INTELIGÊNCIA ARTIFICIAL

APLICAÇÕES

1

CICLO FIDELIDADE — CULTURGEST

Conferências e Debates ×

Mário Figueiredo 05
Inteligência artificial:
o que é e de onde veio?

Luísa Coheur 13
Aplicações em Tecnologias da
Linguagem: no mundo dos agentes
conversacionais

Bruno Martins 19
Inteligência artificial
no combate a fraudes

Milind Tambe 23
Inteligência artificial e sistemas
multiagentes para o bem social:
aprender e planear num pipeline de
implantação de dados de ponta a ponta

Glossário 28

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 ciclo de conferências *Inteligência Artificial: Aplicações, Implicações e Especulações*, realizado em parceria científica com o Instituto Superior Técnico – Universidade de Lisboa (IST–UL) e em parceria mecenática com a Fidelidade – Companhia de Seguros, foi concebido para ampliar o conhecimento público e o campo de debate desta áreaa de produção tecnológica, cada vez mais presente no nosso quotidiano e, por isso, com implicações em quase todas as esferas da vida contemporânea. O programa divide-se em três sessões e, em cada sessão, será publicada uma brochura de textos da autoria dos oradores ou de investigadores convidados especificamente para estas edições.

Esta primeira brochura, intitulada *Inteligência Artificial: Aplicações*, coincide com o primeiro debate do ciclo, pensado para apresentar e pôr à discussão o estado atual de algumas das principais aplicações da inteligência artificial, considerando as suas possibilidades e os seus limites. O texto de abertura é da autoria de Mário Figueiredo, doutorado e agregado em Engenharia Eletrotécnica e de Computadores pelo Instituto Superior Técnico – Universidade de Lisboa, onde é professor catedrático, investigador sénior, coordenador de área e líder de grupo no Instituto de Telecomunicações. *Inteligência artificial: o que é e de onde veio?* apresenta tanto uma perspetiva histórica como um panorama acerca da atualidade das aplicações da Inteligência artificial.

Aplicações em Tecnologias da Linguagem: No mundo dos agentes conversacionais, de Luísa Coheur – professora do departamento de Engenharia Informática do Instituto Superior Técnico – Universidade de Lisboa e investigadora no INESC-ID – Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento –, introduz-nos ao universo dos robôs com os quais podemos conversar e às atuais possibilidades de relação entre homem e máquina que decorrem do processamento informático da língua natural, ou seja, da produção de máquinas capazes (ou não) de compreender, produzir e simular a linguagem humana nas suas variadas formas (gestual, escrita e oral).

Em *Inteligência artificial no combate a fraudes*, um debate conduzido por Pedro Bizarro (cofundador e *chief science officer* da Feedzai), Bruno Martins, do departamento de Engenharia Informática do Instituto Superior Técnico – Universidade de Lisboa, discorre sobre uma área em franca expansão: os usos da inteligência artificial nas implementações de sistemas de segurança e no combate de fraudes informáticas, em particular as relacionadas com as transações financeiras.

Milind Tambe, professor Helen N. e Emmett H. Jones de Engenharia na Universidade do Sul da Califórnia (USC), codiretor e fundador do CAIS/USC – Centro de Inteligência Artificial para a Sociedade (centro de investigação centrado na resolução de problemas sociais), aborda o uso destas tecnologias no âmbito da segurança pública, da conservação da vida selvagem e da saúde pública em *IA e sistemas multiagentes para o bem social: Aprender e planear num pipeline de implantação de dados de ponta a ponta*.

Para finalizar, a brochura inclui um breve glossário de conceitos básicos do universo da inteligência artificial. Várias dessas palavras ou expressões caíram já no uso comum, no entanto, por vezes ainda estão sujeitas a alguma incerteza no que diz respeito ao seu significado. Este glossário é uma contribuição de Ana Paiva, Arlindo Oliveira e Mário Figueiredo do IST–UL e consultores científicos deste ciclo.

Estas e outras questões poderão ser respondidas nos debates e conferências que integram o programa deste ciclo. Contamos que estas breves publicações ajudem a prolongar as discussões para além das salas da Culturgest.

INTELIGÊNCIA ARTIFICIAL: O QUE É E DE ONDE VEIO?

MÁRIO FIGUEIREDO

Inteligência artificial; aprendizagem automática: estes termos ainda recentemente exclusivos dos meios académicos e tecnológicos, têm vindo, na última década, a integrar a linguagem pública quotidiana. A inteligência artificial é hoje tema frequente de discussão nos *media* e no discurso político, económico, social, científico e mesmo filosófico. Porquê? Porque estas técnicas têm actualmente um enorme impacto em muitas vertentes do funcionamento das sociedades modernas: do comércio ao entretenimento, dos *media* às ciências, das finanças à política, da saúde à educação. Poucas são as áreas que se mantiveram imunes.

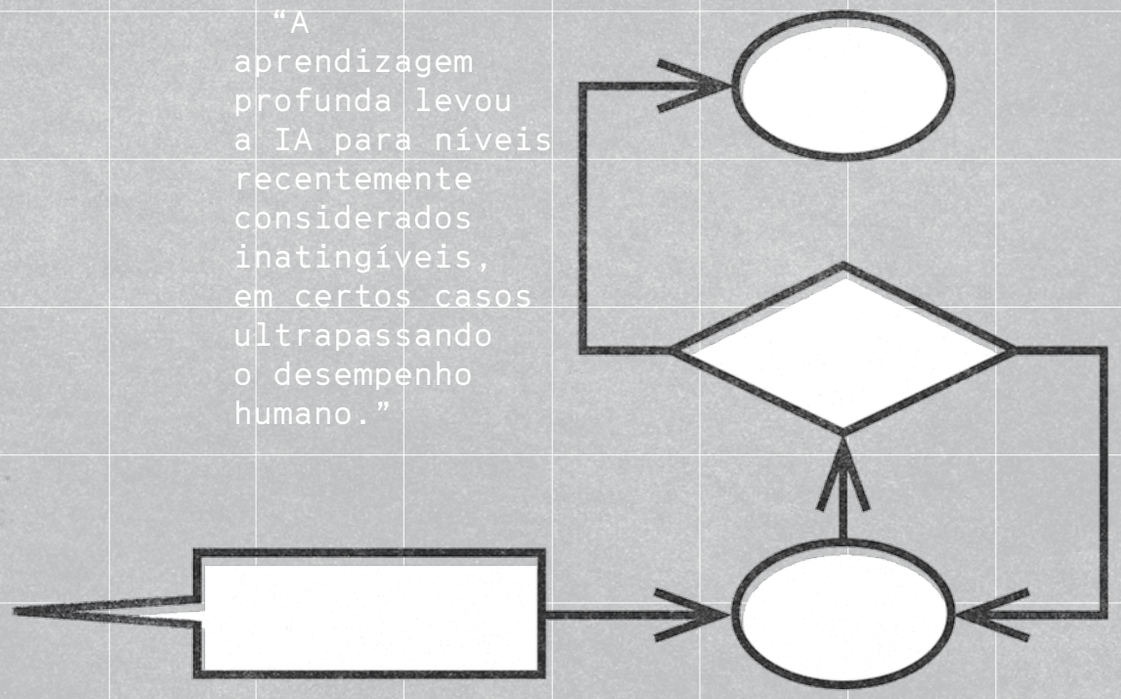
Para a população geral, estes conceitos, embora ouvidos frequentemente, podem ser ainda algo nebulosos, não sendo claro o seu significado preciso, como se relacionam e quais os seus (actuais e potenciais) impactos na sociedade. Este texto explica muito sucintamente o que são a inteligência artificial (IA) e a aprendizagem automática (AA ou *machine learning*). Numa perspectiva histórica, refere-se como o grande desenvolvimento destas técnicas está intimamente ligado à generalização da Internet e a uma mudança radical do modelo de negócio em torno do acesso à informação, bem como à expansão explosiva das redes sociais.

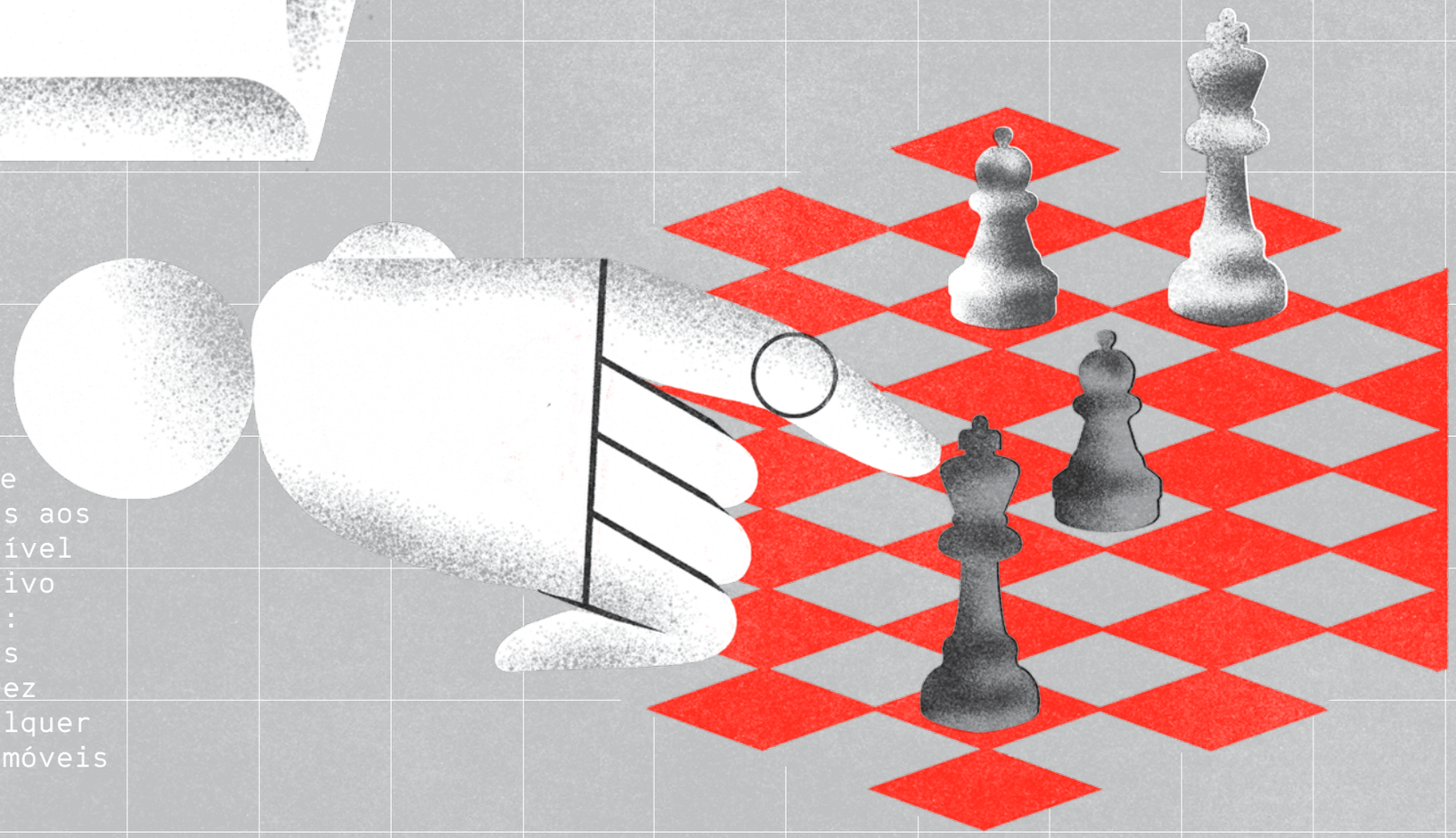
Não é fácil definir IA, pois não existe uma definição consensual de "inteligência"; pode dizer-se que é um daqueles fenómenos que "não sabemos definir bem, mas reconhecemos quando o vemos" (parafraseando o juiz Potter Stewart do Supremo Tribunal dos Estados Unidos da América quando solicitado a definir pornografia). No entanto, numa definição genericamente aceite, IA (termo criado por John McCarthy em 1956) é a exibição de comportamento "inteligente" por máquinas, normalmente computadores (um ou vários). Há muito que a humanidade persegue o objectivo (talvez sonho seja uma melhor palavra) de construir sistemas artificiais capazes de comportamentos "inteligentes". O trabalho académico e científico em IA teve início na década de 1950, podendo considerar-se Alan Turing como o pioneiro da colocação do seu estudo em alicerces científicos sólidos. Passado mais de meio século, existem hoje muitos sistemas aos quais é impossível negar o adjectivo "inteligentes": desde programas que jogam xadrez melhor que qualquer humano, a automóveis autónomos, passando por sistemas auxiliares de diagnóstico médico.

De um sistema de IA, pretende-se que tome decisões (acertadas) ou faça recomendações (boas) baseadas em observações e/ou dados. O acerto ou "bondade" destas decisões ou recomendações é habitualmente formalizado como a maximização e/ou satisfação de um critério de (bom) desempenho; por exemplo, num jogo, o objectivo é ganhar; num veículo autónomo, o objectivo inclui evitar acidentes bem como conduzir rapidamente até ao destino; num sistema de diagnóstico médico, o objectivo é fazer um diagnóstico o mais acertado possível.

Uma das primeiras estratégias para construir sistemas de IA baseou-se em programação explícita por especialistas humanos.

“A aprendizagem profunda levou a IA para níveis recentemente considerados inatingíveis, em certos casos ultrapassando o desempenho humano.”





“Existem hoje muitos sistemas aos quais é impossível negar o adjectivo ‘inteligentes’: desde programas que jogam xadrez melhor que qualquer humano, a automóveis autónomos.”

Um dos pináculos dessa abordagem é o famoso *Deep Blue,* da IBM, um computador que, em 1997, derrotou o então campeão mundial de xadrez, Garry Kasparov. Predominante nos anos 80, essa linha de trabalho levou à criação dos chamados "sistemas periciais" (*expert systems*), cujo objectivo era simular o processo de raciocínio dos peritos de uma dada área (por exemplo, diagnóstico médico), para o automatizar. O ponto frágil desta classe de métodos é a aquisição de conhecimento, o qual tem de ser explicitamente incluído no sistema, exigindo que especialistas da área de aplicação sejam capazes de explicitar (por introspecção) os processos mentais que usam, o que nem sempre é fácil, ou mesmo possível. É normal que um perito não consiga descrever de forma explicita e precisa o processamento de informação visual ou auditiva (não totalmente consciente) subjacente às suas decisões. Por exemplo, a um cardiologista, é difícil ou mesmo impossível descrever de forma utilizável por um computador o motivo pelo qual classifica como sopro cardíaco um determinado som que ouve em auscultação.

A chamada aprendizagem automática ataca directamente o problema da aquisição de conhecimento (aprendizagem), elevando-o à categoria de "o" problema central da IA. O objectivo é desenvolver e estudar sistemas/algoritmos capazes de aprender a tomar decisões a partir de observações, sem que os critérios subjacentes a essas decisões sejam explicitamente programados. Um sistema de AA aprende a partir de muitos exemplos de observações e respectivas decisões pretendidas (os chamados dados de treino ou de aprendizagem), sendo desejável que consiga generalizar, isto é, tomar decisões "parecidas" na presença de observações "parecidas". Por exemplo, os sistemas que nas máquinas fotográficas modernas identificam a existência de caras humanas na imagem, não foram explicitamente "programados" (é impossível explicitar em detalhe como detectar uma cara), mas sim "ensinados" a partir de um grande conjunto de exemplos – pedaços de imagem, cada um previamente etiquetado com a presença ou ausência de uma cara, tarefa fácil (mas fastidiosa) para um ser humano. Este tipo de AA designa-se aprendizagem supervisionada, dada a presença das decisões "certas" nos dados de treino. Existem outros cenários nos quais esta informação está ausente (aprendizagem não supervisionada) ou é apenas parcial (aprendizagem semi-supervisionada).

A aprendizagem supervisionada teve um papel central no recente e notável sucesso da IA/AA em muitas áreas de aplicação. Em particular, na última década, uma classe de métodos parcialmente inspirada na arquitectura neuronal do cérebro, conhecida como "aprendizagem profunda" (*deep learning*, ou *deep neural networks* – redes neuronais profundas), levou o estado da arte em muitas tarefas de IA para níveis recentemente considerados inatingíveis, em certos casos ultrapassando o desempenho humano. Reconhecimento de objectos, caras ou texto em imagens, reconhecimento de fala, tradução automática e vários tipos de diagnóstico médico (nomeadamente a partir de imagens) são casos de sucesso da aprendizagem profunda.

Embora este tipo de abordagem tenha começado a ser desenvolvida nas décadas de 1970 e 1980, o elevado custo computacional e as grandes quantidades de dados-treino necessárias não permitiam, com os computadores dessa época, atacar problemas de dimensão realista, com impacto prático; durante décadas, a aprendizagem profunda esteve confinada a meios académicos. Na última década, o enorme poder de cálculo (milhões de vezes superior, por unidade de custo e energia, ao de há 40 anos), o aumento da capacidade de armazenamento, o acesso eficiente a grandes quantidades de dados, um intenso investimento em investigação e desenvolvimento, permitiram que estas técnicas exibissem o seu extraordinário potencial e impacto.

Para entender alguns dos factores subjacentes a estes desenvolvimentos, considere-se agora uma perspectiva histórica. Até ao início do século XXI, o desenvolvimento da Internet focava-se na disponibilização de conteúdos, sendo o simples acesso a esses conteúdos, muito menos abundantes do que hoje, um valor em si. A indústria cresceu em torno da criação e expansão da infraestrutura para produzir, armazenar e disponibilizar conteúdos, tendo nesse período surgido muitos operadores de rede (fixa e móvel) e portais Web e a presença online dos media tradicionais. Na primeira década deste século, a superabundância de conteúdos (em parte devida ao facto de qualquer pessoa poder hoje criar e disponibilizar conteúdos à escala global) retirou valor económico ao seu simples fornecimento, forçando uma alteração radical do modelo de negócio. O valor transferiu-se do fornecimento de conteúdos para a sua procura, selecção e recomendação, com grande prejuízo para as empresas que disponibilizam conteúdos produzidos profissionalmente, quase sempre com grandes custos (em especial os *media* tradicionais: jornais, televisão). O modelo de negócio passou a basear-se, não no fornecimento dos conteúdos, mas sim na apresentação de publicidade correlacionada com os resultados das procuras ou nas redes sociais, personalizada para cada utilizador. Neste contexto, ganharam preponderância duas tecnologias que foram grandes catalisadores do desenvolvimento explosivo das ciências de dados, nomeadamente da IA e da AA: a procura na Web (*Web search*) e os sistemas de recomendação, núcleos tecnológicos das gigantescas empresas da área (Google, Amazon, Facebook). As empresas suportadas em procura, recomendação e/ou publicidade funcionam essencialmente tentando extrair informação (valiosa) sobre cada um de nós, bem como em capturar a nossa atenção através de diversas formas de interacção. Em particular, as redes sociais são essencialmente sofisticadas máquinas para converter atenção humana num bem transacionável de grande valor comercial.

É neste contexto de intensa competição pela nossa atenção e por informação dos hábitos e das preferências de cada um que se expandiram as tecnologias que permitem realizar estas tarefas, estimuladas por avultadíssimos investimentos em infraestruturas e recursos humanos para investigação e desenvolvimento em IA e AA. Ao contrário do que é tradicional noutras indústrias, estas

“IA é um daqueles fenómenos que ‘não sabemos definir bem, mas reconhecemos quando o vemos’.”

“As redes sociais são sofisticadas máquinas para converter atenção humana num bem transacionável de grande valor comercial.”

empresas optaram por uma grande liberalização do acesso às ferramentas e recursos por elas desenvolvidas. Ainda há cinco, 10 anos, as ferramentas de IA/AA mais sofisticadas e poderosas eram usadas quase exclusivamente por especialistas. Hoje, as grandes empresas que dominam tecnologicamente a área (nomeadamente Google e Facebook) disponibilizam gratuitamente as suas próprias ferramentas, havendo também várias empresas que dão acesso às suas poderosas infraestruturas computacionais (*cloud computing*), a custos muito modestos ou mesmo gratuitamente. É hoje possível, mesmo sem formação académica na área, aceder gratuitamente às mais avançadas ferramentas de IA e AA e aprender a usá-las, explorando a enorme quantidade de recursos de formação disponível na Internet. Por outro lado, uma utilização lúcida e consciente destas técnicas, do seu potencial e das suas limitações, exige formação sólida nos fundamentos das mesmas.

Desde a última década do século XX que se verifica uma crescente utilização de IA/AA para abordar problemas difíceis de análise de informação/dados em inúmeras áreas de actividade. Esta tendência iniciou-se em algumas ciências tradicionais (por exemplo, na biologia, tendo levado ao aparecimento da bioinformática), mas rapidamente alastrou a outras áreas, incluindo as ciências sociais e humanas, a medicina, a política, a economia e finanças, e mesmo as artes. Nos últimos anos, a democratização atrás referida do acesso a sofisticadas e poderosas ferramentas de IA/AA (em particular de aprendizagem profunda) acelerou extraordinariamente a expansão destas técnicas. Esta penetração da IA/AA e, em geral, das ciências de dados, em muitíssimos ramos da actividade humana, sugere que urge dotar a sociedade com “literacia” suficiente nessas técnicas para que a população (e não apenas os especialistas) possa entender os seus impactos e consequências. Tal como não é necessário ser profundo conhecedor de mecânica e termodinâmica para conduzir automóveis e para ter consciência de todo o impacto que têm na sociedade, não é necessário que todos conheçam em detalhe o funcionamento dos métodos de IA/AA, mas é crucial que a sociedade esteja informada e consciente, não só do enorme potencial positivo, mas também dos possíveis riscos que o seu uso generalizado pode implicar. Para além das questões de privacidade, para as quais começa (lentamente) a verificar-se uma consciencialização pública, a utilização massiva de técnicas de IA/AA para capturar a atenção das pessoas e influenciar as suas decisões (não só comerciais mas também políticas e outras) pode ter consequências difíceis de prever e que requerem reflexão. A compreensão das implicações sociais e políticas do uso generalizado de IA/AA não é da responsabilidade exclusiva de um conjunto de especialistas da área, mas sim uma responsabilidade da sociedade em geral.

O autor escreve segundo o Antigo Acordo Ortográfico.

APLICAÇÕES EM TECNOLOGIAS DA LINGUAGEM: NO MUNDO DOS AGENTES CONVERSACIONAIS

LUÍSA COHEUR

Um dos grandes objetivos da inteligência artificial é criar máquinas capazes de compreender e produzir linguagem humana na sua forma oral, escrita e/ou gestual. A área do saber que, numa perspetiva computacional, se dedica ao estudo da língua é a das Tecnologias da Linguagem (Humana), sendo responsável pelo desenvolvimento de muitas aplicações usadas pelo grande público, no seu dia a dia. Exemplos são o Google Tradutor ou o corretor ortográfico (e sintático) do Microsoft Word. Outros exemplos são os sistemas de sumarização, os geradores de poesia, os analisadores de sentimentos, os sistemas de pergunta e resposta e os reconhecedores de fala; há ferramentas que tentam detetar *cyberbullying*, identificar traços de personalidade no que escrevemos (muito útil à área de Linguística Forense, por exemplo), ou assinalar as tão faladas (e preocupantes) *fake news*; são também aplicações da área os *chatbots* e as assistentes virtuais, como a Siri da Apple, a Cortana da Microsoft ou a Alexa da Amazon.

O primeiro agente conversacional foi a famosa Eliza, desenvolvida nos anos 60, com o objetivo de simular uma psicoterapeuta. Esta aplicação teve um sucesso estrondoso na época, a ponto de deixar o seu criador preocupado, pois baseando-se em simples regras, não fazia mais que "enganar" o seu interlocutor, levando-o a falar sobre si ("Por favor, continue...", "O que o faz pensar que a sua sogra não o estima?"). Esta ideia de usar características do personagem virtual para explicar o rumo ou mesmo falhas na conversa foi e continua a ser muito usada (Edgar Smith, um mordomo virtual que operava recentemente no palácio de Monserrate, em Sintra, dizia que ouvia mal quando não percebia uma pergunta). Outro conceito que nasceu com os primeiros *chatbots* (quando ainda não se usava esta palavra) e que ainda hoje é explorado é o de "aprender conversando". Por exemplo, se numa interação perguntarmos a um *chatbot* "O que achas dos Beatles?" e este não souber responder, vai dizer qualquer coisa mais ou menos acertada, mas regista a pergunta. Na próxima interação que tiver com um humano, o *chatbot* far-lhe-á essa mesma pergunta e obterá uma resposta. Claro que esta abordagem pode correr mal. Um caso bastante mediático foi o de um *chatbot* que se dispunha a aprender por interação com os humanos e, poucas horas depois de estar em utilização, era racista, nazi e ordinário, e teve de ser desligado.

Mas pondo de lado alguns casos mais caricatos, há um interesse renovado no desenvolvimento de *chatbots* e assistentes virtuais. Estes podem poupar muito dinheiro a uma organização, fazendo, por exemplo, o apoio a clientes. Supondo-os capazes, estes agentes são perfeitos: estão disponíveis 24 horas por dia e não se sentem perturbados perante clientes *sui generis*.

Existem atualmente inúmeras plataformas para a criação de *chatbots*. Tal como em toda a área de Tecnologias da Linguagem, estas plataformas vivem de pré-processamentos específicos (por exemplo, separar uma frase dos seus constituintes ou eliminar palavras com pouca relevância) e alimentam-se de técnicas de aprendizagem automática. Estas são usadas para detetar as intenções do utilizador (o cliente quer pedir uma pizza, uma bebida

ou fazer uma reclamação), bem como identificar as entidades de uma pergunta (por exemplo: TAP, Londres, lugar 2D). Os dados que vão ser usados em todos os processos de aprendizagem ficam a cargo de quem está a desenvolver o *chatbot* e não há grandes milagres: podem-se tentar aproveitar dados de outros domínios, mas todos têm as suas especificidades e, por isso, mais cedo ou mais tarde, serão necessários dados especializados.

Talvez por se estar à espera de um processo simplesmente *plug and play*, há alguma frustração aquando da criação de *chatbots*, que aumenta quando se torna óbvio que o agente virtual não será capaz de manter uma conversa como um agente humano. Na verdade, apesar dos grandes avanços que se têm verificado recentemente em todas as frentes da inteligência artificial (e também na área das Tecnologias da Linguagem) ainda estamos muito longe de conseguir criar máquinas capazes de realmente compreender e produzir linguagem ao nível dos humanos. Mas porque é que é tão complicado pôr uma máquina a falar como nós?

Dado que conseguimos comunicar com (algum) sucesso, nem nos apercebemos que as sequências de palavras que interpretamos e libertamos a toda a hora são muitas vezes ambíguas e extremamente variáveis, apesar de obedecerem a certas regras de sintaxe. Uma palavra pode ter vários significados (por exemplo: secretária), assim como uma frase ("O João e a Maria casaram-se": casaram-se um com o outro ou cada um com outra pessoa?). Das ambiguidades a nível lexical ou sintático (entre outras) emergem frases com sentidos diferentes. O modo como dizemos algo também influencia a interpretação de uma frase ("Bom dia" pode ser dito num tom desagradável a quem chega atrasado, como quem diz "Finalmente!"). Do mesmo modo, o contexto em que uma palavra ocorre leva a que, por vezes, esta perca o seu significado habitual. É o que se passa nas expressões idiomáticas ("bater a bota", "chover a potes") e, numa escala muito maior em termos de produção, nas chamadas "colocações" – sequências de palavras que aprendemos a usar desde pequeninos, mas que não sabemos realmente porque é que se diz daquela maneira ("apanhar o avião", "pôr a mesa"). Se para um aprendente de uma língua estrangeira a produção destas expressões é um verdadeiro desafio, também o é, certamente, para um sistema computacional. Para complicar ainda mais, apesar de o contexto ajudar a desambiguar algumas produções, também pode levar a mais interpretações de uma frase ("achei hilariante" pode ser um comentário positivo se estivermos a avaliar um filme cómico, mas não o será se nos referimos a um filme de terror ou a um drama). Existe ainda um conjunto de fenómenos do qual fazem parte o humor e o sarcasmo, que complicam ainda mais as tarefas da máquina ("Aqueles que acreditam em telecinesia levantem a minha mão" – Kurt Vonnegut). Mas as complicações não acabam aqui: outro fator que torna terrivelmente complexo o processamento computacional da nossa língua é a variabilidade linguística, isto é, a nossa capacidade de dizer a mesma coisa de tantas maneiras diferentes ("sim", "certo", "ok", "okay", "okie dokie", "parece bem", "absolutamente", "claro", são apenas algumas maneiras de expressar concordância). Esta variabilidade tem a ver com o nosso domínio da língua que, por sua vez, é influenciado pela época em que vivemos, pela nossa idade, condição social, profissão, região em que nascemos e/ou habitamos, estado emocional, entre outros. Uma aplicação verdadeiramente robusta deverá ser capaz de lidar com todas as nossas produções, por muito distintas que sejam; para além disso, tem de ser dotada de alguma capacidade de raciocínio que, para além dos fatores anteriormente referidos, deverá ter em conta todos os elementos não verbais envolvidos numa conversa (a expressão facial do interlocutor, o cenário, entre outros). E a verdade é que ainda estamos longe de conseguir integrar todas estas variáveis.

Em resumo, a área das Tecnologias da Linguagem encontra-se a dar cartas, constituindo um dos cenários mais desafiantes para os loucos avanços na área de AA. Não será provavelmente para amanhã a criação de um agente capaz de nos "enganar" fazendo passar-se por um humano (a não ser em curtas conversas em domínios muito específicos); no entanto, muito em breve teremos o nosso *chatbot* pessoal, que nos guiará em tarefas rotineiras e que será capaz de nos ajudar em alguns domínios em que é especialista. O futuro (que está por minutos) nos dirá.

“Um mordomo virtual que operava no palácio de Monserrate dizia que ouvia mal quando não percebia uma pergunta.”

“Ainda estamos muito longe de conseguir criar máquinas capazes de compreender e produzir linguagem ao nível dos humanos.”

“Quando capazes, os chatbots são perfeitos: estão disponíveis 24 horas por dia e não se sentem perturbados perante clientes sui generis.”

“Poucas horas depois de estar em utilização, esse chatbot era racista, nazi e ordinário, e teve de ser desligado.”

INTELIGÊNCIA ARTIFICIAL NO COMBATE A FRAUDES

BRUNO MARTINS

Os atos de fraude no contexto de transações financeiras estão atualmente associados a perdas muito significativas para bancos, empresas responsáveis por sistemas de pagamento eletrónico e empresas emissoras de cartões de crédito. *Hackers* de todo o mundo desenvolvem constantemente novas maneiras de cometer fraudes como, por exemplo, novos ataques de *phishing* para roubar dados de consumidores, com vista à realização de transações sem a presença física do cliente ou do seu cartão, ou ataques de *skimming* em que vendedores desonestos e/ou dispositivos de leitura são adulterados para armazenar a informação de cartões de débito ou de crédito dos consumidores, com vista à posterior clonagem dos cartões.

Para as entidades do sistema financeiro, confiar exclusivamente em sistemas periciais, baseados em regras e convencionalmente programados para detetar transações fraudulentas, não é a solução mais adequada, dado o contexto de constante evolução associado a este problema. Nesta área de aplicação concreta, técnicas de inteligência artificial e de aprendizagem automática podem ter um papel muito importante, com vista a evitar fraudes de uma forma acessível e transparente para os consumidores. O investimento em tecnologia de IA, como forma de abordar o problema da deteção de transações fraudulentas, tem efetivamente aumentado muito nos últimos anos, sendo que também em Portugal são já vários os institutos de investigação e empresas com atividade na área. De uma forma transparente para os clientes encontramos muitas vezes algoritmos sofisticados de IA, por detrás de mensagens de texto ou de notificações em aplicações de gestão de cartões de crédito, solicitando a verificação de uma transação ou informando de potenciais utilizações ilegítimas.

De um ponto de vista técnico, problemas como a deteção de fraude podem ser abordados como tarefas de classificação automática, em que o objetivo é prever uma classe (por exemplo, a classe “fraude” *versus* a classe “não fraude”) dada uma observação (por exemplo, uma transação bancária, representada por um conjunto de características descritivas da própria transação e do seu contexto de execução). Recorrendo a técnicas de AA, este problema de classificação envolve a criação de modelos com suficiente inteligência para discriminar corretamente quais as transações legítimas das fraudulentas, ou seja, modelos capazes de estimar qual a probabilidade de uma determinada transação ser fraudulenta, balanceando a decisão com base na otimização da contribuição de características descritivas como o montante associado à transação, a diferença entre esse montante e os montantes de transações anteriores, o comerciante associado à transação ou a sua localização. Quando usadas corretamente, técnicas de aprendizagem automática supervisionada – técnicas em que dados do passado são usados para inferir os parâmetros dos modelos a serem usados para a classificação de transações futuras – são extremamente eficazes na deteção de transações ilegítimas, adaptando-se ao longo do tempo a novos padrões de comportamento fraudulento, através da adaptação dos parâmetros associados aos modelos.

Um dos principais desafios associados à modelação do problema de deteção de fraudes como uma tarefa de classificação relaciona-se com facto de, em dados do mundo real, a grande maioria das transações financeiras são efetivamente legítimas. Considerando que há um enorme volume de transações para analisar, apenas uma fração muito reduzida corresponde a transações fraudulentas, o que levanta vários problemas à maioria dos algoritmos de AA do atual estado da arte, baseados na análise de padrões em conjuntos de dados pré-existentes que representem o domínio do problema. Alguma da investigação na área foca-se no problema da otimização de métodos para fazer a classificação em cenários envolvendo a supervisão com dados não balanceados, por exemplo, através de abordagens de amostragem e/ou de recombinação dos dados a serem usados na inferência dos modelos.

Outro desafio importante relaciona-se com a necessidade de representar as instâncias a classificar (ou seja, as transações financeiras) através de características descritivas adequadas ao problema. O desenvolvimento de modelos eficazes e sua adaptação constante a novos padrões de fraude envolve um grande esforço de engenharia associado ao levantamento das características descritivas mais adequadas. Neste contexto, as principais empresas na área, e também os esforços de investigação mais recentes, têm procurado abordar o problema com técnicas modernas de AA baseadas em redes neuronais, que procuram substituir o levantamento explícito de quais as características descritivas mais interessantes, por modelos mais poderosos e capazes de processar diretamente a informação em bruto relativa à execução das transações.

A crescente sofisticação das abordagens usadas neste domínio tem levado a que alguns dos modelos mais recentes sejam, eles próprios, construídos e/ou otimizados com o auxílio de ferramentas de IA (por exemplo, através de técnicas de AA "automatizada", em que a própria estrutura dos modelos de classificação é otimizada automaticamente). Outra tendência observável nos desenvolvimentos recentes envolve o recurso à combinação de técnicas de classificação supervisionada com outros algoritmos para a deteção de anomalias, por forma a mais rapidamente fazer a adaptação a novos padrões de fraude. No entanto, apesar dos muitos desenvolvimentos na área, também os ataques crescem em sofisticação, sendo espectável que continuemos a assistir ao aparecimento de novos métodos e à sua integração nas ferramentas e serviços usados diariamente pelos consumidores.

"Quando usadas corretamente, técnicas de aprendizagem automática supervisionada são muito eficazes na deteção de transações ilegítimas."

INTELIGÊNCIA ARTIFICIAL E SISTEMAS MULTIAGENTES PARA O BEM SOCIAL

MILIND TAMBE

Com o significativo progresso alcançado em IA, existem grandes oportunidades para direcionar esses avanços para o benefício social. A minha investigação centra-se no subcampo dos sistemas multiagentes (SMA) – sistemas onde múltiplos agentes inteligentes interagem uns com os outros. Direcionar os avanços em SMA para fazer face aos atuais desafios das sociedades levaram-me a soluções que beneficiaram as aplicações relevantes, mas também revelaram desafios de investigação fundamentais em IA que tenho abordado no meu trabalho.

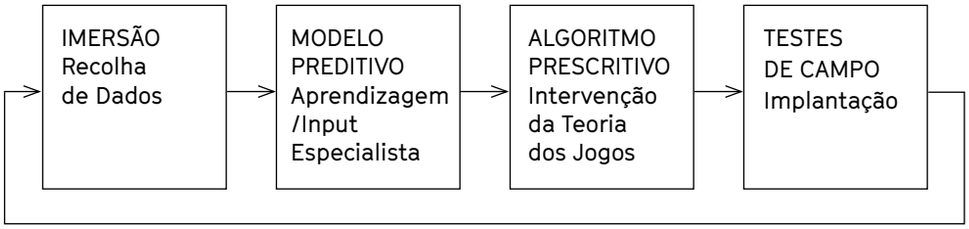
A fim de ser breve, irei focar-me principalmente nos meus últimos 12 anos de investigação, em que trabalhei em três áreas de uso da IA e SMA: segurança e proteção pública, conservação da vida selvagem e saúde pública. Tendo em conta estes desafios societais, concentrei-me numa abrangente maratona de investigação multiagente transversal às três áreas: "Como otimizar o uso de recursos de intervenção limitada perante a sua interação com outros agentes". Ao abordar este desafio de investigação, obtive avanços algorítmicos e de modelação para a teoria dos jogos computacional enquanto abordagem fundamental de solução. Para além disso, o meu trabalho tem contribuído para a modelação do comportamento (adversário) humano, o raciocínio nas redes sociais e os tópicos de sistemas multiagentes relacionados com esta matéria.

Desenvolvi toda esta investigação no contexto do *pipeline* de implantação de dados de ponta a ponta. Este *pipeline* envolve avanços no raciocínio da teoria dos jogos / sistemas multiagentes com conjuntos de dados incertos e com abordagens de aprendizagens automáticas que tratam conjuntos de dados esparsos. A imersão nos domínios como primeiro passo é fundamental para conseguir obter um entendimento crítico dos problemas, das limitações e dos conjuntos de dados. Para esse fim, no contexto da conservação da vida selvagem, por exemplo, patrulhámos florestas na Malásia com agências de conservação da vida selvagem de modo a saber quais são as suas limitações. A seguir a um entendimento aprofundado por via da imersão, o nosso próximo passo é construir um modelo preditivo usando a aprendizagem automática ou o contributo de peritos em domínios. Tal modelo preditivo pode, por exemplo, prever quais os casos de alto *versus* baixo risco numa população. Em seguida, a fase do algoritmo prescritivo é aquela em que, geralmente, nos focamos no uso da teoria dos jogos para planear intervenções com conhecimento dos riscos tendo em conta recursos limitados. O nosso trabalho centra-se com frequência em domínios em que é difícil ter acesso a dados (comunidades com poucos recursos ou países de mercados emergentes), e daí o desafio muitas vezes ser o de planear intervenções apesar de dados incertos ou esparsos.

Finalmente, ao contrário de muitos outros grupos de investigação em IA, também estamos bastante interessados no passo final dos testes de campo e implantação, não só pelo impacto social, mas porque são muitas vezes a forma como aprendemos sobre as limitações-chave dos nossos modelos e algoritmos, o que origina geralmente novos desafios de

investigação fundamentais para abordar. Esta investigação requer urgentemente parcerias interdisciplinares para imersão e testes de campo. Com frequência, estas parcerias dão origem a novas investigações fora do âmbito de qualquer uma das disciplinas.

Enquanto olhamos para o futuro, vejo que a IA possui um potencial tremendo na melhoria da sociedade e na luta contra a injustiça social. Para atingir esse objetivo é necessário trazer a IA aos que ainda não beneficiaram dela, como aqueles que vivem no sul global. Este objetivo possibilita novos temas de investigação e exige avanços fundamentais em IA, mas também obriga a um trabalho interdisciplinar com outros cientistas (na assistência social ou na biologia de conservação, por exemplo). Também implica sair do laboratório para o terreno, não só porque estamos interessados no impacto social, mas porque nos dá uma perceção dos pressupostos certos ou errados nos modelos e algoritmos e, consequentemente, leva a novos rumos de investigação.

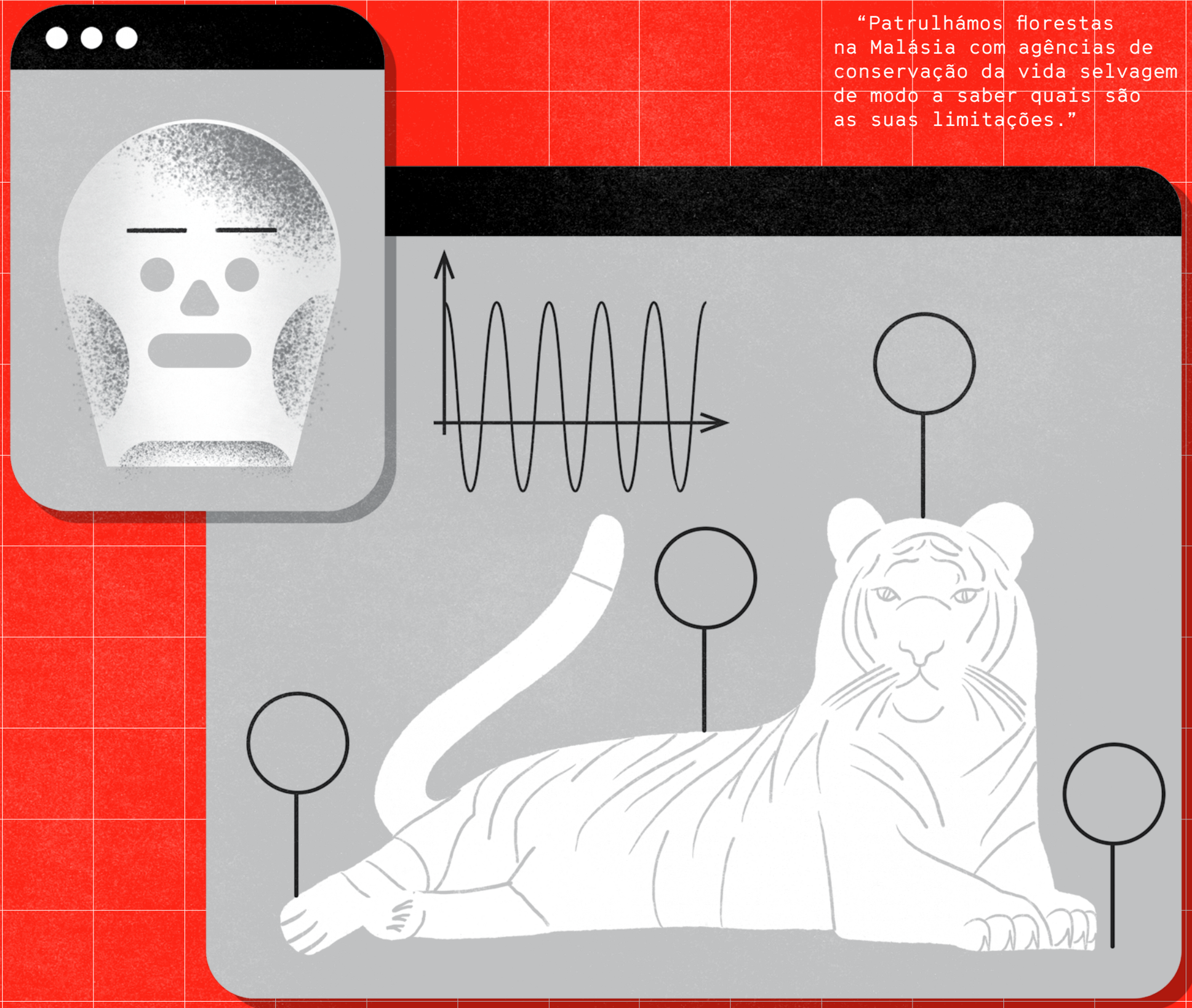


Pipeline de implantação de dados de ponta a ponta.

“A IA tem um potencial tremendo na melhoria da sociedade e na luta contra a injustiça social.”

“É necessário trazê-la aos que ainda não beneficiaram dela, como os que vivem no sul global.”

“Patrulhámos florestas na Malásia com agências de conservação da vida selvagem de modo a saber quais são as suas limitações.”



Inteligência Artificial
[Artificial Intelligence / AI]
Embora não exista uma definição formal consensual, inteligência artificial é a 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, por exemplo: “A IA que conduz o meu automóvel autónomo é competente”.

Inteligência Artificial Geral
[Artificial General Intelligence / AGI]
Também conhecida por inteligência artificial forte, refere-se ao conceito, hoje 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 a segunda metade do século XX, 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.

Aprendizagem Automática
[Machine Learning / ML]
Conjunto de técnicas (com suporte em várias áreas da matemática e da computação) para 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 de AA: supervisionada (supervised learning), não supervisionada (unsupervised learning) e aprendizagem por reforço (reinforcement learning).

Aprendizagem Supervisionada
[Supervised Learning]
Classe de métodos nos quais a aprendizagem automática se baseia num conjunto de pares e/ou exemplos de observação-decisão; ou seja, a supervisão consiste no fornecimento da decisão certa/desejada para cada observação no conjunto de exemplos. Exige a definição de um critério de qualidade de cada decisão produzida pelo sistema (que pode ser simplesmente se esta está certa ou errada, mas pode ser 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 professor/supervisor dizer qual o melhor lance para cada posição (o que seria aprendizagem supervisionada), apenas indica, no fim do jogo, o resultado do mesmo. Compete aos métodos de aprendizagem por reforço identificar as decisões (jogadas, no caso do xadrez) que conduzem ao desfecho desejado. Este tipo de aprendizagem tem numerosas áreas de aplicação (nomeadamente, robótica móvel, veículos autónomos, jogos, sistemas de recomendação), em situações onde um sistema está a aprender a melhor estratégia ao longo do tempo.

Conjunto de Treino
[Training Set]
Também referido como conjunto de aprendizagem. Conjunto de dados usados para treinar um sistema de aprendizagem automática. Poderá consistir num conjunto de pares observação-decisão (aprendizagem supervisionada), ou de apenas observações (aprendizagem não supervisionada).
As observações podem ter muitas formas: tabelas, imagens, vídeos, audio, textos, registos de interação, entre muitos outros. As decisões (quando existem) também podem tomar muitas destas formas ou, no caso mais simples, serem apenas rótulos das observações a serem aprendidos pelo sistema.

Conjunto de Teste
[Test Set]
Conjunto de dados usados para testar o comportamento e o desempenho de um sistema de aprendizagem. Poderá consistir num conjunto de pares observação-decisão (aprendizagem supervisionada) ou num conjunto de observações (aprendizagem não supervisionada). A forma das observações e respetivas decisões, quando existem, é a mesma que no conjunto de treino usado pelo sistema.

Representação do Conhecimento
[Knowledge Representation]
Capacidade das máquinas de guardar e manipular modelos (informação) do mundo para resolver e executar tarefas. Há vários modelos de representação do conhecimento, como as redes semânticas, que representam explicitamente relações entre conceitos (por exemplo, entre os conceitos semânticos “cão” e “animal” existe a relação “é um”; entre “cão” e “pelo” existe a relação “tem”).

Raciocínio Automático
[Automated Reasoning]
Capacidade das máquinas de inferir com base em dados e conhecimento representados de forma explícita (por exemplo, numa rede semântica).

Procura e Planeamento
[Search and Planning]
Técnicas para procurar soluções e planear ações em situações complexas que

conduzam a um dado resultado. Estas técnicas e os problemas subjacentes foram extensivamente estudados nas primeiras décadas da inteligência artificial e conduziram a muitos métodos usados em sistemas que fazem otimização logística, jogam jogos de tabuleiro (como xadrez) e demonstram teoremas matemáticos.

Processamento de Língua Natural

[Natural Language Processing]
Conjunto de técnicas que permitem às máquinas reconhecerem 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, por exemplo, um email ou uma avaliação de um produto, quanto ao sentimento que exprime: positivo, negativo, neutro), a classificação do tópico de um texto (por exemplo, se uma notícia é sobre desporto, negócios, política, ...).

Conexionismo

[Connectionism]
Classe de métodos que se baseiam na utilização de unidades de processamento simples, interligadas entre si, cada uma inspirada no funcionamento de um neurónio biológico. Estes métodos são uma alternativa às abordagens simbólicas, usadas na inteligência artificial clássica, uma vez que usam representações que são normalmente designadas por sub-simbólicas.

Redes Neuronais Artificiais

[Artificial Neural Networks]
Classe de métodos de aprendizagem automática baseado 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, ou seja, de quão plausível é que se assemelhem, mesmo que apenas aproximadamente, ao processo de aprendizagem nos cérebros biológicos.

Redes Neuronais Profundas

[Deep Neural Networks]
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.

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.

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. Por exemplo, 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.

Sistemas Multiagente

[Multiagent Systems]
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. Por exemplo, um conjunto de drones ou robôs tem que se coordenar e negociar as suas ações de forma distribuída para atingir um objetivo comum.

Interpretabilidade

[Interpretability/ Explainability]
Possibilidade de interpretação (ou explicação), 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 que seja possível 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.

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.

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.

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

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–UL)

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: Gráfica Maiadouro
Tiragem: 800 exemplares
Depósito legal: 455042/19

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.