

A MÁQUINA FINAL

Posttra

Amostra

A Dos autômatos mecânicos ao despertar
da Inteligência Artificial Geral

MÁQUINA FINAL

Ricardo Murer



ALTA BOOKS

GRUPO EDITORIAL

Rio de Janeiro, 2025

A Máquina Final

Copyright © 2025 STARLIN ALTA EDITORA E CONSULTORIA LTDA.

Copyright © 2025 Ricardo Murer

ISBN: 978-85-508-2861-9

Alta Cult é um selo do Grupo Editorial Alta Books (Starlin Alta Editora e Consultoria LTDA.).

Produção Editorial: Grupo Editorial Alta Books

Impresso no Brasil — 1ª Edição, 2025 — Edição revisada conforme o Acordo Ortográfico da Língua Portuguesa de 2009.

Dados Internacionais de Catalogação na Publicação (CIP)

M986m

Murer, Ricardo

A máquina final / Ricardo Murer. - 1.ed. - Rio de Janeiro: Alta Books, 2025.

296 p.; il.; 16x23 cm.

ISBN 978-85-508-2861-9

1. Inteligência artificial. 2. Inteligência artificial geral. 3. Ética da tecnologia.
I. Título.

CDD 006.3

Índice para catálogo sistemático:

1. Inteligência artificial : 006.3

Todos os direitos estão reservados e protegidos por Lei. Nenhuma parte deste livro, sem autorização prévia por escrito da editora, poderá ser reproduzida ou transmitida. A violação dos Direitos Autorais é crime estabelecido na Lei nº 9.610/98 e com punição de acordo com o artigo 184 do Código Penal.

O conteúdo desta obra fora formulado exclusivamente pelo(s) autor(es).

Marcas Registradas: Todos os termos mencionados e reconhecidos como Marca Registrada e/ou Comercial são de responsabilidade de seus proprietários. A editora informa não estar associada a nenhum produto e/ou fornecedor apresentado no livro.

Material de apoio e erratas: Se parte integrante da obra e/ou por real necessidade, no site da editora o leitor encontrará os materiais de apoio (download), errata e/ou quaisquer outros conteúdos aplicáveis à obra. Acesse o site www.altabooks.com.br e procure pelo título do livro desejado para ter acesso ao conteúdo.

Suporte Técnico: A obra é comercializada na forma em que está, sem direito a suporte técnico ou orientação pessoal/exclusiva ao leitor.

A editora não se responsabiliza pela manutenção, atualização e idioma dos sites, programas, materiais complementares ou similares referidos pelos autores nesta obra.

Produção Editorial: Grupo Editorial Alta Books

Diretor Editorial: Anderson Vieira

Editor da Obra: J.A. Ruggeri

Vendas Governamentais: Cristiane Mutús

Produtora Editorial: Rita Motta



Rua Viúva Cláudio, 291 — Bairro Industrial do Jacaré

CEP: 20.970-031 — Rio de Janeiro (RJ)

Tels.: (21) 3278-8069 / 3278-8419

www.altabooks.com.br — altabooks@altabooks.com.br

Ouvidoria: ouvidoria@altabooks.com.br





Agradecimentos

AOS MEUS PAIS, LYDIA e Mario Murer (*in memoriam*), agradeço minha paixão por aprender, por ensinar e minha biblioteca. Minha eterna gratidão por terem me colocado, desde cedo, na trilha infinita da busca por conhecimento.

Agradeço de coração à minha esposa, Fabiana, mulher admirável e de inteligência sensível, por sua paciência e compreensão durante as longas noites e finais de semana que dediquei a este livro. Seu apoio incondicional e seu incentivo constante foram fundamentais para que eu mantivesse o fôlego e a determinação, mesmo nos momentos mais desafiadores. Esta conquista também é sua.

Meus sinceros agradecimentos aos amigos e colegas da área digital que, com suas críticas construtivas e reflexões valiosas, contribuíram significativamente para tornar este livro ainda melhor: Antonio Higino Viegas, Ismael Rodrigues, Prof. Andre Queiroz, Arnaldo Leite e João Rocha. Nossas conversas são sempre fonte de inspiração e motivação para minha contínua jornada em busca de conhecimento na área de IA. Agradeço, de forma especial, ao meu editor J.A. Ruggeri, que acreditou neste projeto desde o início e, com suas palavras de incentivo e apoio constantes, foi fundamental para que esta obra se tornasse realidade.

Este livro também não seria possível sem as sugestões generosas e as conversas produtivas que tive com minha estimada amiga da área de IA, Profa. Dra. Roseli Romero, Titular junto ao Departamento de Ciências de Computação e vice-coordenadora do MBA em Inteligência Artificial e

Big Data do ICMC-USP, cuja escuta atenta e contribuições técnicas foram fundamentais nesta trajetória.

Muito obrigado!

Ricardo Murer

Agosto 2025

Amostra

Sumário

Prefácio	1
Prólogo	3
PARTE I – GÊNESE	29
Origens	31
Ética e IA	57
PARTE II – CORPO	71
Inspiração	73
Corpo-máquina	88
Corpo-interface	105
Estética	118
Reflexões	127
PARTE III – CÉREBRO	131
Lar do pensamento	133
Inteligência	167
Como máquinas aprendem	181
Como seres humanos aprendem	195
Reflexões	202

PARTE IV – AGÊNCIA	211
Evolução	213
Reflexão	240
Epílogo	243
Termos e definições	253
Referências	267

Amostra



Prefácio

NESTE LIVRO, O LEITOR encontrará uma exploração abrangente sobre a jornada histórica, científica e tecnológica rumo à criação de uma Inteligência Artificial Geral (IAG).

Ricardo Murer utiliza uma linguagem acessível, mas, ao mesmo tempo, técnica e reflexiva. Ele combina explicações científicas e tecnológicas com reflexões filosóficas e históricas, tornando o conteúdo compreensível para leitores que não são especialistas em áreas como tecnologia, neurociência ou computação. Além disso, o texto é estruturado de forma didática, com exemplos práticos, citações de especialistas e referências a mitos e lendas, o que enriquece a narrativa e facilita a compreensão dos conceitos complexos relacionados à Inteligência Artificial Geral (IAG).

São abordados desde os mitos e lendas que inspiraram a criação de seres artificiais até os avanços modernos em IA, como redes neurais, modelos generativos e robótica. O livro é estruturado em quatro partes principais:

1. **Gênese:** Explora os mitos e o desejo humano de criar seres artificiais, refletindo sobre o equilíbrio entre progresso e aniquilação.
2. **Corpo:** Discute a evolução dos autômatos e robôs humanoides, destacando a importância de um corpo físico para a interação com o mundo.

3. **Cérebro:** Analisa o desenvolvimento de redes neurais e modelos de IA, incluindo os desafios e limitações atuais para alcançar a IAG.
4. **Agência:** Investiga os desafios éticos e técnicos de criar máquinas autônomas capazes de tomar decisões baseadas em valores morais.

Além disso, o livro apresenta um framework conceitual que integra os domínios de Corpo, Cérebro e Agência, e reflete sobre os impactos sociais, culturais e éticos da convivência com máquinas inteligentes. Segundo Murer, a construção da IAG enfrenta desafios profundos que se distribuem entre aspectos técnicos, éticos e conceituais. No plano técnico, destaca-se a necessidade de replicar a complexidade da cognição humana, com autonomia, intencionalidade, integração corpo-cérebro, eficiência energética e superação das limitações dos modelos de linguagem atuais. Do ponto de vista ético, uma IAG exigirá fundamentos morais claros para tomar decisões alinhadas aos valores humanos, além de levantar questões sobre impactos sociais e riscos existenciais. Por fim, no plano conceitual, persistem incertezas sobre a definição de inteligência, a possibilidade de cognição em sistemas não biológicos e os caminhos para integrar abordagens como a IA corporificada e os modelos de mundo, elementos ainda em construção, mas considerados essenciais para avançar em direção a uma IAG funcional e segura.

É uma obra que combina história, ciência e filosofia, convidando o leitor a refletir sobre os desafios e possibilidades de um futuro com IAG. Murer conseguiu reunir de forma ímpar, precisa e objetiva os vários desafios existentes, mostrando que, apesar dos avanços, a IAG ainda está distante e exige novos modelos teóricos, técnicas computacionais e uma compreensão mais profunda do cérebro humano.

Profa. **Dra. Roseli Romero**

Titular junto ao Departamento de Ciências de Computação e
vice-coordenadora do MBA em Inteligência Artificial e Big Data do
ICMC-USP



Prólogo

IMAGINE UMA ENTIDADE ARTIFICIAL dotada de inteligência¹ tão vasta e adaptável quanto a humana, com potencial para igualar nossa habilidade de pensar, aprender e criar. Isto não é o enredo de uma obra de ficção científica, mas um objetivo audacioso, fundamentado em pesquisas sérias, o qual envolve milhares de pesquisadores ao redor do planeta. O nome dado a este novo graal da ciência da computação é Inteligência Artificial Geral (IAG).

Para que uma IAG alcance um nível comparável à inteligência humana, precisamos reconhecer que este é um desafio colossal. Não se trata apenas de criar sistemas ou máquinas capazes de executar tarefas mecânicas e previsíveis, mas de desenvolver entidades artificiais que possuam uma capacidade de aprender qualquer tarefa intelectual a partir da experiência, com compreensão de contexto, sensibilidade às nuances sociais e culturais, saiba transferir conhecimento entre domínios e possua memória de longo e curto prazo. Além disso, será preciso que esta IAG seja capaz de tomar decisões de forma realmente autônoma e que estas decisões sejam fruto de uma reflexão interna, uma forma de “autorreflexão artificial”. Isto não significa avaliar próprios estados internos baseados em programação pré-definida, mas sustentar juízos de valor, avaliar dilemas éticos e conflitos de intenção. Significa desenvolver sistemas racionais e

1. A definição de inteligência varia significativamente entre os diferentes campos do conhecimento. Na PARTE III — CÉREBRO, entraremos em mais detalhes.

lógicos, uma inteligência que possa agir com agência autêntica, sensível aos desdobramentos de suas ações sobre seus semelhantes, os seres humanos e o ambiente no qual estiver inserida. Para uma IAG ser verdadeiramente comparável à inteligência humana, ela deverá dominar áreas que estão além da ciência da computação e robótica, avançando de forma plena para medicina, direito, filosofia e sociologia, por exemplo. Esses fatores, por si só, já deixam claro que ainda estamos décadas, ou até séculos, do despertar da IAG. Não porque ela seja impossível, mas porque justamente precisa chegar “ao nível humano” de inteligência.

A abordagem adotada neste livro será na possibilidade de desenvolver uma IAG com capacidades comparáveis às da inteligência humana e não necessariamente superiores. Não será discutido o conceito de “superinteligência” ou ideia semelhante. Afinal, igualar-se à complexidade, adaptabilidade e profundidade do pensamento humano já representa, por si só, um dos maiores desafios científicos e tecnológicos de todos os tempos.

Apesar da distância que estamos da IAG, há atualmente uma multiplicidade de avanços e esforços promissores que buscam enfrentar esses desafios. Neste livro, vamos explorar não somente a busca pela IAG na perspectiva da ciência da computação, mas também as pesquisas, modelos teóricos e soluções que estão surgindo na filosofia, robótica, neurociências e psicologia cognitiva. Em conjunto, podem ajudar a superar essas barreiras atuais para chegarmos à IAG. E que forma tomará esta IAG? Será uma máquina, um corpo robótico de aparência humanoide, compartilhando nosso espaço social? Ou agentes inteligentes distribuídos em rede, operando colaborativamente, realizando tarefas complexas sem que seja necessária nossa intervenção? Ou uma presença etérea, distribuída em servidores na nuvem, acessível em tela, em qualquer lugar, uma espécie de mente difusa, onipresente como a própria internet? Estas questões têm alimentado um intenso debate na academia e na indústria, e ao longo deste livro, vamos analisar cada uma destas possibilidades.

Entretanto, o caminho para IAG começa pelas pesquisas em andamento na área de inteligência artificial (IA). Nesse sentido, há uma perspectiva poderosa que podemos associar à IA, sabendo que ela não é apenas mais uma inovação, mas uma poderosa tecnologia habilitadora. Como bem observou Andrew Ng, professor da Universidade de Stanford e um dos nomes mais lúcidos da área:

“Assim como a eletricidade transformou quase tudo há quase 100 anos, hoje eu realmente tenho dificuldade em pensar em alguma indústria que a IA não transformará nos próximos anos.”

A eletricidade potencializou a indústria, movendo máquinas, iluminando cidades e abrindo caminho para invenções que transformaram radicalmente a sociedade, os atuais modelos de IA, por sua vez, potencializam nossa capacidade cognitiva, automatizando decisões, interpretando padrões, fazendo previsões e até mesmo gerando novas ideias e conteúdos. A eletricidade está relacionada com a matéria, aquilo que podemos tocar. A IA com o intangível de nossos pensamentos.

Diante do esforço humano atualmente dedicado à busca pela IAG, podemos compará-lo à magnitude da chegada do homem à Lua. Afinal, poucos feitos humanos podem se igualar em ousadia e impacto à criação de uma entidade artificial capaz de simular todos os aspectos da inteligência humana. Pisar em solo lunar redefiniu nossa posição no universo e mostrou que podemos ir além das fronteiras de nosso planeta. Alcançar a IAG irá redefinir a fronteira entre o natural e o artificial, e expandir os limites do que a mente humana pode criar.

Como veremos neste livro, o que impulsionou essa nova era da IA, e alimenta a busca pela IAG, surgiu a partir de uma técnica inovadora no

campo do aprendizado de máquina²: as Redes Adversárias Generativas ou *Generative Adversarial Networks* (GANs), propostas por Ian Goodfellow em 2014 (Goodfellow et al., 2014). Com o uso das GANs, tornou-se possível criar dados realistas, tais como imagens, sons e textos, a partir do zero. Pela primeira vez, máquinas passaram não apenas a reconhecer padrões, mas a gerar conteúdo original, abrindo caminho para uma nova era da IA. Um exemplo é o site *This Person Does Not Exist*³, que gera rostos hiper-realistas de pessoas que nunca existiram.

Outro marco fundamental foi o artigo *Attention is All You Need* (Vaswani et al., 2017), que apresentou ao mundo a arquitetura *Transformer*⁴, atualmente a base tecnológica dos grandes modelos de linguagem (*Large Language Models*) ou LLMs, tais como GPT⁵ e BERT. Essa arquitetura se tornou um dos pilares centrais da revolução que testemunhamos hoje na IA. Estes modelos de linguagem estão dentro da

-
2. O termo “aprendizado de máquina” refere-se a uma subárea da inteligência artificial que desenvolve algoritmos capazes de aprender a partir de dados, sem serem explicitamente programados para cada tarefa. Uma explicação mais detalhada pode ser encontrada no capítulo “Termos e definições”.
 3. O site <https://www.thispersondoesnotexist.com> utiliza o modelo StyleGAN, desenvolvido pela NVIDIA (Karras, Laine & Aila, 2019), uma variação de redes adversárias generativas (GANs) voltada para o controle estilístico da imagem gerada. Cada vez que a página é recarregada, um novo rosto é sintetizado a partir do zero, sem qualquer base em pessoas reais.
 4. O *Transformer* é uma arquitetura de rede neural projetada para processar sequências de dados, como texto com alta eficiência. Sua principal inovação é o mecanismo de “atenção”, que permite ao modelo focar nas partes mais relevantes da entrada capturando relações complexas entre palavras, mesmo que estejam distantes entre si. Essa arquitetura é a base de modelos capazes de responder a comandos textuais (prompts), gerando respostas contextualizadas com alto grau de coerência.
 5. O nome GPT carrega os três pilares do modelo de fundação: G – *Generative* (Gerativo): Capacidade de gerar novo conteúdo, a partir de *prompts* do usuário; P – *Pre-trained* (Pré-treinado): Treinamento anterior em grande escala que fornece base de conhecimento; T – *Transformer*: Arquitetura baseada em autoatenção, que compreende relações entre palavras e contexto com alta precisão.

área chamada de IA generativa, a qual começou com um LLM chamado ChatGPT, lançado pela OpenAI. Em apenas dois meses,⁶ ele alcançou mais de 100 milhões de usuários, o crescimento mais veloz já registrado por um serviço digital na história.

Um dos grandes segredos do sucesso do ChatGPT está na simplicidade da sua interface: uma única caixa de texto, onde o usuário digita sua pergunta, o chamado *prompt*. De fato, o termo mais utilizado hoje é: engenharia de prompts. Isto porque é mais do que uma simples pergunta. A engenharia de prompts consiste em um conjunto de técnicas voltadas à elaboração de instruções textuais ricas em contexto, com o objetivo de orientar o modelo de linguagem ou LLM, a gerar respostas mais precisas, relevantes e de alta qualidade. Essa interação entre o usuário e o modelo, entretanto, ocorre em linguagem natural, no idioma nativo do usuário, sem necessidade de tradutores. Sem dúvida, também podemos creditar ao fato do ChatGPT responder ao usuário em linguagem natural em seu idioma nativo, uma das razões de seu sucesso astronômico. Até então, o acesso à IA era algo exclusivo de cientistas da computação e de dados.

Mas não era somente a interface e o idioma nativo que fizeram do ChatGPT um sucesso. Diferente das abordagens anteriores, focadas em reconhecimento de padrões e previsões estatísticas, os LLMs demonstraram uma capacidade incrível de criar novos dados a partir de seu treinamento inicial. Isto causou um espanto e fascínio entre os usuários, criando um volume de comentários nas mídias sociais, nunca visto antes. Um estudo que analisou mais de 49 mil frases em artigos de notícias

6. Segundo dados da empresa de análise Similarweb e reportagens publicadas em diferentes canais especializadas, o ChatGPT atingiu a marca de 100 milhões de usuários ativos mensais em apenas dois meses após seu lançamento em novembro de 2022, tornando-se o serviço digital de crescimento mais rápido da história (Reuters, 2023).

mostrou que, nos seis meses após o lançamento, a atenção da mídia sobre a IA aumentou dez vezes⁷.

Se no início esses modelos eram capazes apenas de gerar texto a partir de texto, o que já causou um forte impacto nos usuários, atualmente expandiram seu alcance e são capazes de produzir imagens, sons, vídeos, código-fonte de programação e até sugerir novas moléculas e estruturas proteicas desde que treinados com dados adequados. Uma capacidade criativa que, até recentemente, parecia pertencer exclusivamente à mente humana. Algumas perguntas começam a flutuar no ar: será que os LLMs e seus modelos criativos são suficientes para nos levar à IAG? Infelizmente, LLMs não serão capazes deste feito, e as diversas razões que nos levam a esta conclusão, vamos compreender melhor ao longo deste livro.

Nesta jornada para chegarmos à IAG vamos nos deparar com múltiplas perspectivas teóricas e práticas. Isto porque ainda não há um consenso de como, e até mesmo se é possível, a cognição emergir em sistemas não biológicos. Temos uma visão computacionalista, com pesquisadores que veem a mente como um sistema computacional, seja por meio de métodos lógico-simbólicos, redes neurais ou combinados. Essa perspectiva tem suas raízes filosóficas no funcionalismo, defendido por filósofos como Hilary Putnam e Jerry Fodor. Neste livro, vou apresentar dois conceitos dentro desta perspectiva computacionalista. *World models* ou modelos de mundo e *Embodied AI* ou IA corporificada.

Yann LeCun, cientista chefe de IA Meta, e um dos mais conceituados pesquisadores da área de IA, tem defendido o conceito de modelos de mundo. Nesta abordagem promissora, sistemas de IA desenvolvem representações internas do mundo. Em vez de exigir tentativa e erro direto, para compreender os diferentes aspectos do mundo real, o agente ou sistema aprende a partir de um modelo preditivo comprimido, uma forma de “simulação mental artificial”, que pode usar esta simulação para

7. Did ChatGPT make the media love or fear AI more? Disponível em: https://www.aimodels.fyi/papers/arxiv/how-chatgpt-changed-medias-narratives-ai-s-emi?utm_source=chatgpt.com

imaginar estados futuros, planejar ações e tomar decisões. Entraremos em mais detalhes na PARTE IV — AGÊNCIA.

A *Embodied AI*⁸, ou IA corporificada, sustenta que a inteligência genuína só pode emergir de um organismo que possua um corpo situado dentro do mundo real. Nesta linha de pesquisa, estão pensadores tais como Maurice Merleau-Ponty, Francisco Varela e Andy Clark, que argumentam que a cognição está enraizada na experiência sensório-motora, que inteligência e corpo estão interligados profundamente. O professor e roboticista Rodney Brooks, uma das vozes mais influentes dessa abordagem, afirmou:

“Acredito que mobilidade, visão aguçada e a capacidade de realizar tarefas relacionadas à sobrevivência em um ambiente dinâmico constituem uma base necessária para o desenvolvimento da verdadeira inteligência” (Brooks, 1991).

Para Brooks, a existência de um corpo, que permita ao ser interagir com o ambiente, não somente observando passivamente, mas agindo neste ambiente de forma autônoma, é fator essencial para que exista inteligência. Assim, não é possível chegar à inteligência artificial (IA) verdadeiramente avançada, até mesmo a IAG, por meio de sistemas isolados capazes de criar simulações da realidade ou treinados com milhões de dados.

Neste livro, vamos explorar uma perspectiva que considero mais promissora e vem ganhando cada vez mais reconhecimento e adesão na comunidade científica: a integração de duas abordagens. De um lado, sistemas de IA dotados de modelos de mundo e representações internas sofisticadas e dinâmicas da realidade; de outro, a IA corporificada, em

8. Há duas áreas próximas do conceito de *Embodied AI* ou IA corporificada: 1. Cognição Corporificada (*Embodied Cognition*), amplamente discutida por outros pesquisadores a qual sustenta que a mente emerge da interação entre corpo e ambiente. 2. IA Situada (*Situated AI*), proposta por Rodney Brooks, enfatiza que sistemas inteligentes devem operar em contextos reais, reagindo diretamente ao ambiente em vez de depender de representações simbólicas abstratas.

que essas representações não existem de forma isolada, mas operam em conjunto com um corpo igualmente artificial, equipado com sensores capazes de perceber o ambiente, aprendendo e reagindo a partir de *inputs* do mundo físico. É nessa convergência entre representação e encarnação, na qual acredito que reside uma base conceitual e tecnológica mais efetiva para chegarmos à IAG.

A convicção de que a incorporação é indispensável para alcançarmos uma IAG genuína orientou toda a linha de pesquisa e a abordagem temática deste livro. Não por acaso, há capítulos inteiros dedicados a apresentar como os primeiros engenheiros conceberam seus autômatos, de como, ao longo da história, fomos capazes de evoluir tecnicamente para, ainda que parcialmente, reproduzir as habilidades físicas humanas em seres artificiais. Essa escolha não é meramente histórica ou estética: ela reflete minha crença de que, sem um corpo, sem termos uma máquina capaz de compreender e viver a experiência direta do mundo físico, nenhuma IA atingirá o patamar de uma verdadeira IAG. Ao longo deste livro, vamos encontrar diversas menções à IA corporificada, e como consequência direta, sobre robótica e robôs humanoides.

Cada capítulo oferece uma lente para ampliar sua compreensão não apenas sobre os avanços tecnológicos, mas também sobre os desdobramentos sociais, culturais e éticos que emergem à medida que nos aproximamos de uma sociedade híbrida, onde máquinas e humanos estarão compartilhando o mesmo espaço social. Para Ignatyev e Privalov, já vivemos neste momento: “O processo de formação da sociedade híbrida já começou e está em andamento. Hoje, as tecnologias computacionais se aproximaram daqueles processos que definem a essência da humanidade. Essas tecnologias realizam ações humanas em ambientes humanos e entre as pessoas” (Ignatyev: Privalov, 2019). Mas será que estamos realmente preparados para aceitar um futuro em que máquinas dotadas de IA avançada venham a conviver conosco no trabalho, circulem em espaços públicos e até compartilhem a intimidade dos nossos lares? E quais são os reais desdobramentos que um ser artificial capaz de pensar e agir por conta própria pode provocar em nossa sociedade? Como veremos em diferentes capítulos deste livro, a demanda por embarcar senso ético,

transparência e juízo de valor em máquinas e sistemas dotados de IA, os quais já apresentam um certo grau de autonomia, é mais urgente do que possa parecer.

PARTE I — GÊNESE

Neste capítulo, faremos uma jornada ao passado para revisitar mitos e lendas como os de Pigmalião e Galateia, Talos e Pandora, narrativas ancestrais que revelam a persistente fascinação humana pela criação de seres artificiais. Ao explorarmos o papel dos deuses nessas histórias, perceberemos que estas entidades não são apenas frutos da imaginação criadas para entretenimento, mas representações simbólicas de nosso desejo de dominar a vida, moldá-la e replicá-la. Dotados de força, inteligência e uma beleza arrebatadora, esses seres artificiais funcionam como espelhos do impulso humano de imitar os deuses em sua capacidade de criar. No entanto, os mitos não oferecem apenas inspiração: eles também operam como advertências. Junto ao fascínio criador, paira uma dúvida inquietante: o que acontece quando aquilo que criamos passa a agir por conta própria? E quando possui inteligência e autonomia, como podemos controlar estas criaturas não humanas? Como observa Adrienne Mayor, professora de Stanford, em entrevista: “É quase como se os mitos dissessem que é ótimo ter essas coisas artificiais no céu usadas pelos deuses. Mas, uma vez que elas interagem com os humanos, temos caos e destruição” (Shashkevich, 2019).

Vou usar os mitos como ferramentas poderosas para promover uma reflexão sobre valores morais, ética, responsabilidade e a dupla natureza da tecnologia, quando esta chega aos limites de nosso conhecimento e foge de nosso controle. Nossos antepassados, em suas narrativas míticas, colocaram nas mãos dos deuses o poder de criar seres artificiais, criaturas dotadas de forma, função e, muitas vezes, vontade própria. Nessas histórias, já estavam presentes tanto as alegrias quanto os infortúnios que surgem desta convivência. Hoje, vivemos um novo capítulo dessa antiga tensão. Milhares de sistemas baseados em IA, muitos deles integrados às plataformas de mídias sociais, já são capazes de gerar alegria e tristeza,

prazer e dor em milhões de pessoas ao redor do mundo. Além disso, estamos testemunhando a chegada de agentes de IA mais sofisticados, com capacidade de diálogo contínuo, personalização emocional e autonomia crescente. Neste livro, apresento uma descrição e reflexão sobre os princípios da IA Confiável, estabelecidos em 2019 pela União Europeia.

PARTE II — CORPO

A busca por criar máquinas à nossa imagem e semelhança remonta aos primeiros autômatos⁹ mecânicos e se intensificou com os avanços da robótica moderna. Ao longo dos séculos, artesãos e engenheiros da antiguidade buscaram transformar lendas e mitos em realidade, dando vida a autômatos que imitavam a forma e a função humanas, antecipando, de maneira rudimentar, a jornada que hoje nos levou à cibernética e finalmente à robótica. Entre estes mestres engenheiros da antiguidade estão Herão de Alexandria, professor, matemático e engenheiro, um precursor na construção de objetos dotados de movimento e Ismail Al-Jazari, um engenheiro e inventor muçulmano do século XII, reconhecido como o “pai da robótica” devido às suas contribuições pioneiras no desenvolvimento de autômatos e dispositivos mecânicos automatizados.

No século XVII, René Descartes, filósofo francês, comparou o corpo humano a uma máquina, argumentando que suas funções poderiam ser explicadas em termos mecânicos. Essa perspectiva mecanicista abriu caminho para a compreensão do corpo como um sistema automatizado, ajudando no avanço dos autômatos mecânicos e da robótica. A visão dualista de Descartes, entretanto, começou a ser questionada a partir de novas descobertas das ciências cognitivas, biologia e filosofia da mente, que

9. A palavra “autômato” origina-se do termo grego “αὐτόματον” (automaton), que significa “agir por si mesmo”. Ela é derivada de “αὐτός” (autos), que significa “si mesmo”, e “μάτον” (maton), que significa “pensar” ou “querer”, referindo-se a algo que se move ou opera por conta própria, sem influência externa. Um autômato é composto por uma série de dispositivos mecânicos projetados para imitar a ação humana ou animal, funcionando de maneira independente, ou seja, sem intervenção humana direta.