INTELIGÊNCIA ARTIFICIAL

3 ESPECULAÇÕES

CICLO FIDELIDADE — CULTURGEST

Conferências e Debates ×

INTELIGÊNCIA ARTIFICIAL: APLICAÇÕES, IMPLICAÇÕES, ESPECULAÇÕES

APLICAÇÕES Luísa Coheur, Pedro Bizarro, Milind Tambe	ABR	17 QUA	16:00
APLICAÇÕES (AS BOAS E AS MÁS) Mário Figueiredo			18:30
IMPLICAÇÕES Luís Moniz Pereira, Manuel Dias, Virginia Dignum	MAI	15 QUA	16:00
A ASCENSÃO DOS ROBÔS Martin Ford			18:30
ESPECULAÇÕES Ana Paiva, André Martins, Arlindo Oliveira	JUN	05 QUA	16:00
INTELIGÊNCIA ARTIFICIAL HUMANO COMPATÍVEL Stuart Russell			18:30

O campo das práticas associadas à inteligência artificial gera uma grande especulação em relação ao futuro e por isso fechamos este ciclo olhando para essa conjuntura que todos os dias vemos surgir.

Esta publicação inicia-se com uma entrevista a Stuart Russell, conduzida por Tiago Domingos, professor do Instituto Superior Técnico da Universidade de Lisboa. Entre outras questões, conversam sobre a grande atenção mediática que ultimamente tem recaído sobre a inteligência artificial e a forma como é entendida a relação entre a mesma, o corpo e as emoções humanas. Discutem as eventuais contribuições da inteligência artificial para o cumprimento dos Objetivos para o Desenvolvimento Sustentável, apontados pelas Nações Unidas, e terminam com uma pergunta: considerando a hipótese de se criarem agentes autónomos controlados por pessoas ou entidades com más intenções, como garantir que estas máquinas mais poderosas do que a inteligência humana não terão poder sobre o Homem?

André Martins, em *Inteligência Artificial e linguagem natural: para além das grandes pirâmides,* reflete sobre a posição real e a imaginação ficcional da inteligência artificial – em referência à popularização do imaginário do filme *2001: Odisseia no Espaço.* Convida-nos também a imaginar uma forma de inteligência para as máquinas que não tenha como referência a inteligência do ser humano e termina lembrando que os perigos futuros da chamada superinteligência das máquinas são um problema mais humano do que tecnológico: "não me parece plausível que os perigos mais iminentes da inteligência artificial advenham da sua superinteligência: pelo contrário, advirão da nossa impreparação e do mau uso que faremos dessas tecnologias se sobrevalorizarmos as suas capacidades e não conseguirmos entender as suas falhas e os seus enviesamentos."

Em *Inteligência Artificial Geral: ficção científica ou futura realidade?,* Arlindo Oliveira aborda a analogia humano-máquina no que diz respeito a comportamentos e à inteligência, referindo que neste momento a inteligência artificial ainda está no domínio da especialização em realizar determinadas tarefas. Defende que há um longo e incerto caminho a percorrer até as máquinas poderem lidar com toda a complexidade e ambivalência do mundo real. A superinteligência poderá não vir a concretizar-se, no entanto, toda a especulação filosófica em torno deste tema é cada vez mais relevante porque, caso venha a acontecer, implicará profundas mudanças que importam ser pensadas.

Ana Paiva também aborda este tema, colocando a tónica na forma como a continua delegação da autonomia humana pode levar a uma desresponsabilização social provocada pela falta de empatia e compaixão. Considerando um contexto social e cultural onde domina a exploração dos seres humanos e da Terra, onde os extremismos políticos crescem e com eles um mundo excessivamente polarizado, Ana Paiva põe em causa os contratos sociais trazidos pela modernidade – mas não terá sido essa mesma modernidade que criou as bases para esta polarização e exploração?

Serão os perigos e as potencialidades do futuro da inteligência artificial uma questão de ordem tecnológica ou de ordem política – podendo a tecnologia exponenciar os nossos comportamentos humanos que estão "sob a sombra ameaçadora da desresponsabilização e da desumanidade?" Poderá a inteligência artificial estar a colocar-nos perante uma lente de aumento do que temos sido? E será que desejamos uma sociedade mais "transparente, responsável e natural, inspirada na forma como nós, humanos, interagimos e vivemos"? Ou será que este projeto só pode ser realizado se a sociedade como um todo se responsabilizar por fazer este caminho?

INTELIGÊNCIA ARTIFICIAL HUMANO COMPATÍVEL

Entrevista conduzida por
Tiago Domingos, professor
do Instituto Superior Técnico
da Universidade de Lisboa.

STUART RUSSELL

P: Na sua opinião, porque é que atualmente as pessoas se interessam tanto por inteligência artificial (IA)? É justificado ou estamos apenas no pico de uma propaganda exagerada?

R: São várias as razões para este foco atual. Em primeiro lugar, a criação de um mundo *online* e a digitalização de grande parte da atividade comercial global trouxeram muitas oportunidades para a aplicação da IA que não existiam antes. Em segundo lugar, os avanços em IA – especialmente na aprendizagem profunda – tornaram possíveis novas aplicações que nunca poderiam ter existido antes. A tradução direta da voz em tempo real, e com grande qualidade, é apenas um exemplo. Em terceiro lugar, muitas pessoas olham para momentos como a vitória de Alpha Go sobre Lee Sedol [sistema desenvolvido pela empresa DeepMind, adquirida pela Google, que venceu o campeão mundial do jogo Go] enquanto prova de que estamos subitamente muito mais perto de ultrapassar a inteligência humana num sentido amplo.

Penso que esta última razão é, provavelmente, um engano e que, de facto, poderá haver um nível desadequado de otimismo a curto prazo que pode ter levado a alguns maus investimentos. Em algum momento irá haver uma reestruturação, mas não um "rebentar da bolha", porque existem tantas aplicações no mundo real que já são bem-sucedidas e muitas outras irão ser desenvolvidas, mesmo que haja um longo tempo de espera até que surja a próxima grande descoberta.

P: O que será necessário para que a IA atinja um entendimento do mundo? Quais são as limitações das abordagens atuais da IA neste contexto?

R: O mundo é um lugar incrivelmente complexo, como todos nós sabemos. Em particular, contém muitas coisas. Assim, a grande maioria das línguas têm substantivos, pronomes e quantificadores como "todos" e "alguns". Estes permitem-nos aprender e falar sobre coisas individuais, bem como sobre algumas teorias e regras que se aplicam a muitas coisas.

Os lógicos desenvolveram linguagens formais com as mesmas capacidades, especialmente a lógica de primeira ordem. As linguagens de programação, tais como Java ou Python, também possuem essas capacidades. Infelizmente, as abordagens da aprendizagem profunda são baseadas em circuitos sintonizáveis muito grandes que não têm a mesma capacidade para representar o conhecimento sobre as coisas. De forma muito real, os modelos de aprendizagem profunda não *sabem* nada. Por isso, precisamos de algum tipo de linguagem

formal que combine as capacidades representacionais e de conhecimento da lógica de primeira ordem com a modelação flexível de dados e capacidades de aprendizagem dos sistemas de aprendizagem profunda. As chamadas "linguagens de programação probabilísticas" são um dos candidatos e, hoje em dia, representam uma das áreas mais ativas de investigação.

P: Um "corpo" é essencial para a inteligência artificial geral (IAG) ou é possível uma IA incorpórea?

R: Suponho que IA "corpórea" significa ter *inputs* e *outputs* sensório-motores que sejam coerentes com aqueles produzidos por um corpo físico mais ou menos análogo aos corpos de animais ou humanos. Um robô moderno bem equipado com sensores tácteis, proprioceção, câmaras, etc., neste sentido teria um corpo. Que esse corpo seja essencial para a cognição é uma hipótese empírica, mas não parece ser apoiada por muitas provas. Em primeiro lugar, quase todo o progresso que ocorreu até agora na perceção, reconhecimento de voz, etc. é incorpóreo. Em segundo lugar, muitos humanos não possuem um ou vários dos sentidos convencionais e penso que iriam contestar fortemente se lhes dissessem que eles não tinham capacidades cognitivas reais. Geralmente, aquilo que se passa nas mentes humanas não parece variar muito com aquilo que são os sentidos do corpo. Para além disso, adaptamo-nos muito rapidamente a novos tipos de corpos tais como exoesqueletos, com o *input* visual a vir de câmaras localizadas na cabeça de outras pessoas e por aí.

P: Será que uma IAG bem-sucedida irá exigir agentes autónomos dotados de emoções e preferências autónomas?

R: Vamos falar de emoções primeiro. Atualmente, não temos nenhuma ideia de como criar emoções *reais* que são sentidas do mesmo modo que as emoções humanas. Isto relaciona-se com o difícil problema dos Qualia e da consciência – que, na visão de muitos filósofos e quase todos os investigadores em IA, e apesar de todas as tentativas para o resolver, permanece sem solução. Aquilo a que podemos chamar emoções *funcionais* seriam processos que produzem os mesmos efeitos funcionais em máquinas tal como acontece nos humanos; por exemplo, a forma como o medo altera os processos de pensamento e comportamento de uma forma consistente, e de um modo geral, é bastante diferente, digamos, de chegar à inevitável conclusão lógica de que existe, de facto, um monstro debaixo da cama. Imagino que as emoções funcionais possam ser úteis ao permitirem às máquinas gerarem um comportamento mais eficaz em algumas circunstâncias, enquanto outras formas de cognição mais deliberadas seriam demasiado lentas. Não sei se existem muitos indícios disto, e é importante compreender que o *hardware* das máquinas é bastante diferente do *hardware* humano e, por isso, estas talvez nem precisem de usar este tipo de "atalho".

A ideia de preferências autónomas é, na verdade, bastante difícil de compreender. Em que sentido são as preferências humanas autónomas? Podemos dizer "decidi que quero ser astronauta", mas quanto disso é autónomo? É principalmente uma consequência da interação entre as influências culturais e alguns dos complexos e inatos sistemas de recompensa do cérebro. Também penso que seria extremamente perigoso criar sistemas de IAG com alguma capacidade para sonhar com novos objetivos a alcançar, em situações onde o novo objetivo nada tem a ver com a satisfação das preferências que queremos que a máquina satisfaça.

P: Os objetivos do desenvolvimento a médio prazo (até 2030) para a espécie humana podem ser vistos como que codificados pelas Metas de Desenvolvimento Sustentável (MDS) da ONU. Quais pensa serem as principais contribuições da IA neste contexto?

R: As conferências *AI for Social Good* (IA para o Bem Social), que decorrem anualmente em Genebra desde 2017, foram criadas precisamente para abordar este tema. Segundo algumas estimativas, o progresso já alcançado em 14 das 17 MDS pode ser consideravelmente acelerado pela IA. Por exemplo, as metas relacionadas com o clima, a fome e o desenvolvimento urbano podem ser apoiadas substancialmente através da análise contínua de dados por satélite. Hoje em dia, seria necessário ter 30 milhões de pessoas a trabalharem a tempo inteiro para ver todas as imagens de satélite que estão a ser produzidas, enquanto um único sistema de IA poderá vir a ser capaz de fazer isso. Este é um objetivo para o qual muitos de nós trabalhamos.

P: Considera que o maior perigo da IA para a humanidade são os agentes de IA autónomos poderem assumir o controlo ou os danos causados por agentes de IA controlados por humanos mal-intencionados?

R: Existem dois tipos de perigo. O primeiro indica que devemos estar seriamente atentos ao problema do *controlo:* como assegurar que as máquinas que são, de forma muito real, mais poderosas que os humanos, nunca tenham poder sobre os mesmos? Penso que

existem maneiras de assegurar isto, mas requer a substituição da atual abordagem técnica da IA por uma nova. O meu próximo livro, *Human Compatible* [Humano Compatível], é sobre isso.

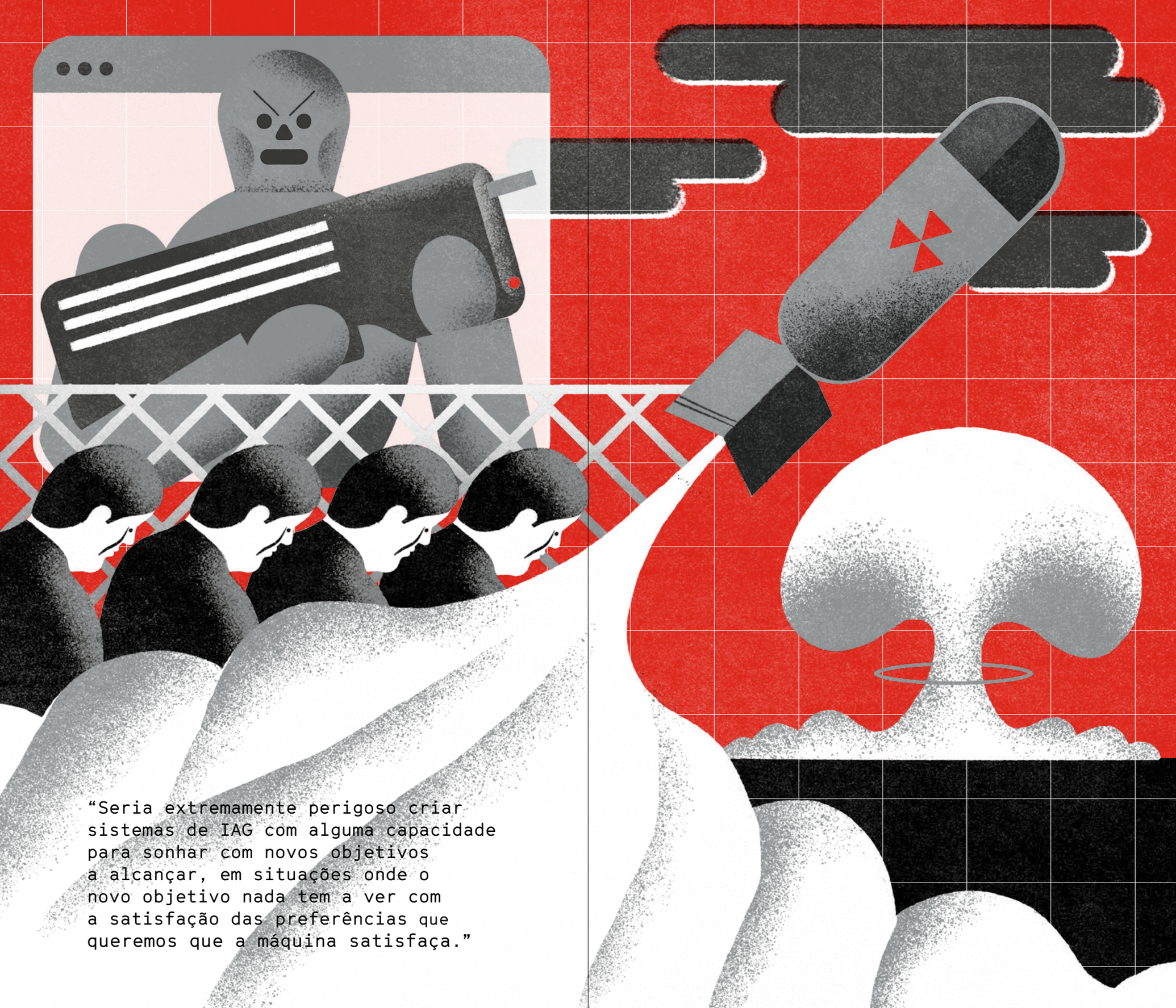
O segundo tipo de perigo – o uso indevido da IA para o mal por parte de pessoas ou grupos – existe para outras tecnologias, como é o caso das armas nucleares. Temos vindo a desenvolver continuadamente um esforço tremendo para assegurar que isso não aconteça, através de regulação, policiamento, serviço de inteligência e medidas militares. Se olharmos para a versão inicial do mau uso da IA, que são os vírus nos computadores e os cibercrimes, temos, muito honestamente, uma resposta não existente a isto, com algumas nações a conspirarem ativamente para impedirem soluções convincentes. Por isso, pedia para resolvermos o problema agora, porque só vai piorar à medida que os sistemas de IA se tornam mais eficazes. Sinceramente, estou mais otimista em relação à possibilidade de encontrarmos uma solução para o primeiro problema do que para o segundo.

Penso existir ainda um terceiro tipo de perigo: o enfraquecimento. E. M. Forster deixou um aviso relativamente a isto no conto *The Machine Stops* [publicado em 1909], que recomendo vivamente. Em breve poderemos ter a opção de entregar a gestão da nossa civilização a máquinas. Quando isso acontecer, os humanos irão perder o incentivo primário de aprender, com cada nova geração, o conhecimento acumulado e as competências da nossa civilização. Assim, acabaremos como no futuro mostrado em *Wall-E* [filme de animação da Disney-Pixar, de 2008], onde somos apenas passageiros mimados num cruzeiro dirigido por máquinas, um cruzeiro que nunca termina.

P: A IA vai dominar o mundo? Será que a IA significa o fim da espécie humana?

R: Vamos separar isso um pouco. Não é bem uma questão de antevisão mas de orientação. Identificar potenciais pedras e barreiras e contorná-las. Poderemos vir a ter alguns desastres que nos irão relembrar da importância de redobrar esforços para evitar males maiores. Por exemplo, o efeito dos algoritmos de seleção de conteúdo das redes sociais em semear a discórdia na nossa sociedade – possivelmente arruinando a UE, a NATO e a democracia no processo – foi provavelmente não intencional, mas resulta da otimização do objetivo errado, nomeadamente as taxas de clique. Isto é um aviso sobre o que acontece quando as máquinas, que conseguem ter um impacto à escala global, otimizam um objetivo fixo. Algumas pessoas, como é o caso de Danny Hillis, defendem que o mesmo se passa com empresas que otimizam o lucro a curto prazo e negligenciam externalidades como as alterações climáticas; estas empresas são, na verdade, uma forma de IA (se bem que com elementos humanos) a otimizar um objetivo fixo e mal desenhado.

Penso que é possível contornar esta barreira – podemos desenhar as máquinas de forma diferente, para que garantam a longo prazo a promoção dos objetivos que nós queremos mesmo. Isto não é fácil mas é possível. Se fizermos isso, e se também evitarmos os obstáculos relativos à má utilização da IA e o enfraquecimento, existe uma boa possibilidade de que a IA possa ajudar a trazer uma era dourada para a humanidade.



“Seria extremamente perigoso criar sistemas de IAG com alguma capacidade para sonhar com novos objetivos a alcançar, em situações onde o novo objetivo nada tem a ver com a satisfação das preferências que queremos que a máquina satisfaça.”

INTELIGÊNCIA ARTIFICIAL E LINGUAGEM NATURAL: PARA ALÉM DAS GRANDES PIRÂMIDES

ANDRÉ MARTINS

Nos últimos anos, assistiu-se a um enorme progresso tecnológico em inteligência artificial (IA), com ênfase nas áreas do processamento de linguagem natural, visão computacional e robótica. Estes avanços devem-se sobretudo ao sucesso de métodos de aprendizagem automática (AA) os quais permitem a uma máquina aprender a partir de dados e melhorar o seu desempenho com a experiência. Mas será que estes avanços tecnológicos nos colocam mais perto de reproduzir a inteligência humana? O que podemos esperar para os próximos anos? E quais serão os desafios seguintes?

ESPECULAÇÕES DE ONTEM ACERCA DO DIA DE HOJE

Na célebre obra *2001: Odisseia no Espaço,* de Arthur C. Clarke e Stanley Kubrick, o computador HAL 9000 compreende a linguagem natural dos humanos e rebela-se ao desconfiar que a sua própria existência corre perigo. Esta obra de 1968 – pouco mais de 50 anos atrás – prevê viagens interplanetárias, comunicação perfeita homem-máquina e máquinas capazes de tomar decisões complexas no mundo real – nenhuma das quais existe em 2019.

O computador HAL 9000 é o arquétipo da IA: uma máquina dotada de faculdades comparáveis aos humanos. Consegue entender a nossa linguagem, elaborar estratégias para alcançar objetivos, recolher dados interagindo com o ambiente exterior e tomar decisões com base nestes dados. O HAL 9000 foi imaginado nos anos 60, uma época particularmente otimista, poucos anos após a seminal conferência de Dartmouth em 1956, que juntou Allen Newell, Herbert Simon, John McCarthy, Marvin Minsky e Ray Solomonoff (entre outros), marcando o início da IA enquanto área de investigação científica.

ONDE ESTAMOS HOJE?

Apesar do otimismo inicial, a história da IA tem sido atribulada. Pensava-se que poucos anos bastariam para desenvolver tecnologia capaz de reconhecer pessoas, entender a fala humana e traduzir entre quaisquer línguas, mas estas expectativas exageradas acabaram por conduzir a um longo "inverno" *(AI winter),* caracterizado por cortes radicais de financiamento.

Décadas depois, este "inverno" foi superado e o otimismo regressou. Hoje usamos quotidianamente algoritmos de IA. Por exemplo, quando fazemos uma pesquisa na internet, quando usamos um tradutor *online,* ou quando um livro nos é recomendado. As transações bolsistas são realizadas por algoritmos em milissegundos. A análise de imagens médicas é cada vez mais feita por algoritmos de reconhecimento de padrões. Grandes empresas como a Google, a Facebook, a Microsoft, a Amazon e a Uber estão a desenvolver veículos autónomos, assistentes pessoais digitais, sistemas de diálogo e tradutores automáticos, armazenando grandes quantidades de dados e

recorrendo a técnicas de AA. Estamos a assistir a uma autêntica "corrida ao ouro", com grandes investimentos estratégicos em IA por parte de vários estados (sobretudo os Estados Unidos da América, a China, o Canadá, a França e a Europa como um todo), com o objetivo de acelerar este progresso.

ESPECULAÇÕES DE HOJE ACERCA DO DIA DE AMANHÃ

Uma das utopias mais antigas em IA é a tradução automática: a capacidade de uma máquina traduzir entre quaisquer línguas, eliminando a barreira da linguagem e mediando a comunicação entre humanos. Esta área tem tido uma evolução notável nos últimos anos, graças a técnicas de aprendizagem profunda com redes neuronais. Embora não seja ainda possível traduzir automaticamente um livro com a qualidade de um tradutor humano, a tradução de certos conteúdos, tais como notícias ou e-mails, é hoje muito superior à de há cinco anos. Nos próximos anos, é expectável que os avanços no processamento de linguagem natural (incluindo o reconhecimento e síntese de fala, a extração de informação semântica e os sistemas de diálogo) sejam integrados em assistentes personalizados: dispositivos capazes de comunicar connosco, gerir a nossa agenda diária e procurar informação *online*. Estes dispositivos saberão tudo sobre os nossos gostos e preferências e tornar-se-ão rapidamente imprescindíveis.

Uma das características desejáveis de uma IA é a capacidade para tomar decisões complexas. Esta capacidade foi objeto de estudo nos primeiros trabalhos realizados por Herbert Simon (nobel da Economia em 1978), a quem se deve o princípio da racionalidade limitada *(bounded rationality)*, segundo o qual um processo de decisão deve ter em conta a limitação da informação disponível, a limitação cognitiva para processar essa informação e o tempo limite para decidir. Atualmente, observamos sucessos em ambientes controlados, tais como jogos com regras bem definidas: o sistema AlphaGo, através de técnicas de aprendizagem por reforço, venceu os melhores humanos no jogo do Go, um marco histórico que se julgava a décadas de distância. Um desafio bem mais difícil é sair destes ambientes simulados e criar máquinas capazes de tomar decisões num sistema aberto a partir das observações que recolhem do mundo real. Quando isto for possível, teremos, para além da robótica industrial que já conhecemos, um conjunto alargado de profissões que podem passar a ser desempenhadas por máquinas: médicos, engenheiros, juízes e analistas financeiros. É expectável que isso aconteça nas próximas décadas. Os recentes avanços em controlo e robótica, patente nos veículos autónomos, revolucionarão em breve o setor dos transportes, tornando obsoleto o automóvel particular. O próximo passo é a exploração espacial: dotados da capacidade de tomar decisões em situações difíceis, os robôs serão a tecnologia adequada para explorar regiões inóspitas deste e de outros planetas.

DEPOIS DE AMANHÃ

Para conjeturar acerca de um futuro mais longínquo, precisamos de uma visão mais abrangente – menos antropomórfica – de "inteligência". Será a inspiração biológica condição necessária para se criar uma IA? Em geral, tendemos a encarar o futuro da IA à luz daquilo que conhecemos sobre a inteligência humana, mas será esta a única forma possível de "inteligência"? Tomemos o exemplo da aerodinâmica: apesar do voo das aves ter servido de inspiração para a criação de engenhos voadores, os aviões não batem as asas como os pássaros. De igual modo, pode ser possível construir máquinas "inteligentes" sem tentar replicar os mecanismos do cérebro.

Formas de "comportamento inteligente" podem emergir em sistemas com múltiplos agentes: perante a necessidade de colaborar para resolver um problema, estes agentes desenvolvem automaticamente protocolos de comunicação para trocar informação entre si. Que linguagem falam estas máquinas? O que haverá em comum entre esta linguagem artificial e a humana? Que linguagem vingará como a mais propícia para obter comportamento inteligente, uma linguagem simbólica como a nossa ou "representações contínuas", para nós incompreensíveis? Será possível fazer a mediação entre estas representações internas e a linguagem humana no sentido de obter interpretabilidade?

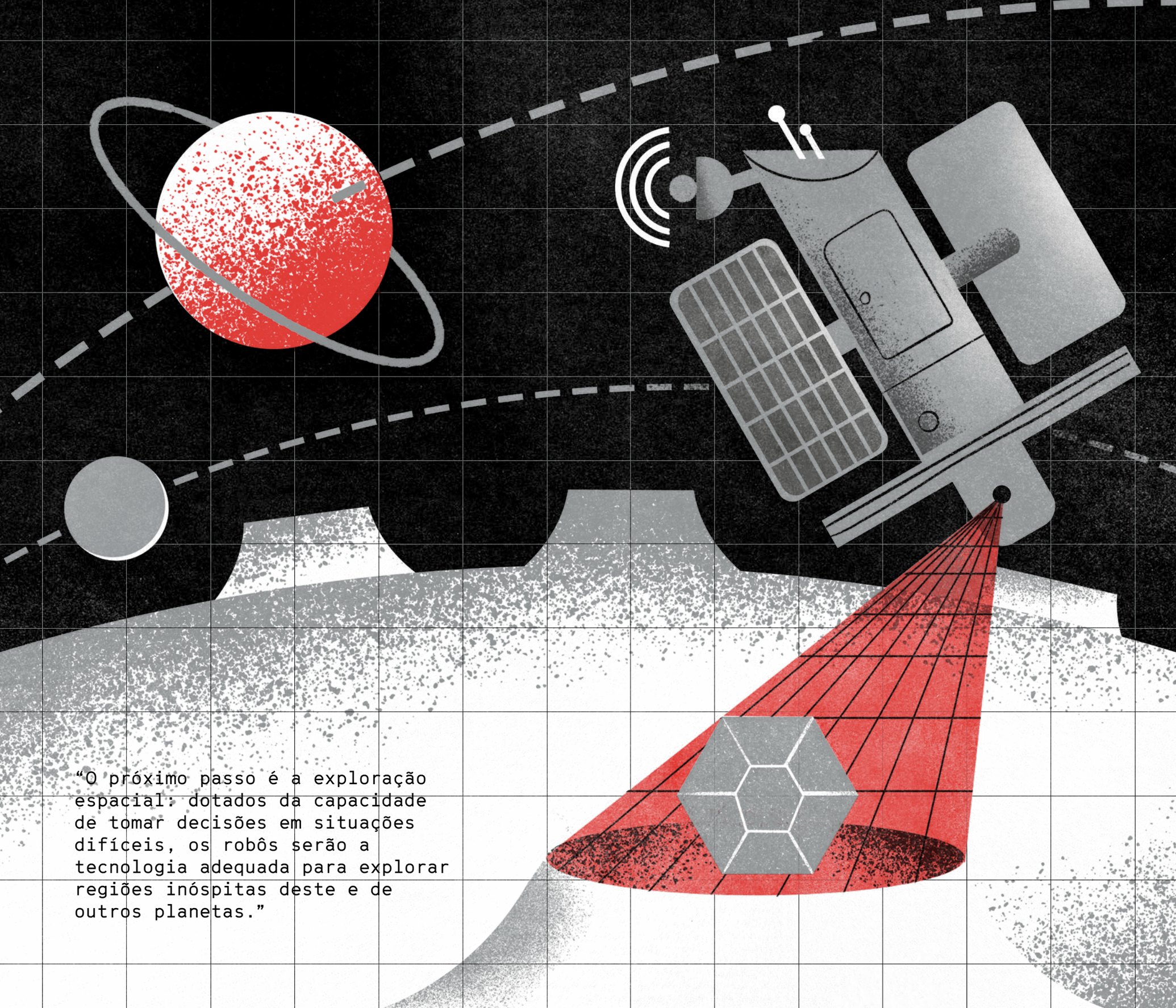
Estas questões colocam em evidência as diferentes perspetivas de duas correntes historicamente antagónicas: o conexionismo (subjacente às redes neuronais artificiais) e o computacionalismo (que procura representar o raciocínio lógico através de operações formais sobre símbolos discretos). Embora a corrente conexionista seja a mais promissora, os modelos atuais são incapazes de responder a questões importantes, tais como a interpretabilidade das decisões – fundamental na interação entre homens e máquinas – e a capacidade de raciocinar com senso comum, uma qualidade intrínseca da inteligência humana. Estas duas correntes não são necessariamente exclusivas, como o demonstram modelos recentes como as máquinas de Turing neuronais, os quais abrem a porta à inclusão de algumas ideias "computacionalistas" em redes neuronais.

Em 1948, Claude Shannon publicou o célebre artigo *A Mathematical Theory of Communication,* revolucionando as telecomunicações e o processamento de linguagem natural. Ao invés, a IA carece atualmente de um alicerce matemático sólido – não existe ainda uma Teoria Matemática da Inteligência que nos dê a conhecer as suas possibilidades e limites teóricos e forneça novas ferramentas para obter avanços disruptivos. Os sucessos recentes em IA, embora impressionantes, têm resultado de um processo essencialmente empírico, caracterizado por um aumento de escala computacional, de dados e de engenheiros. Por outras palavras, a IA acabou de sair da sua fase pré-histórica e está agora em plena Idade Antiga: uma época caracterizada por empreendimentos coletivos extraordinários, tais como as grandes

pirâmides de Gizé (cerca de 2500 a.C.), mas também por uma relativa rudimentaridade técnica. Os historiadores estimam que as grandes pirâmides de Gizé foram construídas (freneticamente) por cerca de dez mil trabalhadores em turnos de três meses ao longo de 30 anos. O número de cientistas e engenheiros que hoje trabalham em IA certamente ultrapassa esse número. O esforço computacional, medido pelo número de *teraflops* e pela energia dissipada em gigantescos centros de processamento de dados, decerto rivaliza com o esforço humano despendido para empilhar os blocos de pedra das pirâmides. No entanto, as técnicas hoje utilizadas em IA parecem igualmente rudimentares. Tal como os dez mil construtores das pirâmides em 2500 a.C. não teriam a maquinaria e o engenho necessários para poder erigir o Empire State Building (construído em 1931 por 3500 trabalhadores em apenas um ano e 45 dias), também não temos hoje a tecnologia necessária para erigir uma inteligência artificial geral (IAG) que vá além da automação de um número de tarefas específicas.

Não creio por isso que tão cedo as máquinas se tornem mais inteligentes do que nós ou que esteja sequer para breve uma comunicação tão fluida como a exibida pelo HAL 9000. Apesar dos alarmismos expressados pelo cientista Stephen Hawking (1942-2018) e pelo magnata Elon Musk, que veem na IA "a mais séria ameaça à sobrevivência da espécie humana", não me parece plausível que os perigos mais iminentes da IA advenham da sua superinteligência: pelo contrário, advirão da nossa impreparação e do mau uso que faremos dessas tecnologias se sobrevalorizarmos as suas capacidades e não conseguirmos entender as suas falhas e os seus enviesamentos.

"A IA acabou de sair da sua fase pré-histórica e está agora em plena Idade Antiga: uma época caracterizada por empreendimentos coletivos extraordinários, tais como as grandes pirâmides de Gizé."



“O próximo passo é a exploração espacial: dotados da capacidade de tomar decisões em situações difíceis, os robôs serão a tecnologia adequada para explorar regiões inóspitas deste e de outros planetas.”

INTELIGÊNCIA ARTIFICIAL GERAL: FICÇÃO CIENTÍFICA OU FUTURA REALIDADE?

ARLINDO OLIVEIRA

A ideia de criar sistemas que exibam inteligência comparável à do ser humano já é antiga. Em 1843, Ada Lovelace colocou a questão se seria possível criar sistemas destes e concluiu pela negativa, argumentando que os computadores nunca poderiam criar nada de novo mas apenas executar os programas que são definidos pelos programadores. Mais de um século depois, em 1950, Alan Turing abordou a mesma questão e concluiu que deveria ser possível, em princípio, criar máquinas inteligentes, que exibissem comportamentos indistinguíveis dos comportamentos humanos, pelo menos em determinadas circunstâncias. O teste que propôs para avaliar o comportamento inteligente de máquinas veio a ficar conhecido como o teste de Turing.

Apesar disso, sete décadas de investigação não permitiram desenvolver sistemas que apresentem inteligência comparável à do ser humano. Foi possível desenvolver sistemas que desempenham tarefas específicas tão bem ou mesmo melhor que uma pessoa, como a identificação de objetos e pessoas em imagens, a execução de diagnósticos médicos, o domínio de diversos jogos de tabuleiro e a análise de contratos legais, entre muitas outras. Por esta razão, nos últimos anos tem sido muito discutida a ideia de inteligência artificial (IA) e poderá pensar-se que estes sistemas exibem algum tipo de inteligência. Todavia, nenhum é verdadeiramente inteligente, na aceção normal da palavra. São uma espécie de *idiots savants,* muito bons numa tarefa específica mas completamente incapazes de lidar com a complexidade do mundo real. Um sistema que faça diagnósticos de cancro a partir de ressonâncias magnéticas apenas sabe fazer isso, não percebe as consequências dos diagnósticos que faz, não tem empatia com os pacientes que diagnostica e não se importa de ser desligado ou destruído. De facto, não sente, não pensa, não raciocina. A única coisa que faz é apresentar um diagnóstico do que "vê" numa imagem médica, diagnóstico este que foi aprendido a partir de milhares de outros diagnósticos que lhe foram apresentados e usados para treiná-lo nesta tarefa. A outros sistemas "inteligentes" aplica-se exatamente a mesma análise. Executam uma e uma só tarefa, de forma repetida, eficaz, coerente e totalmente inconsciente. Estamos assim longe de desenvolver sistemas verdadeiramente inteligentes, que possam adaptar-se a novas condições e problemas, que percebam a importância das tarefas que estão a desempenhar, que tenham motivações próprias e, claro, que sejam conscientes. Apesar de toda a investigação feita nesta área, não sabemos ainda como desenvolver sistemas que exibam a flexibilidade, a criatividade e a adaptabilidade da inteligência humana, muito menos a perceção do mundo e a consciência que caracteriza cada um de nós.

Mais significativamente, não é sequer óbvio o caminho que deve ser trilhado para atingir tal objetivo nem tampouco é óbvio que se deva perseguir esse objetivo. Pode argumentar-se que, embora seja importante desenvolver sistemas que desempenhem tarefas específicas, não existe qualquer interesse em desenvolver sistemas que exibam o comportamento complexo proveniente da multifacetada inteligência humana. Por outro lado, muitos

cientistas acreditam que uma inteligência comparável à humana – complexa, multidimensional e relacionada com o mundo físico – apenas pode existir quando existe um corpo que interage com o meio ambiente, o que significa que não pode ser reproduzida num computador (embora o possa ser num robô).

Existem, assim, reservas muitos fortes à possibilidade de um dia ser possível desenvolver sistemas com IA comparável à do ser humano, a chamada inteligência artificial geral (IAG). É perfeitamente possível que tais sistemas nunca venham a existir e que, nas próximas décadas e nos próximos séculos, conheçamos apenas sistemas com competências específicas muito significativas mas incapazes da flexibilidade, da criatividade e da adaptabilidade que caracteriza os seres humanos. Muitos filósofos, cientistas e investigadores desta área partilham este ceticismo e concentram-se na resolução de problemas concretos, ignorando a questão da IAG.

Porém, milhares de cientistas e investigadores continuam a desenvolver algoritmos que permitam a sistemas inteligentes configurar-se, aprender com o mundo real e, possivelmente, evoluir no sentido de exibirem IAG. A empresa DeepMind é a face mais visível do que pode ser visto como um grande projeto da humanidade que pretende "resolver o problema da inteligência". Os sucessivos desafios que têm sido enfrentados com sucesso, usando técnicas de redes neuronais e de aprendizagem por reforço, são vistos como degraus na longa escada que conduzirá, por fim, a sistemas verdadeiramente inteligentes, capazes de resolver qualquer problema ou desafio tão bem como um ser humano. A verdade é que poderá de facto vir a ser possível desenvolver métodos que permitam a uma máquina aprender a compreender o mundo de forma tão profunda como nós, a ter ideias próprias que podem ser exploradas e desenvolvidas, e a criar conhecimento até então inexistente.

Se agora é impensável que uma máquina descobrisse (ou redescobrisse) a Teoria da Relatividade, escrevesse *Os Lusíadas* ou compusesse a *Sinfonia N.º 9* de Beethoven, é impossível garantir que estes feitos estarão, para sempre, longe do alcance de sistemas de IA. O facto de não sabermos como se projetam sistemas capazes de criar estes avanços científicos e culturais não é uma garantia absoluta de que nunca seremos capazes de os conceber. Há apenas 100 anos, não sabíamos o que eram transístores, bombas atómicas, telemóveis, internet e redes sociais. Nem sequer existiam os conceitos necessários para discutir estes avanços tecnológicos. Ninguém pode garantir que os mecanismos usados pelo cérebro humano para se desenvolver e aprender, assim como os mecanismos usados pela evolução para criar esse mesmo cérebro, não serão um dia mais bem compreendidos e usados para criar IAG, num computador. Ainda mais forte é o argumento de que nem sequer precisamos compreender exatamente os mecanismos do cérebro humano, uma vez que poderá ser possível chegar aos mesmos resultados usando abordagens muito diferentes. O cérebro humano, com os seus neurónios e impulsos elétricos, poderá não ser o único suporte possível para uma inteligência evoluída, capaz de criar arte, de fazer descobertas científicas e de desenvolver tecnologia.

As discussões sobre a IAG são, neste momento, puramente filosóficas. Não estamos muito mais avançados do que em 1950, quando Alan Turing discutiu esta mesma questão e concluiu que a possibilidade existe, em princípio, contudo o objetivo está muito distante e é difícil de atingir. Mas serem filosóficas não faz com que sejam inúteis. Podemos nunca vir a conseguir criar sistemas tão inteligentes como nós. Porém, a hipótese, ainda que remota, de que isso um dia venha a acontecer justifica uma atenção muito especial, porque a criação de uma IA comparável à humana teria consequências muito profundas na sociedade e poderia causar uma enorme disrupção nos mecanismos sociais e económicos que são o sustentáculo do mundo moderno. A ideia de existir uma explosão de inteligência, quando uma máquina inteligente conseguir projetar uma máquina ainda mais inteligente que ela própria, proposta em 1965 por Irving John Good, não é disparatada. Tal processo poderá, em princípio, conduzir a uma superinteligência, uma inteligência muito mais poderosa e geral do que a humana. Com a tecnologia de hoje, tal poderá parecer um acontecimento impossível. Mas a tecnologia evolui e os conhecimentos da humanidade, como um todo, avançam rapidamente. O que hoje é impossível pode ser possível daqui a uma década e tornar-se uma realidade daqui a duas décadas. A história está repleta de previsões feitas sobre o desenvolvimento de diversas tecnologias que se revelaram redondamente erradas, umas vezes por serem demasiado otimistas e outras por serem demasiado pessimistas. A área da IA não é exceção e muitos problemas cuja resolução foi considerada, já neste século, impossível ou muito distante, foram resolvidos nos últimos anos. Outros problemas, que foram considerados relativamente simples há mais de meio século, continuam por resolver. Como reza o aforismo, apocrifamente atribuído a Niels Bohr: "fazer previsões é difícil, especialmente sobre o futuro".

INTELIGÊNCIA ARTIFICIAL DE AMANHÃ: DA INDIVIDUALIDADE À PROSSOCIALIDADE NA IA PARA O BEM SOCIAL

ANA PAIVA

A sociedade atravessa atualmente uma fase de mudança acentuada, sendo a inteligência artificial (IA), sem dúvida, uma das principais causas dessa transformação. Nos anos 50, a IA nasceu com a promessa de criar máquinas capazes de aprender, comunicar, criar abstrações, conceitos e resolver qualquer problema, até então, solucionados apenas por humanos. Exemplos clássicos de desafios colocados foram o de dotar máquinas da capacidade de jogar xadrez ao nível de um grão-mestre, de traduzir automaticamente textos entre qualquer dialeto, ou mesmo levar a cabo com exatidão diagnósticos em medicina. Durante mais de 60 anos de pesquisa, certamente com altos e baixos, a área de IA evoluiu significativamente e superou grande parte dos seus desafios iniciais. Hoje dispomos de algoritmos cada vez mais complexos e rápidos que permitem resolver problemas especializados, de forma automatizada e altamente eficiente. Essa evolução alastrou-se a todos os ramos da nossa sociedade. O nosso quotidiano está repleto de sistemas que classificam imagens, recomendam produtos ou reconhecem e traduzem textos em tempo real. As aplicações de IA são omnipresentes, ainda que nem sempre visíveis. Abundam notícias sobre os grandes feitos da IA, desde derrotar grandes mestres no jogo Go, passando por conduzir de forma autónoma, ou mesmo aprender a jogar videojogos, sem intervenção humana e recorrendo apenas às observações de jogos passados. Estes sucessos da IA influenciam o que fazemos no dia a dia, e centram-se essencialmente em técnicas de processamento de dados adquiridos para aprender a tomar decisões e prever o futuro. Estas máquinas, cada vez mais "inteligentes", entraram em nossa casa prontificando-se a decidir quais as notícias mais importantes, qual o melhor caminho para o trabalho ou que música mais iremos apreciar. Quando procuramos preços de uma viagem de avião, a nossa pesquisa é condicionada pelos nossos comportamentos passados, informação essa usada para antecipar as nossas preferências e intenções. A IA oferece-nos respostas personalizadas: o que nos é apresentado difere do que é oferecido a pessoas com um passado de interação e de pesquisas distinto. A solução parece inovadora e eficiente. Mas estaremos prontos a delegar a totalidade das nossas decisões em algoritmos?

A mudança a que assistimos com a entrada da IA na nossa sociedade está intrinsecamente ligada com noção de "autonomia" das máquinas, isto é, a capacidade de tomarem decisões de forma informada e sem coerção ou controlo humano. Contudo, para essa autonomia ser efetiva, as máquinas têm não só que tomar decisões percecionadas como "certas", como também os humanos têm que aceitar as decisões das máquinas, delegando a responsabilidade nas mesmas. Esta delegação permite um aumento significativo da eficiência na decisão e uma melhoria dos processos em empresas, possibilitando que estas capitalizem e automatizem o trabalho facilmente. Com a crescente ubiquidade desta "delegação", aumenta também o risco de desresponsabilização dos humanos face às decisões tomadas pelas máquinas. Esta ideia é satirizada na popular série

britânica *Little Britain,* com a expressão "Computer says no..." explorando a forte dependência dos humanos quando delegam tarefas em computadores, tornando-se subservientes das suas decisões. Assim, há quem argumente que, com o aumento da autonomia das máquinas, estamos a assistir a uma subsequente diminuição da autonomia humana. Por outro lado, com o aumento da complexidade dos algoritmos usados pela IA para tomadas de decisão, a compreensão dos seres humanos sobre as motivações subjacentes a tais sistemas autónomos é reduzida, levando a uma consequente perda de responsabilidade dos humanos.

A possibilidade de usar aplicações de IA como forma de espiar decisões (des)humanas é perigosamente oportuna. Vários indicadores salientam que o mundo nos últimos anos tem vindo a exibir uma generalizada falta de empatia e compaixão, sendo constatados diariamente exemplos de extremismo, polarização e tribalismo que julgávamos extintos das sociedades modernas. Vários decisores políticos revelam sinais claros de desrespeito pela condição humana, pelo ambiente e pelas gerações futuras. Concomitantemente, normas sociais estabelecidas na sociedade revelam-se insuficientes para gerir corretamente bens públicos e recursos comuns, resultando em situações indesejáveis como a resistência a antibióticos, o aumento de casos de doenças perfeitamente evitáveis por vacinação, a mitigação dos efeitos das alterações climáticas, ou a sobre-exploração de recursos naturais.

Como desenvolver então IA e, consequentemente, "máquinas autónomas", sob a sombra ameaçadora da desresponsabilização e da desumanidade? Para responder à questão, é necessária uma nova corrente em IA, onde esta deixe de ser centrada em otimizar os processos usando critérios de utilidade e racionalidade baseados em eficiência, e considere a sociedade no geral, com humanos e máquinas a coexistirem de uma forma simbiótica para melhorar a nossa maneira de viver e de definir o nosso futuro. A tarefa não é simples. Compreender a inter-relação entre máquinas e sociedade envolve formalizar as regras do comportamento humano – tarefa milenar e naturalmente inacabada. Uma das visões dominantes da tomada de decisão humana é baseada no princípio da maximização da utilidade *(homo economicus).* De facto, inspirada nos humanos, esta é também a espinha dorsal de várias abordagens para modelar o comportamento em máquinas autónomas. No entanto, exemplos provenientes da Psicologia Social ou da Economia comportamental mostram que as pessoas se comportam de forma prossocial e aparentemente irracional, sendo solidárias, empáticas e justas. A IA deve usar estes exemplos de forma positiva para, além de otimizar o proveito imediato, conseguir potenciar a prossocialidade humana.

Algum trabalho recente em IA considera que há necessariamente que passar por tornar a mesma mais social, transparente, responsável e natural, e inspirada na forma como nós, humanos, interagimos e vivemos em sociedade. Desde o início dos anos 80, diversos investigadores em IA afirmam que o sucesso desta área requer um foco no cérebro humano

"Com o aumento da complexidade dos algoritmos usados pela IA para tomadas de decisão, a compreensão dos seres humanos sobre as motivações subjacentes a tais sistemas autónomos é reduzida, levando a uma consequente perda de responsabilidade dos humanos."

e em aspetos como emoções, moralidade e comportamento social. Existe uma componente emocional e social na essência do que é inteligência. Por outro lado, há que considerar que as máquinas não vão funcionar em isolamento, mas terão de interagir com pessoas e outras entidades (os chamados agentes) fazendo parte de um universo habitado por outros. Consequentemente, a inteligência não será só a capacidade de executar uma tarefa simples da forma mais eficaz possível, uma vez que esta visão leva a uma IA que pode ser entendida como "egoísta" já que é apenas baseada em fatores de utilidade, eventualmente aprendidos com os dados fornecidos. Alternativamente, a IA deverá transformar-se numa inteligência coletiva e social oriunda da combinação das competências dos diversos agentes num ecossistema com auto-organização e com padrões de comportamento emergentes.

No futuro, a IA terá que dar o salto e passar de uma "IA egoísta" para uma "IA social" e "prossocial", onde os critérios para tomar decisões são estabelecidos coletivamente e em benefício da própria sociedade. Mas como podemos construir máquinas autónomas (ou robôs) que vão ser integrados em populações híbridas (humanos e máquinas), de forma a promover ações coletivas em situações onde o conflito e o individualismo surgem como o comportamento natural? As máquinas devem ser capazes de estimar que ações são instrumentalmente "boas" para seus objetivos, mas também devem ser capazes de aferir o seu impacto na sempre complexa ecologia das decisões humanas, distinguindo as ações social e moralmente benéficas daquelas que são prejudiciais tanto do ponto de vista individual como coletivo. Só assim as máquinas serão capazes de responderem emocional e socialmente aos humanos e adaptarem-se ao seu contexto ambiental, social e moral.

É necessário, portanto, mais investigação, mas também a consciencialização de que a IA tem que ser desenvolvida para os humanos. A IA transformar-se-á brevemente numa área de charneira na investigação interdisciplinar, combinando os esforços de áreas aparentemente díspares, da engenharia às ciências humanas e do comportamento, criando com isso novos desafios às universidades e empresas. Movimentos recentes como *Human-Centered Artificial Intelligence* [IA centrada no humano] ou *AI for Social Good* [IA para o Bem Social] procuram dar os primeiros passos nessa direção, considerando que os algoritmos têm que ser postos à disposição da humanidade e dirigidos a melhorar a sociedade. Organizações como as Nações Unidas em colaboração com a Fundação XPRIZE estão cientes destes desafios, tendo promovido diversos eventos como o *The AI for Good Global Summit,* onde se destacaram áreas como as cidades sustentáveis, lidar com a resposta a desastres, abordar o impacto da desigualdade ou melhorar a saúde pública.

Para que tal aconteça, acredito que é urgente dotar as máquinas (agentes e robôs) com capacidades de perceberem as sociedades e os humanos, colocarem-se no seu lugar (terem empatia), colaborarem, formarem equipas, aprenderem e, em geral, integrarem-se interagindo com humanos de forma natural, flexível e transparente. Estes objetivos estão longe de ser atingíveis num futuro próximo. Podemos especular e tentar traçar o futuro, mas temos que estar conscientes do desafio que é ter uma sociedade onde máquinas e humanos têm relações duradouras, simbióticas e estáveis. Nessa altura a IA poderá de facto tornar-se um dos pilares da construção de uma sociedade melhor.

“Compreender a inter--relação entre máquinas e sociedade envolve formalizar as regras do comportamento humano – tarefa milenar e naturalmente inacabada.”

“Algum trabalho recente em IA considera que há necessariamente que passar por tornar a mesma mais social, transparente, responsável e natural, e inspirada na forma como nós, humanos, interagimos e vivemos em sociedade.”

Inteligência Artificial (IA)
[Artificial Intelligence]
Área de estudo que se dedica à conceção de sistemas artificiais (geralmente computadores, conjuntos de computadores ou robôs) capazes de exibir comportamentos que um ser humano informado interpreta como inteligentes. Habitualmente, estes comportamentos consistem em tomar decisões, com base em observações, a fim de atingir um certo objetivo. Por exemplo, um sistema de jogar xadrez decide qual o próximo movimento que vai efetuar com base na observação do estado presente do tabuleiro, com o objetivo de ganhar a partida. Também é comum usar-se a expressão IA para referir um sistema específico: "A IA que conduz o meu automóvel autónomo é competente".

Inteligência Artificial Geral (IAG)
[Artificial General Intelligence]
Refere-se ao conceito, puramente teórico, de um sistema que exiba uma inteligência com a flexibilidade, adaptabilidade e competência equivalente à do ser humano. Em contraste, um sistema concebido para um problema específico designa-se IA estreita [narrow AI]. Não existem neste momento métodos para a construção de sistemas deste tipo mas há investigação que tem IAG como objetivo.

Inteligência Artificial Clássica
[Good Old-Fashioned Artificial Intelligence – GOFAI]
Conjunto de técnicas desenvolvidas desde meados dos anos 1950, baseadas na manipulação de símbolos, em técnicas de procura e planeamento e na representação explícita e simbólica de conhecimento. Até há uma década atrás, a expressão IA referia-se quase exclusivamente a este tipo de técnicas.

Superinteligência
[Superintelligence]
Conceito, especulativo, de que poderá vir a existir uma inteligência artificial geral superior à inteligência humana.

Singularidade
[Singularity]
Acontecimento hipotético no qual uma superinteligência provoca uma aceleração do progresso tecnológico que ultrapassa a capacidade de compreensão e previsão dos seres humanos.

Aprendizagem Automática (AA)
[Machine Learning]
Conjunto de técnicas (com suporte em várias áreas da matemática e da computação) que visam dotar um sistema artificial da capacidade de aprender a tomar decisões a partir de um conjunto de exemplos, sem que para tal seja explicitamente programado. A maioria dos sistemas de IA modernos baseiam-se em AA. Existem três tipos principais: supervisionada, não supervisionada e aprendizagem por reforço.

Aprendizagem Supervisionada
[Supervised Learning]
Classe de métodos nos quais a aprendizagem automática se baseia num conjunto de exemplos de observação--decisão; ou seja, a supervisão consiste no fornecimento da decisão certa/desejada para cada observação nesse conjunto de exemplos. Este tipo de aprendizagem exige a definição de um critério de qualidade de cada decisão produzida pelo sistema (simplesmente certo ou errado, ou algo bastante mais complexo); o processo de aprendizagem consiste em otimizar o sistema para maximizar este critério, avaliado sobre os exemplos fornecidos. As duas principais classes de problemas de aprendizagem supervisionada são: regressão – quando a decisão tem caráter quantitativo/numérico (por exemplo, o valor da temperatura máxima do ar no dia seguinte); classificação – quando a decisão tem caráter categórico (por exemplo, se uma dada imagem contém ou não uma cara, ou a identidade da pessoa que surge numa dada imagem).

Aprendizagem Não Supervisionada
[Unsupervised Learning]
Classe de métodos de aprendizagem que permitem a um sistema identificar regularidades nos dados, tais como agrupamentos (clusters), exceções/anomalias e relações entre variáveis. Distingue-se da aprendizagem supervisionada por se basear num conjunto de observações e não em pares observação-decisão. O utilizador tem de especificar que tipo de regularidades pretende identificar, sendo as técnicas usadas para os vários tipos bastante diferentes.

Aprendizagem Por Reforço
[Reinforcement Learning]
Classe de métodos onde a decisão pretendida é fornecida ao sistema apenas no fim de uma série de observações. Por exemplo, num sistema que aprenda a jogar xadrez, ao invés do supervisor dizer qual o melhor lance para cada posição (aprendizagem supervisionada), apenas indica o resultado do mesmo. Compete aos métodos de aprendizagem por reforço identificar as decisões que conduzem ao desfecho desejado. Este tipo de aprendizagem tem numerosas áreas de aplicação (robótica móvel, veículos autónomos, jogos, sistemas de recomendação), em situações onde um sistema está a aprender a melhor estratégia ao longo do tempo.

Redes Neuronais Artificiais
[Artificial Neural Networks – ANN]: classe de métodos de aprendizagem automática baseada numa abordagem conexionista, onde redes de neurónios artificiais são configuradas para desempenhar certas funções, aprendendo a mapear as observações nas decisões pretendidas. Cada neurónio implementa um modelo matemático muito simplificado de neurónios biológicos e o processo de aprendizagem consiste em ajustar os parâmetros das interligações entre os neurónios artificiais para maximizar um dado critério de ótimo desempenho. Existem vários tipos de métodos de aprendizagem automática para redes neuronais artificiais, com diferentes graus de plausibilidade fisiológica (mesmo que apenas aproximadamente) relativamente ao processo de aprendizagem nos cérebros biológicos.

Redes Neuronais Profundas
[Deep Neural Networks – DNN]
Redes neuronais artificiais cuja estrutura se organiza numa sequência de camadas, correspondendo a primeira às próprias observações e a última à saída da rede na qual é produzida a decisão. Cada camada processa a informação proveniente da anterior e alimenta a camada seguinte, caracterizada por um conjunto de parâmetros que são ajustados durante o processo de aprendizagem. A designação "profunda" refere--se a um número elevado de camadas.

Aprendizagem Profunda
[Deep Learning]
Classe de métodos de aprendizagem automática associados às redes neuronais profundas. Uma característica habitual (mas não obrigatória) destes métodos (ou algoritmos) é o facto de processarem, em cada passo, apenas uma pequena fração do conjunto de treino (em casos extremos, apenas uma observação), o que os torna adequados para lidar de forma eficiente com grandes volumes de dados.

Algoritmo
[Algorithm]
Derivado do nome do matemático, geógrafo e astrónomo persa al-Khwarizmi, que viveu nos séculos VIII e IX. Um algoritmo é um conjunto de instruções explícitas que permite a uma máquina resolver uma classe de problemas. Exemplo: o procedimento simples usado para multiplicar manualmente dois números inteiros com vários dígitos é um algoritmo – um conjunto de passos que permite obter a solução pretendida (o resultado da multiplicação), a partir dos dados de entrada (os números a multiplicar). A formalização do conceito de algoritmo tem um papel central na teoria e na prática da IA e da AA, nas ciências da computação e na matemática.

Retropropagação
[Backpropagation]
Um dos componentes mais importantes da técnica matemática (ou algoritmo) usada para treinar redes neuronais artificiais, em particular as redes profundas. Permite determinar, em cada passo do algoritmo, qual a variação que deve sofrer cada parâmetro da rede para reduzir os erros observados na saída da mesma.

Sistemas Multiagente (SMA)
[Multiagent Systems]
Designação de sistemas em que a inteligência está distribuída por vários sistemas (agentes). A resolução de problemas por vários, em vez de uma só entidade, requer a capacidade de coordenação, negociação e execução de planos conjuntos.

Procura e Planeamento
[search and planning]
Técnicas para procurar soluções e planear ações em situações complexas que conduzam a um dado resultado.

Árvores de Decisão
[Decision Trees]
Classe de métodos de aprendizagem automática nos quais se usam representações em árvore para descrever o conjunto de testes que permitem tomar uma decisão a partir de uma observação. Uma árvore de decisão poderá determinar a sequência de testes médicos que se devem realizar para obter um diagnóstico, especificando a escolha de cada teste em função do resultado do teste anterior.

Interpretabilidade
[Interpretability/ Explainability]
Possibilidade de interpretação, em moldes compreensíveis por um ser humano, das decisões produzidas por um sistema de IA. Em algumas aplicações (por exemplo, diagnóstico médico) a interpretabilidade é muito importante para justificar uma dada decisão tomada por um sistema. No extremo da não-interpretabilidade encontram-se as redes neuronais profundas, cujas decisões resultam de sequências de operações matemáticas complexas, com números de parâmetros que podem ascender a dezenas de milhões. As árvores de decisão, em contraste, devido à explícita sequência de testes que suportam a decisão, são sistemas de elevada interpretabilidade.

Big Data
Termo genérico, habitualmente usado para referir cenários de aprendizagem automática ou análise de dados nos quais o volume de dados é suficientemente grande para forçar a utilização de técnicas especificamente desenhadas para o efeito.

Robô
[Robot]
Sistema físico, tipicamente eletromecânico, que interage com o meio exterior, deslocando-se nele e/ou manipulando objetos. O sistema de controlo do robô pode ou não ser dotado de IA, dependendo da sua natureza e das tarefas para o qual é concebido. Também se designam por robôs os programas de computador que percorrem a internet, adquirindo, processando e organizando informação publicamente disponível.

Chatbot
Abreviatura de chatter robot [robot que conversa], expressão usada para referir sistemas, habitualmente baseados em IA, capazes de estabelecer uma conversa/diálogo através de fala ou por mensagens de texto com seres humanos. São usados em muitos contextos para automatizar o diálogo entre uma organização ou um dispositivo e seres humanos, nomeadamente no apoio a clientes, entretenimento, educação, comércio, assistentes pessoais.

Processamento de Língua Natural
[Natural Language Processing]
Conjunto de técnicas que permitem às máquinas reconhecer padrões na língua natural, interpretando os mesmos e respondendo também em língua natural ou tomando decisões com base na observação de textos. Exemplos clássicos incluem a tradução automática, a análise de sentimento (com o objetivo de classificar um texto, um e-mail ou um produto quanto ao sentimento que exprime: positivo, negativo, neutro) ou a classificação do tópico de um texto (se é sobre desporto, negócios, política, ...).

Nuvem
[Cloud]
Termo usado para referir a utilização, através da internet, de recursos de computação e/ou armazenamento em computadores localizados remotamente. Permite uma grande flexibilidade, podendo o utilizador adequar os recursos que adquire às suas necessidades de cálculo e/ou armazenamento. Outro aspeto muito importante é que dispensam os utilizadores das tarefas de aquisição, manutenção e gestão (nomeadamente, backups) dos recursos informáticos, sendo todas estas tarefas garantidas pelo fornecedor do serviço.

CULTURGEST

Conselho Diretivo

PRESIDENTE
José Ramalho
ADMINISTRADORES
Manuela Duro Teixeira
Mark Deputter
SECRETÁRIA DE ADMINISTRAÇÃO
Patrícia Blázquez

Programação

ARTES PERFORMATIVAS
Mark Deputter
ARTES VISUAIS
Delfim Sardo (assessor)
CONFERÊNCIAS E DEBATES
Liliana Coutinho (assessora)
MÚSICA
Pedro Santos (assessor)
PARTICIPAÇÃO / FAMÍLIAS E ESCOLAS
Raquel Ribeiro dos Santos

Coleção da Caixa Geral de Depósitos

CONSERVADORA
Isabel Corte-Real
ASSISTENTES
Lúcia Marques
Maria Manuel Conceição

Espetáculos

DIREÇÃO DE PRODUÇÃO
Mariana Cardoso de Lemos
PRODUÇÃO
Jorge Epifânio
Clara Troni
Ana Rita Santos

Exposições

DIREÇÃO DE PRODUÇÃO
Mário Valente
PRODUÇÃO
António Sequeira Lopes
Fernando Teixeira
Susana Sameiro (Culturgest Porto)
ASSESSORIA E PRODUÇÃO
Sílvia Gomes
AUXILIAR
Rui Assunção (Culturgest Porto)
LIVRARIA
Rosário Sousa Machado

Participação / Famílias e Escolas

PRODUÇÃO
João Belo
ESTAGIÁRIAS
Antónia Honrado
Carla Monteiro

Atividades Comerciais

DIREÇÃO
Catarina Carmona
ASSISTENTE
Sofia Fernandes

Equipa Técnica

DIREÇÃO TÉCNICA
José Rui Silva
DIREÇÃO DE CENA
José Manuel Rodrigues
TÉCNICOS AUDIOVISUAIS
Américo Firmino (coordenador)
Ricardo Guerreiro
Suse Fernandes
ILUMINAÇÃO
Fernando Ricardo (chefe)
Vítor Pinto
MAQUINARIA
Nuno Alves (chefe)
Artur Brandão
TÉCNICO DE PALCO
Vasco Branco
AUXILIAR
Nuno Cunha

Comunicação

DIREÇÃO DE COMUNICAÇÃO
Catarina Medina
ESTAGIÁRIA
Liliana Vaz

Assessoria e Serviços de Comunicação

CONTEÚDOS E MATERIAIS PROMOCIONAIS
Maria João Santos
DESIGN GRÁFICO
Studio Maria João Macedo
ASSESSORIA DE IMPRENSA
Helena César

VÍDEO
Pedro Gancho
Sara Morais

Arquivo e Contéudos

Paula Tavares dos Santos

Serviços Administrativos e Financeiros

DIREÇÃO
Cristina Nina Ferreira
ASSISTENTES
Paulo Silva
Teresa Figueiredo

Frente de Casa e Bilheteira

DIREÇÃO
Rute Sousa
BILHETEIRA
Manuela Fialho
Edgar Andrade

PARCERIA
Fidelidade – Companhia de Seguros
Culturgest

PARCERIA CIENTÍFICA
Instituto Superior Técnico da Universidade de Lisboa (IST)

CONSULTORES CIENTÍFICOS
Arlindo Oliveira (IST)
Ana Paiva (IST)
Mário Figueiredo (IST)

CURADORIA
Arlindo Oliveira
Ana Paiva
Liliana Coutinho
Mário Figueiredo

REVISÃO E EDIÇÃO DE CONTEÚDOS
Maria João Santos

DESIGN GRÁFICO
Studio Maria João Macedo

ILUSTRAÇÃO
Mantraste

Impressão: Maiadouro
Tiragem: 800 exemplares

© da publicação:
Culturgest, 2019

A inteligência artificial impõe-se cada vez mais na realidade das sociedades contemporâneas. Novos desenvolvimentos tecnológicos nascem todos os dias mas raramente o seu impacto é devidamente refletido na esfera pública. Assumindo a importância de conhecer e discutir esta realidade, este ciclo de debates promove o olhar e a reflexão sobre as aplicações atuais da inteligência artificial, as suas implicações sociais nas mais variadas dimensões (da saúde à privacidade, à empregabilidade e outras) e a forma como se imagina o futuro neste novo paradigma.

Culturgest FIDELIDADE