processo estocástico na estatística ou a um sistema de fluídos na mecânica; uma alteração mesmo que pequena em um dos elementos poderia, em consequência da relação ética, influenciar de diferentes maneiras todos os três, inclusive retroagindo sobre aquele próprio elemento, com o passar do tempo.

Essa definição evidencia o *ethos* como relação constitutiva da vivência humana, sempre presente a partir de seus elementos. Também evidencia o seu estudo como interdisciplinar (quicá, como admite a autora em questão, transdisciplinar), de modo que uma área de estudo concentrada apenas em um dos elementos tende a observar com clareza apenas um recorte de toda a relação – para o presente trabalho isso tende a ser um problema reconhecido, mas não tratado.

Nesta perspectiva, quando um novo comportamento é realizado por uma pessoa dentro de um grupo do qual ela faz parte e há uma aceitação quanto à tal comportamento, ele é positivamente reforçado, tanto do ponto de vista subjetivo do próprio agente quanto, através da aprendizagem observacional, das outras pessoas que o cercam, o que pode aumentar sua probabilidade de ocorrer novamente, seja a partir do mesmo agente, seja a partir de qualquer uma das pessoas integrantes daquele grupo; mesmo aquelas pessoas que talvez tivessem algum receio ou desaprovação quanto a adotar aquele comportamento, podem aos poucos modificar esse julgamento a respeito, o que pode ser visto como um impacto (influência; modificação; moldagem) sobre a subjetividade daquela pessoa; mais ainda, se aquele comportamento tem algum impacto sobre o lugar onde aquelas pessoas estão – como jogar lixo no chão de uma rua, por exemplo -, o lugar está sendo diretamente modificado também. De diferentes modos, a relação ética – o *ethos* – está sempre se realizando, está acontecendo, atualizando, como nesse exemplo.

Com a inserção de uma inteligência artificial num contexto humano, a tecnologia pode afetar tanto o comportamento humano quanto a subjetividade e o lugar, de acordo com seu funcionamento e natureza específicos; e, através do *ethos*, da *relação ética*, há sempre a possibilidade de atingir a todos os três elementos a partir do momento em que atinja pelo

menos um deles. Portanto, há que se considerar a IA de um ponto de vista ético, ou seja, que questões éticas surgem a partir da inserção da IA em uma situação real qualquer.

Retomando o exemplo dado no capítulo 5, sobre recrutamento de pessoas em empresas através de uma IA que estime o tempo de estadia na empresa para cada pessoa candidata, podem ser previstos impactos sobre os três fatores:

- Se homens costumam ficar mais tempo na empresa e, devido a isso, são selecionados pela tecnologia como candidatos mais adequados, a decisão está sendo transferida de uma pessoa para uma máquina e, mesmo que anteriormente fossem selecionados números similares de homens e mulheres, a população selecionada passa a ser majoritariamente masculina, impactando diretamente os comportamentos das pessoas que realizam o recrutamento, sem que talvez estas sequer percebam; mais grave aqui do que apenas haver um impacto sobre os comportamentos é a perspectiva de que a maior estadia de homens na empresa se deve a fatores discriminatórios e a gestão da empresa pode não ter clareza a respeito disso; ou seja, pensando-se apenas no elemento comportamental o impacto social possível já é nocivo e grave;
- Se a tecnologia é utilizada com a proposta de “selecionar as melhores pessoas para a empresa” - embora, como comentado originalmente, o critério aplicado é que sejam selecionadas as pessoas com maior tempo estimado de estadia na empresa -, há um descasamento entre o resultado da tecnologia e sua interpretação; esse descasamento pode, contudo, incentivar uma noção de que homens sejam de fato pessoas melhores para trabalhar na empresa, onde a definição de “pessoa melhor para trabalhar naquele lugar” não tem uma definição clara e uniforme entre as pessoas ali presentes; isso pode então incentivar percepções sexistas entre as pessoas dali, impactando sua subjetividade através de suas ideias, opiniões e imagens;
- Por último, há dois recortes em que podemos analisar o impacto sobre o lugar; numa perspectiva menor, observando apenas aquela empresa, ocorre uma mudança na população de pessoas empregadas, com ela se tornando cada vez mais majoritariamente masculina ao longo do tempo; já numa perspectiva maior, quando consideramos a inserção daquela empresa dentro de sua sociedade, temos que a priorização de contratação de candidatos homens impacta, mesmo

que numa pequena escala a princípio, as probabilidades de mulheres e homens conseguirem trabalho, o que por consequência intensifica as diferenças socioeconômicas entre pessoas de diferentes gêneros.

Ou seja, uma tecnologia inserida desavisadamente numa empresa que já possuía uma cultura organizacional discriminatória pode acabar agravando sua situação, desde perspectivas individuais até as mais globais.

O detalhamento desse exemplo serve para evidenciar que a inserção da inteligência artificial numa situação qualquer pode, através do ethos tal como definido por Ribeiro (2018; 2019), ter de fato consequências éticas – e, ao mesmo tempo, que é possível que esse esquema para observação e interpretação do ethos seja um caminho para tentarmos prever com maior clareza tais consequências, de modo a tomar decisões mais previdentes quanto ao emprego da inteligência artificial tendo em vista, em qualquer cenário, o critério da beneficência.

### **Três cenários possíveis para a IA**

Considerando as limitações e impactos da inteligência artificial discutidos até aqui, é possível reconhecer diferentes contextos para sua aplicação, alguns onde a tecnologia realiza tarefas exatamente da mesma forma como humanos a realizariam e outros em que não é bem esse o caso, como em operações que envolvam em humanos alguma questão afetiva, moral ou criativa.

Pode ser válida, para viabilizar avaliações de aplicações de IA assim como futuras discussões, a proposta de uma nomenclatura para esses possíveis cenários. Levando em conta portanto os temas agrupados ao longo do trabalho, temos ao menos três deles a princípio:

- Emulação: realização artificial da operação através de passos análogos àqueles que um ser humano seguiria para realizá-la, sem nenhuma perda cognitiva perceptível; tende a ser possível em operações dependentes estritamente de raciocínio lógico e algébrico, sem envolvimento necessário de elementos como afetividade, moralidade e criatividade para cumprir seu objetivo - mesmo levando em conta que, como comentado anteriormente, a afetividade está frequentemente entrelaçada à inteligência natural humana, e esse aspecto não seria replicado artificialmente; exemplos podem abranger desde problemas matemáticos mais simples como equações de primeiro grau até outros problemas

bem mais complexos como a automação da contabilidade de uma empresa; é até possível afirmar que, nesse cenário, há a possibilidade de máquinas serem capazes de realizar a operação mais rapidamente e com menor chance de erro do que um humano o faria, devido à natureza lógica e algébrica da máquina possuir os requisitos necessários para a realização e não ser influenciada pela afetividade que, no caso desse tipo de problema, não é uma prioridade;

- Simulação: realização artificial da operação através de passos distintos daqueles seguidos por um ser humano, podendo haver um resultado suficiente porém qualitativamente inferior em algum aspecto; um exemplo é o reconhecimento de emoções em pessoas através de suas imagens, como descrito anteriormente, em que uma IA pode chegar a um resultado quantitativamente suficiente, mas por meio de um método sujeito a falhas às quais um humano capacitado e disponível não estaria; outro exemplo é a seleção de candidatos através de uma métrica que em teoria representa “a pessoa mais adequada a trabalhar naquela empresa”, onde essa métrica é um modelo numérico consistente e com um fundamento financeiro - a redução de custos -, mas que ignora outros temas criticamente relevantes para a psicologia organizacional e sujeita o negócio a uma alienação de sua própria cultura; em contextos onde a afetividade, a moral ou a criatividade estejam intrinsecamente envolvidas no problema a ser resolvido provavelmente não há uma possibilidade de emulação ser construída, mas podem haver alternativas para simulação; elas requerem, contudo, um maior cuidado quanto à sua construção e introdução no meio humano, pois o seu funcionamento diferente do natural pode ocasionar algum impacto através do ethos;
- Abstenção: incapacidade de realizar uma operação artificialmente, devido às limitações tecnológicas ou situacionais; são os casos em que não há uma alternativa de emulação ou simulação com resultados satisfatórios à necessidade e às condições de contorno, resultando na decisão de não se aplicar inteligência artificial àquela situação.

Note-se que há situações em que são possíveis múltiplas abordagens do problema, podendo essas abordagens estarem em cenários diferentes dentre os três descritos. Isso é justificado pelos cenários se referirem aos caminhos selecionados e não somente às situações



em que estão inseridos e nas quais se baseiam - uma mesma situação pode ter tanto emulações quanto simulações possíveis, cabendo uma escolha humana sobre qual alternativa implementar. Se considerarmos os impactos éticos que cada um dos três cenários pode ter, podemos tomar uma decisão mais consciente e socialmente responsável, inclusive a decisão de renunciar à aplicação de inteligência artificial sempre que se verifique que, na operação inteligente natural dos seres humanos para a resolução de um determinado problema, estão implicados aspectos comportamentais e/ou subjetivos – em particular a afetividade, a moralidade e a criatividade – que não sejam passíveis, por limitações tecnológicas que podem ser temporárias ou definitivas, de emulação e/ou simulação.

### **Uma proposta de avaliação precaucionária**

Por fim, tomando o esquema para interpretação do ethos humano e os três cenários propostos para a aplicação de inteligências artificiais, podemos construir um método inicial para avaliação do impacto sobre o ethos que cada possível aplicação seria capaz de causar, de modo a antecipá-lo e levá-lo em consideração na decisão sobre onde e como utilizar, ou não, a tecnologia.

O processo se resume a considerar os três elementos conectados pelo ethos - comportamento, subjetividade e lugar - junto da compreensão sobre a partir de que a tecnologia opera e como se dá sua atuação para, de modo recursivo, mapear a possível alteração causada pela aplicação.

A realização deste método requer alguma compreensão das questões teóricas delineadas acima acerca do ethos humano, seus elementos componentes e as relações entre eles, bem como alguma experiência quanto às situações, vicissitudes e processos típicos da interação humana, portanto deve ocorrer a partir de pessoas familiarizadas com estes assuntos. Mais do que compreender definições desses elementos, é imprescindível compreender a complexidade em que estão inseridos, pois possuem interações e nuances frente a múltiplas temáticas que podem estar envolvidas em cada situação. Nesse sentido, esses três elementos podem ser tomados mais como pontos de partida do que como objetivos. O despreparado da pessoa que utilizar essa ferramenta pode levar a perigosas simplificações analíticas de situações e dos riscos nelas presentes.

Considerando a criticidade desse embasamento teórico para uma plena utilização dessa ferramenta, surge um desafio prático: quem deveria realizar essa avaliação? É comum que empresas possuam áreas dedicadas ao acompanhamento de temas como governança e compliance, assim como centros de pesquisa tenham entidades como comitês de ética ou conduta. Isso não significa que nesses locais se encontrem pessoas capacitadas a utilizar esta ferramenta, mas sim pessoas encarregadas de avaliar iniciativas em andamento a fim de tentar garantir algum padrão qualitativo. E a própria definição desse padrão pode ser, por si só, um segundo desafio a ser explorado. De todo modo, esses dois desafios parecem ser intrinsecamente codependentes.

Quando se discute legalmente a questão dos dados pessoais, como por exemplo no caso da Lei Geral de Proteção de Dados Pessoais no Brasil, eventos já ocorridos de utilização desses dados são levados à discussão, sejam eles “desejáveis” ou não. O passado é, nesse contexto, um referencial concreto, que pode até ser revisto e reinterpretado a partir de novas informações, mas que não cabe ignorar a princípio. Possibilidades e interesses sobre o futuro podem ser também levados em consideração, mas cumprindo um papel diferente - um papel próprio de possibilidades, condicional, incerto e em aberto. Essa mesma dinâmica pode se aplicar à construção dos chamados códigos de conduta ou de ética, seja em empresas ou centros de pesquisa, assim como pode se aplicar à definição de um padrão qualitativo para a ferramenta aqui em discussão.

Embora a princípio pudesse se imaginar que o próprio código de conduta ou de ética viesse a servir como um referencial para o padrão qualitativo na aplicação dessa ferramenta em um lugar, há o risco de que tais documentos levem apenas em consideração uma perspectiva do que deve ou não acontecer dentro daquele lugar, sem necessariamente considerar a inserção desse lugar em um outro, maior, ou o impacto do que acontece ali sobre outros lugares. Ou seja, as pessoas de uma empresa podem seguir à risca as normas de governança desta empresa, mas essa afirmação não implica que haja ou não - ou qual seja - algum impacto sobre pessoas fora da empresa ou sobre o meio ambiente, por exemplo. Mais ainda, a classificação de um impacto como “aceitável” ou não é mais complexa do que seria possível desenvolver neste trabalho.

Com isso, temos que pode ser necessária a definição de um novo papel nas discussões sobre implementar inteligências artificiais - o papel de avaliador. Levando em conta que o conhecimento necessário para uso adequado da ferramenta foge dos conhecimentos

envolvidos nas atividades de desenvolvimento computacional e de governança ou compliance, parece inevitável que uma terceira pessoa realize a avaliação. E, considerando que a dificuldade - talvez, até mesmo, infactibilidade - de se definir algum padrão qualitativo a partir do qual se possa “aprovar” a implementação da tecnologia, é importante que o embasamento e a conclusão dessa avaliação sejam contextualizados, detalhados e explicados como possível, a fim de se manter a maior transparência possível.

Cumpra deixar claro que o objetivo do processo avaliativo aqui proposto não é realizar uma predição exata do impacto efetivo, mas uma previsão razoável, um mapeamento de impactos possíveis, cuja identificação antecipada viabilize uma intervenção orientada no sentido de evitar malefícios e promover benefícios ao pleno desenvolvimento humano individual e coletivo.

Etapa 1: descrever ou ilustrar a situação inicial e o problema que se considera resolver através de inteligência artificial; se possível, tentar categorizar as informações incluídas como itens de um dos elementos do ethos - comportamento, subjetividade ou lugar (em todas as dimensões abarcadas por tais categorias); a essa figura inicial chamaremos “retrato 0”, ou  $R_0$ .

Podemos descrever  $R_0$  ontologicamente como a seguir:

$$R_0 = \{C_0, S_0, L_0\}$$

$$C_0 = \{c_{0,1}, \dots, c_{0,x}\}$$

$$S_0 = \{s_{0,1}, \dots, s_{0,y}\}$$

$$L_0 = \{l_{0,1}, \dots, l_{0,z}\}$$

Nessa definição, temos que:

- $c_{n,a}$  é o a-ésimo elemento da descrição dos comportamentos daquele contexto no momento  $n$ ; os elementos  $S_{n,a}$  e  $L_{n,a}$  são análogos, mas referentes à subjetividade e ao lugar, respectivamente;
- $C_n$  é o conjunto de elementos  $c$  no momento  $n$ , descrevendo por completo o comportamento no momento  $n$ ; os conjuntos  $S_n$  e  $L_n$  são análogos, mas referentes à subjetividade e ao lugar, respectivamente;
- $R_n$  é o retrato completo no momento  $n$ , dado pela composição dos conjuntos  $C_n$ ,  $S_n$  e  $L_n$ ; no caso, o conjunto  $R_0$  anterior se refere ao momento 0.

Etapa 2: a partir do retrato 0, considerar uma possível solução para o problema e inserir seu embasamento e sua atuação dentro do esquema, criando uma nova figura à qual chamaremos retrato 1 ou  $R_1$ ; onde:

- o embasamento nesse contexto se refere àquilo que for tomado como entrada para a inteligência artificial; se for utilizado um histórico de dados contendo decisões tomadas por pessoas quanto àquele mesmo problema, então os comportamentos humanos do passado irão influenciar as decisões da máquina e isso é bem relevante para essa análise, devendo ser incluído no novo retrato como um novo elemento  $c_{x+1}$ ; caso o espaço físico seja uma fonte de dados para um robô que irá se movimentar através dele, então o lugar é também um embasamento relevante, podendo ser um novo elemento  $l_{x+1}$ ;
- a atuação dentro do esquema se refere a classificar a solução dentro de uma das quatro formas de IA definidas por Russell & Norvig (1995), principalmente na dicotomia pensar/agir, ou seja, avaliar se a tecnologia resulta em uma resposta dada a uma pessoa dentro do sistema ou se ela por si só toma uma ação sobre o sistema a partir de sua conclusão; essa distinção é bastante relevante para a identificação dos impactos e suas possíveis contenções.

Para o retrato 1 temos então:

$$R_1 = \{C_1, S_1, L_1\}$$

$$C_1 = \{c_{1,1}, \dots, c_{1,m}\}$$

$$S_1 = \{s_{1,1}, \dots, s_{1,n}\}$$

$$L_1 = \{l_{1,1}, \dots, l_{1,o}\}$$

Cabe comentar que os números  $m$ ,  $n$  e  $o$  de elementos nos conjuntos  $C_1$ ,  $S_1$  e  $L_1$  podem ser iguais ou não aos números  $x$ ,  $y$  e  $z$  de elementos nos conjuntos  $C_0$ ,  $S_0$  e  $L_0$ . Não precisa haver restrição quanto a isso.

Etapa 3: comparando os retratos 0 e 1, considerar as diferentes formas como a tecnologia poderia afetar os três elementos do ethos a partir de sua inserção, utilizando perguntas como:

- A atuação (seja pensamento ou ação) da solução pode reforçar, punir ou extinguir algum comportamento, ou gerar, modelar, um novo comportamento na situação?
- A atuação da solução pode alterar as motivações e os focos de atenção seletiva dos envolvidos na situação? De que modo isto poderia repercutir na senso-percepção, conhecimento ou opinião das pessoas sobre algum tema?
- A atuação da solução altera de algum modo o lugar, seja física ou simbolicamente?
- A resolução do problema tem impacto direto sobre algum dos três fatores? Como? Por exemplo, a solução alteraria a forma de disseminação (e o alcance, e a seletividade deste alcance) dos estímulos reforçadores e/ou punitivos, das informações, dos conhecimentos e opiniões? Alteraria a dinâmica de poder das pessoas envolvidas? Alteraria, no caso de uma aplicação de tecnologia em uma empresa, o conteúdo dos cargos e funções de cada “funcionário”? Isto poderia ter, na subjetividade, algum impacto sobre a motivação, o senso de responsabilidade, a autorrealização, ou outros fenômenos?

As respostas a tais perguntas, e outras cuja conveniência pode ser definida *ad hoc*, devem ser incluídas na montagem de um retrato 2. E não precisam nem devem se restringir aos elementos já mapeados nos retratos 0 e 1, sendo possível encontrar novos itens relevantes justamente por conta das perguntas sendo feitas. Dado isso, é razoável incluir novos itens pertinentes aos retratos 0 e 1, o que ajuda a evidenciar a moldagem ao longo do tempo. O retrato 2 é, em linhas gerais, uma representação dos impactos diretos da inserção da solução naquele meio situacional.

Ontologicamente temos que as diferenças entre os retratos 0 e 1 podem ser representadas como:

$$\begin{aligned}
 D_{0,1} &= \{C_{0,1}, S_{0,1}, L_{0,1}\} \\
 C_{0,1} &= (C_0 - C_1) \cup (C_1 - C_0) \\
 S_{0,1} &= (S_0 - S_1) \cup (S_1 - S_0) \\
 L_{0,1} &= (L_0 - L_1) \cup (L_1 - L_0)
 \end{aligned}$$

Definindo  $D_{0,1}$  dessa forma ficam contemplados quaisquer elementos que estejam presentes em apenas um dos dois momentos. A finalidade dessa definição é facilitar a análise a seguir:

$$R_{n+1} = E(R_n, D_{n-1,n})$$

$$R_2 = E(R_1, D_{0,1})$$

$$R_2 = \{C_2, S_2, L_2\}$$

$$C_2 = \{c_{2,1}, \dots\}$$

$$S_2 = \{s_{2,1}, \dots\}$$

$$L_2 = \{l_{2,1}, \dots\}$$

A relação  $E(R_n, D_{n-1,n})$  aqui apresentada é uma analogia à moldagem do ethos, com possibilidade de modificação sobre todos os elementos que compõem  $R_{n+1}$ . Não se trata de uma função no sentido algébrico formal, mas apenas de uma relação no sentido epistemológico: a partir de  $R_n$  e  $D_{n-1,n}$  se gera uma imagem de resultados possíveis, sendo  $R_{n+1}$  essa imagem e não apenas um de seus resultados.

Etapa 4: avaliar se o retrato 2, levando em conta tanto a transformação esperada pela resolução do problema quanto a moldagem gerada através do ethos, é preferível em relação ao retrato 0; caso haja um padrão qualitativo adotado para ajuizamento dos resultados da ferramenta - levando em conta as problemáticas apontadas no início desta seção - pode contribuir para esta avaliação; com isso, temos três casos:

- caso a resposta seja positiva, seguir para a etapa 5;
- caso a resposta seja negativa, mas havendo uma outra solução tecnológica possível para o problema em questão, deixar de lado os retratos do 1 adiante, mantendo apenas o retrato 0, e retornar à etapa 1;
- caso não haja outra solução possível para o problema, deve-se optar pelo cenário de abstenção quanto a ele, não fazendo uso de inteligência artificial.

Etapa 5: repetir a etapa 3, mas levando em consideração os últimos dois retratos criados por esse método ( $R_{n-1}$  e  $R_n$ , sendo  $n$  o número do retrato mais recente antes da etapa atual) e

também a possibilidade de que as diferenças entre tais retratos ( $D_{n-1,n}$ ) possam ocasionar novas moldagens em algum dos fatores; ou seja, enquanto na etapa 3 se consideram apenas os impactos diretos da resolução do problema e da forma de atuação da tecnologia, aqui se consideram as últimas moldagens ocorridas, independentemente de sua origem, permitindo que se observem os impactos dos três elementos uns sobre os outros; o resultado dessa etapa deve ser um retrato  $n+1$ .

$$\begin{aligned}
 D_{n-1,n} &= \{C_{n-1,n}, S_{n-1,n}, L_{n-1,n}\} \\
 C_{n-1,n} &= (C_{n-1} - C_n) \cup (C_n - C_{n-1}) \\
 S_{n-1,n} &= (S_{n-1} - S_n) \cup (S_n - S_{n-1}) \\
 L_{n-1,n} &= (L_{n-1} - L_n) \cup (L_n - L_{n-1})
 \end{aligned}$$

$$R_{n+1} = E(R_n, D_{n-1,n})$$

$$\begin{aligned}
 R_{n+1} &= \{C_{n+1}, S_{n+1}, L_{n+1}\} \\
 C_{n+1} &= \{c_{n+1,1}, \dots\} \\
 S_{n+1} &= \{s_{n+1,1}, \dots\} \\
 L_{n+1} &= \{l_{n+1,1}, \dots\}
 \end{aligned}$$

Etapa 6: deve ser feita uma avaliação quanto ao cenário a fim de decidir sobre a continuidade do método, como a seguir:

- caso os retratos  $n$  e  $n+1$  sejam idênticos e  $D_{n-1,n}$  seja vazio, ou seja, não havendo nenhuma nova moldagem prevista entre os últimos retratos, temos que  $R_{n+1} = R_n$  e o sistema chegou a um estado de equilíbrio, onde novas moldagens não seriam mais previsíveis; pode se prosseguir à etapa 7;
- caso os últimos  $x$  retratos sejam idênticos aos outros  $x$  retratos anteriores a estes, ou seja,  $(R_{n-2x+2}, \dots, R_{n-x+1}) = (R_{n-x+2}, \dots, R_{n+1})$  para algum  $x$  natural, temos uma convergência periódica no método, o que indica que moldagens continuariam ocorrendo, mas dentro de um certo ciclo sem fim observável; pode se prosseguir à etapa 7;
- caso o retrato  $n+1$  já seja claramente menos adequado que o retrato 0 e haja outra solução possível, deixar de lado os retratos do 1 adiante, mantendo apenas o retrato 0, e retornar à etapa 1;

- caso o retrato  $n+1$  não seja claramente menos adequado que o retrato 0, retornar à etapa 5 e continuar o processo.

Etapa 7: o método deve ser encerrado e os retratos devem ser analisados; deve se avaliar se os últimos retratos gerados são adequados - considerando o padrão qualitativo adotado, caso exista, ou os critérios declarados pela pessoa conduzindo a avaliação -, tendo por fim os cenários a seguir:

- caso a resposta seja positiva, seguir com a solução, encerrando esse processo, mas tendo os devidos cuidados para que o impacto seja direcionado;
- caso contrário, mas havendo outra solução possível, deixar de lado os retratos do 1 adiante, mantendo apenas o retrato 0, e retornar à etapa 1;
- caso não haja outra solução possível para o problema, deve-se optar pelo cenário de abstenção quanto a ele, não fazendo uso de inteligência artificial.

Cabe observar que, no cenário de emulação, é possível que a inserção da máquina não represente uma grande mudança para o sistema como um todo, pois a operação estaria sendo realizada de maneira bem similar por uma pessoa ou uma máquina. Os maiores impactos tendem, portanto, a ocorrer num cenário de simulação.

Esses passos, como se pode notar, viabilizam uma análise dos comportamentos, subjetividades e lugares ao longo do tempo dentro de cenários hipotéticos mapeados para o futuro. Uma aplicação adequada deste método pode exigir um conhecimento prévio sobre o funcionamento da tecnologia assim como conhecimentos em psicologia social e nos temas abarcados relevantes para a análise. Apesar da dificuldade, seu sucesso pode resultar em uma antecipação de impactos negativos do uso de inteligência artificial ou em uma maior confiança quanto a esse uso, permitindo uma integração da tecnologia à sociedade humana mais consciente sobre as consequências possíveis.



## Conclusão

Ao longo do estudo sobre a temática da inteligência artificial, muitas questões são possíveis quanto ao seu efeito sobre a sociedade e a psique humana. As duas tomadas inicialmente como propostas neste presente trabalho puderam ser respondidas, de fato.

Múltiplas aplicações de inteligência artificial já existem hoje em nosso dia a dia e muitas outras são pensadas para o futuro, mas são classificáveis dentro de quatro formas propostas por Russell & Norvig de acordo com seu funcionamento e sua atuação sobre o meio, assim como pode ser feita uma analogia para com as estruturas de inteligência de Piaget. Tais classificações podem nos dar uma visão geral sobre o já existe hoje em termos de IA, assim como as noções de afetividade, moral e criatividade aparecem como os maiores limites para a tecnologia, indicando um subdomínio de operações que a IA, na forma como existe hoje, seria incapaz de emular.

Quanto aos interesses que direcionaram a história da IA, ficaram definidos quatro meios pelos quais a tecnologia passou, cada um dotado de seus próprios interesses: Cultura & Públicos, Ciências & Alquimia, Empresas e Governos. Os interesses específicos foram se alterando para cada um destes meios, mas há uma consistência dentro de cada um deles ao longo do tempo.

Além destas respostas, surgem três instrumentos úteis para o desenvolvimento e a utilização de inteligência artificial na atualidade.

O primeiro é um esquema de seis dimensões para acompanhar a evolução da IA ao longo do tempo - captação, representação, armazenamento, velocidade, modelagem de soluções e capacidade de ação mecânica. Esse esquema também pode ser utilizado por um grupo de pesquisa ou empresa para indicar onde exatamente está cada avanço tecnológico obtido.

O segundo é uma classificação de cenários onde podemos aplicar IA, podendo ser uma emulação caso a operação seja estritamente lógico-algébrica, uma simulação caso a máquina não disponha dos atributos necessários para executar a operação de maneira análoga a um ser humano ou uma abstenção caso não seja razoável aplicar tecnologia àquele contexto. Esses termos podem servir de base para uma compreensão mais clara sobre o que a tecnologia realmente está realizando ou não em cada cenário.

O terceiro é um método para tentar antecipar o impacto ético causado pela inserção de uma tecnologia no contexto humano, a partir de onde podemos decidir uma melhor abordagem para a solução de um problema através da IA, entendendo-a como uma emulação ou simulação, ou decidir que o cenário mais adequado é nos abstermos de aplicar IA sobre aquele contexto.

Temos assim uma visão geral sobre o histórico dessa tecnologia assim como novos instrumentos para apoiar no direcionamento do seu futuro. Mais do que temer o que pode vir a acontecer pela utilização inadequada da tecnologia, podemos nos apoderar mais dela a fim de garantir o melhor impacto possível: um que contribua de fato ao desenvolvimento humano.

## Bibliografia

1. ASIMOV, Isaac. **Liar!**, em **Astounding Science Fiction**, edição de maio de 1941. Disponível em [https://archive.org/stream/Astounding\\_v27n03\\_1941-05#page/n41/mode/2up](https://archive.org/stream/Astounding_v27n03_1941-05#page/n41/mode/2up). Último acesso em 08/07/2020.
2. BORGES, Jorge Luis. **El Otro, El Mismo**. Em “Borges: Obras Completas”. Buenos Aires: Emecé Editores, 1974.
3. BUCHANAN, Bruce. **A (Very) Brief History of Artificial Intelligence**, em **AI Magazine**, edição 26. 2005.
4. **Data Protection Laws of the World**, por **DLA Piper**. Disponível em <https://www.dlapiperdataprotection.com/>. Último acesso em 10/06/2021.
5. DREYFUS, Hubert. **Being-in-the-World: A Commentary on Heidegger’s Being and Time, Division I**. Cambridge: MIT Press, 1991.
6. DREYFUS, Hubert. **What Computers Can’t Do: The Limits of Artificial Intelligence**. 1972. Disponível em <https://archive.org/details/whatcomputerscan017504mbp/page/n7/mode/2up> . Último acesso em 08/07/2020.
7. DREYFUS, Hubert. **Why Heideggerian Artificial Intelligence failed and how fixing it would require making it more Heideggerian**, em **Philosophical Psychology**. Londres: 2007.
8. DURAND, Gilbert. **Champs de l’Imaginaire**. Ellug, 1996. Tradução por Maria João Batalha Reis.
9. DUSKA, Ronald & WHELAN, Mariellen. **O desenvolvimento moral na idade evolutiva: um guia a Piaget e Kohlberg**. São Paulo: Loyola, 1994.

10. FAUR, Mirella. **Ragnarök: o Crepúsculo dos Deuses**. São Paulo: Cultrix, 2011.
11. FEIGENBAUM, & FELDMAN. **Computers and Thought**. New York: McGraw-Hill, 1963.
12. FLEMING, Nic. **Machine that can taste coffee invented**, no jornal **The Telegraph**. 2008. Disponível em [telegraph.co.uk/news/science/science-news/3324864/Machine-that-can-taste-coffee-invented.html](http://telegraph.co.uk/news/science/science-news/3324864/Machine-that-can-taste-coffee-invented.html) . Último acesso em 08/07/2020.
13. FOOT, Philippa. **The Problem of Abortion and the Doctrine of the Double Effect**, na revista **Oxford Review**, número 5 Oxford: Basil Blackwell. 1967.
14. GOMES, Rodrigo Benevides Barbosa. **Hubert Dreyfus e Martin Heidegger: Representação e Cognição**, na revista **Kinesis**, volume X, em 22/07/2018. Disponível em <https://revistas.marilia.unesp.br/index.php/kinesis/article/view/8070>. Último acesso em 08/07/2020.
15. HEADRICK, Daniel. **Technology: A World History**. Oxford: Oxford University Press, 2009.
16. HEIDEGGER, Martin. **Ser e Tempo**. Rio de Janeiro: Editora Vozes, 2005. Publicação original em 1927.
17. IDEL, Moshe. **Golem: Jewish Magical and Mystical Traditions on the Artificial Anthropoid**. Albany: State University of New York Press, 1990.
18. JUNG, Carl. **Alchemical Studies**, em **The Collected Works of C. G. Jung**. Princeton: Princeton University Press, 1968.

19. KIRAM, Prem. **Three Top AI Trends To Watch This Year**, na revista **Forbes**. 2021. Disponível em <https://www.forbes.com/sites/forbestechcouncil/2021/03/16/three-top-ai-trends-to-watch-this-year/>. Último acesso em 10/06/2021.
20. LA TAILLE, Yves de. **O Sentimento de Vergonha e suas Relações com a Moralidade, em Psicologia: Reflexão e Crítica**, edição 15, 13-25. Porto Alegre: Springer, 2002.
21. MACEDO, Lino de. **As estruturas da inteligência, segundo Piaget: ritmos, regulações e operações**. Rio de Janeiro: 1980.
22. MARR, Bernard. **The Top 10 Artificial Intelligence Trends Everyone Should Be Watching In 2020**, na revista **Forbes**. 2020. Disponível em <https://www.forbes.com/sites/bernardmarr/2020/01/06/the-top-10-artificial-intelligence-trends-everyone-should-be-watching-in-2020/>. Último acesso em 10/06/2021.
23. MCCORDUCK, Pamela. **Machines Who Think**. Natick: A. K. Peters Ltd., 2004. Publicação original em 1979.
24. MCCULLOCH, Warren S. & PITTS, Walter H.. **A Logical Calculus of the Ideas Immanent in Nervous Activity**, em BODEN, Margaret A., **The Philosophy of Artificial Intelligence**. Oxford: Oxford University Press, 1990. Publicação original de 1943.
25. MINSKY, Marvin. **Semantic Information Processing**. Cambridge: MIT Press, 1968.
26. NEUMANN, John von. **First Draft of a Report on the EDVAC**. University of Pennsylvania, 1945.
27. NEWMAN, Daniel. **4 AI Trends Set To Accelerate In 2021**, na revista **Forbes**. 2020. Disponível em <https://www.forbes.com/sites/danielnewman/2020/11/18/4-ai-trends-set-to-accelerate-in-2020/>. Último acesso em 10/06/2021.

28. NOF, Shimon. **Handbook of Industrial Robotics**. Hoboken: Wiley, 1999.
29. O'NEIL, Cathy. **Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy**. New York: Broadway Books, 2016.
30. PALMA, Amanda, & PACHECO, Clarissa. '**O policial já foi com a arma na cabeça dele', diz mãe de rapaz confundido por reconhecimento facial**, no jornal **Correio 24 Horas**, em 05/01/2020. Disponível em <https://www.correio24horas.com.br/noticia/nid/o-policial-ja-foi-com-a-arma-na-cabeca-dele-diz-mae-de-razap-confundido-por-reconhecimento-facial/>. Último acesso em 10/06/2021.
31. PIAGET, Jean. **Inteligencia y afectividad**. Buenos Aires: Aique, 2005.
32. PIAGET, Jean. **Psychology of Intelligence**. New Jersey: Littlefield, Adams & Co., 1960.
33. PIAGET, Jean, & INHELDER, Bärbel. **A Psicologia da criança**. São Paulo: Difel, 1974.
34. RIBEIRO, Sandra Maria Patrício. **Lições preliminares para o estudo do *ethos* humano contemporâneo**. São Paulo: 2018.
35. RIBEIRO, Sandra Maria Patrício. **Uma perspectiva para a compreensão do *ethos* humano**, em **Ethos Humano e mundo contemporâneo**. São Paulo: Baracoa, 2019.
36. RIBEIRO, Sandra Maria Patrício, & ARAÚJO, Alberto Filipe. **Paisagem, imaginário e narrativa: olhares transdisciplinares e novas interrogações da psicologia social**. São Paulo: Zagodoni, 2015.
37. ROBERT, Enrico, & LOPES, Marcelo Frullani. **Quando um carro autônomo atropela alguém, quem responde?**, no jornal **El País**, em 16/04/2018. Disponível em [https://brasil.elpais.com/brasil/2018/04/16/tecnologia/1523911354\\_957278.html](https://brasil.elpais.com/brasil/2018/04/16/tecnologia/1523911354_957278.html). Último acesso em 10/06/2021.

38. ROSENBLATT, Frank. **The Perceptron - a perceiving and recognizing automaton.** Buffalo, Cornell Aeronautical Laboratory, Inc., 1957.
39. ROSENBLATT, Frank. **The Perceptron - a probabilistic model for information storage and organization in the brain,** na revista **Psychological Review**, volume 65, número 6. 1958.
40. RUSSELL, Stuart Jonathan & NORVIG, Peter. **Artificial Intelligence - A Modern Approach.** New Jersey: Prentice-Hall, 1995.
41. SCHANK, Roger & ABELSON, Robert. **Scripts, Plans, Goals, and Understanding.** Hillsdale: Erlbaum, 1977.
42. SCHOETTE, Derek. **Five Data Trends That Will Transform Cloud And AI In 2018,** na revista **Forbes.** 2018. Disponível em <https://www.forbes.com/sites/forbestechcouncil/2018/03/08/five-data-trends-that-will-transform-cloud-and-ai-in-2018/>. Último acesso em 10/06/2021.
43. SHANAHAN, Murray. **The Technological Singularity.** Cambridge: MIT Press, 2015.
44. SPANGLER, Teresa. **Five Big Trends Leveraging Artificial Intelligence To Watch In 2019,** na revista **Forbes.** 2018. Disponível em <https://www.forbes.com/sites/forbestechcouncil/2018/11/29/five-big-trends-leveraging-artificial-intelligence-to-watch-in-2019/>. Último acesso em 10/06/2021.
45. SOMERVILLE, Heather. **Motorista de carro autônomo do Uber estava assistindo "The Voice" antes de acidente fatal, diz polícia,** no jornal **Reuters**, em 22/06/2018. Disponível em <https://www.reuters.com/article/tech-uber-acidente-autonomo-idBRKBN1J12BB-OBRIN>. Último acesso em 10/06/2021.

46. SUPPES, Patrick. **Introduction to Logic**. New York: Dover Publications Inc., 1999.
47. TASSARA, Eda, & TASSARA, Marcello. **Chronos, o Tempo. Da Mitologia Grega ao Relógio**, em RIBEIRO, Sandra Maria Patrício, & ARAÚJO, Alberto Filipe, **Paisagem, Imaginário e Narratividade: Olhares transdisciplinares e novas interrogações da Psicologia Social**. São Paulo: Zagodoni, 2015.
48. THOMPSON, Judith. **Killing, letting die, and the trolley problem**, na revista **Monist: An International Quarterly Journal of General Philosophical Inquiry**, volume 59, páginas 204-217.. 1976.
49. TURIEL, E. **The development of social knowledge: Morality and convention**. Cambridge: Cambridge University Press, 1983.
50. TURING, Alan M.. **Computing machinery and intelligence**, na revista **Mind**, número 49. Oxford: Oxford University Press, 1950.
51. WALSH, Toby. **Machines That Think: The Future of Artificial Intelligence**. Amherst: Prometheus Books, 2018.
52. WEIZENBAUM, Joseph. **ELIZA - A Computer Program for the Study of Natural Language Communication Between Man and Machine**. Cambridge: Massachusetts Institute of Technology: 1965.
53. WEIZENBAUM, Joseph. **Computer Power and Human Reason**. San Francisco: W. H. Freeman, 1976.
54. WIENER, Norbert. **Cybernetics: or Control and Communication in the Animal and the Machine**. Cambridge: MIT Press, 1965. Publicação original em 1948.



55. WINOGRAD, Terry. **A Procedural Model of Language Understanding**, em SCHANK, Roger & COLBY, Kenneth, **Computer Models of Thought and Language**. San Francisco: W. H. Freeman, 1973.

56. **The Mechanical Brain**, no jornal **The Times**, edição de 11 de junho de 1949, página 4. Disponível em <https://www.thetimes.co.uk/archive/article/1949-06-11/4/15.html>. Último acesso em 08/07/2020.