GuiA

Guia para a Inteligência Artificial

GUIA PARA UMA INTELIGÊNCIA ARTIFICIAL ÉTICA, TRANSPARENTE E RESPONSÁVEL NA ADMINISTRAÇÃO PÚBLICA

AGÊNCIA PARA A MODERNIZAÇÃO ADMINISTRATIVA
tic.gov.pt
Cofinanciado por: COMPETE 2020 PORTUGAL 2020 UNIÃO EUROPEIA Fundo Social Europeu

Com a **Inteligência Artificial** deveremos ter a ambição de superar o ser humano no que diz respeito à ética, à transparência e à responsabilidade, de forma a contribuir para um **mundo mais igualitário e inclusivo**.

SIGLAS E ACRÓNIMOS

AI-HLEG	High-Level Expert Group on Artificial Intelligence
AP	Administração Pública
AMA	Agência para a Modernização Administrativa
CE	Comissão Europeia
DL	*Deep Learning*
FCT	Fundação para a Ciência e a Tecnologia, I.P.
G20	Grupo dos 20, formado pelos Ministros das Finanças e Chefes dos Bancos Centrais das 19 maiores economias do mundo e UE
IA	Inteligência Artificial
iAP	Interoperabilidade na Administração Pública
IoT	*Internet of Things*
I&D	Investigação e Desenvolvimento
LGBTQ+	Lésbicas, Gays, Bissexuais, Travestis, Transexuais, Transgéneros, Queer e outras entidades de género
ML	*Machine Learning* ou Aprendizagem Automática
NLP	*Natural Language Processing*
OCDE	Organização para a Cooperação e Desenvolvimento Económico
ONU	Organização das Nações Unidas
RGPD	Regulamento Geral sobre a Proteção de Dados
RNID	Regulamento Nacional de Interoperabilidade Digital
SAMA	Sistema de Apoio à Modernização Administrativa
UE	União Europeia
UNESCO	Organização das Nações Unidas para a Educação, Ciência e Cultura

ÍNDICE

SUMÁRIO EXECUTIVO

A crescente escolha de soluções de Inteligência Artificial (IA), com elevado impacto sobre inúmeros setores da sociedade, suscita um conjunto de desafios aos quais urge dar resposta. Neste âmbito, enfatizam-se as questões associadas à ética, justiça, transparência, responsabilidade e explicabilidade destes sistemas.

Embora a evolução dos sistemas de IA se faça acompanhar de um significativo aumento de investigação académica e científica, a rapidez com que estas tecnologias proliferam e o modo como se relacionam com a sociedade, exige um espaço para dúvidas e discussão.

Por outro lado, é necessário identificar o que pode estar em causa, no uso de IA, por exemplo se estamos a avaliar candidatos para um emprego ou para o acesso a uma instituição de ensino, se estamos a decidir sobre atribuição de crédito ou um apoio social, se estamos perante diagnósticos médicos ou decisões judiciais, ou veículos autónomos, todos estes casos constituem alguns exemplos onde a questão do impacto ganha uma outra dimensão de discussão e importância.

É importante lembrar que a IA é concebida por pessoas e deve centrar-se na criação de benefícios para as pessoas, e que as questões éticas associadas aos sistemas de IA estão intrinsecamente relacionadas com todos aqueles que estão envolvidos na sua conceção e utilização, desde aqueles que desenvolvem os algoritmos, aos decisores, e aos governos. Em igual medida, é pertinente considerar o que está sob o controlo humano e o grau de autonomia destas soluções.

A Agência para a Modernização Administrativa (AMA) vem uma vez mais assumir o seu papel enquanto instituição pública responsável pela promoção e desenvolvimento da modernização administrativa em Portugal propondo, através do projeto IA Responsável, a exploração de caminhos para se conseguir uma IA que considere estas questões. Esta disponibiliza documentos e instrumentos com o intuito de dar suporte e de promover a adoção progressiva e gradual de uma IA ética, transparente e responsável.

Faz igualmente parte da visão deste projeto, contribuir para o equilíbrio entre uma discussão filosófica/teórica e a discussão prática, aprofundando os conceitos de responsabilização, transparência, explicabilidade, justiça e ética. Conceitos estes que contêm, a título de exemplo, a problemática do viés, muito associada aos algoritmos com impacto social. A partir de uma reflexão que considere ambas as discussões, pretende-se garantir a proteção da democracia, do Estado de direito e dos direitos fundamentais, com a materialização destes conceitos no modo como os serviços de IA são pensados, desenhados e providenciados, quer no setor público quer no setor privado.

Na sua estrutura, o guia procura definir o que é a IA e o seu funcionamento, mostrar como está presente na sociedade e identificar alguns dos efeitos decorrentes da sua utilização. Procura ainda informar sobre o enquadramento da IA no Mundo e em Portugal, e identifica

o ecossistema e os principais atores no contexto nacional. Por outro lado, e considerando a importância dos dados para o desenvolvimento e alimentação destes sistemas, faz também referência ao ecossistema de dados na Administração Publica (AP) e aos princípios que lhe devem estar subjacentes.

Este guia pretende trazer para a discussão publica a necessidade de estabelecer os pilares de regulação, supervisão, liderança e governação, para a elaboração de um código de ética, fomentar regulamentação e leis que forneçam orientações e suporte aos desenvolvimentos tecnológicos.

O guia enuncia também um conjunto de valores e princípios em linha com a lista de direitos humanos, e explora o tema da inclusão, da igualdade, do desenvolvimento sustentável e do bem-estar.

Em contraponto, são abordados os efeitos perniciosos associados a sistemas de IA, com alguns exemplos, e é reforçada a importância de se criarem mecanismos rigorosos de monitorização, auditoria, proteção e segurança.

Este trabalho fornece também recomendações do ponto vista mais amplo e genérico, e identifica um conjunto de barreiras e desafios que devem ser considerados aquando da construção e implementação de sistemas de IA Responsáveis.

Na dimensão prática deste projeto foi desenvolvida uma ferramenta de avaliação do risco, construída em linha com as orientações do guia. Este instrumento possibilita a analise da suscetibilidade de sistemas de IA, associada às cinco dimensões subjacentes a uma IA Responsável, referidas no guia, e dá recomendações de ações e sugestões de leituras, em função do nível de maturidade dos atores.

Em última instância, o guia e a ferramenta permitem estruturar o processo de construção e de implementação de sistemas inteligentes para que sejam mais responsáveis, éticos e transparentes, por via da compreensão e adoção de conceitos, e da mudança comportamental. Desse modo, ambos constituem um recurso importante na antecipação e mitigação de riscos em sistemas de IA nas cinco dimensões para uma IA Responsável, proporcionando-se soluções mais éticas tanto na Administração Pública como no Setor Privado.

O guia é sintetizado em três documentos de leitura simples e rápida, focados: nos valores, princípios e recomendações; nas dimensões de avaliação; e na ferramenta de avaliação do risco. É nossa ambição que as versões referentes aos conteúdos teóricos e à ferramenta, não sejam estáticas e possam sofrer atualizações ao longo do tempo.

Nas referências bibliográficas que fundamentaram o desenvolvimento deste projeto, incluíram-se as de autoria de Organizações Intergovernamentais, da Comissão Europeia, do Setor Privado, de Developers e de Consultoras.

1. ENQUADRAMENTO

1.1. ÂMBITO E ESTRUTURA DO GUIA

O projeto "Guia Responsável" tem como objetivo a elaboração de um guia para a utilização da Inteligência Artificial (IA) na Administração Pública (AP). O mesmo pode servir ainda de referência para o setor privado.

No âmbito deste projeto foi também desenvolvida uma aplicação de avaliação de risco, disponível em https://ia.tic.gov.pt, com duas vertentes: identificação e mitigação. Esta Ferramenta permite dar apoio a políticas públicas relacionadas com *Data Science, Big Data,* ML e IA, nomeadamente divulgar melhores práticas e estabelecer critérios de avaliação que possam suportar pareceres prévios e candidaturas de financiamento.

Numa altura em que as tecnologias emergentes têm cada vez mais impacto na sociedade e que as implicações éticas das mesmas começam a ser amplamente discutidas, este Guia dá continuidade às bases criadas pelo Regulamento Geral sobre a Proteção de Dados e pela Estratégia Nacional de Inteligência Artificial para uma Inteligência Artificial mais transparente, ética e responsável em Portugal.

Dada a existência de normas de *guidelines* em áreas como finanças, indústria farmacêutica, aviação, produção de dispositivos médicos, proteção do consumidor e proteção de dados, o Guia foi estruturado de forma a complementar as recomendações já existentes. Visa-se assim, complementar os conhecimentos especializados existentes e ajudar as autoridades na monitorização e supervisão das atividades das organizações que utilizam sistemas com IA e possuem produtos e serviços baseados em IA.

Este Guia, ao qual se associa uma Ferramenta de Avaliação de Risco online, é pensado e escrito na encruzilhada do florescimento da IA em Portugal e da procura mundial por uma IA Responsável. O Guia inicia com a apresentação geral do contexto e conceitos associados à IA, prossegue com a exemplificação de aplicações de IA e finaliza com a explicação dos princípios associados a uma IA Responsável e com recomendações de ações para o seu cumprimento.

Embora tenha o intuito específico de orientar projetos na AP e ser referência para o Setor Privado e Academia dirige-se também a todas as pessoas e entidades com a intenção de aprender sobre IA Responsável.

Garantir uma IA ética implica não só os responsáveis pela conceção, implementação e monitorização de sistemas de IA, mas também o envolvimento de toda a comunidade. Ao remeter a um ecossistema mais abrangente, este Guia procura aproximar indivíduos melhor informados, capazes de compreender e discernir sobre a conduta ética dos produtos de IA que utilizam ou que são correntes no quotidiano, e de decidir se são confiáveis e os aprovam.

Não se trata de uma obra científica, mas um documento orientador de apoio a projetos de IA Responsável, para consulta de forma sistemática e pedagógica.

Entre os vários desafios atuais, destaca-se a garantia da conservação do papel constitucional do Estado na vigilância e no controlo da implementação e da utilização da IA em soluções na AP.

O Guia foi escrito com base nas melhores práticas internacionais, em *guidelines* e *frameworks* de IA, nas diretrizes da UE para uma IA de Confiança, na Estratégia Nacional para a Inteligência Artificial e no Regulamento Geral sobre a Proteção de Dados.

Ao longo do Guia são indicadas várias referências a partir das quais o leitor pode explorar os temas abordados para maior detalhe.

Alguns dos termos associados à IA foram mantidos em inglês, uma vez que a sua tradução conduziria à definição incompleta do seu significado.

O caminho a realizar pela AP portuguesa no domínio da IA impacta em todos os cidadãos e, por essa razão, deve ser iniciado o quanto antes com assertividade, conhecimento e responsabilidade.

O GUIA TEM COMO OBJETIVOS:

- **Contextualizar** os riscos que decorrem da emergência da IA;
- **Apresentar** os princípios e o quadro conceptual metodológico para a implementação de projetos de IA Responsável;
- **Explicar** a Ferramenta de Avaliação de Risco aplicável a projetos com IA.

1.2. GLOSSÁRIO

Para melhor compreensão do Guia apresenta-se, de forma simplificada, a definição de alguns conceitos de base.

ALGORITMO: sequência lógica e finita de instruções que visam atingir um determinado propósito.

COMPUTAÇÃO: busca de solução para um problema na forma de um resultado que foi obtido processando informação de entrada.

LINGUAGEM COMPUTACIONAL: conjunto de instruções para implementação de algoritmos, com uma sintaxe e semântica definidas, que permite que os recursos computacionais gerem as saídas desejadas.

PARTILHA DE DADOS: ocorre quando as bases de dados, parcialmente ou em complementaridade, ou os dados são copiados e/ou utilizados por outros sistemas (computacionais).

PROCESSAMENTO: execução de um algoritmo.

SAÍDAS: relatórios, gráficos, tabelas, ecrãs e qualquer outro resultado obtido por processamentos de programas computacionais ou algoritmos.

SISTEMA INTELIGENTE: sistema computacional que tem alguma capacidade de aprender e consequentemente exibir comportamentos adaptativos.

TECNOLOGIAS EMERGENTES: aplicações de conhecimento científico que, na sua maioria, só se tornaram possíveis com os avanços da computação inteligente.

UTILIZADOR: alguém que faz uso da computação, normalmente um ser humano, mas que pode ser, alternativa ou complementarmente, um ou mais computadores.

VIESES: algum comportamento observável durante o processamento e saída dos sistemas computacionais que não tenha sido programado e que não reflita o conjunto de valores e princípios éticos da sociedade que alberga os sistemas.

1.3. O QUE É A IA?

A IA é uma disciplina de proeminente utilização cuja definição é ainda difícil de compreender, pois a possibilidade de ser abordada numa perspetiva filosófica e técnica confere-lhe um cariz complexo.

Na década de 50, a IA foi definida como ciência e engenharia capaz de gerar máquinas que desencadeiam processos e respostas que anteriormente necessitavam da inteligência humana. A sua definição evoluiu para um conceito mais analítico, onde se refere que a inteligência não é o atributo de concretizar tarefas, mas a capacidade de um sistema se adaptar e improvisar ações num novo contexto, vulgarizar conhecimento e aplicá-lo a cenários desconhecidos, traduzindo-se em eficiência de aprendizagem e na aquisição de novas competências.

Focando o conceito de IA num sentido técnico e associado à própria tecnologia, pode-se afirmar que a IA reúne ciências, teorias e técnicas (referem-se a lógica matemática, a estatística, a probabilidade, a neurobiologia computacional e a ciência da computação) para conseguir a mimetização das capacidades cognitivas de um humano por uma máquina.

Os sistemas de IA demonstram um subconjunto das seguintes operações, as quais são figurativas de comportamentos gerados pela inteligência humana: aprendizagem, adaptação, interação, raciocínio, resolução de problemas, representação de conhecimento, previsão e planeamento, autonomia, perceção, movimento e manipulação.

De um modo geral, a computação inteligente fundamenta-se numa programação adaptativa, a qual está centrada na aprendizagem. Distingue-se pela flexibilidade, capacidade de fazer generalizações e resolução de problemas.

1.4. COMO FUNCIONA A IA?

A IA utiliza algoritmos computacionais para resolver problemas complexos. O seu funcionamento baseia-se em redes neurais artificiais, computação evolucionária, sistemas especialistas, entre outros. Esses traduzem-se em máquinas com capacidade de aprendizagem automatizada, aprendizagem profunda (DL), análise de dados em tempo real, processamento de linguagem natural e visão computacional.

Com o intuito de responderem a um objetivo complexo, os sistemas com IA atuam na esfera física ou digital reunindo dados, por meio de sensores, câmaras ou outros recetores, interpretando-os e processando a informação deles derivada, para perceção do ambiente e decisão da melhor ação face ao objetivo inicialmente definido. Alguns sistemas podem ainda adaptar o seu comportamento em função do impacto/efeito desencadeado por ações tidas anteriormente ou como resultado do *feedback* direto de utilizadores ou operadores.

As ações podem ser executadas digitalmente, quando integradas num sistema de tecnologias de informação, ou serem uma solução física, como ocorre em robótica.

1.5. IA NA SOCIEDADE E AS SUAS IMPLICAÇÕES

O poder de transformação que caracteriza a IA deve estar ao serviço das pessoas e do planeta visando a sustentabilidade e a melhoria do ambiente. A IA deve ser entendida como um meio potencial para a reestruturação de sociedades, permitindo otimizar a economia, contribuir para o bem-estar, ajudar na elaboração de previsões e apoiar a tomada de decisões.

Paralelamente, é acompanhada de um sentimento de dúvida e falta de confiança pela comunidade, originando ansiedade e preocupações éticas, nomeadamente em relação a questões de equidade e de privacidade. É, por isso, essencial promover o seu desenvolvimento orientado para a transparência, responsabilidade e para um bem global. O relacionamento de questões técnicas, éticas e legais deve viabilizar o alinhamento de normas e códigos de conduta que garantam a interoperabilidade de leis e regulamentos.

A IA pode ter um impacto significativo nas políticas e na disponibilização de serviços públicos. Entre outros benefícios destaca-se: o potencial de reduzir o tempo necessário para executar tarefas pelo ser humano, criando disponibilidade para a realização de trabalho de alto valor; o aumento de produtividade e eficiência nas ações, conseguindo maior consis-

tência que o ser humano; a capacidade de interpretar e processar grandes quantidades de dados, identificando e relacionando padrões; a projeção de melhores e mais sustentadas políticas e decisões; a simplificação da comunicação e o envolvimento dos cidadãos; a rapidez e a melhoria da qualidade dos serviços públicos; e a criação de emprego.

Mencionam-se alguns exemplos da utilização prática da IA:

No Setor da **SAÚDE:**

- Reconhecimento de padrões imagiológicos com relevância clínica, nomeadamente em oncologia, através de visão computacional;
- Deteção de padrões microbianos em diagnósticos por imagem, para auxílio na construção de diagnósticos diferenciais sólidos e prescrição de antibióticos mais adequados;
- Construção de modelos para prever a viabilidade de vacinas em toda a cadeia de abastecimento e garantir a sua entrega eficaz;
- Eliminação de barreiras associadas à inacessibilidade a instalações de saúde, comum nas zonas rurais;
- Identificação precoce de pandemias;
- Análise de registos médicos para fornecer serviços de saúde mais personalizados, melhores e mais rápidos;
- Apoio na projeção de planos de tratamento personalizados;
- Desenvolvimento de cuidados de saúde baseados em precisão e genómica;
- Projeção de novos medicamentos e terapias médicas;
- Melhoria de processos burocráticos do Sistema Nacional de Saúde, como o agendamento e o atendimento ao utente;
- Redução nos custos de saúde por meio de melhor programação e otimização de ativos de saúde e força de trabalho;
- Auxílio em trabalhos repetitivos, como a higienização ou testes de laboratório;
- Controlo de qualidade de alimentos e de medicamentos.

No Setor da **EDUCAÇÃO**:

- Tradução de sinais de linguagem gestual para a linguagem corrente;
- Construção de pareceres imediatos sobre a escrita dos alunos, permitindo que revisem os seus trabalhos e melhorem rapidamente as suas competências, através de sistemas de NLP e *Deep Learning* (DL);
- Recomendação de formação profissional, por meio de *Intelligent Tutoring Systems* (ITS);
- Seleção das candidaturas ao ensino não tendenciosa e igualitária;
- Ensino interativo.

No Setor das **INFRAESTRUTURAS E CIDADES INTELIGENTES:**

- Monitorização da condição das infraestruturas, por via de sensorização e análise preditiva;

- Navegação autónoma;
- Determinação de itinerários para deslocações mais rápidas e sem constrangimentos através de ML;
- Otimização da mobilidade dos transportes, assim como, da experiência dos utilizadores de transportes públicos;
- Redução do congestionamento de tráfego por meio de melhores serviços de informação, gestão de semáforos e planeamento de obras viárias;
- Otimização da utilização e da segurança de condução nas estradas;
- Controlo de entrada/saída de turistas nas cidades;
- Gestão de multidões.

No Setor do **AMBIENTE E DA AÇÃO CLIMÁTICA:**

- Rastreamento e previsão de padrões de poluição do ar por meio de sensores, para conseguir melhores medidas de intervenção na qualidade do ar;
- Identificação de pontos de vulnerabilidade nas reservas naturais, protegendo os animas da caça ilegal;
- Previsão de horários e locais de risco de deslizamentos de terra, criando um sistema de alerta para minimizar o impacto de desastres naturais através de sistemas de DL;
- Monitorização bioacústica e tecnologia móvel para rastrear a saúde das florestas e detetar ameaças através de sistemas de DL;
- Medição e modelação de variáveis relacionadas com as alterações climáticas;
- Redução do consumo de energia e água;
- Armazenamento de energia de sistemas de energia renovável, por meio de redes inteligentes;
- Gestão de recursos naturais.

No Setor da **AGRICULTURA**:

- Recolha e processamento de dados climáticos e agrícolas, com a finalidade de melhorar os sistemas de irrigação utilizados por agricultores com poucos recursos;
- Planeamento bem sustentado dos processos agrícolas, desde a preparação dos terrenos à colheita de alimentos;
- Resolução de desafios associados à procura de alimentos, à escassez de irrigação e ao uso inapropriado de pesticidas;
- Rastreamento e análise de medidas de controlo de pragas, de modo a ter intervenções mais oportunas e localizadas, para estabilizar a produção agrícola e reduzir o uso de pesticidas, utilizando sistemas de visão computacional.

No Setor da **JUSTIÇA**:

- Recolha e confronto de informações relevantes em documentos relacionados em casos judiciais, permitindo que advogados pesquisem e defendam casos com maior eficácia, utilizando NLP e ML.

No Setor da **GESTÃO ADMINISTRATIVA**:

- Disponibilização de interfaces de conversação automatizadas com funcionários virtuais para automatizar cenários de atendimento ao cidadão e às empresas;
- Reforço da segurança e privacidade nos sistemas informáticos.

No Setor **SOCIAL E DA SOLIDARIEDADE**:

- Ajuda de refugiados na tradução das suas habilitações para apresentação no mercado de trabalho europeu e recomendação de trabalhos relevantes face a essa informação;
- Determinação do nível de risco de suicídio de jovens LGBTQ+, para melhoria na resposta dos serviços aos indivíduos que procuram ajuda, utilizando NLP;
- Identificação de casos de dependência, como problemas com o jogo e o consumo de drogas, que careçam de acompanhamento e ajuda;
- Identificação, medição e indagação das causas subjacentes da desigualdade;
- Construção inteligente capaz de tornar a habitação mais acessível;
- Determinação justa de apoios sociais.

No Setor da **CULTURA**:

- Combate às notícias falsas, através de *blockchain*;
- Sugestão de atividades de lazer.

No Setor **ECONÓMICO E FINANCEIRO**:

- Criação de processos de negócio inovadores e mais eficientes, através de sistemas de ML;
- Automação de processos transacionais, como pagamentos e faturação;
- Redução do crédito mal parado (vencido) para cidadãos e empresas;
- Deteção de fraude;
- *Credit scoring*;
- Algoritmos para *trading*;
- Processos automatizados de *pricing*;
- Controlo de práticas abusivas para o consumidor;
- Impedimento da prosperidade de monopólios, por exemplo, através da monitorização de transações;
- Otimização da experiência do utilizador, fornecendo sugestões personalizadas, navegação baseada em preferências e pesquisa de produtos suportada em imagens, para o setor de vendas a retalho. Ainda nesse setor, antecipação da procura pelos clientes, gestão de existências e de entregas;
- Automatização de processos na indústria, com impactos na engenharia, na cadeia de fornecimento, na gestão de produto, nos custos de produção, na manutenção, na garantia de qualidade e na logística e armazenamento.

1.6.
IA NO MUNDO

A elaboração de Estratégias Nacionais para a criação de condições que incitem o crescimento sustentado da IA tem sido uma medida prioritária. Muitos países, entre os quais o Canadá, a Finlândia e a Itália, possuem já Estratégias Nacionais aprovadas exclusivamente dirigidas à IA. Outros países, como Portugal, Espanha, Alemanha, México e China, incluíram a IA em estratégias governamentais mais amplas. Os EUA destacam-se dos demais por apostarem numa estratégia onde a IA é colocada ao serviço do setor privado.

Para além das estratégias nacionais, governos e empresas têm criado outros instrumentos com carácter orientador e/ou vinculativo a uma IA responsável. Aqui enquadram-se estratégias, recomendações, guias profissionais e códigos de conduta, e ainda leis, regulamentos e ordens executivas, de cariz regulatório.

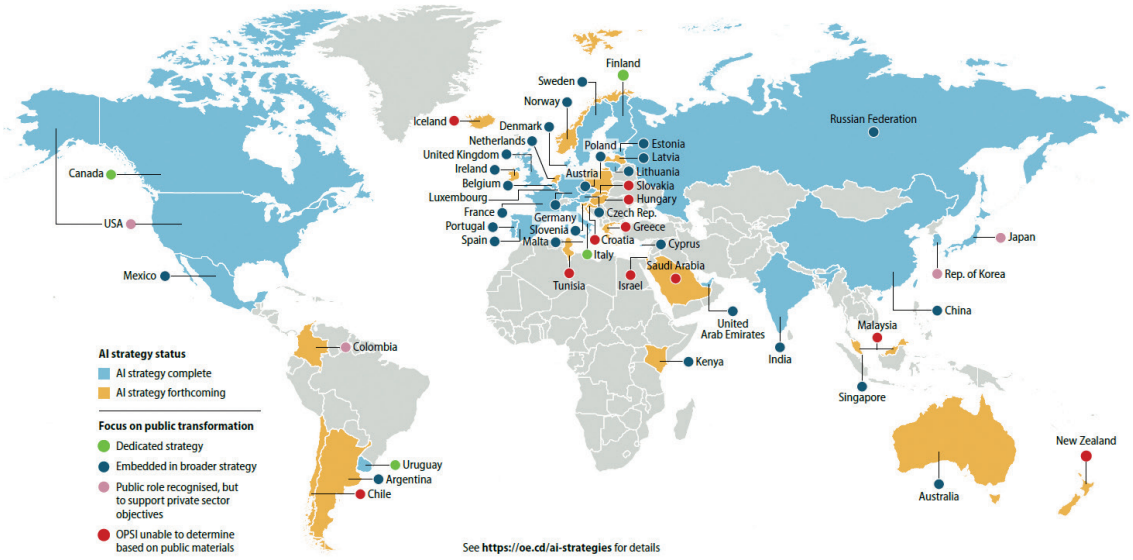

FIGURA 1: ESTRATÉGIAS DE IA NO PROCESSO DE TRANSFORMAÇÃO PÚBLICA.
FONTE: OPSI ANALYSIS OF NATIONAL STRATEGIES AS OF 15 NOVEMBER 2019.

Desde 2012 que se assiste, a nível mundial, à publicação de um número crescente de artigos científicos que remetem para os temas da interpretação, explicação e responsabilidade dos sistemas inteligentes. Ao longo deste percurso, de quase uma década, emergiram definições de "*Interpretable AI*", "*Explainable AI*" e mais recentemente "*Responsible AI*", progressivamente marcadas pelo pendor acentuado da transparência e da capacidade de mitigação ética dos sistemas inteligentes.

Atenta a esta evolução, a Comissão Europeia (CE) criou o grupo de peritos peritos, o *High-Level Expert Group on Artificial Intelligence* (AI-HLEG), e a União Europeia publicou, entre 2019 e 2020, três instrumentos relevantes para a discussão das questões éticas nos sistemas inteligentes: *Ethics Guidelines for Trustworthy Artificial Intelligence*, o Livro Branco sobre a IA e a Lista de Avaliação para IA de Confiança.

De acordo com este quadro conceptual considera-se que os sistemas inteligentes são confiáveis quando:

- Existe uma IA legal, ética e robusta;
- Se concretizam quatro princípios éticos: respeito pela autonomia humana, prevenção de danos, equidade e explicabilidade;
- Se asseguram sete requisitos: controlo e supervisão humana, segurança e robustez técnica, privacidade de governação dos dados, transparência, diversidade, não discriminação e justiça, bem-estar social e ambiental, e responsabilização;
- É possível aplicar uma avaliação de fiabilidade, como seja o caso da ALTAI (*Assessment List for Trustworthy AI*) desenvolvida pelo AI-HLEG da CE e oficialmente publicada no final de 2020.

Com vista a promover a excelência no domínio da IA, a CE definiu como aposta para o período de 2019 a 2024:

- A criação de uma nova parceria público-privada para a IA e robótica;
- O reforço e integração dos centros de excelência em investigação;
- A criação de plataformas nacionais de inovação digital especializadas em IA;
- O reforço do financiamento destinado ao desenvolvimento e utilização de IA;
- A aplicação da IA na contratação pública, tornando os processos mais eficientes;
- O incentivo à aquisição de sistemas de IA por parte dos organismos públicos.

Mundialmente, verificam-se grandes variações de maturidade, de referências legais, de iniciativas de estruturação, de objetivos de investigação e de atores/grupos. Alguns países iniciaram cedo o caminho da IA e seguem-no no sentido de favorecer uma IA que

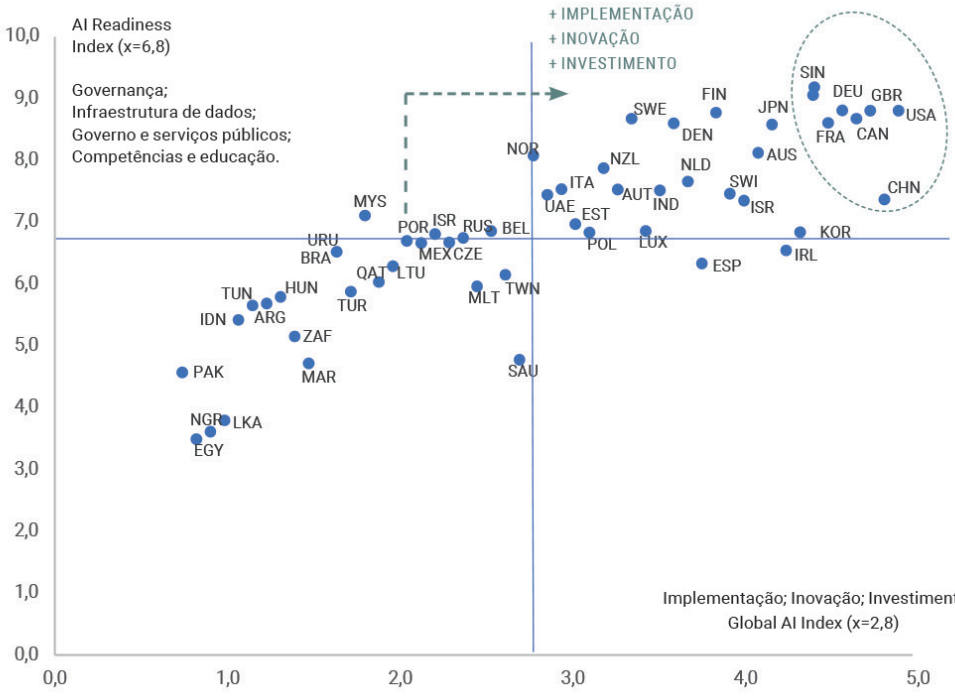

FIGURA 2: POSICIONAMENTO DE PORTUGAL NOS ÍNDICES MUNDIAIS DE IA.
FONTE: BASEADO NO AI READINESS INDEX (2019) E NO GLOBAL AI INDEX (2019)

garanta uma sociedade mais justa e que privilegie o bem-estar. Uns deram prioridade à utilização da IA no setor privado, outros para fortalecer o Governo e outros para colocar os seus cidadãos como o centro do tema (*human-centric*).

O cruzamento do índice de Resposta Governamental para IA e do Índice Global de IA evidencia a liderança mundial em IA pelos seguintes países: Singapura, Alemanha, França, Canadá, Reino Unido, EUA e China (Figura 2). O posicionamento de Portugal no mundo em 2019, avaliado a partir dos dois índices, indica progressos significativos e oportunidades para mais implementação, inovação e investimento em IA.

2.
IA EM PORTUGAL

2.1.
ESTRATÉGIA NACIONAL

A "Estratégia Nacional de Inteligência Artificial" (*AI Portugal 2030*), é um marco importante para a IA em Portugal. Trata-se de um projeto integrado no programa INCoDe.2030, que pretende promover a investigação e a inovação nesta área específica, em prol do seu desenvolvimento e aplicação em campos como a AP, o ensino, a formação e as empresas. Esta estratégia está em linha com as diretivas da CE que pretendem levar os Estados-Membro a promover o desenvolvimento e a utilização da Inteligência Artificial na Europa. Mais informações em https://www.incode2030.gov.pt/ai-portugal--2030

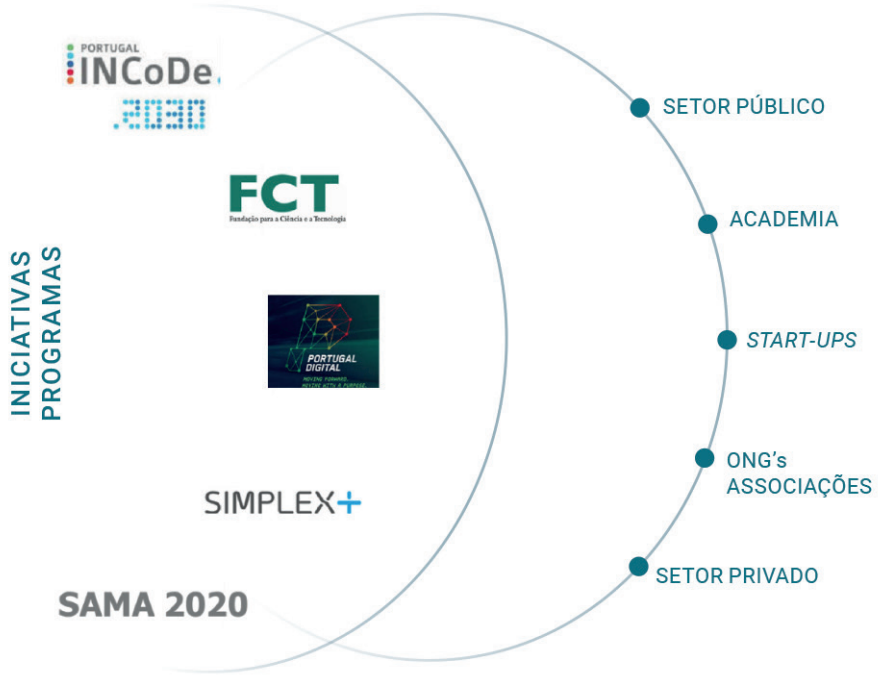

FIGURA 3: ECOSSISTEMA DE IA EM PORTUGAL.

A *AI Portugal 2030* constitui um processo coletivo para a construção de um mercado de trabalho intensivo em conhecimento. Mobilizando cidadãos, em geral, e principais *stakeholders*, em particular, tem a intenção de criar uma forte comunidade de empresas de vanguarda que produzem e exportam tecnologias de IA, apoiada por comunidades de desenvolvimento e inovação envolvidas em investigação de alto nível.

A Estratégia Nacional de Inteligência Artificial, oficialmente publicada em 2019, definiu cinco linhas de ação:

5 LINHAS DE AÇÃO

- **Especialização** em áreas com impacto internacional;
- **Modernização** da AP;
- **Disseminação** generalizada do conhecimento sobre IA;
- **Novos desenvolvimentos tecnológicos** (supercomputação, materiais e computação quântica);
- E **resposta** a desafios sociais associados à IA, nomeadamente aspetos relacionados com a ética e a segurança.

A Estratégia Nacional de Inteligência Artificial reconhece que os sistemas com IA devem ser transparentes, auditáveis, éticos (incluindo privacidade e justiça) e responsabilizáveis.

Reconhece também que os sistemas com IA tomam decisões críticas e importantes de forma autónoma, podendo pôr em risco princípios e valores da sociedade.

Mesmo antes desta Estratégia, o Regulamento Geral sobre a Proteção de Dados, publicado em 2016, alertava, entre outros aspetos, para a necessidade da transparência no tratamento dos dados, da segurança, do processamento de forma adequada e da responsabilidade, a assegurar pelo correto nível de tratamento e proteção dos dados pessoais.

PROGRAMA INCODE.2030

O INCoDe.2030 assume-se como uma iniciativa integrada de política pública dedicada ao reforço das competências digitais. Visa aumentar os conhecimentos, qualificações e competências da população, bem como, melhorar o posicionamento e competitividade de Portugal no contexto internacional. De modo a atingir estes objetivos foram definidas cinco linhas de ação: inclusão, educação, qualificação, especialização e investigação.

O programa INCoDe.2030, dedicado ao reforço das competências digitais, trouxe consigo vários temas de extrema importância para o posicionamento de Portugal na Sociedade da Inteligência.

A Estratégia Nacional para a Inteligência Artificial, bem como a Estratégia Nacional para a Computação Avançada, fazem parte da linha de ação de investigação que tem como objetivo garantir as condições para a produção de novos conhecimentos e a participação ativa em redes e programas internacionais de Investigação e Desenvolvimento (I&D)..

A FCT vem promovendo uma série de concursos desde 2018, cujo objetivo central é o apoio a atividades e projetos de I&D na área de *Data Science* e IA na AP que se enquadram na Iniciativa Nacional em Competências Digitais e.2030, Portugal INCoDe.2030 (designadamente no Eixo 5 – Investigação). Mais informações em https://www.incode2030.gov.pt/

SAMA 2020

SAMA 2020

No âmbito do SAMA 2020, o objetivo central para estes temas é promover a adoção por parte da AP de técnicas de análise e de modelos associados às áreas de IA ou *Data Science*, como por exemplo, análise preditiva, NLP, análise de padrões e ML.

Pretende-se promover a implementação de algoritmos e modelos de análise de dados, implementados pelo menos em protótipos funcionais, que permitam a demonstração e utilização experimental por parte das entidades da AP, e a implementação generalizada de sistemas completos e preparados para operar em ambiente real.

Os impactos esperados visam o aumento da eficiência e eficácia dos processos internos, contribuir para a implementação de políticas públicas e melhorar a prestação de serviços aos cidadãos e empresas.

Mais especificamente, pretende-se desenvolver soluções experimentais, em estreita colaboração com a comunidade científica, promovendo assim a transferência de conhecimento e a adoção de técnicas avançadas de IA e *Data Science* na AP.

No contexto destes programas, a Agência para a Modernização Administrativa (AMA) foi uma das entidades que financiou projetos nas áreas de *Data Science* e IA, no valor global de 10M de Euros, envolvendo a AP, as Empresas e a Academia.

2.2. ECOSSISTEMA E ATORES

O ecossistema abrange todas as organizações e indivíduos envolvidos ou afetados por sistemas com IA, direta ou indiretamente.

Os atores de IA são aqueles que desempenham um papel ativo no ciclo de vida do sistema de IA, incluindo organizações e indivíduos que implementam ou operam IA. Destes atores é esperado que garantam a colaboração multidisciplinar e a diversidade de pontos de vista em todo o ciclo de vida da IA para maximizar os benefícios e minimizar os danos potenciais.

Um ecossistema de IA deve:

- Dinamizar um diálogo público moderado pelo governo, bem informado e interativo, incluindo todas as partes interessadas, para melhorar a compreensão da IA, debater oportunidades e desafios relacionados com a IA para a economia, a sociedade e o mundo do trabalho, e informar os formuladores de políticas em todos os setores;
- Promover uma IA responsável na educação e investigação, por meio do intercâmbio de conhecimentos e das melhores práticas, da orientação para uma conduta organizacional responsável e de incentivos para transformar o conceito de IA Responsável numa vantagem competitiva;
- Elaborar, adotar e disseminar políticas que articulem os seus compromissos com os valores centrados no homem e na justiça e que se alinhem com instrumentos internacionais, como por exemplo as Diretrizes preconizadas pela OCDE, UE, UNESCO, ONU, etc;
- Encorajar a aliança entre o setor académico e o setor público para conjuntamente promoverem a inovação, o crescimento e o desenvolvimento humano numa ótica sustentável; e reforçar os mecanismos de partilha de dados entre estes setores para melhoria da qualidade dos algoritmos e redução/eliminação de vieses.

Para desenvolver sistemas de IA confiáveis, é indispensável consultar os interessados e a população alvo que podem ser afetados direta ou indiretamente pelo sistema ao longo do seu ciclo de vida. É benéfico solicitar *feedback* regular mesmo após a implantação e configurar mecanismos de longo prazo para a participação dos *stakeholders*, por exemplo, garantindo aos visados informações, consulta e participação em todo o processo de implementação de sistemas de IA nas organizações.

Os benefícios dos sistemas com IA são imensos e a UE quer garantir que estão disponíveis para todos. Isso requer uma discussão aberta e o envolvimento dos parceiros sociais e das partes interessadas, incluindo o público em geral. Muitas organizações já contam com painéis para discutir o uso de sistemas com IA e análise de dados. Esses painéis podem incluir vários membros, como especialistas jurídicos, especialistas técnicos, especialistas em ética, representantes dos consumidores e trabalhadores. Procurar ativamente a participação e o diálogo sobre o uso e impacto dos sistemas de IA permite a avaliação de resultados e das abordagens, podendo ser útil principalmente em contextos com maior complexidade.

2.3.
ECOSSISTEMA DE DADOS NA GÉNESE DA IA NA AP

Na génese da IA na AP encontra-se um ecossistema de dados essencialmente vinculado a três dimensões: *Big data* e dados abertos; pilares nacionais da governança de dados – o Regulamento Geral sobre a Proteção de Dados e o Regulamento Nacional de Interoperabilidade Digital; e princípios que suportam a sustentabilidade deste ecossistema, principalmente no setor público (Figura 4).

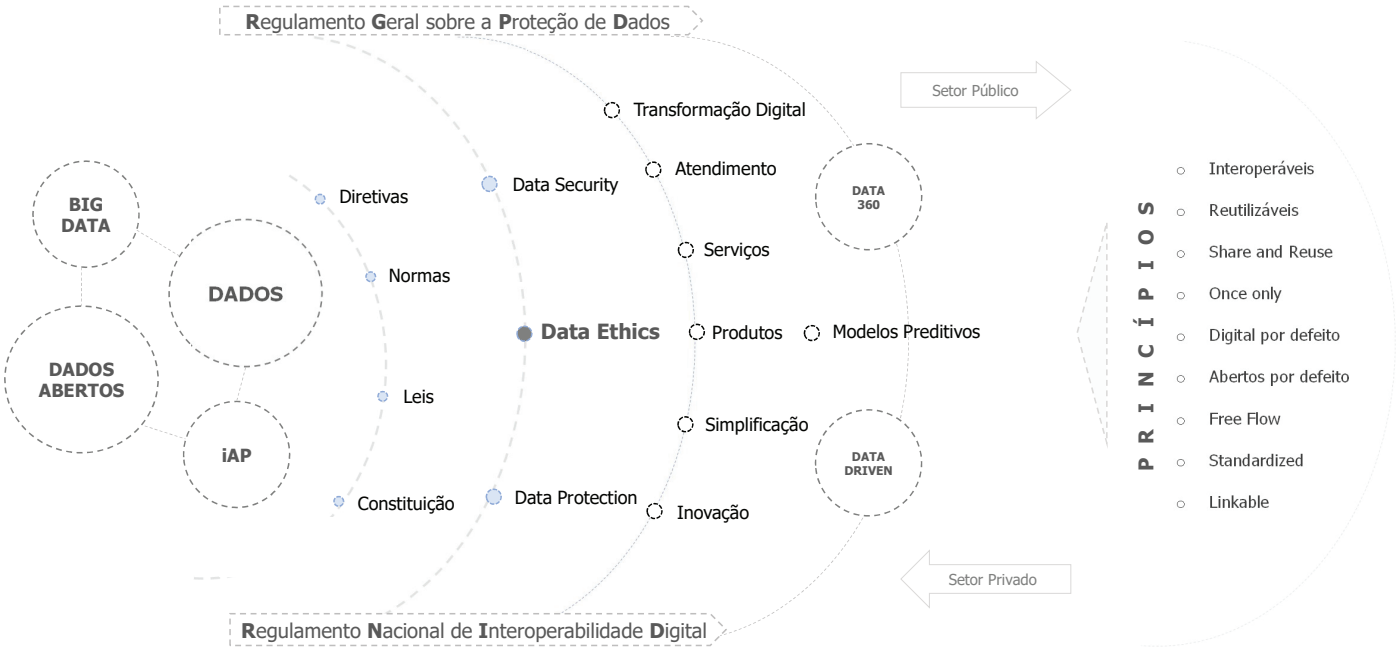

FIGURA 4: ECOSSISTEMA DE DADOS NA ADMINISTRAÇÃO PÚBLICA

A ética associada aos dados (*data ethics*), a segurança e a proteção dos dados (*data security and privacy*) são garantias fundamentais no reforço da confiança em abordagens *data 360,* dirigidas ou orientadas por dados. De seguida, descrevem-se alguns conceitos essenciais.

DATA ETHICS

Os dados assumem um papel essencial para a IA, uma vez que os algoritmos que a suportam são alimentados por dados. É, por isso, fundamental que a construção de um caminho para a ética na IA se sustente na ética, no acesso, na partilha e utilização de dados.

Uma das estratégias primordiais é o desenvolvimento de estruturas que permitam orientar e promovam o acesso, a utilização e reutilização da crescente quantidade de evidências, estatísticas e dados relativos a operações, processos e resultados, no sentido de aumentar a sua abertura e transparência. Esse crescimento do universo de dados deve, em consonância, ser acompanhado de instrumentos que lhes atribuam confiança. O envolvimento público na formulação de políticas, na criação de valor para a sociedade e no

desenvolvimento de serviços consolida o *framework* ético dos dados e gera uma comunidade mais orientada aos dados.

A ética dos dados é firmada pelo já vigente conjunto de leis e regulamentos que assegura os direitos e liberdades humanos, bem como por outros pressupostos que lhes são adicionados, dos quais se destacam os seguintes princípios éticos:

PRINCÍPIOS ÉTICOS

- A utilização dos dados respeita e serve o interesse público e cumpre o fim expectável;
- A finalidade de uma dada utilização é indicada com clareza e especificidade;
- São explicitados os limites à sua utilização;
- Os dados são utilizados com princípios de integridade;
- Os conjuntos de dados são compreensíveis, transparentes e o seu uso responsabilizável;
- Os dados devem ser abertos;
- Os cidadãos têm controlo sobre os seus dados pessoais;
- Os dados são utilizados para combater a discriminação e apoiar a inclusão.

DATA SECURITY & PRIVACY

A segurança e a privacidade dos dados são elementos basilares à proteção de informação e à sua confidencialidade, integridade e disponibilidade.

A segurança de dados é o processo através do qual se protegem dados de acessos não autorizados ou de ataques perniciosos e abusos de exploração. Os métodos utilizados incluem a monitorização de atividades, diferentes modos de criptografia e o controlo de acessos com autenticação segura por meio de chaves.

A privacidade dos dados relaciona-se com o modo como os dados são recolhidos, processados, armazenados, utilizados e eliminados, e garante que todos os processos são adequados e estão em conformidade com os direitos e liberdades dos indivíduos, com respeito às suas informações pessoais. A garantia da privacidade dos dados recai sobre a gestão de políticas, a aplicação de regulamentos ou leis e a coordenação com terceiros.

A cibersegurança, de dimensão digital, refere-se às medidas capazes de proteger os ativos individuais digitais de eventos deletérios, como erros técnicos ou humanos ou acesso por utilizadores não autorizados.

A abordagem dos riscos de segurança e privacidade dos cidadãos deve ser uma das prioridades de um governo aberto. O cumprimento desse fim pode ser conseguido através da modernização do quadro político e legislativo, apoiando, em simultâneo, uma estratégia de promoção da utilização de dados. As políticas, requisitos e ferramentas escolhidas têm de assegurar que os dados e informações são seguros, fiáveis e confiáveis. Em paralelo, a implementação de sistemas de avaliação e gestão de risco permitirá a identificação de ameaças emergentes no mundo digital e a formulação de respostas no sentido de proteger e mitigar potenciais impactos na segurança cibernética.

Em relação ao tratamento de informações pessoais, é importante garantir:

- Princípios de legalidade e justiça;
- A utilização dos dados apenas para a finalidade definida;
- A qualidade e a veracidade dos dados;
- A fiabilidade e nível de atualização dos dados;
- A retenção dos dados num formato que permita a identificação estritamente necessária do seu titular;
- A segurança e a privacidade.

Em Portugal, o conjunto de leis, diretivas, regulamentos, medidas e princípios que preservam os dados governamentais e os serviços digitais que garantem a adequada proteção das informações pessoais dos cidadãos estão transcritos nos documentos referenciados na tabela 1.

OBJETO	LEGISLAÇÃO	ÂMBITO
Aprova os princípios gerais em matéria de dados abertos e de reutilização de informação do setor público	Lei n.º 68/2021, de 26 de agosto; transpõe a Diretiva (UE) 2019/1024, do Parlamento Europeu e do Conselho, alterando a Lei n.º 26/2016, de 22 de agosto	Nacional
Relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados	Lei n.º 58/2019 de 8 de agosto - assegura a execução, na ordem jurídica nacional, do Regulamento (UE) 2016/679 do Parlamento e do Conselho, de 27 de abril de 2016	Nacional
Relativa aos dados abertos e à reutilização de informações do setor público (reformulação)	Diretiva (UE) 2019/1024 do Parlamento Europeu e do Conselho, de 20 de junho de 2019	Europeu
Estabelece os direitos de autor e direitos conexos no mercado único digital	Diretiva (UE) 2019/790 do Parlamento Europeu e do Conselho de 17 de abril de 2019	Europeu
Estabelece o regime para o livre fluxo de dados não pessoais na União Europeia	Regulamento (UE) 2018/1807 do Parlamento Europeu e do Conselho de 14 de novembro de 2018	Europeu
Define orientações técnicas para a Administração Pública em matéria de arquitetura de segurança das redes e sistemas de informação relativos a dados pessoais	Resolução do Conselho de Ministros n.º 41/2018	Nacional
Regulamento Nacional de Interoperabilidade Digital (RNID)	Resolução do Conselho de Ministros n.º 2/2018	Nacional
Aprova o regime de acesso à informação administrativa e ambiental e de reutilização dos documentos administrativos	Lei n.º 26/2016, de 22 de agosto; transpõe a Diretiva 2003/4/CE, do Parlamento Europeu e do Conselho, de 28 de janeiro, e a Diretiva 2003/98/CE, do Parlamento Europeu e do Conselho, de 17 de novembro	Nacional
Regulamento Geral sobre a Proteção de Dados (RGPD) - Relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados	Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016	Europeu
Estabelece a reutilização de informações do setor público	Diretiva 2013/37/UE do Parlamento Europeu e do Conselho, de 26 de junho de 2013	Europeu

TABELA 1: LEIS, DIRETIVAS E REGULAMENTOS DE PRESERVAÇÃO DOS DADOS GOVERNAMENTAIS E DOS SERVIÇOS DIGITAIS, EM PORTUGAL.

CONTINUAÇÃO ⟶

(CONTINUAÇÃO DA TABELA) OBJETO	LEGISLAÇÃO	ÂMBITO
Relativa ao tratamento de dados pessoais e à proteção da privacidade no setor das comunicações eletrónicas,	Lei n.º 46/2012, de 29 de agosto; transpõe a Diretiva n.º 2009/136/CE, na parte que altera a Diretiva n.º 2002/58/CE, do Parlamento Europeu e do Conselho, de 12 de julho	Nacional
Estabelece a adoção de normas abertas nos sistemas informáticos do Estado	Lei n.º 36/2011, de 21 de junho	Nacional
Relativa à conservação de dados gerados ou tratados no contexto da oferta de serviços de comunicações eletrónicas publicamente disponíveis ou de redes públicas de comunicações	Portaria n.º 694/2010, de 16 de agosto; procede à terceira alteração da Portaria n.º 469/2009, de 6 de maio	Nacional
Relativa ao serviço universal e aos direitos dos utilizadores em matéria de redes e serviços de comunicações eletrónicas, ao tratamento de dados pessoais e à proteção da privacidade no sector das comunicações eletrónicas e o à cooperação entre as autoridades nacionais responsáveis pela aplicação da legislação de defesa do consumidor	Diretiva 2009/136/CE do Parlamento Europeu e do Conselho, de 25 de novembro de 2009; altera a Diretiva 2002/22/CE, a Diretiva 2002/58/CE e o Regulamento (CE) no 2006/2004	Europeu
Relativa à conservação de dados gerados ou tratados no contexto da oferta de serviços de comunicações eletrónicas publicamente disponíveis ou de redes públicas de comunicações	Lei n.º 32/2008, de 17 de Julho; transpõe para a ordem jurídica interna a Diretiva n.º 2006/24/CE, do Parlamento Europeu e do Conselho, de 15 de Março	Nacional
Relativa à conservação de dados gerados ou tratados no contexto da oferta de serviços de comunicações eletrónicas publicamente disponíveis ou de redes públicas de comunicações	Diretiva 2006/24/CE do Parlamento Europeu e do conselho de 15 de março de 2006	Europeu
Relativa ao acesso do público às informações sobre ambiente	Diretiva 2003/4/CE do Parlamento Europeu e do Conselho, de 28 de janeiro de 2003	Europeu

TABELA 1: LEIS, DIRETIVAS E REGULAMENTOS DE PRESERVAÇÃO DOS DADOS GOVERNAMENTAIS E DOS SERVIÇOS DIGITAIS, EM PORTUGAL.

Referem-se ainda as seguintes **normas e plano estratégico**:

- Norma ISO/IEC 27000 - Princípios e Vocabulário: define a nomenclatura utilizada nas normas internacionais;
- Norma ISO/IEC 27001- Tecnologia da Informação: aborda técnicas de segurança e sistema de gestão de segurança da informação;
- Norma ISO/IEC 27002 - Tecnologia da Informação: aborda técnicas de segurança e código de prática para controlos de segurança da informação;
- Estratégia Nacional de Segurança do Ciberespaço: visa aprofundar a segurança e a proteção das redes e dos sistemas de informação e potenciar uma utilização livre, segura e eficiente do ciberespaço, por parte de todos os cidadãos e das entidades públicas e privadas.

DATA 360

O termo *Data 360* ou visão 360 graus do cliente, cidadão, consumidor ou empresa, designa todas as informações disponíveis e significativas, recolhidas por uma organização com o propósito de fornecer o atendimento e serviço mais personalizado e eficiente. O conceito é amplamente utilizado por entidades que implementam uma abordagem centrada no cliente para a sua atividade.

A importância da visão de 360 graus do cliente não pode ser exagerada. Ela melhora a eficácia de todos os esforços efetuados pelos clientes, prevê a procura potencial de serviços por parte dos clientes e ajuda na identificação de soluções integradas. A visão de 360 graus permite que as organizações forneçam a melhor experiência ao cliente, aumentando a fidelização e a satisfação.

BIG DATA

Big Data é um termo que descreve um grande volume de dados estruturados semiestruturados e não estruturados que são gerados a cada momento.

A emergência de novas tecnologias, como as redes móveis, as redes sociais e a IoT, revelou um aumento na criação de dados. Hoje, a conexão usual de carros, eletrodomésticos, *wearable devices*, e outros, veio gerar ainda mais dados passíveis de serem processados e transformados, e de criarem conhecimento e informações úteis.

O que distingue o *Big Data* está precisamente relacionado com a possibilidade e a oportunidade de cruzar esses dados provenientes de diversas fontes para obtermos *insights* mais rápidos e mais preciosos. A exigência dos cidadãos com os serviços públicos e também como consumidores e o aumento da competitividade obriga a inovar e dar respostas baseadas em elevados padrões de qualidade.

Quanto mais dados são gerados, maior será o esforço de processamento para gerar informações e conhecimento. Desta forma, a rapidez em obter informação faz parte de todo o potencial que o *Big Data* pode proporcionar no setor público e no setor privado.

Por último, de referir os vários V's que estão associados à definição do *Big Data*. São eles, o valor, a variabilidade, a variedade, a velocidade, a veracidade e o volume.

FIGURA 5: OS "V'S" DO *BIG DATA*

DADOS ABERTOS

Com o lançamento do portal dados.gov.pt em 2011, Portugal assumiu-se como um país pioneiro em compromissos com questões relativas a dados abertos e partilha de informação do setor público na Europa.

Nos últimos anos tem-se assistido a um crescimento exponencial de movimentos e plataformas associadas aos dados abertos. Têm sido lançados dezenas de portais nacionais, regionais ou locais em todo o mundo, o que contribuiu para vários desenvolvimentos a nível de protocolos, *standards* ou tecnologias.

Ao mesmo tempo, a nível nacional, o dados.gov.pt tem cumprido o seu papel como portal nacional de dados abertos da AP. A 26 de Agosto de 2021, a Lei n.º 68/2021 reconheceu-o juridicamente como o catálogo central de dados abertos em Portugal, atribuindo-lhe a função de agregar, referenciar, publicar e alojar dados abertos de diferentes organismos e setores da AP. Hoje, é um serviço partilhado de alojamento de dados provenientes de

vários organismos públicos, que permite disponibilizar a informação para sua reutilização através de mecanismos automatizados (*webservices*) ou de ficheiros para *download*.

O dados.gov.pt disponibiliza dados de diferentes domínios e sistemas, constituindo-se como o catálogo central de dados abertos em Portugal. Agrega mais de 4.834 *datasets* (conjuntos de dados) e 9.506 recursos, e conta com mais de uma centena de organizações e mais de mil utilizadores (dados de Setembro de 2021). Aloja informação utilizada em plataformas públicas, tais como o Mapa do Cidadão ou o Portal da Transparência Municipal. É hoje uma das peças centrais na estratégia de *open government* em Portugal, contando com cerca de 110.000 visitas nos últimos 2 anos.

Os dados abertos representam um subconjunto muito importante do vasto domínio de informação do setor público e são parte das políticas dedicadas ao Governo Aberto, combinando princípios da transparência, democracia, participação e colaboração e contribuindo para uma maior eficiência dos serviços governamentais e medição do impacto das políticas.

O portal dados.gov.pt ambiciona ser não só um repositório, mas um ponto de troca de informação em tempo real, onde os dados possam ser utilizados para a criação de valor para a sociedade em geral, através da produção de conhecimento, produtos e serviços.

Os dados gerados na AP congregam em si um potencial incomensurável de utilização, de criação de conhecimento e de desenvolvimento para a sociedade. A transformação digital e as tecnologias emergentes vieram contribuir para este valor, que lhe é intrínseco, ao contribuírem para uma crescente quantidade de dados gerados e aí centralizados.

O âmbito de reutilização destes dados pela AP, Academia e Empresas é muito vasto, sendo um exemplo recente, a criação de apps com base em dados georreferenciados.

Neste sentido, procura-se implementar um conjunto de melhorias contínuas e estabelecer uma evolução constante do portal, de forma a que a qualidade e a quantidade dos *datasets* disponibilizados possam criar mais oportunidades e desafios à sua reutilização.

A dinamização da comunidade em torno dos dados abertos e a reutilização desses mesmos dados, contribui para reunir evidências que possam apoiar a formulação de políticas de futuro melhor informadas, sustentadas e mais ajustadas, bem como, alcançar impactos sociais e económicos significativos.

Os dados abertos são um excelente contributo de conhecimento sobre políticas, estratégias e iniciativas e permitem apoiar o desenvolvimento de metodologias para avaliar o impacto e a criação de valor económico, social e de boa governança.

Alguns exemplos de **benefícios gerados a partir dos dados abertos**:

- Tempo poupado para os cidadãos, entidade públicas e empresas;
- Melhores decisões;
- Sustentabilidade/eficiência energética e ganhos para o ambiente;
- Novos empregos criados;
- Redução de custos para o setor público;
- Ganhos de eficiência e ganhos de produtividade;
- Desenvolvimento de tecnologia.

As estimativas feitas no contexto da visão da UE de construir uma economia europeia de dados, sublinham o potencial que o livre fluxo de dados tem para o crescimento económico em toda a Europa. Estima-se que o valor derivado da sua reutilização em 2030, atinja um valor de 194 mil milhões de euros.

No âmbito da Estratégia TIC 2020 (CTIC) – Eixo II Inovação e competitividade, foi dado cumprimento à Medida 6: Transparência e participação, alargando a divulgação e utilização de dados abertos através do portal dados.gov.pt.

Neste sentido, considera-se importante propor um plano transformador que incorpore soluções, que visem contribuir de forma relevante e impactante para a promoção dos dados abertos em Portugal, constituindo um acelerador para a disponibilização e reutilização de *datasets*, e que tenha a intenção de:

- Providenciar serviços inovadores incorporando soluções que possam ser percecionadas pelos *stakeholders* como facilitadoras e inovadoras;
- Eliminar barreiras associadas à escassez de recursos humanos e técnicos;
- Beneficiar um elevado número de entidades e cidadãos;
- Generalizar e replicar conhecimento pela comunidade;
- Definir *standards* para metadados (qualidade, estruturação e normalização);
- Definir elementos que permitam construir conhecimento imediato com base nos dados publicados;
- Promover a partilha do conhecimento gerado;
- Medir o impacto da abertura dos dados;
- Eliminar silos.

DATA DRIVEN

Uma abordagem orientada a dados, isto é *data driven*, significa que a gestão e tomada de decisões estratégicas se baseia na análise e interpretação de dados verificáveis e confiáveis. É um modo de otimizar recursos e tornar projetos, programas e serviços mais eficientes, efetivos e assertivos.

De acordo com a OCDE, a gestão orientada a dados é uma das dimensões prioritárias da transformação digital dos governos. Nessa perspetiva, através da utilização de dados, um governo ou empresa consegue com maior facilidade identificar prioridades e gerar valor onde este é essencial, antecipando tendências sociais e necessidades dos cidadãos e empresas. A partir de dados é possível gerar conhecimento que informe, avalie e permita a compreensão do impacto de políticas e serviços públicos. Deste modo, a governação de um país ou a gestão de uma entidade torna-se mais sustentável e resiliente, capaz de se transformar e adaptar.

Para tal ser possível, é necessária uma cultura e maturidade analítica que requer a formação das equipas e organizações, orientada às melhores práticas de gestão baseada em dados.

Na gestão orientada a dados, um dos maiores desafios é a capacitação e estímulo do estudo e da interpretação dos dados, de modo a conseguir respostas e soluções para problemas que não estejam assentes em suposições ou intuições. Para suprir esta lacuna utiliza-se a análise de dados para encontrar tendências e antecipar cenários.

Desde 2005, a estratégia nacional tem-se baseado na plataforma de Interoperabilidade da AP (iAP), como ambiente preferencial de governança e circulação de dados. A estratégia foca-se em seguir estruturas de fontes de informação autênticas, assumindo padrões de relevância para a interoperabilidade na AP, por meio de guias de boas práticas, catálogos de classificação e uma macroestrutura funcional. A visão é continuar a construir a Estratégia de Governança de Dados alinhada com a Estratégia de Transformação Digital.

Outro aspeto importante é o tema da interoperabilidade e a identificação de princípios que reforcem a interoperabilidade entre sistemas e que possam ser também geradores de dados de elevado valor.

No contexto da modernização administrativa, da desmaterialização e melhoria contínua dos processos da AP, com foco no serviço prestado aos cidadãos e empresas, a iAP proporciona um método fácil e integrado de disponibilização de serviços eletrónicos transversais, tornando-se uma peça fundamental no processo de modernização administrativa do Estado.

A iAP é uma plataforma comum, orientada a serviços, com o objetivo de disponibilizar à AP ferramentas para interligação entre sistemas. Esta permite a composição e disponibilização de serviços eletrónicos multicanal mais próximos das necessidades do cidadão e empresas, de forma ágil e com economia de escala, e promove a reutilização, a partilha e normalização de recursos.

Destina-se a organismos e entidades da AP, extensível através de suporte legal, ao setor privado, e segue os princípios *once only, share and reuse, standards* abertos, segurança e disponibilidade.

A iAP neste momento liga 123 entidades, gerando benefícios significativos no que diz respeito a poupanças, tempo poupado aos cidadãos e sustentabilidade ambiental, encontrando-se perto de suportar anualmente mais de meio milhar de milhão de mensagens de negócio. Ver mais em: https://www.iap.gov.pt/web/iap/iAP-em-numeros

2.4. INOVAÇÃO EM PORTUGAL

O aparecimento das tecnologias emergentes está relacionado com a exponencial disponibilidade de dados vinda da digitalização de serviços com origem no advento da internet. O processamento dos dados que estas tecnologias permitem motiva o modo como os cidadãos veem, agem e se envolvem com o que os rodeia.

No setor público, através da integração de informação, tecnologia e inovação conseguiu-se melhorar as operações e os serviços prestados. Esta abordagem integrada permite a compreensão da comunidade e, como resultado, a avaliação mais acurada das situações e a tomada de decisões ou de respostas mais rápidas, eficazes e adequadas. As tecnologias emergentes, designadamente as inteligentes, facilitam a inovação, a sustentabilidade e a competitividade, o que pode significar uma melhoria nos cuidados de saúde, na resposta às alterações climáticas, nos apoios sociais, na gestão das obras públicas e na educação.

Atualmente, a aplicação de tecnologias emergentes em cidades possibilita a análise de tendências de estacionamento, a gestão de energia, a monitorização dos níveis de poluição do ar, a otimização da recolha de resíduos e a disponibilização de serviços mais diferenciados e ajustados às necessidades de cada cidadão, como nos casos dos transportes públicos.

Aceder a essas oportunidades pressupõe que a gestão de dados no setor público seja acompanhada de recursos digitais, seja por meio de recursos humanos habilitados e/ou de ferramentas tecnológicas. No âmbito das tecnologias emergentes, é essencial que os funcionários compreendam o que estes sistemas podem trazer.

Um estudo realizado em 2020 pela Microsoft e pela Ernst & Young Global Limited confirmou que a maioria das entidades do setor público tem limitadas estruturas de IA e metade das entidades avaliadas que afirmou ter as estruturas de IA, ainda não as implementou. Dos setores inquiridos, o da saúde é o que tem a maior taxa de implementação de IA.

Apresentam-se de seguida, alguns casos de algoritmos e tecnologias utilizados para promover a inovação no setor público em Portugal, até ao ano 2021:

Setor Social - A Tarifa de Energia Social Automática é o produto da utilização de tecnologia para dar proteção social a famílias em situação de carência socioeconómica. Esta tarifa é atribuída de modo totalmente automático através da plataforma de Interoperabilidade na AP, que permite a identificação dos consumidores contemplados pelo cruzamento de dados ilegíveis da Segurança Social e da Autoridade Tributária.

Setor da Saúde - O Centro Hospitalar Universitário São João, integrado num projeto Europeu de ajuda aos serviços de Radiologia mais ocupados por pacientes com COVID-19, utiliza atualmente um sistema com IA capaz de analisar tomografias computadorizadas e de desencadear um aviso caso se verifiquem sinais sugestivos de infeção por este agente, assim como, delinear um prognóstico subsequente.

Setor da Segurança Social - O Instituto de Informática implementou e usa tecnologias ML num projeto de ChatBot piloto, na área interna de Gestão de Informação.

Setor da Justiça - O Tribunal de Contas utiliza a IA para conseguir um fluxo processual totalmente gerido por aplicativos digitais, através da digitalização, desmaterialização e automatização de processos. Uma das funcionalidades práticas é o apoio na tomada de decisões dos juízes. Nesse caso, existe uma plataforma dirigida por algoritmos de Aprendizagem Automatizada capaz de estabelecer ligações entre sentenças e processos judiciais, enquanto se adapta ao contexto português, e que gera informação mais eficiente e precisa passível de ser consultada rapidamente pelos órgãos judiciais.

Setor Agrícola - O Centro de Engenharia e Desenvolvimento (CEiiA) tem em curso um projeto de implementação de veículos aéreos não tripulados com integração de IA, Aprendizagem Automática e Inteligência Visual, para a monitorização e gestão de ativos agrícolas.

3.
IA ÉTICA, RESPONSÁVEL E TRANSPARENTE

3.1.
BASES SÓLIDAS DE GOVERNAÇÃO E LIDERANÇA

À medida que a IA se firma na sociedade e assume uma função proeminente na construção de decisões automatizadas, surgem paralelamente desafios éticos. Embora com evidente potencial de orientar serviços e aumentar a eficiência e eficácia das instituições governamentais, ao aumentar a sua capacidade de resposta às necessidades da sociedade, a consciencialização de riscos inerentes à sua implementação é perentória. Tal é indissociável de uma aplicação responsável e ética, ponderada em função do contexto e abordada metodicamente. Nesse sentido, deve ser parte da estratégia de implementação de IA, a criação de estruturas consolidadas de governação, nomeadamente políticas públicas, regulamentos e um código de ética.

A crescente aplicabilidade da IA a diferentes áreas na sociedade torna necessária uma abordagem mais geral, previsível e transparente, que permita a compreensão da sua complexidade e priorize com clareza o potencial em criar valor e o papel e a proteção do indivíduo e dos bens comuns à sociedade.

Idealmente, uma estratégia governamental em IA deve ser desenvolvida em cooperação com o setor científico, empresarial, público e a sociedade civil, envolvendo também associações, organizações e instituições com ação nacional. As medidas definidas devem procurar garantir sistemas com IA confiáveis, seguros, protegidos, transparentes, responsáveis e rastreáveis.

O código de ética permitirá a projeção de um conjunto de valores, princípios e diretrizes que acompanhem os desenvolvimentos tecnológicos, bem como, os elementos sociais e políticos associados. Um grupo de trabalho de ética e sociedade de IA deve ser estabelecido para investigar coletivamente os impactos éticos, investigar questões, definir diretrizes para as melhores práticas e publicar os conhecimentos adquiridos. Este grupo de trabalho deve estar alinhado com outros organismos internacionais, particularmente The *Partnership on AI*.

Por sua vez, na definição da estrutura regulatória e de limites legais à implementação da IA, um **governo deve procurar garantir**:

- O respeito pela privacidade, a inviolabilidade dos direitos humanos e o princípio da equidade social;
- O planeamento de mecanismos de segurança para proteger os sistemas de erros, como distorções, discriminação, manipulação e utilização indevida de dados;
- A mitigação de riscos associados à sua implementação e à mudança;
- A regulamentação da utilização dos dados;
- A conformidade com um código de ética;
- A eficiência e a sustentabilidade das tecnologias ao mesmo tempo que permite criarem benefícios para o cidadão, a sociedade, o ambiente, a economia e o país;
- A participação individual, a inclusão social, a liberdade de ação e a autodeterminação de cada cidadão em relação à IA;
- O incentivo ao investimento em investigação e desenvolvimento de IA;
- A promoção do potencial das tecnologias emergentes;
- O crescimento do tecido empresarial, incluindo as pequenas e médias empresas;
- A criação de valor na AP e sociedade;
- A integração de uma política de emprego que facilite a transição.

No âmbito destas temáticas é necessária uma estrutura de governança europeia sob a forma de um quadro de cooperação das autoridades nacionais competentes, para evitar a fragmentação de responsabilidades, aumentar a capacidade nos Estados-Membros e assegurar que a Europa se dote progressivamente da capacidade necessária para testar e certificar produtos e serviços baseados em IA. Para isso, deverá contar com uma rede de autoridades nacionais, bem como redes sectoriais e autoridades reguladoras, a nível nacional e comunitário. Adicionalmente, um comité de peritos poderá prestar assistência à Comissão.

3.2. DIMENSÕES

A previsão de uma crescente utilização dos sistemas de IA para assistência à sociedade ou delegação de decisões exige que haja uma concordância entre a sua aplicação e os valores que definem uma sociedade. É imperativo, para conseguir uma IA confiável, entender como apoiar o desenvolvimento, a implementação e a utilização desta tecnologia garantindo que melhore e defenda uma cultura democrática, o Estado de direito e os direitos fundamentais de uma sociedade. A reflexão para uma Inteligência Artificial ética assume assim um papel fundamental que se deve alicerçar nos conceitos de: Responsabilização, Transparência, Explicabilidade, Justiça e Ética.

Na concetualização do Guia e da Ferramenta procurou-se:

- Assegurar a complementaridade dos princípios suprarreferidos, nomeadamente nos pares transparência-explicabilidade e justiça-ética, em que o primeiro releva a existência e o segundo a efetividade;
- E criar as bases para a adoção progressiva e gradual de uma IA Responsável.

Apesar de não ser obrigatória a sequência de princípios proposta, a abordagem adotada para o Guia e para a Ferramenta possibilita a passagem progressiva para IA Responsável.

IA RESPONSÁVEL

A IA Responsável materializa-se quando sistemas baseados em técnicas adaptativas, i.e. sistemas inteligentes:

- Cumprem a execução de responsabilidades e possibilidade plena de auditoria/ inspeção (Responsabilização);
- Asseguram a visualização das suas componentes e dos procedimentos aplicados (Transparência);
- Explicam o funcionamento e as implicações decorrentes das funções computacionais (Explicabilidade);
- Incorporam garantias e salvaguardas aos utilizadores e beneficiários (Justiça);
- Asseguram mecanismos efetivos de mitigação de vieses inesperados, diminuindo desta forma possíveis riscos éticos (Ética).

Cinco questões de base devem ser colocadas para uma **IA Responsável**:

- **Responsabilização:** Os sistemas geram responsabilização, são seguros e passíveis de auditoria?
- **Transparência:** Os sistemas são transparentes?
- **Explicabilidade:** Os sistemas são facilmente compreendidos através da explicação facultada?
- **Justiça:** Os sistemas são justos e não discriminam?
- **Ética:** Os sistemas oferecem mitigações para tratar vieses éticos?

Ao assumirem estes cinco princípios (Responsabilização, Transparência, Explicabilidade, Justiça e Ética) os sistemas de IA tornam-se confiáveis.

Um sistema com IA Responsável garante que utilizadores, empresas e governos não incorram ou sofram consequências associadas a vieses produzidos por sistemas inteligentes que de alguma forma instanciem comportamentos diferentes dos planeados e, assim, respeitem os valores morais da sociedade que os alberga.

3.2.1. RESPONSABILIZAÇÃO

Uma das dimensões da IA que gera maior controvérsia é a responsabilidade. Os maiores obstáculos surgem quando se procura responder a questões como: Quem está no controlo? Como sabemos se um sistema de IA cumpre o propósito que realmente procuramos? Os dados são interpretados com precisão? Como podemos codificar um sistema de IA para cumprir obrigações legais como a Declaração Universal de Direitos Humanos ou a Lei de Privacidade? Quem ou o que é responsável pelos efeitos gerados da utilização dessa tecnologia, sejam benéficos ou perversos? O que será uma IA responsável?

É o conceito que mais contribui para a sustentabilidade de um *framework* ético de um sistema com IA.

A análise dos efeitos das decisões ou ações baseadas na IA demonstra que resultam de um conjunto multifatorial de interações em que estão implicados *designers*, programadores, utilizadores, *software* e *hardware*, etc. A responsabilização integra os indivíduos responsáveis pelas partes que desempenham num fluxo de trabalho complexo num sistema com IA, desde a conceção à sua implementação. E daí advém uma responsabilidade distribuída onde, numa cadeia de responsabilidade, o operador humano do sistema é responsável pelas decisões do algoritmo.

A garantia de uma IA ética carece assim de mecanismos que assegurem a responsabilização pela criação, desenvolvimento e/ou utilização de sistemas de IA, e minimizem os efeitos da parcialidade e opacidade inerentes às redes de autonomia artificiais.

Tais mecanismos estão intimamente relacionados com a gestão de risco e o rastreamento completo de procedimentos e resultados de modo transparente, de modo a poderem ser explicados e auditados por terceiros. Isto significa ter em conta a origem e a utilização de dados, modelos, interfaces de programação de aplicações e outros componentes estruturais desse sistema com IA. Inclui-se igualmente os responsáveis pela monitorização da solução, pelas revisões ao sistema na sua fase evolutiva, pelo suporte ao utilizador/beneficiário e pela supervisão e gestão de riscos éticos, antes e após estar em operação.

Uma das ferramentas à tomada de decisões de IA responsável é a definição de limites e regulamentos, por legisladores e autoridades, para os temas de recolha e utilização de dados e de certificados elegíveis para a produção de sistemas que utilizam IA.

3.2.2. TRANSPARÊNCIA E EXPLICABILIDADE

A transparência e a explicabilidade são duas dimensões complementares, uma vez que a primeira consolida o cumprimento da segunda.

A transparência num sistema com IA é garantida pela possibilidade de interpretação desse mesmo sistema, isto é, pela clareza e inteligibilidade do seu conteúdo ou da sua explicabilidade. A criação de algoritmos e sistemas de IA transparentes permite-nos explicar, inspecionar e reproduzir as decisões e a utilização dos dados por esses sistemas.

Quando não é possível explicar os processos técnicos implícitos e o raciocínio que fundamentou tais decisões ou previsões do sistema (*black boxes*), o recurso a outras medidas pode ainda assim garantir a sua transparência. Neste sentido, a apropriada documentação de dados e processos que geram as decisões do sistema de IA, incluindo informação relativa à recolha, ao armazenamento e à utilização de dados, ao permitir a rastreabilidade do ciclo de procedimentos de desenvolvimento desse sistema, aumenta a sua transparência. A comunicação clara das capacidades e nível de precisão do sistema, bem como das suas limitações é outro modo ainda de contribuir para a sua transparência.

Por outro lado, a recolha, o armazenamento e a utilização dos dados pessoais ou sensíveis devem ter expressa autorização dos seus detentores e toda a documentação deve estar disponível, de preferência em repositórios de livre acesso, quer internamente, quer para os utilizadores/beneficiários da solução que utiliza IA.

No final, mais que conhecer questões técnicas específicas, como os algoritmos, trata-se de ser capaz de explicar como uma decisão foi tomada por um modelo de IA e entender as implicações dos resultados/impactos decorrentes, esperados ou não.

A partir da transparência de informação inerente aos sistemas de IA, mecanismos de explicabilidade permitem ao utilizador e aos indivíduos que podem ser afetados por tais confirmar que o sistema funcionou como esperado e prever o que é possível desencadear a partir de diferentes ações computacionais. Por outro lado, reduz a resistência à utilização e aceitação da IA, ao consolidar a confiança da sociedade em relação à tecnologia.

Um sistema com IA transparente e passível de explicação torna os valores de base da entidade que o desenvolveu explícitos e conduz à responsabilização na tomada de decisões.

3.2.3. JUSTIÇA

A justiça é um conceito que se refere ao estado ideal de ausência de preconceito ou favoritismo direcionado a um indivíduo ou grupo, com base nas suas características herdadas ou adquiridas. Este engloba os ideais de equidade, imparcialidade, igualitarismo e não discriminação.

É necessário que o sistema com IA, em particular o algoritmo, garanta:

- Os direitos individuais (p.e., liberdade de expressão e proteção dos dados);
- Os direitos coletivos (p.e., direito à saúde, à educação, à cultura, à vacinação e à justiça);
- E o cumprimento de normativos.

A abertura à inclusão e à diversidade nos sistemas com IA, aproximam a comunidade através do aumento de confiança nestas tecnologias. É intuito de uma IA ética dirigir-se a todos os utilizadores, permitindo que todos possam utilizar os seus produtos e serviços. A acessibilidade está intimamente ligada ao sentido de justiça.

3.2.3.1.
VIÉS

A introdução ou criação de viés em sistemas de IA está intimamente ligada ao princípio de justiça. É essencial ter presente que pode haver inclusão inadvertida de histórias tendenciosas, incompletas e imprecisas, e modelos de má governança que incorrem no sentido oposto ao objetivo inicial do sistema. A perpetuação inconsciente destas tendências pode conduzir ao preconceito e à discriminação não intencional.

Um sistema com IA tendencioso pode criar, através do algoritmo que lhe foi imputado, resultados injustos, ao descrever erros sistemáticos e repetíveis num sistema de computador. Estes podem ser consequência, por exemplo, do algoritmo desenhado ou das escolhas relacionadas com a codificação, recolha, seleção ou utilização dos dados para alimentar o algoritmo. Elementos tendenciosos podem entrar no sistema algorítmico de IA por arquétipos culturais, sociais ou institucionais pré-existentes, por limitações técnicas ou porque são utilizados por indivíduos que não foram considerados na fase conceptual de desenvolvimento. Em todos os sistemas existem estes elementos e os impactos podem variar, no entanto a sua deteção exige que sejam removidos. Com frequência essa remoção tem de ser feita nas fases mais prematuras do processo, ou seja, na recolha de dados.

Algumas estratégias para gestão da justiça na IA foram já partilhadas por entidades líderes em IA, a exemplo a Google, a IBM (*International Business Machines Corporation*) e a Microsoft. É, contudo, essencial o desenvolvimento de procedimentos padronizados que verifiquem preconceitos ou outros objetos de discriminação adaptados ao número crescente de dados disponíveis e da celeridade desse mesmo crescimento.

No processo de construção de um sistema com IA, algumas questões podem ser formuladas com a finalidade de identificar elementos tendenciosos, entre as quais:

- Consideraram-se a diversidade e representatividade dos utilizadores finais e / ou sujeitos nos dados?
- Testaram-se grupos-alvo específicos ou casos de utilização problemáticos?
- Utilizaram-se ferramentas técnicas disponíveis publicamente, de última geração, para melhorar a compreensão dos dados, modelo e desempenho?
- Implementaram-se metodologias de testagem e monitorização para todo o processo do sistema com IA?

3.2.4. ÉTICA

À medida que a Inteligência Artificial se firma na sociedade e assume uma função proeminente na construção de decisões automatizadas, desafios éticos surgem paralelamente. Embora com evidente potencial de orientar serviços e aumentar a eficiência e eficácia das instituições governamentais, ao aumentar a sua capacidade de resposta às necessidades da sociedade, a consciencialização de riscos inerentes à sua implementação é perentória. Tal é indissociável de uma aplicação responsável e ética, ponderada em função do contexto e abordada metodicamente.

A confiança no florescimento de sistemas com IA implica uma abordagem previsível e transparente, que permita a compreensão da sua complexidade e priorize o benefício, o aumento de poder individual e a proteção do indivíduo e dos bens comuns à sociedade. Neste sentido, uma estrutura legal robusta não é suficiente *per se*, deve ser acompanhada de um código de ética. Este permitirá a projeção de um conjunto de valores, princípios e diretrizes que acompanhem os desenvolvimentos tecnológicos, bem como, os elementos sociais e políticos associados. A sua aplicação deve estar assegurada em todas as fases do ciclo de um sistema com IA e deve ser cumprida por todos os indivíduos associados.

Os princípios éticos da IA estão em consonância com os princípios éticos para qualquer iniciativa de dados. Releva-se o respeito pelo homem, o respeito pelos direitos humanos, a participação e a responsabilidade pelas escolhas e decisões.

—

O ciclo de valor dos dados abrange as etapas de recolha e criação, de armazenamento, proteção e processamento, de partilha, organização e publicação, e de utilização e reutilização. O sucesso da aplicação da IA resulta do sucesso de todas as etapas.

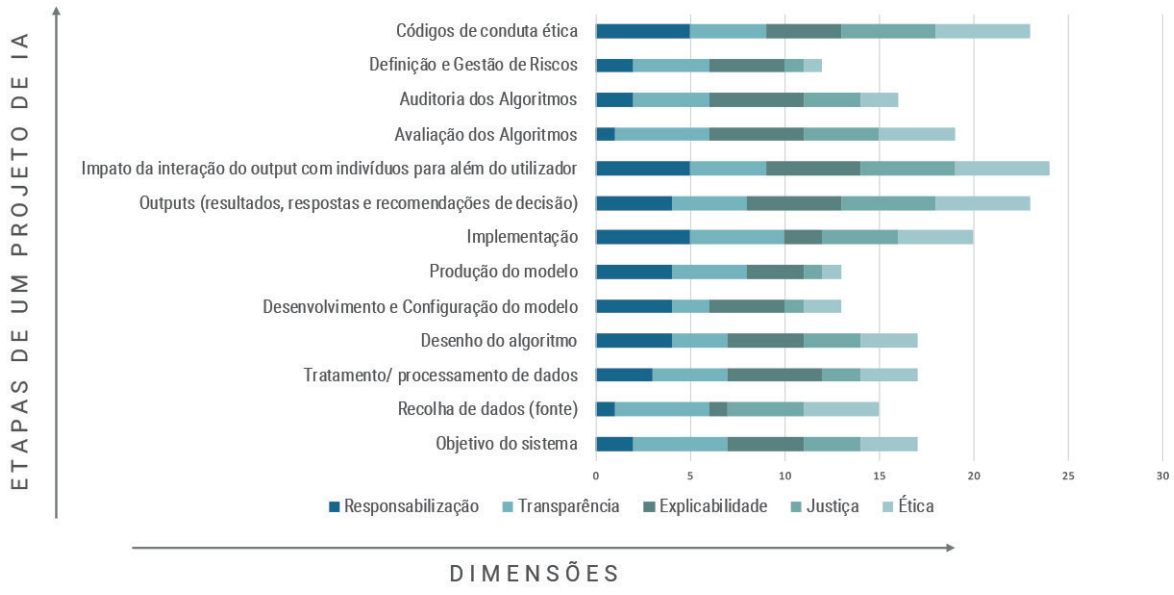

FIGURA 6: PONTOS CRÍTICOS À VULNERABILIDADE DAS DIMENSÕES EM FUNÇÃO DAS ETAPAS DE UM PROJETO COM IA.

As cinco dimensões referidas apresentam diferentes vulnerabilidades em função das etapas de desenvolvimento de um projeto com IA. É particularmente importante dar atenção ao impacto da interação do *output* com indivíduos para além do utilizador, aos *outputs* (resultados, respostas e recomendações de decisão) e aos códigos de conduta ética (Figura 6).

3.2.5.
DIREITOS HUMANOS

A IA afeta quase todos os direitos humanos reconhecidos internacionalmente, direta ou indiretamente, pelo modo como se interrelacionam e são interdependentes. Na tabela 2, indicam-se quais os direitos em situação de vulnerabilidade para alguns sistemas com IA já produzidos ou idealizados.

SISTEMAS DE INTELIGÊNCIA ARTIFICIAL	ARTIGOS DA DECLARAÇÃO UNIVERSAL DOS DIREITOS HUMANOS											
Recolha e análise de dados usando sistemas de IA								2	7	12	18	19
Previsão da conduta sexual por ML e reconhecimento facial				1	2	3	7	11	12	16	18	23
Modelos ML preditivos baseados em dados de localização – estimar idade, género, ocupação, estado civil e mobilidade						12	13	18	19	20	27	30
Difusão da cultura (música, cinema, arte) por sistemas de AI - principais produtores China e Ocidente								1	2	19	23	27
Modelos de ML para identificação de conteúdos pornográficos, violentos ou considerados politicamente sensíveis						10	12	18	19	20	23	30
Algoritmos de indexação de conteúdos em motores de pesquisa							18	19	21	26	27	30
Algoritmos que determinam o conteúdo do *feed* de notícias de um utilizador e com quem o conteúdo é compartilhado								18	19	21	26	27
Moderação de conteúdo online - cumprimento de padrões				2	3	12	18	19	21	23	27	30
Software de assistência à escrita de informação de disseminação – preparação de novas histórias ou outros conteúdos									18	19	23	30
Sistemas de classificação de conteúdo e de criação e reforço de filtros-bolha							12	18	19	21	27	30
Anúncios personalizados, baseados no comportamento dos utilizadores								2	18	19	21	27
Deep fakes								18	19	21	27	30
Chatbots de influência									18	19	20	21
Plataformas de comunicação social com algoritmos que estimam pontos de vista/opiniões que receberão maior visibilidade								2	18	19	21	27
Programas governamentais de monitorização de meios de comunicação social		2	3	7	11	12	18	19	20	21	27	30
Sistemas de IA para sinalização de publicações relacionadas com atos terroristas, discursos de ódio, *fake news*		2	3	7	11	12	18	19	20	21	27	30
Sistemas de vigilância de grupos – de grande importância em regimes ditatoriais							2	3	18	19	20	23
Vigilância com reconhecimento facial em locais de votação							2	18	19	21	27	30
Sistemas de vigilância com reconhecimento facial nas fronteira – controlo de entradas								12	13	18	19	20
Biometria no registo de refugiados e recomendação de decisões	1	2	6	9	12	13	15	18	22	23	29	30

TABELA 2: EXEMPLOS DE SISTEMAS DE IA E DIREITOS HUMANOS POR ELES IMPACTADOS.

CONTINUAÇÃO

SISTEMAS DE INTELIGÊNCIA ARTIFICIAL	ARTIGOS DA DECLARAÇÃO UNIVERSAL DOS DIREITOS HUMANOS											
Rodovias inteligentes e sistemas de transporte público com marcação biométrica - IoT aplicadas a infraestruturas											12	13
Vigilância com reconhecimento facial – drones de vigilância								12	13	14	27	
Sistemas de reconhecimento facial e dados de localização a partir de telemóveis e imagens de satélite										13	14	
Sistemas de decisão de julgamento jurídico					1	2	3	7	9	10	11	12
Reconhecimento facial – fins judiciais, sistemas governamentais centralizados para reconhecimento de ameaças				1	2	3	7	10	12	18	19	20
Modelos de IA projetados para classificar e filtrar, categorizando indivíduos, para aplicação em justiça criminal					1	2	3	7	9	10	11	12
Justiça criminal - *scoring* de risco de reincidência					1	2	3	7	9	10	11	12
Software de ML preditivo para identificar linguagem ou comportamentos que mostram uma propensão para a violência								1	2	3	7	10
Armas automatizadas						2	3	4	7	9	10	11
Acesso ao sistema financeiro - *scoring* de crédito					2	7	12	19	20	23	25	26
Cuidados de saúde – *Robots* em substituição de médicos							1	2	3	12	23	25
Triagem de saúde geral e reprodutiva								1	2	12	16	23
Testes genéticos								1	2	12	16	23
Atribuição de seguros de saúde								2	3	12	22	25
Previsão de prognósticos e recomendações em caso de doença									1	2	3	25
Controlo de pandemais - com dados de saúde e de vivilância digital									2	3	12	25
Admissão de candidaturas nas escolas - previsão de sucesso										1	2	26
Ensino por robots											23	26
Educação – correção automática de redações							1	12	19	23	25	26
Recursos humanos - recrutamento e contratação								2	12	19	20	23
Automatização de funções									1	2	23	25
Assistentes virtuais												23
Tradutores virtuais												23
Assistência social - p.e., ajuda à habitação para pessoas em situação de sem-abrigo									1	2	12	23

TABELA 2: EXEMPLOS DE SISTEMAS DE IA E DIREITOS HUMANOS POR ELES IMPACTADOS.

LEGENDA: 1. Igualdade de dignidade e direitos; **2**. Direitos e liberdades sem discriminação; **3.** Direito à vida, à liberdade e à segurança pessoal; **4.** Proibição da escravatura ou servidão; **5**. Ser livre de tortura ou tratamentos cruéis, desumanos ou degradantes; **6.** Direito ao reconhecimento da sua personalidade jurídica; **7.** Igualdade perante a lei; **8.** Direito a recurso efetivo para as jurisdições nacionais competentes; **9.** Ser livre de prisão, detenção ou exílio arbitrário; **10.** Direito a um julgamento público equitativo, independente e imparcial; **11.**Direito a ser considerado inocente até que a sua culpabilidade fique legalmente provada; **12.** Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; **13.** Direito à circulação livre e entrada e saída do seu país; **14.** Direito a asilo num outro país, em perseguição; **15.** Direito a uma nacionalidade e liberdade à sua mudança; **16.** Direito ao casamento e a constituir família; **17.** Direito à propriedade; **18.** Liberdade de pensamento, de consciência e de religião; **19.** Liberdade de opinião, de expressão e à informação; **20.** Liberdade de reunião e de associação pacíficas; **21.** Direito de acesso às funções públicas de um país e liberdade de voto no poder público; **22.** Direito à segurança social; **23.** Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos; **24.** Direito ao repouso e aos lazeres; **25.** Direito a um nível de vida adequado; **26.** Direito à educação; **27.** Direito de participação na vida cultural da comunidade; **28.** Direito a uma ordem, no plano social e no plano internacional, articulada com o presente nesta Declaração; **29.** Deveres comunitários essenciais ao desenvolvimento completo e livre dos princípios das Nações Unidas; **30.** Nenhum Estado, agrupamento ou indivíduo poderá destruir os direitos e liberdades enunciados.

3.3.
VALORES E PRINCÍPIOS

Os valores e princípios devem ser respeitados por todos os atores durante o ciclo de vida dos sistemas com IA, devem ser promovidos mediante uma avaliação e evolução contínua das leis, dos regulamentos e das várias diretrizes internacionais existentes, nomeadamente em relação aos direitos humanos, e estarem alinhados com os objetivos de sustentabilidade social, política, ambiental, educacional, científica e económica.

Os valores desempenham um papel importante como ideais que motivam a orientação de medidas políticas e normas jurídicas. Deste modo a identificação de valores associados à IA, permite inspirar comportamentos desejáveis e representa os fundamentos dos diversos princípios, que por sua vez, revelam os valores subjacentes de forma mais objetiva, de modo que os valores possam ser mais facilmente percecionados e incorporados em políticas dirigidas aos cidadãos e às empresas.

As pessoas devem poder confiar que os sistemas com IA oferecem benefícios que podem ser compartilhados por toda a sociedade, ao mesmo tempo que tomam medidas adequadas para mitigar os riscos. Um requisito essencial para aumentar a confiança é que em todo o seu ciclo de vida, os sistemas com IA estejam sujeitos a monitorização do governo, empresas privadas, sociedade civil e outras partes interessadas independentes.

Os sistemas computacionais clássicos, pelo fato de executarem apenas algoritmos simples, raramente foram alvo de reflexões sobre princípios e valores durante o seu desenvolvimento e operação.

Com o crescimento da dimensão das bases de dados, das novas tecnologias e dos mecanismos automáticos de recolha e partilha de dados, as primeiras iniciativas para a proteção e privacidade de dados ganharam relevo. No caso europeu, entrarem em vigor em maio de 2018.

As novas tecnologias inteligentes contribuíram para o aumento do raciocínio indutivo e abdutivo, que conduz inevitavelmente à geração de novas informações, muitas das quais frequentemente desconhecidas dos cidadãos, das empresas e dos governos. Deste modo, novos avanços na legislação e na proteção de dados precisam ser implementados, no alinhamento com os valores e princípios reconhecidos.

Pensar numa IA com valores e princípios é acima de tudo pensar numa Inteligência Artificial Responsável.

No caso de Portugal estes valores e princípios devem considerar:

A Constituição Portuguesa, nomeadamente:

- A dignidade da pessoa humana;
- Uma sociedade livre, justa e solidária.

A Constituição Europeia, nomeadamente:

- Os direitos individuais;
- As liberdades individuais.

A Declaração Universal dos Direitos Humanos, nomeadamente:

- O direito à vida;
- O direito à segurança.

São valores e princípios basilares a uma IA responsável e transparente:

- Inovação de forma responsável;
- Promoção de um ecossistema digital;
- Cooperação entre organismos para uma IA de confiança;
- Incorporação de *feedback* do sector privado, da indústria, de universidades e de organismos da AP;
- Promoção da partilha de dados em dados.gov.pt;
- Identificação de consequências negativas não intencionais que sistemas e soluções podem ter sobre os indivíduos e as comunidades a que se dirigem;
- Partilha de modelos de IA transparentes e explicáveis;
- Acesso aos consequentes benefícios;
- Identificação clara dos serviços que os fornecedores de IA disponibilizam;
- Promoção de um governo orientado para o utilizador, a abertura, a colaboração e a acessibilidade;
- Utilização inovadora e responsável das novas tecnologias;
- Disponibilização de ferramentas aos serviços públicos;
- Alinhamento dos sistemas de IA com elevado grau de autonomia com os valores humanos em toda a sua operação;
- Proteção prioritária dos valores sociais, de justiça e o interesse público;
- Benefício e capacitação do maior número de pessoas possível;
- A distribuição de prosperidade económica criada pela IA;
- Respeito e melhoria dos processos sociais e cívicos;
- Objeção às tecnologias e sistemas de IA utilizados para a tomada de decisões e/ou criação de serviços e produtos antagónicos aos valores prevalecentes na nossa sociedade, a exemplo, para o fabrico de armas, comércio de droga, prostituição, exploração infantil, tráfico humano, pornografia, e outros.

Os benefícios da IA devem colocar em primeiro lugar a Humanidade, em segundo lugar o Estado e, por último, a Organização. Benefícios universais devem sobrepor-se a benefícios organizacionais (Figura 7).

FIGURA 7: ELEMENTOS A TER EM CONTA NA GARANTIA DE UMA IA INCLUSIVA E QUE CONSIDERE TODOS NA SOCIEDADE.

3.4.
INCLUSÃO, IGUALDADE, DESENVOVIMENTO SUSTENTÁVEL E BEM-ESTAR

A consciencialização da precariedade social e material de uma comunidade cria espaço e estímulo para a ação dos setores industriais e empresariais, organizações não-governamentais, comunidade científica e tecnológica e autoridades locais, no sentido de combater carências e disfuncionalidades e progredir para um desenvolvimento mais sustentável. A IA surge enquanto ferramenta com grande potencial para criar efeitos benéficos e alinhados com os objetivos de um desenvolvimento sustentável.

As áreas onde a IA pode ser utilizada para a promoção de um desenvolvimento sustentável vão desde a saúde, à agricultura, ao abastecimento de água, à energia, ao ambiente, à equidade e ao transporte. É possível combinar maior produtividade para a economia, melhores condições e qualidade de vida e conservação da natureza.

A nível empresarial, verifica-se que a IA pode contribuir também para o aumento do seu impacto positivo nas dimensões social e ambiental ao melhorar a eficiência dos seus serviços, desenvolver novos produtos ou, ainda, promover a equidade e inclusão dos trabalhadores.

O bem-estar individual e coletivo deve ser um fim da IA. Sistemas com IA podem ser utilizados para favorecer a habitação, cultura, educação e saúde, promover a inclusão de populações sub-representadas, reduzir as desigualdades económicas, sociais e de género, e ainda produzir resultados benéficos para o planeta, contribuir para a ação climática e conservar os ambientes naturais.

O tratamento igualitário e justo de todos os indivíduos e grupos necessita que a conceção dos sistemas com IA e o tratamento de dados considere conscientemente a autonomia individual, grupos vulneráveis, princípios de não discriminação e um sentido de solidariedade.

Alguns dos conceitos que devem estar presentes nos preâmbulos do planeamento de um sistema com IA, assim como, na monitorização de efeitos e na medição de impactos estão ilustrados na figura 8.

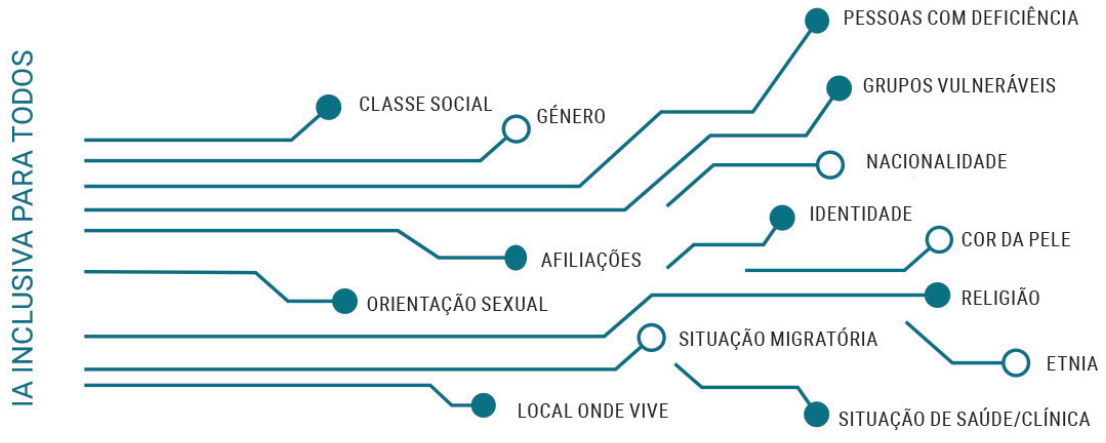

FIGURA 8: A IA DEVE SER INCLUSIVA E CONSIDERAR TODOS NA SOCIEDADE.

4.
IA MALICIOSA

Embora o valor da IA seja inquestionável pelo seu potencial de gerar benefícios para a sociedade, a utilização destes sistemas pode também desencadear resultados indesejáveis significativos para os indivíduos, organizações e sociedade. Quando se pensa em sistemas de IA complexos, capazes de aprender e se adaptar a desafios generalizados, prevê-se um aumento da incerteza e da dimensão dos efeitos gerados, sejam inadvertidos e imprevisíveis ou de intenção maliciosa.

A origem de tais efeitos danosos pode estar associada a sistemas programados para uma função que desencadeará resultados benéficos, mas cuja metodologia para atingir o objetivo pretendido foi destrutiva, ou a sistemas programados a *priori* para uma resposta perversa. Sem mecanismos rigorosos de controlo, proteção e segurança, os algoritmos podem ser corrompidos ou sistemas de IA podem ser criados com intuito malicioso.

Considerando que algumas dessas consequências foram já verificadas e ocorrem de modo limitado, e que o avanço tecnológico pode exacerbar os seus efeitos, a transição de uma posição de avaliação retrospetiva para uma preventiva torna-se assim essencial. A compreensão dos riscos, quais as suas interdependências e causas subjacentes, permitirá a sua identificação e priorização. Por sua vez, a identificação de riscos ocultos, incompreendidos ou *apriori* não identificados, otimizará a sua deteção nos modelos mesmo antes de serem implementados.

A aproximação aos elementos de risco, que podem conduzir a resultados maliciosos, deve ser efetuada em diferentes perspetivas que se possam complementar.

Deste modo, poder-se-á observar o risco como produto da:

- expansão de ameaças já identificadas e descritas, decorrente da eficiência, escalabilidade e facilidade de difusão destes serviços;
- introdução de novas ameaças;
- alteração completa das características das ameaças típicas, de modo a torná-las mais difíceis de prever.

Do ponto de vista da construção de um sistema com IA, a análise de risco deve compreender todas as fases, desde a conceção aos resultados. Assim sendo, podem-se identificar como pontos críticos de controlo:

- A fase de conceptualização – fundamentado por casos não éticos;
- A gestão de dados – adição de dados incompletos ou incorretos, dados não seguros e incumprimentos regulatórios;

- O desenvolvimento do algoritmo – escolha de dados não representativos, resultados tendenciosos ou discriminatórios e instabilidade do modelo/ baixo desempenho;
- A implementação do modelo – erros de implementação, projeto de *design* tecnológico fraco e treino e habilitações insuficientes;
- A utilização do modelo e o processo de decisão – mau funcionamento, lenta deteção e/ou resposta a questões de desempenho, ameaça de cibersegurança e falhas na interface humano-máquina.

Numa perspetiva de análise global dos sistemas podemos identificar como principais pontos problemáticos ao incremento de riscos nestes sistemas:

- **Os dados** – Problemas na estruturação dos dados, erros de escrita, lapsos na gestão de dados e julgamento erróneo na fase de treino do modelo podem perpetuar as divisões na sociedade, ao refletirem preconceitos sociais ou informação tendenciosa, comprometer a justiça, a privacidade, a segurança e a conformidade;
- **Problemas tecnológicos**;
- **Obstáculos à segurança**;
- **Mau comportamento dos modelos** – Resultados tendenciosos, modelos instáveis ou que levam a conclusões para as quais não há recurso acionável para aqueles afetados pelas suas decisões. Os sistemas com IA são inerentemente tendenciosos, na medida em que são desenvolvidos por humanos, o que reforça a dificuldade em alinhar o comportamento dos sistemas atuais com os objetivos definidos na sua conceção;
- **Interações Máquina-Homem** – os contextos mais desafiantes são os transportes, a manufaturação e as infraestruturas, onde já se registaram acidentes e lesões humanas.

A compreensão das semelhanças entre riscos específicos de utilização nefasta, não só maliciosa, permitirá direcionar esforços para sua prevenção e mitigação, com análises e decisões mais claras e informadas. A fim de ilustrar de que modo a IA pode assumir um caráter pernicioso, referem-se algumas potenciais e relevantes consequências:

- **Diminuição de emprego** – por substituição do trabalho automatizado; a criação de emprego através da IA prevê a necessidade de qualificação específica, pelo que muitos desses postos de trabalho serão inacessíveis a pessoas com menores níveis de escolaridades, de perfil correspondente àqueles que perderão o emprego pela IA;
- **Desigualdade socioeconômica**;
- **Discriminação**;
- **Decisões tendenciosas**;
- **Ameaças à segurança digital** - que comprometeriam a confidencialidade, integridade e disponibilidade dos sistemas digitais, nomeadamente:
    - Violação de privacidade;
    - *Phishing, chatbots* e criação de *malware;*

  - Exploração de vulnerabilidades humanas (por exemplo, síntese de voz), de *software* (por exemplo, *hacking* automatizado) ou de sistemas com IA (por exemplo, "envenenamento" de dados);
  - Automatização da procura de vulnerabilidade de códigos e *passwords;*
  - *Hacking;*
  - Negação de serviços através de personificação;
  - Identificação eficiente de vítimas;
  - Reaproveitamento terrorista de sistemas comerciais de IA.
- **Ameaças à segurança física** - dirigidos a humanos ou a infraestruturas; as armas autónomas assumem um papel preponderante por serem capazes de desestabilizar nações, subjugando populações e matando seletivamente indivíduos ou grupos. Alguns exemplos:
  - Ataques com drones e outros sistemas físicos (armas autónomas);
  - Subversão de sistemas ciberfísicos (acidentes com veículos autónomos);
  - Envolvimento de sistemas físicos que seria inviável dirigir remotamente (ataques enxame);
  - Ataques de grande escala;
  - Ataques rápidos e coordenados;
  - Ataques dissociados no tempo, espaço e de um ator.
- **Ameaças à segurança política** – Bloqueio de discussões verdadeiras, livres e produtivas sobre questões de importância pública, incapacitando a comunidade de decidir e implementar legitimamente políticas justas e benéficas; comprometimento da segurança nacional, pela introdução de informação num sistema com IA nacional, por adversários. Alguns exemplos:
  - Vigilância de plataformas para suprimir dissidentes;
  - Formação de perfis falsos;
  - *Deepfakes* de áudio e vídeo, para manipulação;
  - Manipulação social;
  - Analise mais eficiente de comportamentos, humores e crenças humanos para persuasão;
  - Campanhas de desinformação automatizadas e hiperpersonalizadas;
  - Campanhas de influência;
  - Manipulação da disponibilidade de informações;
  - Circulação de opiniões tendenciosas e dados falsos.
- **Perda de vida humana**, por exemplo, consequente de um erro num algoritmo médico de um sistema de IA.
- **Crimes**, como tráfico, venda, compra e posse de drogas proibidas
  - planeamento e navegação autónoma;
- **Atrofia de habilidades humanas**, como efeito de segunda ordem.

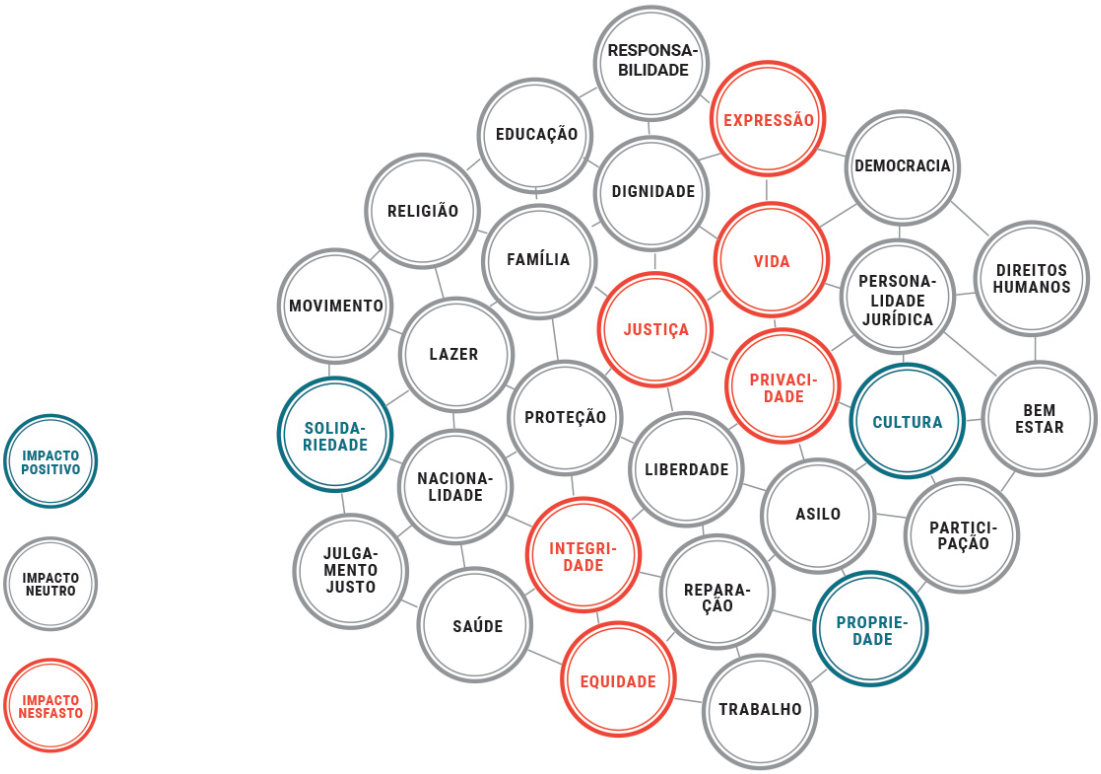

FIGURA 9: IMPACTO DE UMA SOLUÇÃO DE IA NAS VÁRIAS DIMENSÕES SOCIAIS E HUMANAS, TENDO COMO EXEMPLO DE SISTEMA DE IA UM *CHATBOT*.

Em casos de resultados maliciosos desencadeados por sistemas de IA, os detentores dessa tecnologia são paralelamente lesados. Os efeitos maliciosos dos seus produtos conduzirão, pela sua responsabilização, a danos de reputação, perdas de receita, reações regulatórias adversas, investigação criminal e à diminuição da confiança pública.

A multiplicidade de efeitos maliciosos que podem advir da utilização de um sistema de IA fundamenta a necessidade de uma abordagem multidisciplinar à sua prevenção e mitigação, com envolvimento dos dirigentes das companhias de desenvolvimento de modelos de IA e de especialistas das áreas jurídica, da tecnologia de informação, da segurança e da analítica, bem como, de legisladores, funcionários públicos, reguladores, educadores, cientistas e engenheiros de IA.

Tem de estar implícito na conceção de um sistema de IA o reconhecimento das dinâmicas sociais em todas as suas dimensões, a inclusão de diferentes etnias, géneros, culturas e grupos, e a representação dos planos fundadores das comunidades, como o direito, a economia, a psicologia, a medicina, a filosofia, a sociologia e a história. A integração destes conceitos permite cumprir os valores individuais e coletivos de uma sociedade enquanto se garante a evolução conjunta da IA.

No contexto político, é importante desimpedir possíveis entraves ao progresso da investigação científica e sustentar as escolhas políticas com a informação desencadeada da investigação acerca das técnicas e tecnologias disponíveis.

Por sua vez, é importante que os investigadores e engenheiros dedicados à inteligência artificial tenham presente os efeitos duplos do seu trabalho, de modo a conseguirem de-

finir objetivos e orientar conscientemente o seu trabalho, e identificar situações de risco e vulnerabilidade, desencadeando as ações necessárias à sua prevenção ou mitigação.

Como recomendações prioritárias à gestão de risco de implementação de serviços de IA maliciosos, indicam-se:

- Aprendizagem com a comunidade de cibersegurança – formar equipas de verificação formal, divulgar vulnerabilidades da IA, desenvolver ferramentas de segurança e hardware seguro.
- Exploração de diferentes modelos de abertura – criar mecanismos e modelos de avaliação de risco de pré-publicação em áreas técnicas de preocupação especial, de licenciamento de acesso central e de partilha, que promovam a segurança e proteção.
- Promoção de uma cultura de responsabilização - educar, instituir diretrizes e padrões éticos, regulamentos, normas e diretivas.
- Desenvolvimento de soluções tecnológicas e políticas – respostas legislativas e regulatórias de proteção da privacidade, uso coordenado de IA para segurança do bem público, monitorização de recursos relevantes.

A evolução que se antevê no domínio da IA, embora imbuída de um conjunto de incertezas, carrega um conjunto de questões relevantes em relação aos efeitos maliciosos na segurança digital, física e política. A preparação de respostas em cenários de utilização maliciosa, ou com consequências negativas não desejadas ou previstas, é por isso urgente.

5. RECOMENDAÇÕES

No âmbito do cumprimento e preservação de valores e princípios que definirão uma IA responsável, identificaram-se e organizaram-se 10 categorias principais de recomendações, descritas ***infra*** em detalhe:

O Controlo Humano

- Controlo humano da tecnologia;
- Controlo humano dos algoritmos;
- Inclusão de rotinas para treino e validação dos mecanismos de aprendizagem;
- Revisão humana de decisões automatizadas - As pessoas devem ser governadas por pessoas;
- Capacidade de reverter decisões automatizadas.

A Transformação Digital e Tecnológica

- Investimentos em dados e dados abertos;
- Esforços para acelerar a digitalização da AP;
- Inclusão digital;
- Ações da academia, indústria e sociedade civil.

A Cooperação e Envolvimento

- Empresas de tecnologia de IA;
- Fornecedores de IA para a AP;
- Organizações de utilizadores finais do setor privado;
- Consultoria;
- Especialistas em ética aplicada à IA, tecnologias emergentes e em ciências sociais;
- Equipas multidisciplinares com uma variedade de valências;
- Apoio a *start-ups*.

A Justiça e Não discriminação

- Prevenção de preconceitos subjacentes aos dados;
- Prevenção de preconceitos subjacentes aos princípios e pressupostos;

- Inclusão no desenho das soluções;
- Inclusão no impacto das soluções;
- Dados representativos;
- Dados de elevada qualidade;
- Reduzir o impacto negativo para os funcionários e, quando viável, permitir a sua participação na conceção e implementação desses sistemas.

A Privacidade

- Controlo de dados do utilizador;
- Consentimento;
- Recomendação e informação de leis de proteção de dados;
- Capacidade de restringir o processamento;
- Direito à retificação;
- Direito de apagar registo.

A Promoção de valores humanos

- Foco na criação de benefícios para a sociedade;
- Refletir os Valores e a Ética do Setor Público, bem como as obrigações internacionais e direitos humanos;
- Acesso à tecnologia para todos;
- Acesso à informação para todos;
- Replicabilidade por indivíduos nas mesmas circunstâncias.

A Responsabilidade Profissional

- Colaboração entre atores e *stakeholders*;
- *Design* responsável;
- Consideração de efeitos a longo prazo;
- Integridade e excelência científica;
- Precisão por meio de análises aprofundadas em todas as etapas;
- Supervisão externa;
- Qualificar e certificar fornecedores.

A Responsabilização

- Recomendação para novos regulamentos;
- Avaliação de impactos mensuráveis;
- Requisitos para avaliação e auditoria;
- Verificabilidade e replicabilidade;
- Responsabilidade legal;
- Capacidade de intervir;
- Responsabilidade ambiental;
- Criação de órgão de monitorização;
- Correção de decisões automatizadas;

- Explicação satisfatória e auditável da ocorrência de resultados erróneos.

A Segurança e Proteção

- Mecanismos de confiabilidade;
- Mecanismos de previsibilidade;
- Geração de alarmística;
- Colaboração estreita do Governo com os técnicos e investigadores para investigar, prevenir e mitigar os potenciais usos maliciosos de IA.

A Transparência e Explicabilidade

- Análise de dados recorrendo a código aberto;
- Algoritmos de código aberto;
- Notificação dos utilizadores e/ou beneficiários ao interagirem com IA;
- Notificação quando um sistema de IA toma uma decisão sobre um indivíduo ou um grupo de indivíduos;
- Requisito de relatórios regulares ao longo de todo o ciclo de vida da solução de IA;
- Direito à informação por parte dos utilizadores e ou beneficiários;
- Aquisição transparente de tecnologia para o Governo;
- Acesso à explicação das decisões tomadas;
- Comunicação à comunidade das mudanças permitidas pela IA.

Numa perspetiva geral da conceção de soluções de IA, é recomendado às partes interessadas:

1. A criação de um Comité Ético e de um Comité de Especialistas, que inclua profissionais das áreas em que são utilizadas tecnologias de IA (por exemplo, um painel de médicos e um painel de juízes);
2. A escolha de uma metodologia de gestão de projetos adequada, alinhada com uma estratégia de comunicação com as partes interessadas e ajustada às expectativas previstas e à sua mudança;
3. A inclusão de programas de formação/qualificação dos recursos humanos e dos utilizadores/beneficiários;
4. O desenho de um *roadmap* em torno de IA Responsável considerando os seguintes aspetos:

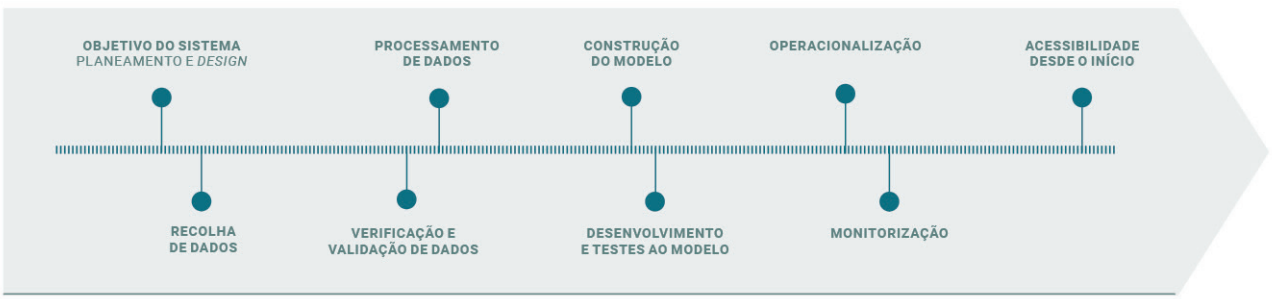

FIGURA 10: ETAPAS DE DESENVOLVIMENTO DE UM SISTEMA DE IA RESPONSÁVEL.

5. O planeamento de um projeto que responda às seguintes questões "Como?":
    - Inovar de forma responsável;
    - Identificar possíveis implicações éticas do uso da IA na organização e na sociedade;
    - Promover um ecossistema digital para IA;
    - Permitir a cooperação entre organismos para promover uma IA de confiança;
    - Incorporar pareceres do setor privado, indústria, academia, organismos da AP e promover a partilha de dados em dados.gov.pt;
    - Identificar consequências negativas não intencionais que os seus sistemas e soluções possam ter sobre os indivíduos e as comunidades que afetam;
    - Partilhar modelos de IA transparentes e explicáveis;
    - Gerar benefícios indiretos e identificá-los;
    - Tornar o Governo mais eficiente e melhor prestador de serviços;
    - Identificar fornecedores de serviços de IA com experiência comprovada e expertise em ética.

6. A consideração dos seguintes elementos, em relação aos algoritmos gerados:
    - Os pressupostos em que são baseados;
    - A atualização sistemática a que deverão estar sujeitos;
    - A avaliação da replicabilidade de soluções desenvolvidas;
    - O grau de supervisão humana sobre as decisões;
    - A capacidade para prever eventos raros;
    - A previsão e mensuração do impacto social resultante da sua aplicação;
    - As certificações necessárias;
    - Os indicadores de fiabilidade;
    - O impacto monetário nos organismos públicos;
    - A identificação de métricas para avaliar o treino e monitorização;
    - A garantia de padrões de excelência científica;
    - O grau de abertura na perspetiva de serem auditados e serem detetados eventuais erros.

6.
BARREIRAS E DESAFIOS

Entre as várias barreiras e desafios que se colocam à IA Responsável destacam-se seis: governança, aspetos organizacionais, aspetos legais, aspetos técnicos, aspetos financeiros e consciencialização (Figura 11).

FIGURA 11: PLANOS CRÍTICOS DE AÇÃO FACE A BARREIRAS E DESAFIOS NA IA.

7. FERRAMENTA DE AVALIAÇÃO DE RISCO

A Ferramenta de Avaliação de Risco traduz os valores e princípios de IA Responsável, detalhados ao longo do Guia. A utilização desta Ferramenta é indispensável à antecipação e mitigação de riscos em sistemas com IA de forma global e nas cinco dimensões: Responsabilização, Transparência, Explicabilidade, Justiça e Ética.

7.1. OBJETIVOS

A Ferramenta de Avaliação de Risco (Ferramenta) foi elaborada com o intuito de:

- Analisar a suscetibilidade de projetos de IA, de sistemas inteligentes ou de algoritmos relativamente às cinco dimensões subjacentes a uma IA Responsável;
- Comparar os resultados obtidos com as avaliações nacional e setorial de referência;
- Recomendar ações em função do nível de maturidade de IA auferido.

As dimensões consideradas transpõem os cinco princípios de IA Responsável adotados:

- Responsabilização (responsabilidade e possibilidade de auditoria/inspeção);
- Transparência (acesso às componentes e procedimentos);
- Explicabilidade (explicação do funcionamento);
- Justiça (proteção e garantias para os utilizadores e beneficiários);
- Ética (mecanismos efetivos de mitigação de vieses inesperados).

7.2. DESTINATÁRIOS

A Ferramenta destina-se a todas as pessoas e entidades que pretendam avaliar riscos em projetos de IA, sistemas inteligentes ou algoritmos que se encontrem numa das seguintes etapas:

- Conceção;
- Planeamento;

- Desenvolvimento inicial;
- Desenvolvimento avançado;
- Testes;
- Protótipo;
- Validação; e Produção.

Procura-se, deste modo, garantir a avaliação ao longo do ciclo do projeto, quer na fase anterior à implementação (*by design*), quer na fase posterior (*by evolution*).

A Ferramenta pode ser preenchida por qualquer pessoa, mesmo que não esteja associada a uma entidade ou a uma equipa específica do projeto. Esta pode inclusive ser utilizada por diferentes pessoas da mesma entidade. Como destinatários incluem-se:

- Pessoas externas à entidade;
- Utilizadores;
- Programadores;
- Analistas/ Engenheiros;
- Consultores de IA;

7.3. BENEFÍCIOS

A Ferramenta auxilia utilizadores e desenvolvedores na construção de sistemas inteligentes mais responsáveis por via da compreensão/ assimilação de conceitos e da mudança comportamental.

A Ferramenta contribui significativamente para a:

- Eliminação do efeito de *black box*;
- Compreensão de como a aprendizagem de máquina pode ser incorporada nas entidades;
- Redução de vieses;
- Proteção de pessoas vulneráveis;
- Não discriminação;
- Interdisciplinaridade;
- Identificação de riscos e impactos nos utilizadores/ beneficiários;
- Monitorização de resultados;
- Melhoria contínua do desempenho dos sistemas, através da aprendizagem;
- Melhoria das políticas e dos mecanismos de mitigação de riscos;
- Eficácia dos processos de inspeção/ auditoria dos sistemas;
- Segurança, qualidade e proteção dos sistemas;

- Compreensão dos resultados no contexto nacional e setorial;
- Sustentabilidade ambiental, social e económica.

7.4. ARQUITETURA

A Ferramenta está estruturada em:

- Conjunto de perguntas do tipo binário, *likert* ou escolha múltipla, validado anualmente;
- Ponderações atribuídas pela AMA e pelo utilizador a cada uma das cinco dimensões, validadas anualmente;
- Pontuação de avaliação;
- E matriz de recomendações associada ao nível de maturidade em que se encontra a entidade.

7.5. UTILIZAÇÃO

A utilização da Ferramenta tem início com a autenticação do utilizador através de Chave Móvel Digital ou do Cartão de Cidadão.

Após a autenticação, procede-se ao registo do utilizador, da entidade e do projeto, solicitando-se as informações explicitas na tabela 13.

Todas as perguntas são de resposta obrigatória.

Após preenchimento das respostas, o utilizador solicita o Relatório de Avaliação, o qual pode ser impresso ou arquivado em formato PDF.

No final da avaliação e, após consentimento prévio, o utilizador pode submeter os seus contactos pessoais para iniciativas desenvolvidas no âmbito da IA Responsável.

UTILIZADOR	ENTIDADE	PROJETO
• Habilitações académicas • Área profissional/ académica • Função	• Setor • Validação do registo associado à entidade • *Upload* do documento da entidade para efeitos de credenciação do utilizador	• Designação • Setor de aplicação • Contexto em que se desenvolve o projeto de IA • Objetivos do projeto • Resultados esperados • Aplicação • Técnica • Etapa do projeto • Grupo-alvo do projeto • Impacto esperado • *Framework*, Ambiente de Desenvolvimento Integrado, Simulador, Linguagem e Biblioteca utilizados no projeto • Tecnologias emergentes aplicadas

FIGURA 12: ELEMENTOS DE REGISTO DO UTILIZADOR, ENTIDADE E PROJETO.

7.6. NÍVEL DE MATURIDADE

O nível de maturidade da entidade está associado ao projeto consignado.

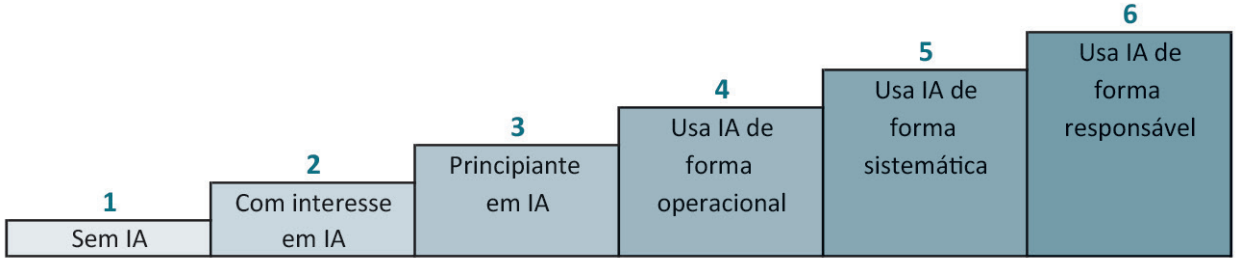

FIGURA 13: NÍVEIS DE MATURIDADE/ ESTÁGIOS DE MATURIDADE.

Este está relacionado com diferentes estágios de maturidade, ilustrados na figura 13:

1. Uma entidade sem IA, que não possui qualquer tipo de sistema baseado em técnicas adaptativas.
2. Uma entidade com interesse em IA, que tem pesquisado sobre o tema e discute internamente alguns assuntos, mas não concretizou nenhuma iniciativa com vista ao desenvolvimento de sistemas baseados em técnicas adaptativas.
3. Uma entidade principiante em IA, que tem ideias e já iniciou, de forma embrionária, alguns projetos de IA. Os membros da equipa são normalmente pluridisciplinares e estão orientados para a experimentação e visualização de resultados, na tentativa de perceber se o algoritmo tem interesse e cumpre a função pretendida.
4. Uma entidade que usa IA de forma operacional, que tem normalmente processos de ML implementados e equipas mais diversificadas em termos de conhecimento. Os algoritmos desenvolvidos são aplicados a casos de utilização concretos e apoiam de forma efetiva os processos de decisão.
5. Uma entidade que usa IA de forma sistemática, que aplica os pressupostos anteriores, suportando-os por princípios de IA, *frameworks* e *standards* internacionalmente reconhecidos.
6. Uma entidade que usa IA de forma responsável, que resulta do desenvolvimento do estágio antecedente, que segue processos e procedimentos que contribuem, no mínimo, para a transparência e para a mitigação dos riscos éticos dos sistemas.

7.7. RECOMENDAÇÕES

Cada pergunta é alvo de uma recomendação gerada em função de um limiar de respostas definido.

A matriz de recomendações considera as seguintes possibilidades:

	ABAIXO DO LIMIAR	ACIMA DO LIMIAR
Sem IA	Sugerimos ler sobre o tema	Sugerimos ler sobre o tema
Interesse em IA	Sugerimos ler sobre o tema	Excelente para a fase em que se encontra
Principiante em IA	Recomendamos ler e planear	Ótimo para a fase em que se encontra
Usa IA de forma operacional	Recomendamos planear e atuar	Bom para a fase em que se encontra
Usa IA de forma sistemática	É essencial atuar	Razoável para a fase em que se encontra
Usa IA de forma responsável	Já deveria ter atuado sobre este aspeto	Siga assim, está no bom caminho

FIGURA 14: MATRIZ DE RECOMENDAÇÕES.

Todas as recomendações são complementadas com sugestões de leituras acessíveis através da internet.

Expõem-se dois casos exemplificativos:

Caso A – Uma entidade que confirma usar IA de forma responsável e está abaixo de um dado limiar, apresenta por isso um risco elevado. Por essa razão "Já deveria ter atuado sobre esse aspeto".

Caso B – Uma entidade que confirma ter interesse em IA, iniciou um projeto piloto e está acima de um dado limiar no aspeto que está a avaliar, recebe a seguinte recomendação: "Excelente para a fase em que se encontra".

7.8. RELATÓRIO DE AVALIAÇÃO

O relatório de avaliação apresenta, por projeto, os seguintes resultados:

- Pontuação relativa ao projeto;
- Pontuação nacional;
- Pontuação setorial;
- Pontuação por dimensão de avaliação;
- Respostas dadas a cada pergunta;
- Recomendações por pergunta.

7.9. PARTICIPAÇÃO DE PARTES INTERESSADAS

A Ferramenta é uma plataforma evolutiva e pretende-se que venha a reunir os contributos dos seus utilizadores.

Caso detete algum erro ou identifique algum elemento que possa beneficiar a mesma, em relação às dimensões de avaliação ou a questões cruciais para a avaliação de riscos de sistemas inteligentes, deve submete-los à AMA através do seguinte email: guia@ama.pt

7.10. PROGRAMA DE AVALIAÇÃO PLURIANUAL

A Ferramenta está associada a um programa plurianual de avaliação das dimensões de IA Responsável.

Pretende-se que as entidades e os projetos a partir da Responsabilização, da Transparência e da Explicabilidade evoluam para sistemas cada vez mais justos e éticos.

8. FONTES COMPLEMENTARES

Access Now. (2018). *Human Rights in the Age of Artificial Intelligence*

Adler S. (2019). *A Quick Guide to Artificial Intelligence (AI), AI defined in a nutshell for corporate enterprise practitioners. Intelligent Automation network.* https://www.intelligentautomation.network/decision-ai/news/a-basic-guide-to-ai

Agência para a Modernização Administrativa. (2018). *ESTRATÉGIA TIC 2020 - Estratégia para a Transformação Digital na Administração Pública*

AI Group of experts. (2019). *AI applications.* OECD. https://www.oecd-ilibrary.org/sites/eedfee77-en/1/2/3/index.html?itemId=/content/publication/eedfee77-en&_csp_=5c39a73676a331d76fa56f36ff0d4aca&itemIGO=oecd&itemContentType=book

Alija A. (2020). *Emerging technologies and open data: ARTIFICIAL INTELLIGENCE.* Gobierno de España, red.es e iniciativa aporta

Altran Group. (2017). *Maximizing Value from AI - The digital transformers' guide.* Tessela

Bird, E., Fox-Skelly, J., Jenner, N. et al. (2020). *The ethics of artificial intelligence: Issues and initiatives.* Panel for the Future of Science and Technology. European Parliamentary Research Service. European Union

Burkhardt, R., Hohn, N. and Wigley, C. (2019). *Leading your organization to responsible AI.* McKinsey and Company. https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/leading-your-organization-to-responsible-ai#

Center for the study of existential risk. (2020). *Risks from Artificial Intelligence.* University of Cambridge. https://www.cser.ac.uk/research/risks-from-artificial-intelligence/

Cheatham, B., Javanmardian, K. and Samandari, H. (2019). *Confronting the risks of artificial intelligence.* McKinsey and Company. https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/confronting-the-risks-of-artificial-intelligence#

Commissioner for Human Rights. (2019). *Unboxing Artificial Intelligence: 10 steps to protect Human Rights.* Council of Europe

Conn, A. (2017). *How Do We Align Artificial Intelligence with Human Values?.* Future of Life Institute. https://futureoflife.org/2017/02/03/align-artificial-intelligence-with-human-values/

Data Flair. (2019). *Beware! Criminals are using AI to steal your personal details.* https://data-flair.training/blogs/how-criminals-use-ai/

Declaração Universal dos Direitos Humanos, Diário da República, I Série A, n.º 57/78, de 9 de Março de 1978

Deloitte AI and Analytics. (2020). *Human values in the loop, Design principles for ethical AI*. Deloitte. https://www2.deloitte.com/uk/en/insights/focus/cognitive-technologies/design-principles-ethical-artificial-intelligence.html

Directorate for Science, Technology and Innovation. (2019). *Report – Roles and Responsibilities of Actors: Governance of Digital Security in Organizations and Security of Digital Technologies.* Committee on Digital Economy Policy. Organization for Economic Co-operation and Development

Español A.G. (2020). *Marco ético para la inteligencia artificial en Colombia.* Consejería Presidencial para asuntos económicos y transformación digital.

European Parliament. (2020). *The ethics of artificial intelligence: Issues and initiatives.* European Union

Google. (n.d.). *Responsible AI practices.* https://ai.google/responsibilities/responsible-ai-practices/

Government Digital Service and Office for Artificial Intelligence. (2019). *Understanding Artificial Intelligence, Ethics and Safety.* GOV.UK. https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety

Government of Canada. (2020) *Responsible use of artificial intelligence (AI).* https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html

Hartwig, B. (2020). *The Impact of Artificial Intelligence on Human Rights. Transforming Data With Intelligence.* https://tdwi.org/articles/2020/06/29/adv-all-impact-of-ai-on-human-rights.aspx

Independent High-Level Expert Group on Artificial Intelligence. (2020). *The Assessment List for Trustworthy Artificial Intelligence (ALTAI), for self-assessment.* European Commission. 2020

Kashav, A. (2020). *Fairness in A.I.* Towards *data science*. https://towardsdatascience.com/fairness-in-a-i-5d3ceaaf649

Leslie, D. (2019). *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector.* The Alan Turing Institute

Manyika, J. (2018). *Automation and the future of work. McKinsey and Company.* https://www.mckinsey.com/mgi/overview/in-the-news/automation-and-the-future-of-work

Microsoft and Ernst & Young LLP. (2020). *Artificial Intelligence in the Public Sector, Portugal - How 213 Public Organizations Benefit from AI*. EY

Musikanski, L., Rakova, B., Bradbury, J. et al. (2020). *Artificial Intelligence and Community Well-being: A Proposal for an Emerging Area of Research*. Int. Journal of Com. WB 3, 39–55.

NoBIAS. (2020). *Artificial Intelligence without Bias.* NoBias. https://nobias-project.eu/

Open Data Institute. (2018). *Emerging tech and AI.* https://theodi.org/topic/emerging-tech

Silberg, J. and Manyika, J. (2019). *Tackling bias in artificial intelligence (and in humans).* McKinsey and Company. https://www.mckinsey.com/featured-in-

sights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans#
The Federal Government. (2018). K*ey points for a Federal Government Strategy on Artificial Intelligence.* Germany

Toreini, E., Aitken, M., Moorsel, A. et.al. (2019). *The relationship between trust in AI and trustworthy machine learning technologies*. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. 272–283

9. ANEXOS

Anexo A (Figura 1)

A figura apresenta um mapa publicado pelo Observatório de Inovação do Setor Público a 15 de Novembro de 2019, onde os países são classificados em função das Estratégias de IA.

Lista de países com Estratégia de IA completa à data de publicação do mapa: Canadá, Estados Unidos da América, México, Uruguai, Portugal, Espanha, França, Alemanha, Luxemburgo, Bélgica, Malta, Áustria, Reino Unido, Holanda, Dinamarca, Suécia, Finlândia, Estónia, Lituânia, Federação Russa, Índia, China, Japão, República da Coreia, Estados Árabes Unidos e China.

Lista de países com Estratégia de IA em elaboração à data de publicação do mapa: Colômbia, Chile, Argentina, Noruega, Irlanda, Polónia, Eslováquia, Hungria, Croácia, Grécia, Tunísia, Arábia Saudita, Quénia, Singapura, Malásia, Austrália e Nova Zelândia.

O mapa classifica ainda as abordagens seguidas por país com base em três classes:

a) **Estratégia de IA dedicada:** Canadá, Finlândia, Itália e Uruguai;

b) **Reconhecimento da importância pública da IA, mas com desenvolvimento da IA focado no sector privado:** Estados Unidos da América, Japão e República da Coreia;

c) **Estratégia de IA inserida numa abordagem mais vasta:** restantes países anteriormente citados.

Anexo B (Figura 2)

A figura, elaborada pelos autores, representa o cruzamento de dois indicadores para o ano 2019:

a) O **índice de prontidão para IA**, composto por variáveis que medem governança, infraestrutura de dados, governo e serviços públicos, competências e educação em IA e

b) O **índice global de IA,** composto por variáveis que medem implementação, inovação e investimento em IA.

Os países estão posicionados num sistema de coordenadas cartesiano, dividido em quatro quadrantes pelas linhas médias dos dois indicadores. A média do índice de prontidão para IA é de 6,8. A média para o índice global de IA é de 2,8.

No primeiro quadrante posicionam-se os países com índices superiores às médias dos dois indicadores. No segundo quadrante, posicionam-se os países com valores acima da média do índice de prontidão para IA, mas com valores abaixo da média do índice global de IA. No terceiro quadrante, os países com índices inferiores às médias dos dois indicadores. Finalmente, no quarto quadrante, os países com valores superiores à média do índice de prontidão para IA, mas com valores abaixo da média do índice global de IA.

A figura destaca os seis países com o índice de prontidão para IA e o índice global de IA mais elevados - Estados Unidos da América, Alemanha, Reino Unido, Canadá, França e Singapura – bem como, o posicionamento de Portugal neste sistema. Portugal está muito próximo da média do índice de prontidão para IA, mas ainda assim, abaixo da média do índice global de IA, o que evidencia a necessidade de mais implementação, mais inovação e mais investimento.

Anexo C (Figura 3)

A figura, elaborada pelos autores, apresenta o ecossistema de IA em Portugal. Este ecossistema resulta de iniciativas e programas governamentais e da atuação de vários atores.

Iniciativas e programas: SAMA 2020, SIMPLEX, Portugal Digital, INDoDE.2030 e projetos financiados pela FCT.

Atores: setor público, academia, *start-ups*, organizações não governamentais, associações e setor privado.

Anexo D (Figura 4)

A figura, elaborada pelos autores, apresenta o ecossistema de dados em Portugal. Este ecossistema é enquadrado, em termos normativos, pelo Regulamento Geral sobre a Proteção de Dados e pelo Regulamento Nacional de Interoperabilidade Digital. No cerne do ecossistema de dados estão quatro elementos: *big data*, dados, dados abertos e Interoperabilidade na Administração Pública. Estes quatro elementos são superentendidos por diretivas, normas, leis e pela Constituição; são partes integrantes de políticas de *data security, data ethics* e *data protection;* e estão sujeitos a processos de transformação digital, atendimento, serviços, produtos, predição, simplificação e inovação numa visão que deve ser *data 360º* e *data-driven*. Para o bom funcionamento deste ecossistema de dados, os seguintes princípios devem ser assegurados: abertos por defeito, digitais por defeito, estandardizados, interoperáveis, *linkable*, livre fluxo, *once only e sharing and reuse*. O ecossistema de dados é dinamizado por entidades do setor privado e do setor público.

Anexo E (Tabela 1)

A tabela sistematiza as leis, diretivas e regulamentos de preservação dos dados governamentais e dos serviços digitais em Portugal. A saber:

- **Lei n.º 58/2019 de 8 de agosto** - assegura a execução, na ordem jurídica nacional, do Regulamento (UE) 2016/679 do Parlamento e do Conselho, de 27 de abril de 2016. Esta Lei, de âmbito nacional, é relativa à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados;
- **Diretiva (UE) 2019/1024 do Parlamento Europeu e do Conselho, de 20 de junho de 2019.** Esta Diretiva, de âmbito europeu, regula os dados abertos e a reutilização de informações do setor público (reformulação);
- **Diretiva (UE) 2019/790 do Parlamento Europeu e do Conselho de 17 de abril de 2019.** Esta Diretiva, de âmbito europeu, estabelece os direitos de autor e direitos conexos no mercado único digital;
- **Regulamento (UE) 2018/1807 do Parlamento Europeu e do Conselho de 14 de novembro de 2018**. Este Regulamento, de âmbito europeu, estabelece o regime para o livre fluxo de dados não pessoais na União Europeia;
- **Resolução do Conselho de Ministros n.º 41/2018.** Esta Resolução, de âmbito nacional, define orientações técnicas para a Administração Pública em matéria de arquitetura de segurança das redes e sistemas de informação relativos a dados pessoais;
- **Resolução do Conselho de Ministros n.º 2/2018.** Esta Resolução, de âmbito nacional, estabelece o Regulamento Nacional de Interoperabilidade Digital (RNID);
- **Lei n.º 26/2016, de 22 de agosto**, transpõe a Diretiva 2003/4/CE, do Parlamento Europeu e do Conselho, de 28 de janeiro, e a Diretiva 2003/98/CE, do Parlamento Europeu e do Conselho, de 17 de novembro. Esta Lei, de âmbito nacional, aprova o regime de acesso à informação administrativa e ambiental e de reutilização dos documentos administrativos;
- **Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016. T**rata-se do Regulamento Geral sobre a Proteção de Dados (RGPD). Este regulamento é referente à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados;
- **Diretiva 2013/37/UE do Parlamento Europeu e do Conselho, de 26 de junho de 2013.** Esta diretiva, de âmbito europeu, estabelece a reutilização de informações do setor público;
- **Lei n.º 46/2012, de 29 de agosto** - transpõe a Diretiva n.º 2009/136/CE, na parte que altera a Diretiva n.º 2002/58/CE, do Parlamento Europeu e do Conselho, de 12 de julho. Esta Lei, de âmbito nacional, é relativa ao tratamento de dados pessoais e à proteção da privacidade no setor das comunicações eletrónicas;
- **Lei n.º 36/2011, de 21 de junho**. Esta Lei, de âmbito nacional, estabelece a adoção de normas abertas nos sistemas informáticos do Estado;
- **Portaria n.º 694/2010, de 16 de agosto** - procede à terceira alteração da Portaria n.º 469/2009, de 6 de maio. Esta Portaria, de âmbito nacional, é relativa à conservação de dados gerados ou tratados no contexto da oferta de serviços de comunicações eletrónicas publicamente disponíveis ou de redes públicas de comunicações;
- **Diretiva 2009/136/CE do Parlamento Europeu e do Conselho, de 25 de novembro de 2009** - altera a Diretiva 2002/22/CE, a Diretiva 2002/58/CE e o Regulamen-

to (CE) no 2006/2004. Esta Diretiva, é relativa ao serviço universal e aos direitos dos utilizadores em matéria de redes e serviços de comunicações eletrónicas, ao tratamento de dados pessoais e à proteção da privacidade no sector das comunicações eletrónicas e o à cooperação entre as autoridades nacionais responsáveis pela aplicação da legislação de defesa do consumidor;

- **Lei n.º 32/2008, de 17 de julho** - transpõe para a ordem jurídica interna a Diretiva n.º 2006/24/CE, do Parlamento Europeu e do Conselho, de 15 de março. Esta Lei, de âmbito nacional, é relativa à conservação de dados gerados ou tratados no contexto da oferta de serviços de comunicações eletrónicas publicamente disponíveis ou de redes públicas de comunicações;
- **Diretiva 2006/24/CE do Parlamento Europeu e do conselho de 15 de março de 2006.** Esta Diretiva, de âmbito europeu, é relativa à conservação de dados gerados ou tratados no contexto da oferta de serviços de comunicações eletrónicas publicamente disponíveis ou de redes públicas de comunicações;
- **Diretiva 2003/4/CE do Parlamento Europeu e do Conselho, de 28 de janeiro de 2003.** Esta Diretiva, de âmbito europeu, é relativa ao acesso do público às informações sobre ambiente.

Anexo F (Figura 5)

A figura, elaborada pelos autores, apresenta os seis V associados ao conceito de *big data*: valor, variabilidade, variedade, velocidade, veracidade e volume.

Anexo G (Figura 6)

O gráfico, elaborado pelos autores, apresenta os pontos críticos à vulnerabilidade das dimensões em função das etapas de um projeto com IA.

No eixo das ordenadas encontram-se as etapas de um projeto de IA. No eixo das abcissas, as cinco dimensões da IA Responsável.

Por etapa, as pontuações, numa escala de 1 a 5, são as seguintes:

1) **Objetivos do sistema:** Responsabilização 1; Transparência 5; Explicabilidade 4; Justiça 3; Ética 3;
2) **Recolha de dados (fonte):** Responsabilização 1; Transparência 5; Explicabilidade 1; Justiça 4; Ética 4;
3) **Tratamento/ processamento de dados:** Responsabilização 3; Transparência 4; Explicabilidade 5; Justiça 2; Ética 3;
4) **Desenho do algoritmo:** Responsabilização 4; Transparência 3; Explicabilidade 4; Justiça 3; Ética 3;
5) **Desenvolvimento e configuração do modelo:** Responsabilização 4; Transparência 2; Explicabilidade 4; Justiça 1; Ética 2;
6) **Produção do modelo:** Responsabilização 4; Transparência 4; Explicabilidade 3; Justiça 1; Ética 1;
7) **Implementação:** Responsabilização 5; Transparência 5; Explicabilidade 2; Justiça 4; Ética 4;

8) ***Outputs* (resultados, respostas e recomendações de decisão):** Responsabilização 4; Transparência 4; Explicabilidade 5; Justiça 5; Ética 5;
9) **Impacto da interação do *output* com indivíduos para além do utilizador:** Responsabilização 5; Transparência 4; Explicabilidade 5; Justiça 5; Ética 5;
10) **Avaliação dos algoritmos:** Responsabilização 1; Transparência 5; Explicabilidade 5; Justiça 4; Ética 4;
11) **Auditoria dos algoritmos:** Responsabilização 2; Transparência 4; Explicabilidade 5; Justiça 3; Ética 2;
12) **Definição e gestão de riscos:** Responsabilização 2; Transparência 4; Explicabilidade 4; Justiça 1; Ética 1;
13) **Códigos de conduta ética:** Responsabilização 5; Transparência 4; Explicabilidade 4; Justiça 5; Ética 5.

Anexo H (Tabela 2)

A tabela sistematiza o impacto dos artigos da declaração universal dos direitos humanos em diversos sistemas de inteligência artificial. A saber:

Acesso ao sistema financeiro (*scoring* de crédito): Direitos e liberdades sem discriminação; Igualdade perante a lei; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Liberdade de opinião, de expressão e à informação; Liberdade de reunião e de associação pacíficas; Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos; Direito a um nível de vida adequado; Direito à educação.

Admissão de candidaturas nas escolas (previsão de sucesso): Igualdade de dignidade e direitos; Direitos e liberdades sem discriminação; Direito à educação.

Algoritmos de indexação de conteúdo e decisão dos resultados em motores de pesquisa: Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Direito de acesso às funções públicas de um país e liberdade de voto no poder público; Direito à educação; Direito de participação na vida cultural da comunidade; Nenhum Estado, agrupamento ou indivíduo poderá destruir os direitos e liberdades enunciados.

Algoritmos que determinam o conteúdo do *feed* de notícias de um usuário e com quem o conteúdo é compartilhado: Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Direito de acesso às funções públicas de um país e liberdade de voto no poder público; Direito à educação; Direito de participação na vida cultural da comunidade.

Anúncios personalizados, baseados no comportamento dos utilizadores: Direitos e liberdades sem discriminação; Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Direito de acesso às funções públicas de um país e liberdade de voto no poder público; Direito de participação na vida cultural da comunidade.

Armas automatizadas: Direitos e liberdades sem discriminação; Direito à vida, à liberdade e à segurança pessoal; Proibição da escravatura ou servidão; Igualdade

perante a lei; Ser livre de prisão, detenção ou exílio arbitrário; Direito a um julgamento público equitativo, independente e imparcial; Direito a ser considerado inocente até que a sua culpabilidade fique legalmente provada.

Assistência social (por exemplo, ajuda à habitação para pessoas em situação de sem-abrigo): Igualdade de dignidade e direitos; Direitos e liberdades sem discriminação; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos.

Assistentes virtuais: Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos.

Atribuição de seguros de saúde: Direitos e liberdades sem discriminação; Direito à vida, à liberdade e à segurança pessoal; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Direito à segurança social; Direito a um nível de vida adequado.

Automatização de funções: Igualdade de dignidade e direitos; Direitos e liberdades sem discriminação; Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos; Direito a um nível de vida adequado.

Biometria no registo de refugiados e recomendação de decisões: Igualdade de dignidade e direitos; Direitos e liberdades sem discriminação; Direito ao reconhecimento da sua personalidade jurídica; Ser livre de prisão, detenção ou exílio arbitrário; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Direito à circulação livre e entrada e saída do seu país; Direito a uma nacionalidade e liberdade à sua mudança; Liberdade de pensamento, de consciência e de religião; Direito à segurança social; Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos; Deveres comunitários essenciais ao desenvolvimento completo e livre dos princípios das Nações Unidas; Nenhum Estado, agrupamento ou indivíduo poderá destruir os direitos e liberdades enunciados.

***Chatbots* de influência:** Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Liberdade de reunião e de associação pacíficas; Direito de acesso às funções públicas de um país e liberdade de voto no poder público.

Controlo de pandemias (com dados de saúde e de vigilância digital): Direitos e liberdades sem discriminação; Direito à vida, à liberdade e à segurança pessoal; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Direito a um nível de vida adequado.

Cuidados de saúde (robots de substituição de médicos): Igualdade de dignidade e direitos; Direitos e liberdades sem discriminação; Direito à vida, à liberdade e à segurança pessoal; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos; Direito a um nível de vida adequado.

***Deep fakes*:** Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Direito de acesso às funções públicas de um país e liberdade de voto no poder público; Direito de participação na vida cultural da

comunidade; Nenhum Estado, agrupamento ou indivíduo poderá destruir os direitos e liberdades enunciados.

Difusão da cultura (música, cinema, arte) por sistemas de IA (principais produtores China e Ocidente): Igualdade de dignidade e direitos; Direitos e liberdades sem discriminação; Liberdade de opinião, de expressão e à informação; Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos; Direito de participação na vida cultural da comunidade.

Educação (correção automática de redações): Igualdade de dignidade e direitos; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Liberdade de opinião, de expressão e à informação; Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos; Direito a um nível de vida adequado; Direito à educação.

Ensino por *robots*: Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos; Direito à educação.

Justiça criminal (*scoring* de risco de reincidência): Igualdade de dignidade e direitos; Direitos e liberdades sem discriminação; Direito à vida, à liberdade e à segurança pessoal; Igualdade perante a lei; Ser livre de prisão, detenção ou exílio arbitrário; Direito a um julgamento público equitativo, independente e imparcial; Direito a ser considerado inocente até que a sua culpabilidade fique legalmente provada; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação.

Modelos de IA projetados para classificar e filtrar, categorizando indivíduos, para aplicação em justiça criminal: Igualdade de dignidade e direitos; Direitos e liberdades sem discriminação; Direito à vida, à liberdade e à segurança pessoal; Igualdade perante a lei; Ser livre de prisão, detenção ou exílio arbitrário; Direito a um julgamento público equitativo, independente e imparcial; Direito a ser considerado inocente até que a sua culpabilidade fique legalmente provada; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Direito a um julgamento público equitativo, independente e imparcial; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Liberdade de reunião e de associação pacíficas; Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos; Nenhum Estado, agrupamento ou indivíduo poderá destruir os direitos e liberdades enunciados; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Direito à circulação livre e entrada e saída do seu país; Liberdade de pensamento, de consciência e de religião; Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Liberdade de reunião e de associação pacíficas; Direito de participação na vida cultural da comunidade; Nenhum Estado, agrupamento ou indivíduo poderá destruir os direitos e liberdades enunciados.

Moderação de conteúdo online (cumprimento de padrões): Direitos e liberdades sem discriminação; Direito à vida, à liberdade e à segurança pessoal; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e

na correspondência, e ataques à honra e reputação; Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Direito de acesso às funções públicas de um país e liberdade de voto no poder público; Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos; Direito de participação na vida cultural da comunidade; Nenhum Estado, agrupamento ou indivíduo poderá destruir os direitos e liberdades enunciados.

Plataformas de comunicação social com algoritmos que estimam pontos de vista/ opiniões que receberão maior visibilidade: Direitos e liberdades sem discriminação; Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Direito de acesso às funções públicas de um país e liberdade de voto no poder público; Direito de participação na vida cultural da comunidade.

Previsão de conduta sexual por ML e reconhecimento facial: Igualdade de dignidade e direitos; Direitos e liberdades sem discriminação; Direito à vida, à liberdade e à segurança pessoal; Igualdade perante a lei; Direito a ser considerado inocente até que a sua culpabilidade fique legalmente provada; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Direito ao casamento e a constituir família; Liberdade de pensamento, de consciência e de religião; Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos; Igualdade de dignidade e direitos; Direitos e liberdades sem discriminação; Direito à vida, à liberdade e à segurança pessoal; Direito a um nível de vida adequado.

Programas governamentais de monitorização de meios de comunicação social: Direitos e liberdades sem discriminação; Direito à vida, à liberdade e à segurança pessoal; Igualdade perante a lei; Direito a ser considerado inocente até que a sua culpabilidade fique legalmente provada; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Liberdade de reunião e de associação pacíficas; Direito de acesso às funções públicas de um país e liberdade de voto no poder público; Direito de participação na vida cultural da comunidade; Nenhum Estado, agrupamento ou indivíduo poderá destruir os direitos e liberdades enunciados.

Recolha e análise de dados usando sistemas de IA: Direitos e liberdades sem discriminação; Igualdade perante a lei; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação.

Reconhecimento facial (fins judiciais, sistemas governamentais centralizados para reconhecimento de ameaças): Igualdade de dignidade e direitos; Direitos e liberdades sem discriminação; Direito à vida, à liberdade e à segurança pessoal; Igualdade perante a lei; Direito a um julgamento público equitativo, independente e imparcial; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Liberdade de reunião e de associação pacíficas.

Recursos humanos (recrutamento e contratação): Direitos e liberdades sem dis-

criminação; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Liberdade de opinião, de expressão e à informação; Liberdade de reunião e de associação pacíficas; Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos.

Rodovias inteligentes e sistemas de transporte público com marcação biométrica (IoT aplicadas a infraestruturas): Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Direito à circulação livre e entrada e saída do seu país;

Sistema de reconhecimento facial e dados de localização a partir de telemóveis e imagens de satélite: Direito à circulação livre e entrada e saída do seu país; Direito a asilo num outro país, em perseguição.

Sistemas de classificação de conteúdo e de criação e reforço de filtros-bolha: Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Direito de acesso às funções públicas de um país e liberdade de voto no poder público; Direito de participação na vida cultural da comunidade; Nenhum Estado, agrupamento ou indivíduo poderá destruir os direitos e liberdades enunciados.

Sistemas de decisão do julgamento jurídico: Igualdade de dignidade e direitos; Direitos e liberdades sem discriminação; Direito à vida, à liberdade e à segurança pessoal; Igualdade perante a lei; Ser livre de prisão, detenção ou exílio arbitrário; Direito a um julgamento público equitativo, independente e imparcial; Direito a ser considerado inocente até que a sua culpabilidade fique legalmente provada; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação.

Sistemas de IA para sinalização de publicações relacionadas com atos terroristas, discursos de ódio, *fake news*: Direitos e liberdades sem discriminação; Direito à vida, à liberdade e à segurança pessoal; Igualdade perante a lei; Direito a ser considerado inocente até que a sua culpabilidade fique legalmente provada; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Liberdade de reunião e de associação pacíficas; Direito de acesso às funções públicas de um país e liberdade de voto no poder público; Direito de participação na vida cultural da comunidade; Nenhum Estado, agrupamento ou indivíduo poderá destruir os direitos e liberdades enunciados.

Sistemas de vigilância com reconhecimento facial nas fronteiras (controlo de entradas): Direito à circulação livre e entrada e saída do seu país; Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Liberdade de reunião e de associação pacíficas.

Sistemas de vigilância de grupos (de grande importância em regimes ditatoriais): Direitos e liberdades sem discriminação; Direito à vida, à liberdade e à segurança pessoal; Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Liberdade de reunião e de associação pacíficas; Direito ao trabalho desejado, em condições equitativas e à fundação e

filiação a sindicatos.

***Software* de assistência à escrita de informação de disseminação (preparação de novas histórias ou outros conteúdos):** Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos; Nenhum Estado, agrupamento ou indivíduo poderá destruir os direitos e liberdades enunciados; Igualdade de dignidade e direitos; Direitos e liberdades sem discriminação; Direito à vida, à liberdade e à segurança pessoal; Igualdade perante a lei; Direito a um julgamento público equitativo, independente e imparcial.

Testes genéticos: Igualdade de dignidade e direitos; Direitos e liberdades sem discriminação; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Direito ao casamento e a constituir família; Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos; Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos.

Tradutores virtuais: Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos.

Triagem de saúde geral e reprodutiva: Igualdade de dignidade e direitos; Direitos e liberdades sem discriminação; Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Direito ao casamento e a constituir família; Direito ao trabalho desejado, em condições equitativas e à fundação e filiação a sindicatos.

Vigilância com reconhecimento facial (*drones* de vigilância): Ninguém sofrerá intromissões arbitrárias na vida privada, na família, no domicílio e na correspondência, e ataques à honra e reputação; Direito à circulação livre e entrada e saída do seu país; Direito a asilo num outro país, em perseguição; Direito de participação na vida cultural da comunidade.

Vigilância com reconhecimento facial em locais de votação: Direitos e liberdades sem discriminação; Liberdade de pensamento, de consciência e de religião; Liberdade de opinião, de expressão e à informação; Direito de acesso às funções públicas de um país e liberdade de voto no poder público; Direito de participação na vida cultural da comunidade; Nenhum Estado, agrupamento ou indivíduo poderá destruir os direitos e liberdades enunciados.

Anexo I (Figura 7)

A figura ilustra os elementos a ter em conta na garantia de uma IA inclusiva e que considere todos na sociedade. Os elementos, por ordem de importância são: Humanidade, Estado e Organização.

Anexo J (Figura 8)

A figura apresenta os aspetos a ter em conta para uma IA inclusiva e que inclua todos na sociedade. A saber: classe social, orientação sexual, género, afiliações, local onde vive, situação migratória, identidade, situação de saúde/ clínica, nacionalidade, grupos vulneráveis, pessoas com deficiência, cor da pele, religião e etnia.

Anexo K (Figura 9)

A figura apresenta um sistema de IA que consiste num *chatbot*, permitindo evidenciar um exemplo de classificação dos possíveis impactos nas várias dimensões sociais e humanas.

Impacto positivo: solidariedade, cultura e propriedade.

Impacto neutro: movimento, julgamento justo, religião, lazer, nacionalidade, saúde, educação, família, proteção, responsabilidade, dignidade, liberdade, reparação, trabalho, asilo, democracia, personalidade jurídica, participação, direitos humanos e bem-estar.

Impacto nefasto: integridade, equidade, justiça, expressão, vida e privacidade.

Anexo L (Figura 10)

A figura apresenta as etapas de desenvolvimento de um sistema de IA Responsável:

- Primeira etapa: objetivo do sistema (planeamento e *design*);
- Segunda etapa: Recolha de dados;
- Terceira etapa: Verificação e validação de dados;
- Quarta etapa: Processamento de dados;
- Quinta etapa: Construção do modelo;
- Sexta etapa: Desenvolvimento e testes ao modelo;
- Décima etapa: Operacionalização
- Décima primeira etapa: Monitorização

A acessibilidade, por defeito, deve ser assegurada em todas as etapas.

Anexo M (Figura 11)

A figura apresenta os planos críticos de ação face a barreiras e desafios na IA:

- **Governança:** estratégia e apoio político e institucional; constrangimentos financeiros; iliteracia digital; baixo investimento em desenvolvimento e investigação em IA; rigidez cultural;
- **Organizacionais:** divergências das políticas de registo e partilha de dados num mesmo setor; sustentabilidade das iniciativas inconsistente; elevado custo dos recursos humanos e materiais; ausência de reforço colaborativo;
- **Legais (legislação, políticas, diretivas e licenças):** ambiguidade das medidas de segurança e privacidade; sistemas jurídicos desatualizados; sustentabilidade das iniciativas inconsistente; regime de propriedade intelectual pouco atraente, para investigadores e investidores nas soluções de IA;
- **Técnicos (infraestruturas, plataformas e tecnologia):** falta de mão de obra experiente e especializada; elevado custo dos recursos humanos e materiais; ausência de ecossistemas de dados habilitadores;

- **Financeiros:** constrangimentos financeiros; baixo investimento em desenvolvimento e investigação em IA; elevado custo dos recursos humanos e materiais; ausência de ecossistema de dados habilitadores;
- **Consciencialização:** regime de propriedade intelectual pouco atraente, para investidores e investigadores nas soluções de IA; ausência de esforço colaborativo; iliteracia digital; rigidez cultural.

Anexo N (Figura 12)

A imagem sistematiza os elementos de registo do utilizador, entidade e projeto na Ferramenta de Avaliação:

- **Utilizador:** habilitações académicas; área profissional/ académica e função;
- **Entidade:** setor; validação do registo associado à entidade e *upload* do documento da entidade para efeitos de credenciação do utilizador;
- **Projeto:** designação; setor de aplicação; contexto em que se desenvolve o projeto de IA; objetivos do projeto; resultados esperados; aplicação; técnica; etapa do projeto; grupo-alvo do projeto; impacto esperado; *framework*, ambiente de desenvolvimento integrado, simulador, linguagem e biblioteca utilizados no projeto e, por fim, tecnologias emergentes aplicadas.

Anexo O (Figura 13)

A imagem esquematiza os níveis de maturidade ou estágios de maturidade existentes nas entidades. Os níveis ou estágios, por ordem crescente são:

- Sem IA;
- Com interesse em IA;
- Principiante em IA;
- Usa IA de forma operacional;
- Usa IA de forma sistemática;
- Usa IA de forma responsável.

Anexo P (Figura 14)

A imagem resume a matriz de recomendações aplicada na Ferramenta de Avaliação. A matriz cruza níveis de maturidade ou estágios de maturidade com recomendações abaixo e acima do limiar.

Recomendações para abaixo do limiar:

- Sem IA: Sugerimos ler sobre o tema;
- Interesse em IA: Sugerimos ler sobre o tema;
- Principiante em IA: Recomendamos ler e planear;

- Usa IA de forma operacional: Recomendamos planear e atuar;
- Usa IA de forma sistemática: É essencial atuar;
- Usa IA de forma responsável: Já deveria ter atuado sobre este aspeto.

Recomendações acima do limiar:

- Sem IA: Sugerimos ler sobre o tema;
- Interesse em IA: Excelente para a fase em que se encontra;
- Principiante em IA: Ótimo para a fase em que se encontra;
- Usa IA de forma operacional: Bom para a fase em que se encontra;
- Usa IA de forma sistemática: Razoável para a fase em que se encontra;
- Usa IA de forma responsável: Siga assim, está no bom caminho.

AGRADECIMENTOS

Das várias pessoas que contribuíram para a elaboração deste documento, queremos destacar, pelos seus contributos e revisões: o Prof. Dr. Fernando Buarque, o Prof. Dr. Luís Janeiro, o Eng.º Mário Nogueira e o Eng.º Raúl Martins.

AGÊNCIA PARA A MODERNIZAÇÃO ADMINISTRATIVA
tic.gov.pt
Cofinanciado por:
COMPETE 2020
PORTUGAL 2020
UNIÃO EUROPEIA
Fundo Social Europeu

Rua de Santa Marta, 55
1150-294 Lisboa | Portugal
guia@ama.pt