INTELIGÊNCIA ARTIFICIAL

2 IMPLICAÇÕES

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

Implicações é o tema da segunda sessão deste ciclo de debates e conferências dedicado à inteligência artificial e é também o mote dos textos publicados nesta brochura – na qual mantivemos, e manteremos também na terceira publicação, o glossário publicado anteriormente, para um melhor acompanhamento dos mesmos.

Em entrevista, Martin Ford, engenheiro e autor de várias publicações sobre inteligência artificial, responde a algumas preocupações relacionadas com o seu impacto no emprego, na economia, na segurança, na educação, e na equidade social e económica, assim como questões sobre os preconceitos de género e raciais que podem ser transferidos e exponenciados através do uso irresponsável da inteligência artificial.

Em *Da moral da máquina à maquinaria moral,* Luís Moniz Pereira – investigador do Centro NOVA LINCS - departamento de Informática da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa e membro do Comité Coordenador Europeu para a Inteligência Artificial –, aludindo à crescente autonomia dos agentes computacionais e ao impacto que têm na vida biológica e nas relações sociais, defende a necessidade de conceber uma ética para as máquinas, considerando a responsabilidade das suas ações num campo de experiência mais vasto que inclui humanos e planeta.

Manuel Dias, da Microsoft Portugal e da DSPA (Data Science Portuguese Association), também professor na NOVA Information Management School, em *Inteligência Artificial – impactos atuais* e futuros acentua o potencial benéfico da inteligência artificial quando é considerada e usada "em prol de um bem maior e não como um fim em si", referindo-se ao imaginário assustador, em parte forjado pela ficção científica, como um sinal, não das suas potencialidades em si, mas do "mal que a natureza humana pode fazer quando dotada destas capacidades".

É também sob a égide da responsabilidade, de quem desenvolve e de quem usa estas tecnologias, que Virginia Dignum, professora de Inteligência Artificial Social e Ética no departamento de Ciência Computacional da Universidade de Umeå (Suécia) e membro do grupo de especialistas de alto nível da Comissão Europeia em inteligência artificial, coloca o seu texto, lembrando que "qualquer que seja o nível de autonomia e consciência social e a capacidade de aprendizagem, os sistemas de inteligência artificial são artefactos construídos por pessoas para cumprir certos objetivos".

Todas as contribuições deixam antever a importância de debater estes assuntos como possibilidade de podermos, enquanto cidadãos (fazendo uso das palavras de Moniz Pereira), "exercer uma cidadania consciente e crítica", influente no rumo que o nosso mundo pode tomar com o crescimento exponencial das tecnologias ligadas à investigação na área da inteligência artificial.

OS DESAFIOS DA INTELIGÊNCIA ARTIFICIAL

Entrevista conduzida por Liliana Coutinho, programadora de Conferências e Debates na Culturgest e curadora do ciclo *Inteligência Artificial: Aplicações, Implicações e Especulações.*

MARTIN FORD

P: Tem vindo a trabalhar nas implicações sociais e económicas do uso de robôs e da inteligência artificial (IA). Uma das áreas que preocupa a opinião pública, tendo em conta essas implicações, é a do emprego – uma apreensão sempre presente na história dos desenvolvimentos tecnológicos. Quais são as suas ideias em relação a este tópico, pensando especificamente nas particularidades que a IA pode trazer?

R: Não há dúvida de que a IA irá tanto destruir como criar emprego. A minha opinião é que o número de empregos destruídos pela tecnologia poderá, no final, ultrapassar o número de empregos criados. Em última análise, a IA pode ameaçar praticamente qualquer emprego ou tarefa – em qualquer nível de qualificação – que seja fundamentalmente rotineira e previsível na sua natureza. Isso pode perfeitamente vir a representar uma ameaça para metade ou mais de metade da nossa força de trabalho. Tenho dúvidas de que sejam gerados novos empregos em número suficiente para absorver todos os trabalhadores atingidos. É importante notar que, na maioria dos casos, a IA e a robótica, mais do que completar trabalhos, irão eliminar tarefas. Á medida que partes significativas das tarefas vão sendo automatizadas, é provável que haja uma consolidação que, de um modo geral, resulte em menos postos de trabalho. Talvez sejam necessários uns 15-20 anos até que seja óbvio o impacto absoluto deste processo. Outra questão é se os novos empregos criados serão acessíveis à maioria dos trabalhadores. À medida que a IA evolui, os empregos criados poderão requerer altos níveis de competência – ou qualidades, como a criatividade. Contudo, não é razoável esperar que a maioria dos trabalhadores, especialmente aqueles que executam tarefas rotineiras e previsíveis, seja capaz de transitar para estas novas áreas.

P: Um dos instrumentos que propôs para fazer face à desigualdade social que pode surgir pela utilização da IA foi a criação de um rendimento básico universal. Seria possível adotar esta solução a nível global?

R: Acredito que uma política semelhante à do rendimento básico universal ou rendimento mínimo garantido seria uma boa forma de lidar com a desigualdade gerada pela IA e pela robótica. Um dia vai ser inevitável seguir nessa direção. Contudo, num futuro próximo, penso que isso dependerá das políticas nacionais. É uma solução que pode funcionar em todos os países do mundo, mas cada país terá de adotar a sua própria política e, claro, o nível de rendimento fornecido irá ser bastante diferente de país para país.

P: Tendo em conta que, neste momento, há uma grande quantidade de sistemas de ensino ainda focados no ser humano como sendo parte de uma cadeia de trabalho e produção, qual o impacto que a IA pode ter na educação? (Estou a pensar, por exemplo, no facto do MIT, quando recruta alunos pós-graduados e investigadores na área da IA, dar importância aos seus interesses nas artes e nas humanidades). As competências pessoais (*soft skills*) e a criatividade serão uma mais-valia?

R: É verdade que as tarefas e os trabalhos mais difíceis de automatizar são os que envolvem criatividade ou requerem o desenvolvimento de relações sofisticadas com outras pessoas. Isto pode aumentar o valor da educação nas artes criativas ou nas humanidades. Também é verdade que, um destes dias, um rendimento básico pode fazer com que as pessoas passem mais tempo a trabalhar em áreas menos valorizadas pelo mercado hoje em dia (como é o caso das artes).

P: Outra grande preocupação é a segurança, nos seus diferentes níveis: dados, privacidade, desenvolvimento de armas, entre outros. É como no mito de Frankenstein – que, não por acaso, é lembrado na introdução de *Disposições de Direito Civil sobre a Robótica*, a primeira resolução emitida, em 2017, pelo Conselho Europeu sobre estas tecnologias. Podemos estar a criar tecnologias que se irão virar contra nós? Ou é um risco apenas porque estamos a projetar parte do nosso comportamento habitual, enquanto seres humanos, nas máquinas?

R: Sim. Existem preocupações óbvias em relação à segurança dos sistemas de IA. Num futuro próximo, a principal preocupação é a IA poder ser usada contra nós por outras pessoas. Isto inclui adversários políticos ou militares mas também *hackers* e cibercriminosos. Por isso, construir sistemas seguros contra os ciberataques e a pirataria é um dos desafios mais importantes, considerando que a IA é cada vez mais utilizada em aplicações críticas. Mais à frente, num futuro longínquo, muitos especialistas temem que a IA possa tornar-se "superinteligente" ou, por outras palavras, muito mais inteligente que o Homem. Para lá desse limite, pode ser muito complicado controlar tal sistema. Algumas pessoas muito inteligentes levam isto muito a sério e estão a trabalhar neste problema. No entanto, neste momento, ainda é ficção científica. Não estamos nem perto de construir um computador que esteja próximo de ter a capacidade da mente humana, e ninguém sabe quanto tempo é que isso pode demorar. Pode levar até 100 anos ou mais.

P: No seu mais recente livro *Architects of Intelligence* [Arquitetos da Inteligência] encontramos entrevistas a alguns dos mais proeminentes investigadores e programadores em IA. Pode resumir as principais preocupações éticas contemporâneas que encontrou nestas conversas?

R: As questões éticas surgiram em quase todas as entrevistas e essas preocupações já estão a ter impacto no mundo real. Uma das mais importantes é o preconceito nos algoritmos, possivelmente com base na raça ou no género. Isto acontece porque os dados usados para treinar os algoritmos são gerados por pessoas e, logo, contêm os preconceitos das pessoas. Encontrámos esse problema nos sistemas de IA usados para avaliar candidaturas a empregos, por exemplo. É um problema bem conhecido na comunidade de IA e há várias pessoas a tentar resolvê-lo. Penso que há boas razões para ter esperança porque, no fundo, corrigir o preconceito num algoritmo é muito mais fácil do que corrigi-lo numa pessoa. Um futuro onde confiamos mais na IA para tomar decisões pode, na verdade, significar menos discriminação. Outra área que colhe grande interesse por parte de muitos investigadores é o uso de tecnologia da IA para criar armas autónomas. Mais de mil investigadores, incluindo muitas das pessoas que entrevistei para o meu livro, assinaram um compromisso de nunca trabalhar nessas armas e grande parte contribuiu para levar a cabo uma iniciativa nas Nações Unidas para banir totalmente as armas autónomas.

P: O que seria necessário para criar um imaginário mais frutífero e pragmático, algo que ajudasse a focar os usos desta tecnologia para lidar com os grandes desafios do planeta – como as alterações climáticas, por exemplo?

R: É a visão de muitas pessoas na comunidade da IA. Por exemplo, Demis Hassabis, presidente executivo da DeepMind (uma das pessoas que entrevistei), defende "resolver primeiro a IA e depois usar isso para resolver tudo o resto". Por outras palavras, a IA vai tornar-se a ferramenta mais poderosa que temos para resolver os maiores problemas que enfrentamos, incluindo as alterações climáticas, as energias limpas, as doenças, a pobreza, etc. Isto já está a acontecer e é a razão mais importante para sermos entusiastas e otimistas acerca do futuro da IA. Contudo, é necessário abordar os desafios que irão surgir com a tecnologia – tais como a desigualdade e o desemprego – para ter a certeza de que podemos impulsionar esses avanços em benefício de todos.

“A minha opinião é que o número de empregos destruídos pela tecnologia poderá, no final, ultrapassar o número de empregos criados.”

“As tarefas e os trabalhos mais difíceis de automatizar são os que envolvem criatividade. Isto pode aumentar o valor da educação nas artes criativas ou nas humanidades.”

“Há boas razões para ter esperança porque, no fundo, corrigir o preconceito num algoritmo é muito mais fácil do que corrigi-lo numa pessoa.”

“Acredito que uma política semelhante à do rendimento mínimo garantido seria uma boa forma de lidar com a desigualdade gerada pela IA e pela robótica.”

DA MORAL DA MÁQUINA À MAQUINARIA MORAL

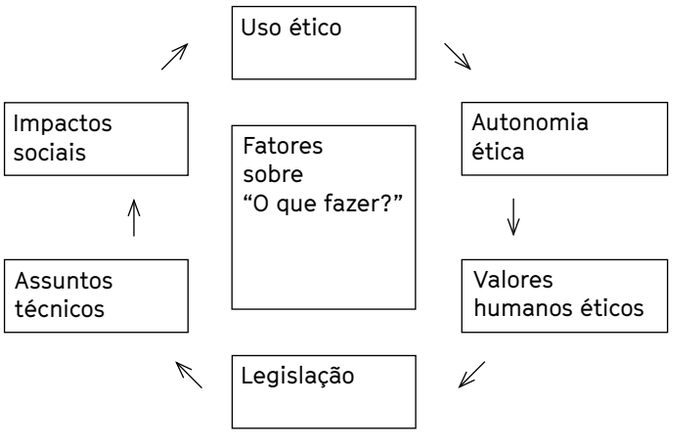
Da moral da máquina à maquinaria moral é o título do novo livro de Luís Moniz Pereira em coautoria com António Lopes, a publicar brevemente pela NOVA-FCT Editorial.

LUÍS MONIZ PEREIRA

Estamos numa encruzilhada entre a inteligência artificial (IA), a ética das máquinas, e os seus impactos sociais. E porquê uma ética para as máquinas? Porque os agentes computacionais tornaram-se mais sofisticados, mais autónomos, atuam em grupo, e formam populações que incluem humanos; porque estes agentes estão a ser desenvolvidos numa variedade de domínios, onde questões complexas de responsabilidade exigem maior atenção, nomeadamente em situações de escolha ética; e porque, uma vez que a sua autonomia está a aumentar, o requisito de que funcionem com responsabilidade, ética e de modo seguro, é uma preocupação crescente.

O surgimento de ferramentas de *deep learning* sobre *big data* permitiu tratar os dados numa quantidade e qualidade até agora impensáveis. Além disso, os algoritmos em geral são cada vez mais capacitados para tomarem decisões autónomas e é agora pensável a implementação dessa tecnologia em robôs com variadas e diversas funções. De facto, o estado atual da IA faz emergir uma problemática incontornável: os seres humanos não serão os únicos agentes autónomos, com capacidade para deliberar sobre aspetos que impactam diretamente na nossa vida. Neste contexto, a deliberação autónoma e criteriosa reclama por regras e princípios de natureza moral aplicáveis à relação entre máquinas e seres humanos e aos impactos da entrada destas máquinas no mundo do trabalho e na sociedade em geral. O atual desenvolvimento da IA, tanto na sua capacidade de elucidação dos processos cognitivos emergentes na evolução, quanto na sua aptidão tecnológica para a conceção e produção de programas informáticos e artefactos inteligentes, constitui-se, na verdade, como o maior desafio intelectual do nosso tempo.

A complexidade destas questões ilustra-se sinteticamente no carrossel apresentado em baixo, que reúne as problemáticas interconexas que compõem os fatores para as decisões sobre a constituição de máquinas éticas. Percebe-se por aí que o tema da moral computacional interessa não apenas às empresas e às instituições públicas, mas também a quem queira exercer uma cidadania consciente e crítica.



Carrossel da maquinaria ética

Do ponto de vista do paradigma sobre o que é a evolução e a cognição, as investigações têm evidenciado uma perspetiva integradora. É possível ver a inteligência como resultado de uma atividade de processamento de informação, traçar uma linha evolutiva que vai dos genes aos memes, e a sua coevolução. Nestes termos, ruturas tradicionais entre o ser humano e os restantes animais, ou entre cultura e natureza, passam a fazer pouco sentido. Toda a vida é um palco evolucionário, onde a replicação, a reprodução e a recombinação genética têm ensaiado soluções para uma cognição e uma ação cada vez mais aprimoradas e distribuídas. A Biologia, dada a sua matriz computacional, instaura sobre a Física uma primeira artificialidade. Assim sendo, o atual estado do conhecimento implica uma redefinição do lugar do ser humano no mundo, lançando desafios a várias áreas do conhecimento. Desde logo a muitas disciplinas da Filosofia, pois problemas como "o que é conhecer", "o que é o Homem", "o que são e como surgiram valores de natureza moral", ganham perspetivas até agora impensáveis.

No que diz respeito ao conhecimento propriamente dito, surge a possibilidade de o mesmo ser simulado em computadores, superando os limites antes impostos por uma especulação que não podia passar da experiência mental, quiçá compartilhada. Quanto ao questionamento antropológico, a tradicional discussão sobre "o que é o Homem?", mercê do cruzamento entre a IA, a engenharia genética e a nanotecnologia, vê-se agora substituída por uma poderosa e desafiante problemática em torno daquilo que pode vir a considerar-se desejável e possível que seja e irá sendo o Homem.

Do ponto de vista dos critérios de ação, a moral alcandorada nos céus do passado confronta-se com uma nova perspetiva sobre os sistemas morais nascentes, estudados no âmbito da Psicologia evolucionária e aprofundados através de modelos testáveis em cenários artificiais, agora permitido pelos computadores. À medida que a investigação avança, podemos conhecer melhor os processos inerentes à decisão moral, ao ponto de poderem ser "ensinados" a máquinas autónomas capacitadas para manifestarem discernimento ético.

No domínio da Economia há toda uma problemática associada ao impacto no trabalho e à dignidade que lhe é inerente, assim como à produção e distribuição da riqueza; ou seja, toda uma reconfiguração das relações económicas que resultará não apenas da automação de atividades rotineiras, mas fundamentalmente da entrada em cena de robôs e *software* que poderão substituir médicos, professores ou assistentes em lares de terceira-idade (para dar nota de profissões às quais o olhar comum não perceciona como facilmente substituíveis). O conhecimento deste contexto é especialmente relevante, exigindo tomadas de posição que sustentarão a necessidade de uma moral social atualizada, um renovado contrato social. A problemática da moral computacional ganha assim uma existência num contexto em que o ecossistema do conhecimento ficará bastante enriquecido, pois terá de incorporar agentes não-biológicos com capacidade para se tornarem intervenientes ativos em dimensões que, até agora, têm sido atribuídas exclusivamente a humanos.

"Os seres humanos não serão os únicos agentes autónomos, com capacidade para deliberar sobre aspetos que impactam diretamente na nossa vida."

"Problemas como 'o que é conhecer','o que é o Homem', 'o que são e como surgiram valores de natureza moral', ganham perspetivas até agora impensáveis."

“Assistimos a um desenvolvimento técnico acelerado sobretudo no campo das redes neuronais, motivado pela proliferação exponencial de dados disponíveis para as treinar.”

INTELIGÊNCIA ARTIFICIAL: IMPACTOS ATUAIS E FUTUROS

MANUEL DIAS

O potencial da informação que produzimos e utilizamos diariamente permite-nos pensar em inúmeras possibilidades de aplicação de inteligência artificial (IA) com um objetivo primordial – uma sociedade melhor. Facilmente pensamos em carros autónomos, que calculam e otimizam o nosso percurso reduzindo o risco de acidente, em assistentes virtuais inteligentes que falam connosco e antecipam as nossas necessidades, ou em sistemas avançados de reconhecimento de imagem com ampla aplicação em vários ramos das ciências médicas, como por exemplo o apoio ao diagnóstico e tratamento do cancro, entre outras doenças.

Apesar disto e fruto da conotação com filmes de pura ficção científica, é fácil ficarmos apreensivos quando ouvimos falar de superinteligência ou, numa nomenclatura mais técnica, de IA geral. Muitos dos receios, pelo menos os mais fundamentados, não são apenas e exclusivamente devidos às capacidades de algoritmos que não compreendemos, mas sobretudo ao que de mal a natureza humana pode fazer quando dotada destas capacidades. Um destes exemplos foi o *chatbot* de conversação que a Microsoft criou – a Tay – que, após várias interações com utilizadores *online,* se tornou racista, ou o sistema de reconhecimento de imagem da Google que identificou um casal de pessoas negras como gorilas. Podemos argumentar que as máquinas aprendem a ser racistas da mesma forma que os humanos, por isso as implicações práticas intencionais ou acidentais merecem uma discussão mais profunda, que extravasa em muito os conceitos matemáticos de aprendizagem automática (AA), onde a ética e a justiça assumem um papel fundamental.

Num estudo recente da Casa Branca sobre os principais riscos na ciência de dados, uma das conclusões primordiais foi a necessidade de auditoria a algoritmos, como forma de garantir a justiça das decisões, o não enviesamento dos dados e, principalmente, a salvaguarda dos valores morais que regem o comportamento humano. É fundamental que o desenvolvimento da IA seja feito segundo normas abertas e visíveis para o público em geral, capaz de ser explicada e entendida por todos e realizada por uma ampla comunidade de investigadores, públicos ou do setor privado.

Ao mesmo tempo que esta discussão ocorre, assistimos a um desenvolvimento técnico acelerado sobretudo no campo das redes neuronais e nas múltiplas arquiteturas de redes neuronais profundas, motivado pela proliferação exponencial de dados disponíveis para as treinar e pela capacidade colossal de computação disponível na *cloud*. O resultado final traduz-se em algoritmos com uma precisão ao nível dos humanos em áreas como o processamento de voz, o reconhecimento de imagens ou a tradução de texto, que depois são integrados nas mais diversas soluções, muitas delas imersas na nossa experiência digital diária.

É por isso relevante perceber e evidenciar os potenciais benefícios da aplicação de IA em prol do ser humano, onde a inovação em campos como a saúde, a educação ou a sustentabilidade do planeta, pode ter implicações radicais na vida de todos nós. Na área médica, um dos exemplos mais importantes

é o tratamento do cancro: atualmente é possível tirar partido do reconhecimento de imagem para o diagnóstico precoce de tumores e, depois de diagnosticado, identificar o tecido maligno para a reconstrução tridimensional do tumor e assim poder aplicar uma terapia altamente orientada e menos lesiva para o corpo. Outra área fortemente beneficiada pela IA é o tratamento personalizado do doente, baseado em algoritmos de AA que, para além do reconhecimento de padrões mais complexos, permitem antecipar e prever alguns tipos de doença. Se extrapolarmos para a genómica, a quantidade de informação disponível no futuro para treinar algoritmos cada vez mais precisos, mudará radicalmente a medicina tal como a conhecemos hoje.

Na área da sustentabilidade ambiental importa mencionar alguns exemplos de aplicação da IA com implicações enormes para o planeta. No que se refere às alterações climáticas, que todos conhecemos, a utilização de algoritmos preditivos cada vez mais precisos poderá reduzir o seu impacto, tanto na segurança, nas infraestruturas e mesmo na saúde. Na agricultura, o processamento de imagens de satélite de alta resolução pode ser usado para o rastreio de zonas florestais, a previsão de incidências, a existência de água, ou na ajuda da preservação de espécies.

Estamos por isso numa época única da nossa história onde, apesar do desenvolvimento da IA estar no seu início, o potencial de benefícios para o ser humano e para as organizações é enorme, mas difícil de quantificar. A sua aplicação generalizada carece, por isso, de uma reflexão aberta e alargada a toda a sociedade, investigadores, universidades, entidades reguladoras, profissionais de ciências de dados e às grandes tecnológicas mundiais, que reúnem um vasto repositório de informação. Só assim conseguiremos progredir mantendo os nossos valores éticos e humanos, utilizando a tecnologia a favor de um bem maior e não como um fim em si.

> “Num estudo recente sobre os principais riscos na ciência de dados, uma das conclusões primordiais foi a necessidade de auditoria a algoritmos como forma de garantir a salvaguarda dos valores morais que regem o comportamento humano.”

A RESPONSA-BILIDADE É NOSSA

VIRGINIA DIGNUM

Atualmente, grande parte da discussão sobre a inteligência artificial (IA) centra-se nas eventuais consequências negativas do uso de tecnologias inteligentes e no facto da revolução em IA, como muitos lhe chamam, estar a acontecer a uma velocidade cada vez maior. Um momento! Somos nós que fazemos a IA acontecer, ela não surge por mão própria! Somos nós que introduzimos a IA, estamos conscientes dos seus potenciais perigos, por isso a responsabilidade em intervir é nossa! Isto significa que temos o controlo e significa também que somos responsáveis por qualquer coisa que a IA desenvolva.

O interesse dos *media* na IA quase nos faz acreditar que se trata de uma nova tecnologia que "repentinamente" está a tomar conta do mundo. Na verdade, tem mais de 60 anos enquanto campo de investigação (o termo surgiu na conferência de Dartmouth, realizada no Dartmouth College, em 1956), ao passo que a ideia de máquinas inteligentes é provavelmente tão antiga como a humanidade e tem sido um dos principais motores da informática. Os avanços em IA e em aprendizagem automática (AA) devem-se a esforços contínuos durante décadas sob situações maioritariamente desfavoráveis. Quem trabalhasse em IA nos anos 1990 ou 2000 era basicamente um falhado, alguém que tinha perdido o contacto com a realidade. Contudo, é o trabalho desse período que tem servido de base aos resultados de hoje.

De facto, há já algum tempo que as máquinas têm vindo a tomar decisões por nós. Os algoritmos decidem qual o melhor percurso para o encaminhamento das nossas comunicações móveis, as portas automáticas de uma estação de comboios dão-nos acesso ao cais com base na informação obtida no nosso crédito de viagem e, do trilião de páginas existentes, o Google decide quais as que temos maior propensão a ler quando fazemos uma pesquisa *online*. Por isso, as decisões tomadas por máquinas não são nada de novo. O que torna a decisão tomada por IA diferente, e assustadora para alguns, é a complexidade cada vez maior e o potencial impacto dessas decisões, muitas vezes resultado de algoritmos que são difíceis, ou mesmo impossíveis, de perceber, e as decisões de IA serem frequentemente feitas sem intervenção humana. Contudo, somos nós que determinamos os objetivos de otimização e as funções de utilidade que estão na base dos algoritmos de AA, e decidimos o que a máquina deve estar a maximizar.

Para assegurar que esses futuros distópicos não se tornam realidade, estes sistemas devem ser introduzidos de modo a inspirar confiança e compreensão, e a respeitar os direitos civis e humanos. A necessidade de considerações éticas no desenvolvimento de sistemas interativos inteligentes está a tornar-se uma das áreas de investigação mais influentes dos últimos anos, convergindo na apresentação de várias iniciativas de investigadores e profissionais, incluindo a iniciativa global da IEEE (Instituto de Engenheiros Eletrotécnicos e Eletrónicos, sediada nos EUA) sobre *Ética dos Sistemas Autónomos* (https://ethicsinaction.ieee.org), e às mais recentes *Diretrizes*

“Estamos a construir IA para otimizar o desempenho das empresas ou para otimizar o rendimento agrícola dos pequenos agricultores? Estamos a pensar no dinheiro ou estamos a agir em prol dos melhores interesses da sociedade?”

éticas europeias para IA confiável (https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai). Sinto-me honrada por fazer parte destas duas iniciativas.

Em todas as áreas de aplicação, o raciocínio da IA deve ser capaz de ter em conta: os valores da sociedade, considerações morais e éticas, pesar as respetivas prioridades de valores mantidas pelas partes interessadas e em múltiplos contextos multiculturais, explicar o seu raciocínio e garantir a transparência. À medida que aumentam as capacidades para tomar decisões autónomas, talvez a questão mais importante seja a necessidade de repensar a responsabilidade. Isto é, a nossa responsabilidade. Qualquer que seja o nível de autonomia e consciência social e a capacidade de aprendizagem, os sistemas de IA são artefactos construídos por pessoas para cumprir certos objetivos. As teorias, os métodos e os algoritmos são necessários para integrar os valores societais, legais e morais nos progressos tecnológicos em IA, em todas as fases de desenvolvimento (análise, *design,* construção, implantação e avaliação). Estas estruturas devem lidar com o raciocínio autónomo da máquina sobre questões que nós consideramos terem um impacto ético, e, ainda mais importante, precisamos de estruturas que guiem as escolhas de *design,* regulem o alcance dos sistemas de IA, assegurem uma gestão apropriada dos dados e ajudem os indivíduos a determinarem o seu próprio envolvimento. As considerações acima mencionadas mostram que a ética e a IA estão interligadas em diferentes níveis:

Ética *pelo* Design: a integração técnica e/ou algorítmica das capacidades de raciocínio ético enquanto parte do comportamento do sistema autónomo artificial;

Ética *no* Design: os métodos regulatórios e de engenharia que apoiam a análise e a avaliação das implicações éticas dos sistemas de IA por estes integrarem ou substituírem as estruturas sociais tradicionais;

Ética *para* o Design: os códigos de conduta, normas e processos de certificação que asseguram a integridade dos criadores e utilizadores à medida que estes investigam, desenham, constroem, empregam e gerem os sistemas de IA.

Inteligência artificial responsável trata-se, no fundo, da responsabilidade humana para o desenvolvimento de sistemas inteligentes, segundo princípios e valores humanos fundamentais, de forma a assegurar o bem-estar e a prosperidade humana num mundo sustentável. É mais do que assinalar itens numa lista de verificação ética num relatório, ou desenvolver características complementares ou botões de desligar nos sistemas de IA. Pelo contrário, a responsabilidade é fundamental para a autonomia, e deveria ser uma das posturas essenciais subjacentes à investigação em IA.

Cabe a nós tomar essa decisão. Estamos a construir algoritmos para maximizar o lucro dos acionistas ou para maximizar a justa distribuição de recursos numa comunidade, fornecendo soluções para a tragédia de situações comuns e assegurando o acesso livre à informação e à educação? Estamos a construir IA para otimizar o desempenho das empresas ou para otimizar o rendimento agrícola dos pequenos agricultores por todo o mundo, fornecendo informação em tempo real sobre os níveis de fertilização, os períodos de plantação e de colheita e as condições do tempo? Estamos a pensar no dinheiro ou estamos a agir em prol dos melhores interesses da sociedade? Estamos a basear os avanços da IA nos valores dos acionistas ou nos direitos e nos valores humanos?

Nós somos responsáveis.

“A ideia de máquinas inteligentes é provavelmente tão antiga como a humanidade e tem sido um dos principais motores da informática.”

“As decisões tomadas por máquinas não são nada de novo. Do trilião de páginas existentes, o Google decide quais as que temos maior propensão a ler quando fazemos uma pesquisa online. O que torna a decisão tomada por IA diferente, e assustadora para alguns, é a complexidade cada vez maior e o potencial impacto dessas decisões.”

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

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.