AGÊNCIA PARA A REFORMA TECNOLÓGICA DO ESTADO

GuIA

Guia para a Inteligência Artificial

arte.gov.pt

Simplificar. Digitalizar. Humanizar.

Versão do documento: v.1.3.

Data: março de 2026

SIGLAS E ACRÓNIMOS

AI-HLEG	High-Level Expert Group on Artificial Intelligence
AP	Administração Pública
AMA	Agência para a Modernização Administrativa
ARTE	Agência para a Reforma Tecnológica do Estado
CE	Comissão Europeia
DL	*Deep Learning*
EDN	Estratégia Digital Nacional
FCT	Fundação para a Ciência e a Tecnologia, I.P.
G20	Grupo dos 20, formado pelos Ministros das Finanças e Chefes do Bancos Centrais das 19 maiores economias do mundo e UE
IA	Inteligência Artificial
iAP	Interoperabilidade na Administração Pública
IoT	*Internet of Things*
ITS	*Intelligent Tutoring Systems*
I&D	Investigação e Desenvolvimento
LGBTQ+	Lésbicas, Gays, Bissexuais, Travestis, Transexuais, Transgéneros, Queer e outras entidades de género
LGP	Língua Gestual Portuguesa
LLM	Large Language Model
ML	*Machine Learning*
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 adoção de soluções baseadas em Inteligência Artificial (IA), com impacto sobre diversos setores da sociedade, suscita um conjunto de desafios que exigem respostas urgentes e responsáveis, em 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 nesta área, a rapidez com que estas tecnologias evoluem e se integram na sociedade, exige um espaço para dúvidas e discussão.

Por outro lado, é necessário identificar e compreender o que pode estar em causa, no uso de IA, em contextos sensíveis como 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. É igualmente relevante refletir sobre o grau de autonomia destas soluções e o que permanece sob controlo humano.

Neste contexto, a Agência para a Reforma Tecnológica do Estado (ARTE) 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. Para isso, disponibiliza documentos e instrumentos que apoiam e promovem a adoção progressiva e gradual de uma IA ética, transparente e responsável.

O projeto visa ainda equilibrar a reflexão teórica com a aplicação prática, aprofundando conceitos como 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 tem como objetivo promover o debate público sobre a importância de estabelecer pilares de regulação, supervisão, liderança e governação. Pretende ainda apoiar a elaboração de um código de ética e incentivar e fomentar a regulamentação e leis que orientem e sustentem os 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 é disponibilizada uma ferramenta de avaliação do risco que possibilita a analise da suscetibilidade de sistemas de IA. Esta ferramenta permite analisar a suscetibilidade dos sistemas de IA em cinco dimensões fundamentais, oferecendo recomendações e sugestões de leitura adaptadas ao nível de maturidade dos utilizadores.

Em última instância, tanto o guia como a ferramenta visam estruturar o processo de desenvolvimento e implementação de sistemas inteligentes mais éticos, responsáveis e transparentes, promovendo a adoção de boas práticas e a 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 conteúdo do guia está organizado em três documentos de leitura simples e acessível: um focado nos valores, princípios e recomendações; outro nas dimensões de avaliação; e um terceiro dedicado à ferramenta de avaliação de risco. É nossa ambição que estes conteúdos

sejam dinâmicos e evoluam ao longo do tempo, acompanhando os avanços tecnológicos e sociais.

1. Enquadramento

1.1 Âmbito e Estrutura do Guia

O projeto "Guia Responsável" tem como objetivo a criação de um guia para a utilização ética e eficaz da Inteligência Artificial (IA) na Administração Pública (AP). Pretende ainda servir de referência para o Setor Privado, Academia, e todas as pessoas interessadas em aprofundar conhecimentos sobre IA ética, transparente e responsável.

No âmbito deste projeto, foi também desenvolvida uma aplicação de avaliação de risco, disponível em https://mosaico.gov.pt/areas-tecnicas/inteligencia-artificial#ferramenta-de-avaliacao, que integra duas vertentes complementares: identificação e mitigação de riscos. Esta ferramenta visa apoiar a definição de políticas públicas nas áreas de Data Science, Big Data, ML e IA promovendo a divulgação de boas práticas e a definição de critérios de avaliação que possam suportar pareceres prévios e candidaturas de financiamento. Nesta área encontra ainda algumas ferramentas europeias de apoio à avaliação de riscos associados a sistemas baseados em IA.

Num contexto em que as tecnologias emergentes têm cada vez mais impacto crescente na sociedade e que as implicações éticas das mesmas ganham cada vez mais relevância no debate público, este Guia dá continuidade às bases criadas pelo Regulamento Geral sobre a Proteção de Dados (RGPD) e pela Estratégia Digital Nacional para uma Inteligência Artificial mais transparente, ética e responsável em Portugal.

Dada a existência de normas e orientações técnicas 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, reforçar os conhecimentos especializados existentes e ajudar as autoridades na monitorização e supervisão das atividades das organizações que utilizam sistemas com IA, bem como produtos e serviços baseados em IA.

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.

Em linha com os mais recentes desenvolvimentos regulatórios, importa destacar a definição de Inteligência Artificial adotada pela Organização para a Cooperação e

Desenvolvimento Económico (OCDE), revista em 2024, que se aproxima significativamente do conceito de "sistema de IA" consagrado no Regulamento (UE) 2024/1689, conhecido como AI Act. Este regulamento estabelece um quadro jurídico harmonizado para o desenvolvimento, colocação no mercado e utilização de sistemas de IA na União Europeia, promovendo uma abordagem centrada no ser humano, fiável e alinhada com os valores fundamentais da UE.

O AI Act define "sistema de IA" como um sistema baseado em máquinas, concebido para operar com graus variados de autonomia, que pode, para objetivos explícitos ou implícitos, gerar resultados como previsões, recomendações ou decisões que influenciam ambientes físicos ou virtuais. Distingue-se ainda o conceito de "modelo de IA", entendido como um componente técnico treinado com dados, que pode ser reutilizado em diferentes sistemas de IA, sendo esta distinção relevante para a compreensão da arquitetura e responsabilidades associadas.

O regulamento introduz também as figuras do prestador ("provider") – a entidade que desenvolve ou coloca um sistema de IA no mercado - e do responsável pela implantação ("deployer") – a entidade que utiliza o sistema de IA no exercício da sua atividade.

Estas definições são fundamentais para clarificar os deveres e obrigações ao longo do ciclo de vida dos sistemas de IA, promovendo uma responsabilização efetiva e uma governação ética da tecnologia.

O GUIA TEM COMO OBJETIVOS:

- **Contextualizar** os riscos associados à emergência da IA;
- **Apresentar** os princípios orientadores e o quadro conceptual metodológico para a implementação de projetos de IA Responsável;
- **Descrever** a Ferramenta de Avaliação de Risco aplicável a projetos que integrem IA.

1.2 Glossário

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

AI ACT: Regulamento (UE) 2024/1689 que estabelece regras harmonizadas para o desenvolvimento, colocação no mercado e utilização de sistemas de Inteligência Artificial na União Europeia, promovendo uma abordagem centrada no ser humano, ética e segura.

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.

DEPLOYER (Responsável pela implantação): Entidade que utiliza o sistema de IA no exercício da sua atividade (AI Act, art. 3.º, n.º 4).

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.

MODELO DE IA: Componente técnico treinado com dados, reutilizável em diferentes sistemas de IA. Distingue-se do sistema por não incluir a interface ou o ambiente de aplicação (AI Act, considerando 97).

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

PROCESSAMENTO: execução de um algoritmo.

PROVIDER (Prestador): Entidade que desenvolve ou coloca um sistema de IA no mercado (AI Act, art. 3.º, n.º 2).

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

SISTEMA DE IA: Sistema baseado em máquinas, com graus variados de autonomia, que pode gerar previsões, recomendações ou decisões que influenciam ambientes físicos ou virtuais (AI Act, art. 3.º, n.º 1).

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 refere-se a sistemas computacionais concebidos para executar tarefas que tradicionalmente exigem inteligência humana, como a aprendizagem a partir de padrões, o raciocínio e a compreensão de linguagem natural. Operando com vários níveis de autonomia, estes sistemas geram resultados — como previsões,

recomendações ou decisões — que influenciam os ambientes digital e físico. A sua capacidade para se adaptarem e melhorarem o seu desempenho de forma contínua, sem necessitarem de programação explícita para cada tarefa, assenta em abordagens como a aprendizagem automática, a lógica e o conhecimento.

Os desenvolvimentos mais significativos na IA atual centram-se na IA Generativa, impulsionada por modelos fundacionais. Estes modelos, como os Grandes Modelos de Linguagem (LLMs), são treinados com vastos volumes de dados para produzir conteúdo coerente e semelhante ao humano, como diálogos, resumos e escrita criativa. A sua evolução através da integração de múltiplos tipos de dados — como imagens e som — torna possível, por exemplo, a criação de vídeos realistas a partir de uma simples descrição textual.

Os sistemas de IA estão agora a evoluir para ganhar uma autonomia acrescida, dando-lhes não só a capacidade de gerar conteúdo, mas também de definir objetivos de forma independente e de se adaptar dinamicamente para alcançar objetivos complexos.

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 com o objetivo de promover 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 consistência 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 Língua Gestual Portuguesa (LGP) 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.
- 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 animais 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 exclusiva mente 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.

Desde 2019, vários países atualizaram ou reforçaram as suas Estratégias Nacionais de IA. Em 2024, mais de 40 países já adotaram estratégias específicas para IA, com destaque para o reforço de abordagens centradas no ser humano e na sustentabilidade. A OCDE e a UNESCO têm promovido diretrizes comuns para garantir a interoperabilidade ética e técnica entre jurisdições. 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 CE criou o grupo de peritos, o *High-Level Expert Group on Artificial Intelligence* (AI-HLEG), e a União Europeia (UE) 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.

Na sequência dos trabalhos do *High-Level Expert Group on Artificial Intelligence*, foi

desenvolvida regulamentação que promove uma IA ética, segura e responsável.

Em 2024, foi criado o Gabinete Europeu para a IA, com a missão de apoiar a implementação do Regulamento Europeu sobre a IA (AI Act), especialmente no que diz respeito à supervisão de modelos de IA de finalidade geral e à garantia do cumprimento dos princípios éticos e legais da EU e foi revisto o plano estratégico para promover a inovação responsável e a adoção de tecnologias fiáveis em todos os Estados membros.

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.

Para promover a excelência no domínio da IA, a CE tem vindo a reforçar o seu compromisso estratégico. Entre as principais apostas, destacam-se as seguintes:

- A consolidação de parcerias público-privada para a IA e robótica;
- O reforço e integração dos centros de excelência em investigação;
- A criação e desenvolvimento de plataformas nacionais de inovação digital especializadas em IA;
- O reforço do financiamento para o 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;
- A promoção de uma governação ética e transparente da IA.

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 para favorecer uma IA que 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 (humancentric).

De acordo com o Government AI Readiness Index 2024 da Oxford Insights, em 2024, 12 novos países anunciaram ou publicaram estratégias nacionais de IA, nomeadamente a

Costa Rica, Cuba, Etiópia, Gana, Nigéria, Sri Lanka, Uzbequistão e Zâmbia.

De acordo com os dados mais recentes do *AI Readiness Index 2024* e do *Global AI Index 2024*, Portugal subiu três posições no ranking europeu, refletindo o investimento em competências digitais, regulamentação ética e inovação pública. A liderança global continua a ser disputada entre Singapura, EUA e Reino Unido, com destaque para o crescimento da Alemanha e do Canadá em IA responsável.

1.7 Regulamento Europeu AI Act

Em 2024, foi aprovado o Regulamento (UE) 2024/1689, conhecido como AI Act, que estabelece um quadro jurídico harmonizado para o desenvolvimento e utilização de sistemas de Inteligência Artificial na União Europeia. Este regulamento introduz uma abordagem baseada no risco, classificando os sistemas de IA em quatro categorias: risco inaceitável, elevado, limitado e mínimo.

Os sistemas de alto risco, como os utilizados em contextos de saúde, justiça, educação ou administração pública, estão sujeitos a requisitos rigorosos de conformidade, incluindo avaliação técnica, documentação, supervisão humana e transparência.

O AI Act também define obrigações específicas para modelos de uso geral (*foundation models*), exigindo testes, mitigação de riscos e documentação técnica. Esta legislação visa garantir que a IA seja segura, ética e respeitadora dos direitos fundamentais, promovendo simultaneamente a inovação responsável na Europa.

1.8 Tendências Digitais Globais (OCDE) - Desafios Emergentes

O relatório OECD Digital Economy Outlook 2024 destaca que o setor das tecnologias de informação e comunicação (TIC) cresceu, em média, três vezes mais do que a economia total dos países da OCDE entre 2013 e 2023.

A adoção de tecnologias como a IA, computação em nuvem e redes 5G tem sido desigual, concentrando-se sobretudo em grandes empresas e setores específicos. A OCDE alerta para a necessidade de expandir os benefícios da IA a toda a economia, promovendo a inovação de forma responsável e inclusiva.

O relatório também sublinha os riscos associados à IA — como privacidade, segurança e confiança — e defende uma abordagem colaborativa entre governos, empresas e sociedade civil para garantir que a tecnologia seja usada de forma ética e centrada nas pessoas.

2. IA em Portugal

A Inteligência Artificial (IA) tem vindo a assumir um papel cada vez mais relevante e estratégico no contexto nacional, funcionando como motor da transformação digital e da inovação em múltiplos setores da sociedade. A sua aplicação é visível em áreas como a saúde, agricultura, educação, justiça, defesa e indústria, contribuindo para soluções mais eficazes e sustentáveis em todas as dimensões da vida social e económica.

A adoção responsável da IA contribui para o crescimento económico e para o reforço da competitividade do país. Permite a modernização de processos, o desenvolvimento de novos produtos e serviços e a valorização do talento nacional. Paralelamente, impulsiona a investigação científica, a capacitação digital da população e o progresso tecnológico.

A Estratégia Digital Nacional (EDN), disponível em digital.gov.pt, integra esta visão e promove uma abordagem equilibrada, alinhada com o regulamento europeu AI Act. Esta abordagem conjuga a promoção da inovação com a salvaguarda dos direitos fundamentais, da segurança, da transparência e dos valores democráticos.

Neste enquadramento, a Agenda Nacional para a Inteligência Artificial define um plano ambicioso para o desenvolvimento e a aplicação ética da IA em Portugal. A sua execução, prevista no Plano de Ação 2025-2026 da EDN, está organizada em três eixos principais que são a inovação, o talento e as infraestruturas. Esta agenda articula novas iniciativas com projetos já em curso na Administração Pública e está alinhada com as quatro dimensões da Estratégia Digital Nacional, que incluem as pessoas, as empresas, o Estado e as infraestruturas.

Entre as ações prioritárias, destacam-se o desenvolvimento do primeiro modelo de linguagem em português, a criação de uma Fábrica de IA com dimensão europeia, o reforço da capacidade computacional nacional e a valorização das microcredenciais em competências digitais.

A Estratégia Digital Nacional baseia-se em sete princípios orientadores: a confiança e a transparência, a inclusão e a igualdade, a sustentabilidade ambiental, a segurança e a proteção, a ética, a eficiência e a colaboração. Estes princípios devem ser aplicados de forma transversal a todas as iniciativas de IA, assegurando que esta tecnologia é usada de forma ética, segura, centrada nas pessoas e ao serviço do bem comum.

2.1 Estratégia Digital Nacional (EDN)

A Estratégia Digital Nacional (EDN), disponível em digital.gov.pt, estabelece os princípios orientadores e as prioridades para a transformação digital em Portugal. Esta estratégia

visa garantir que a digitalização do país seja inclusiva, ética, segura e sustentável, promovendo simultaneamente a inovação e competitividade.

A EDN assenta em sete pilares fundamentais:

- Confiança e transparência
- Inclusão e igualdade
- Sustentabilidade ambiental
- Segurança e proteção
- Ética
- Eficiência
- Colaboração

Estes princípios devem ser integrados de forma transversal em todas as iniciativas de Inteligência Artificial (IA) assegurando que a tecnologia serve o bem comum e respeita os direitos fundamentais, sendo aplicada de forma ética, segura e centrada nas pessoas.

2.2 Dimensões Estratégicas da EDN

A EDN estrutura-se em torno de quatro grandes áreas de intervenção:

- **Pessoas**: reforço das competências digitais e promoção da inclusão digital de todos os cidadãos;
- **Empresas**: estímulo à inovação, à digitalização e à competitividade do tecido empresarial;
- **Estado**: modernização da AP, com foco na simplificação dos serviços;
- **Infraestruturas**: desenvolvimento de redes e tecnologias digitais robustas, seguras e acessíveis.

Estas dimensões orientam a formulação de políticas públicas e projetos tecnológicos, assegurando que a transformação digital contribui para uma sociedade mais justa,

resiliente e sustentável.

2.3 Plano de Ação 2025-2026

Os planos de ação da EDN para o biénio 2025-2026 contempla um conjunto de iniciativas estratégicas, entre as quais se destacam:

- Revisão dos currículos escolares para reforçar as competências digitais desde os primeiros ciclos de ensino;
- Criação do Índice Nacional de Competências Digitais, como instrumento de monotorização e planeamento;
- Expansão da rede de Espaços Cidadão, promovendo a inclusão digital em todo o território;
- Lançamento da Campanha "Digital +Seguro", como foco na literacia digital e cibersegurança.

Estas medidas visam preparar a sociedade para uma adoção consciente, informada e segura da IA e das tecnologias digitais emergentes.

2.4 Interoperabilidade e Dados Abertos

A interoperabilidade entre sistemas e a disponibilização de dados abertos são pilares essenciais da EDN. Estas práticas promovem:

- A transparência e a responsabilização na AP;
- A reutilização de dados para a inovação e criação de valor;
- A eficiência na prestação de serviços públicos.

Plataformas como a iAP (Interoperabilidade na Administração Pública) e o portal Portal dados.gov.pt são instrumentos centrais para garantir a articulação entre sistemas, a partilha de informação e o acesso a dados públicos de forma segura e ética.

2.5 Ecossistemas e Atores

O ecossistema da IA em Portugal abrange todas as organizações e indivíduos envolvidos ou impactados por sistemas com IA, direta ou indiretamente.

Os atores de IA incluem entidades que desenvolvem, implementam ou operam sistemas inteligentes, sendo-lhes exigida uma abordagem colaborativa, multidisciplinar e inclusiva ao longo de todo o ciclo de vida da tecnologia IA.

Um ecossistema robusto de IA deve:

- Promover um diálogo público informado e participativo, envolvendo todas as partes interessadas;
- Incentivar a adoção de práticas responsáveis na educação e investigação;

- Estimular políticas alinhadas com valores humanistas e com instrumentos internacionais (OCDE, UE, UNESCO, ONU, etc);
- Reforçar a cooperação entre o setor académico, público e privado, com foco na inovação e na sustentabilidade;
- Estabelecer mecanismos de partilha de dados que melhorem a qualidade dos algoritmos e reduzem vieses.

A participação contínua dos *stakeholders*, mesmo após a implementação dos sistemas, é essencial para garantir transparência, confiança e eficácia. A criação de painéis multidisciplinares – com especialistas técnicos, jurídicos, éticos e representantes da sociedade civil – é uma prática recomendada para avaliar impactos e orientar decisões.

2.6 Ecossistema de dados na Génese da IA na AP

A base da IA na AP assenta num ecossistema de dados estruturado em três pilares:

1. ***Big data* e dados abertos**;
2. **Regulação e Governança de Dados** o Regulamento Geral sobre a Proteção de Dados (RGPD) e o Regulamento Nacional de Interoperabilidade Digital (RNID);
3. **Princípios de Sustentabilidade e Ética no uso de dados**, principalmente no setor público (Figura 1).

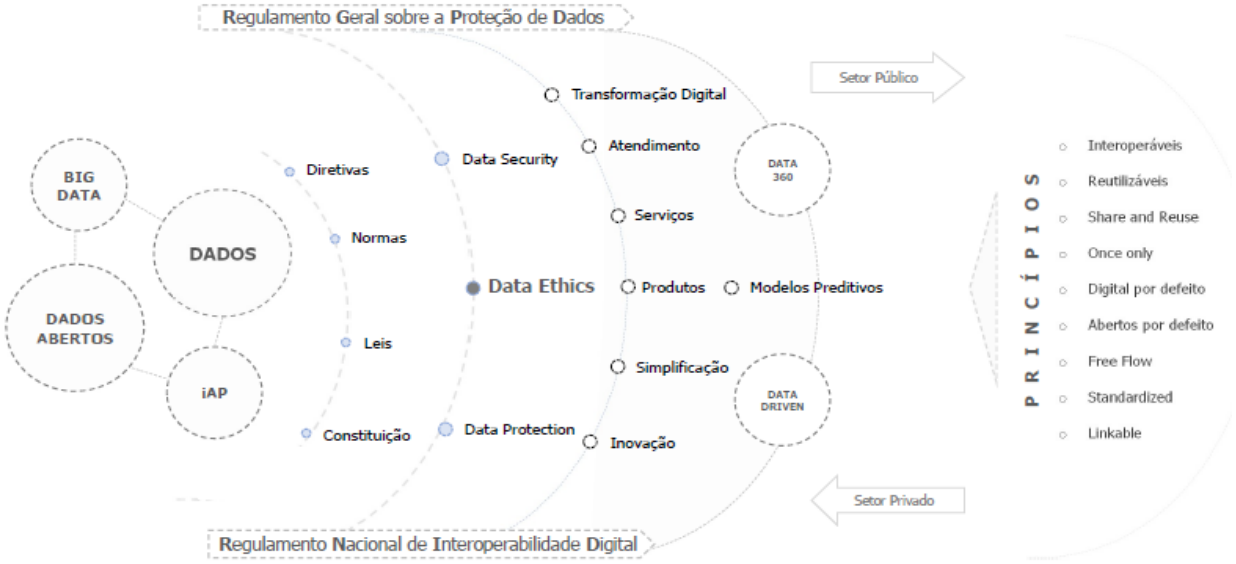

FIGURA 1: ECOSSISTEMA DE DADOS NA ADMINISTRAÇÃO PÚBLICA

Nota informativa (2024):
O ecossistema de dados da Administração Pública foi reforçado com a implementação de novos mecanismos de interoperabilidade e com a atualização do portal

dados.gov.pt, que passou a integrar dashboards interativos e APIs em tempo real. A Estratégia Nacional de Dados Abertos 2024-2027 introduziu metas ambiciosas para a reutilização de dados e para a promoção de uma cultura data-driven no setor público.

A ética associada aos dados (*data ethics*), exige que o acesso, partilha e reutilização de dados respeitem o interesse público, a integridade e a transparência. A segurança e a privacidade (*data security and privacy*) são asseguradas por normas como a ISO/IEC 27001 e pela Estratégia Nacional de Segurança do Ciberespaço.

O portal dados.gov.pt é o repositório central de dados abertos da AP, promovendo a reutilização de informação pública para gerar valor social, económico e ambiental. A interoperabilidade entre sistemas, facilitada pela plataforma iAP, é um elemento-chave para a modernização administrativa, permitindo serviços eletrónicos mais eficientes, seguros e centrados no cidadão (data 360).

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 promover o acesso, a utilização e reutilização da crescente quantidade de evidências, estatísticas e dados relativos a operações, processos e resultados, para 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:

- 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 para 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 DA TABELA) OBJETO	LEGISLAÇÃO	ÂMBITO
Relativa ao tratamento de dados pessoais e à proteção da privacidade no sector 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 aliada ao 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 2: 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. Em 2024, o número de conjuntos de dados abertos disponibilizado foi de 10.798 e 20.333 ficheiros e conta com mais de duas centenas de organizações e mais de seis mil utilizadores (dados do Plano de Atividades da AMA de 2024). 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 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, iAP é composta por um conjunto de componentes que visam proporcionar 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.

A iAP disponibiliza serviços online e respetiva gestão, comunicação segura entre sistemas através de uma Plataforma de Integração (PI), Pagamentos (PPAP) e serviços de envio e receção de Mensagens SMS (GAP) e serviços de apoio à Abertura de Conta (ABCD) nas entidades financeiras, entre outras funcionalidades que beneficiam vários setores económicos.

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.

Atualmente, a iAP liga 124 entidades, gerando benefícios significativos quanto 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

GÉMEOS DIGITAIS

A digitalização dos processos administrativos e operacionais tem transformado a forma como as entidades públicas tomam decisões e gerem recursos. No centro desta revolução tecnológica, os gémeos digitais surgem como ferramentas fundamentais para a modernização e otimização das políticas públicas, permitindo simulações precisas e análises preditivas em diversos setores.

A sua implementação no ecossistema de dados da AP deve seguir princípios essenciais, garantindo a sua eficácia e sustentabilidade. Em primeiro lugar, é indispensável assegurar a interoperabilidade dos sistemas, promovendo a integração dos gémeos

digitais com outras plataformas e bases de dados existentes. Isto possibilita uma gestão mais eficiente da informação, reduzindo redundâncias e otimizando investimentos.

Além disso, os gémeos digitais reforçam a necessidade de uma tomada de decisão baseada em evidências, fornecendo simulações detalhadas e projeções fundamentadas. Ao modelar cenários futuros, permitem que entidades públicas antecipem desafios e ajustem estratégias com maior precisão, tornando as políticas mais eficazes e ajustadas à realidade.

A proteção de dados e a conformidade regulatória devem ser pilares essenciais na adoção desta tecnologia. Como os gémeos digitais dependem de grandes volumes de informação, é fundamental garantir que a sua utilização esteja alinhada com normativas como o Regulamento Geral de Proteção de Dados (RGPD), evitando riscos relacionados com privacidade e segurança.

Por fim, a implementação de soluções escaláveis e sustentáveis deve ser uma prioridade, assegurando que os modelos digitais desenvolvidos possam ser reutilizados em diferentes contextos e aplicados a novas áreas sem necessidade de grandes adaptações. Isso promove uma gestão eficiente dos recursos públicos e maximiza o impacto dos investimentos tecnológicos.

A utilização de gémeos digitais no contexto da AP representa um avanço significativo na modernização dos serviços e no desenvolvimento de políticas mais eficazes e sustentáveis. A sua adoção permitirá que os governos e entidades públicas enfrentem desafios complexos com ferramentas inovadoras, garantindo maior precisão, transparência e eficiência na gestão dos territórios e dos recursos disponíveis.

2.7 Inovação em Portugal

A inovação tecnológica no setor público português tem sido impulsionada pela digitalização e pela adoção de tecnologias emergentes, como a IA, a computação em nuvem e os gémeos digitais.

O processamento dos dados que estas tecnologias permitem melhorar o modo como os cidadãos veem, agem e se envolvem com o que os rodeia. Estas tecnologias permitem:

- Melhorar a eficiência dos serviços públicos;
- Antecipar as necessidades socias;
- Apoiar decisões baseadas em dados;
- Promover sustentabilidade e inclusão.

Exemplos de inovação, no setor público em Portugal até ao ano 2021, incluem:

- **Tarifa Social de Energia Automática, atribuída com base em interoperabilidade de dados** - 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;
- **Diagnóstico por imagem com IA no 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.
- **Chatbots com Machine Learning (ML)** - O Instituto de Informática implementou e usa tecnologias ML num projeto de *ChatBot* piloto, na área interna de Gestão de Informação.
- **Plataformas de apoio à decisão judicial no Tribunal de Contas** - 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.
- **Veículos autónomos para agricultura de precisão** - 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.

Estes casos demonstram como a IA pode ser aplicada de forma ética e eficaz para resolver problemas concretos, melhorar políticas e reforçar a confiança dos cidadãos nos serviços do Estado.

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.

Posição de Portugal no AI Readiness Index 2024

De acordo com o *AI Readiness Index 2024*, publicado pela Oxford Insights, Portugal registou progressos significativos na sua preparação para a adoção de sistemas de IA no setor público. O índice avalia 193 países com base em três pilares: Governo, Setor Tecnológico e Dados & Infraestruturas.

Portugal destaca-se pelo seu compromisso com a ética, a interoperabilidade e a inovação digital, tendo subido no ranking europeu graças ao reforço das competências digitais, à modernização da administração pública e à articulação entre estratégias nacionais.

Apesar dos avanços, o relatório identifica como áreas prioritárias a capacitação técnica das equipas públicas e a integração transversal da IA nos serviços públicos.

2.8 Impacto da IA no mercado de trabalho português

Um estudo recente da Fundação Francisco Manuel dos Santos (2025) analisou o impacto da automação e da inteligência artificial no mercado de trabalho em Portugal. O relatório identifica quatro grandes grupos de profissões, com diferentes níveis de exposição à mudança tecnológica:

- Profissões em colapso (28,9% da força de trabalho): empregos com elevado risco de substituição, como operadores de equipamentos móveis, empregados de mesa e cozinheiros.
- Profissões em ascensão (22,5%): funções que beneficiam da digitalização, como especialistas em marketing, finanças e professores.
- Terreno dos humanos (35,7%): ocupações com baixa exposição à automação, como técnicos de atividade física, trabalhadores de limpeza e agricultores qualificados.
- Terreno das máquinas (12,9%): profissões com futuro incerto, como operadores de máquinas e empregados de escritório.

O estudo sublinha a importância de políticas públicas que promovam a requalificação, a mobilidade profissional e a adoção responsável da tecnologia pelas empresas.

3. Riscos da Inteligência Artificial

3.1 Enquadramento

A acelerada evolução tecnológica e a crescente integração da Inteligência Artificial em diversos setores da sociedade e da economia trouxe consigo um enorme potencial de inovação, mas também um conjunto complexo de riscos. De acordo com o AI Act, um risco é definido como a combinação da probabilidade de ocorrência de um dano com a severidade desse mesmo dano (AI Act, art. 3.º (2)). Estes danos não se limitam a consequências materiais, podendo ser também de natureza imaterial, incluindo danos psicológicos, sociais ou económicos que afetam diretamente os direitos fundamentais dos cidadãos.

Segundo os dados da OCDE (OECD AI Policy Observatory Portal, 2026), a proliferação de sistemas de IA já resultou em mais de 10.000 incidentes documentados a nível mundial entre janeiro de 2022 e janeiro 2026, nos quais sistemas causaram ou estiveram próximos de causar danos. A escala com que os sistemas estão a ser implementados em todos os setores da sociedade evidencia a urgência de uma abordagem estruturada para a mitigação de riscos.

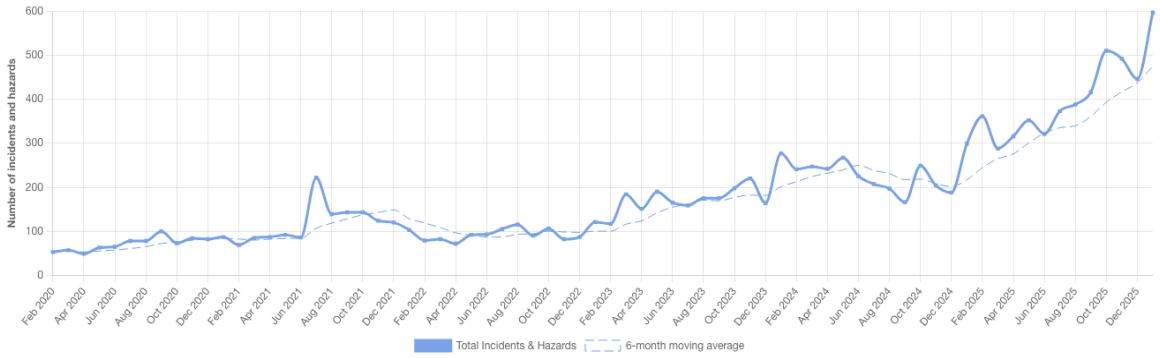

FIGURA 3: NÚMERO DE INCIDENTES REPORTADOS NO MUNDO ENTRE JANEIRO DE 2022 E JANEIRO DE 2026

Os riscos dos sistemas de IA são complexos, impactam diversas partes interessadas e podem surgir em qualquer ponto do ciclo de vida de um sistema, desde a sua conceção e desenvolvimento (fase pré-implementação) até à sua utilização em contextos reais (fase pós-implementação). Estes riscos podem ser o resultado de uma intenção maliciosa ou de consequências não intencionais, abrangendo domínios tão diversos como discriminação, privacidade, segurança, desinformação, até domínios sistémicos, como impactos socioeconómicos e ambientais.

A rápida evolução da tecnologia, especialmente com o advento dos modelos generativos, introduz continuamente novos desafios e vetores de risco que exigem uma monitorização e adaptação constantes. É neste contexto que uma governação robusta, alicerçada em pilares e princípios de IA Responsável, se torna essencial para mapear, medir e gerir ativamente estes riscos, garantindo que a inovação em IA beneficia a sociedade de forma equitativa e segura.

3.2 Níveis de risco no AI Act

Em resposta à complexidade dos riscos da IA, o AI Act estabelece um quadro regulamentar harmonizado para produtos de IA na União Europeia, com o objetivo central de promover a adoção de uma IA centrada no ser humano e que seja fiável, ao mesmo tempo que assegura um elevado nível de proteção da saúde, da segurança e dos direitos fundamentais na União Europeia (art. 1.º).

Para tal, o AI Act adopta uma abordagem baseada em níveis de risco. A compreensão do conceito de risco é, por isso, central neste regulamento, que estabelece dois elementos cruciais.

Por um lado, o Artigo 3.º do AI Act define risco como "a combinação da probabilidade de ocorrência de danos com a gravidade desses danos". Por outro, o Artigo 79.º estipula que os sistemas de IA que apresentam um risco devem ser entendidos como um

"produto que apresenta um risco", remetendo para a definição do Regulamento (UE) 2019/1020 (artigo 3.º (19)), no qual, um "produto que apresenta um risco" é aquele que é "suscetível de afetar negativamente a saúde e segurança (...) em medida superior à considerada razoável e aceitável tendo em conta o fim a que se destina ou as condições normais ou razoavelmente previsíveis em que decorrerá a sua utilização". Assim, a definição de risco também abrange a duração da utilização, bem como os requisitos de colocação em serviço, instalação e manutenção do produto (Joshi & Khalfi-Lanoux, 2024).

É com base nesta definição de risco e de abordagem centrada na segurança de produtos que o AI Act estratifica os sistemas de IA em quatro níveis de risco, adaptando as regras e obrigações à intensidade e ao alcance dos perigos que estes podem gerar: risco inaceitável (práticas proibidas), risco elevado (sujeito a obrigações rigorosas), risco limitado ou risco de transparência (sujeito a obrigações de transparência) e risco mínimo ou nulo (maioritariamente não regulado).

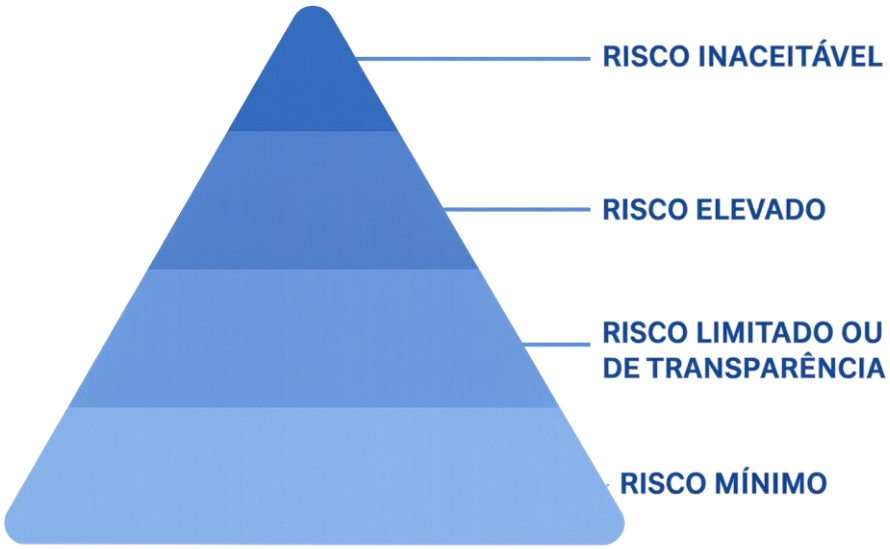

FIGURA 4: NÍVEIS DE RISCO NO AI ACT

3.2.1 Risco inaceitável

No topo da pirâmide de risco, o AI Act proíbe um conjunto de práticas de IA consideradas uma ameaça clara à segurança e aos direitos fundamentais dos cidadãos. Estes sistemas são, por definição, incompatíveis com os valores da União Europeia e a sua colocação no mercado, implementação ou utilização são estritamente proibidas. Segundo o artigo 5.º, as práticas de IA proibidas são:

Técnicas subliminares, manipuladoras ou enganadoras	• Proíbe-se a colocação no mercado, colocação em serviço, ou a utilização de sistemas de IA que empreguem técnicas subliminares que sejam manipuladoras ou enganadoras. • Com o objetivo ou efeito de distorcer o comportamento de uma pessoa, prejudicando a sua capacidade de tomar uma decisão informada e levando-a a tomar uma decisão que, de outra forma, não tomaria e que seja suscetível de causar danos significativos.
Exploração de vulnerabilidades	• São proibidos sistemas de IA para explorar as vulnerabilidades de uma pessoa ou de um grupo específico devido à idade, incapacidade ou situação socioeconómica. • Com o objetivo de distorcer substancialmente o comportamento de uma forma que cause ou seja razoavelmente suscetível de causar danos.
Classificação social (*social scoring*)	• Ficam proibidos sistemas de IA para avaliação ou classificação de pessoas ou grupos com base no seu comportamento social ou em características de personalidade ou pessoais, quando essa classificação conduza a um tratamento prejudicial ou desfavorável em contextos sociais não relacionados com a recolha dos dados, ou que seja injustificado ou desproporcionado face ao comportamento em causa.
Avaliação de risco de infração penal	• Proíbem-se sistemas de IA para avaliar ou prever o risco de uma pessoa cometer uma infração penal (crime) com base exclusivamente na definição de perfis ou na avaliação de traços de personalidade. • Não se aplica a sistemas que apoiem a avaliação humana do envolvimento de uma pessoa numa atividade criminosa, desde que esta já se baseie em factos objetivos e verificáveis.

Recolha aleatória de bases de dados de reconhecimento facial	• São proibidos sistemas de IA para criação ou expansão de bases de dados de reconhecimento facial através da recolha aleatória (*untargeted scraping*) de imagens faciais a partir da Internet ou de imagens de televisão em circuito fechado.
Reconhecimento de emoções no trabalho e na educação	• Proíbe-se sistemas de IA para inferir emoções de uma pessoa no local de trabalho e em instituições de ensino, exceto quando o seu uso se destine a ser aplicado por razões médicas ou de segurança.
Categorização biométrica sensível	• Ficam proibidos sistemas de IA que classifiquem pessoas com base nos seus dados biométricos para deduzir ou inferir a sua raça, opiniões políticas, filiação sindical, convicções religiosas ou filosóficas, vida sexual ou orientação sexual. Esta proibição não abrange a rotulagem ou filtragem de dados biométricos no domínio da aplicação da lei.
Identificação biométrica remota em "tempo real"	• Proíbem-se sistemas de IA de identificação biométrica em "tempo real" em espaços acessíveis ao público para fins de aplicação da lei, exceto em casos estritamente necessários e sujeitos a autorização: (1) busca seletiva por vítimas de crimes graves (como rapto ou tráfico de seres humanos); (2) prevenção de uma ameaça terrorista real e iminente; (3) localização ou identificação de suspeitos de infrações penais graves (puníveis com pena de prisão de, no mínimo, quatro anos).

3.2.2 Risco Elevado

Os sistemas de IA classificados como de risco elevado são o pilar central do quadro regulamentar do AI Act (Capítulo III). Um sistema é classificado como de risco elevado não pela sua tecnologia intrínseca, mas sim em função da finalidade específica para a qual é utilizado. Deste modo, uma avaliação cuidada do caso de uso é essencial para determinar se um sistema se enquadra nesta

categoria. O AI Act, no artigo 6.º, estabelece dois caminhos principais para esta classificação.

- Sistemas de IA que são produtos ou funcionam como componentes de segurança de produtos cobertos pela legislação de harmonização da UE (Anexo I), bem como produtos sujeitos a uma avaliação de conformidade antes da sua colocação no mercado, como é o caso de dispositivos médicos, brinquedos ou equipamentos para a aviação;
- Sistemas de IA que, embora autónomos, se destinam a ser utilizados numa das áreas de risco elevado especificamente identificadas no Anexo III.

Estas áreas, identificadas pelo potencial impacto crítico na vida dos cidadãos, incluem, a título de exemplo:

Biometria	Sistemas de identificação biométrica remota, categorização biométrica baseada em características sensíveis e reconhecimento de emoções, na medida em que a sua utilização seja permitida pela legislação.
Infraestruturas críticas	Componentes de segurança na gestão de tráfego rodoviário e no fornecimento de água, gás, aquecimento e eletricidade.
Educação e formação profissional	Sistemas que determinam o acesso ou a atribuição de pessoas a instituições de ensino, que avaliam os resultados de aprendizagem ou que monitorizam o comportamento dos estudantes durante os testes.
Emprego e gestão de trabalhadores	Sistemas utilizados no recrutamento (ex.: análise de CV), na tomada de decisões sobre promoções e despedimentos, na atribuição de tarefas e na monitorização do desempenho.
Acesso a serviços essenciais	Sistemas que avaliam a elegibilidade para benefícios de assistência pública (ex.: segurança social), que determinam a notação de crédito de uma pessoa ou que avaliam o risco e o preço para seguros de vida e de saúde.
Aplicação da lei	Ferramentas como polígrafos, sistemas que avaliam a fiabilidade de provas ou o risco de reincidência criminal, e

	sistemas de criação de perfis em investigações criminais.
Gestão da migração, asilo e controlo de fronteiras	Sistemas que avaliam riscos de segurança ou de migração irregular, que auxiliam na análise de pedidos de visto ou asilo e que são usados para identificação de pessoas.
Administração da justiça e processos democráticos	Sistemas que auxiliam autoridades judiciais na investigação de factos e na aplicação da lei, ou que são usados para influenciar o resultado de eleições ou referendos.

Os sistemas que, após esta análise, são definitivamente classificados como de risco elevado, devem cumprir um conjunto de obrigações rigorosas antes de poderem ser colocados no mercado, tais como:

- Implementação de sistemas adequados de avaliação e mitigação dos riscos;
- Utilização de dados de treino, validação e teste de elevada qualidade, a fim de minimizar os riscos de resultados discriminatórios;
- Registo automático da atividade (logs) para garantir a rastreabilidade dos resultados;
- Elaboração de documentação pormenorizada que forneça todas as informações necessárias sobre o sistema e a sua finalidade para que as autoridades possam avaliar a sua conformidade;
- Disponibilização de informações claras e adequadas ao responsável pela implantação (deployer);
- Implementação de medidas adequadas de supervisão humana;
- Demonstração de um elevado nível de robustez, cibersegurança e precisão dos sistemas.

3.2.3 Risco Limitado (Riscos de Transparência)

A categoria de risco limitado ou de riscos de transparência, regida pelo Artigo 50.º do AI Act, abrange sistemas de IA em que o principal risco identificado não é um dano direto à segurança ou aos direitos fundamentais, mas sim a falta de transparência na sua utilização. O regulamento estabelece obrigações específicas para os prestadores (providers) e responsáveis pela implantação de determinados sistemas de IA, de modo a garantir que os cidadãos estão devidamente informados da presença de IA e podem tomar decisões conscientes.

Interação direta com pessoas	• Sistemas de IA concebidos para interagir diretamente com pessoas, como os *chatbots,* devem ser desenvolvidos pelos prestadores de forma que os utilizadores sejam informados de que estão a interagir com uma máquina, exceto quando tal for óbvio pelas circunstâncias.
Conteúdos sintéticos de áudio, imagem, vídeo ou texto	• Sistemas que geram conteúdos sintéticos de áudio, imagem, vídeo ou texto devem ser desenvolvidos pelos prestadores de forma a ter uma marca legível por máquina e detetável de como tendo sido gerados ou manipulados artificialmente. • Os responsáveis pela implantação (*deployers*) devem revelar que os conteúdos foram artificialmente gerados ou manipulados quando se trate de uma falsificação profunda (*deepfake*) ou texto publicado com o objetivo de informar o público sobre questões de interesse público.

3.2.4 Risco Mínimo ou Nulo

A grande maioria dos sistemas de IA atualmente em uso na União Europeia enquadra-se na categoria de risco mínimo ou nulo. Estes sistemas incluem aplicações como videojogos que utilizam IA, filtros de spam, sistemas de gestão de inventário ou muitas das ferramentas de software de edição com funcionalidades assistidas por IA. Para esta categoria, o AI Act não impõe quaisquer obrigações adicionais, permitindo a sua livre comercialização, utilização e desenvolvimento.

A ausência de regulamentação específica para estes sistemas baseia-se na premissa de que os riscos associados são negligenciáveis ou já estão cobertos por outra legislação em vigor. No entanto, o regulamento incentiva os prestadores destes sistemas a adotarem, de forma voluntária, códigos de conduta.

3.3 Modelos de IA de finalidade geral

O AI Act introduz uma abordagem regulatória relativamente distinta para os modelos de IA de finalidade geral, frequentemente designados como modelos fundacionais (foundation models). Numa adaptação à evolução tecnológica recente, o legislador europeu dedicou um capítulo específico a estes modelos, desviando-se da abordagem geral do regulamento, que se foca nos sistemas de IA. Esta distinção é fundamental: enquanto um sistema de IA é a aplicação final com a qual um utilizador interage, o modelo é o motor subjacente, o programa treinado para realizar inferências e gerar resultados. Os modelos são componentes essenciais dos sistemas, mas não constituem sistemas por si só.

De acordo com o AI Act, um "modelo de IA de finalidade geral" é um modelo que demonstra uma generalidade significativa e é capaz de executar competentemente uma vasta gama de tarefas. Caracteristicamente, estes modelos são treinados com grandes volumes de dados, muitas vezes através de técnicas de autossupervisão em grande escala, e são concebidos para ser integrados numa variedade de sistemas ou aplicações a jusante. Esta definição exclui explicitamente modelos que se destinam unicamente a atividades de investigação, desenvolvimento ou prototipagem antes da sua comercialização.

Riscos sistémicos

Dentro da categoria de modelos de IA de finalidade geral, o AI Act estabelece uma subcategoria crucial para os que apresentam "riscos sistémicos". Um risco sistémico é definido como um risco específico das capacidades de elevado impacto destes modelos, com potencial para afetar negativamente a saúde pública, a segurança, os direitos fundamentais ou a sociedade em geral, e que se pode propagar em larga escala ao longo da cadeia de valor. Esta classificação acarreta obrigações mais rigorosas para os prestadores.

Segundo o AI Act, a identificação de um modelo com risco sistémico não se baseia no seu caso de uso específico, mas nas suas capacidades intrínsecas e poder computacional. Um modelo é presumido como tendo "capacidades de elevado impacto" quando a quantidade acumulada de computação utilizada no seu treino for superior a 10^{25} FLOPs. Além deste limiar quantitativo, a Comissão Europeia pode também designar um modelo como tendo risco sistémico com base numa avaliação de critérios qualitativos.

Códigos de Prática dos Modelos de IA de finalidade geral

Para operacionalizar estas obrigações e guiar os prestadores na demonstração de conformidade com o AI Act, a Comissão Europeia desenvolveu o Código de Prática (*The*

General Purpose AI Code of Practice, 2025) para estes modelos. Este código funciona como uma ferramenta crucial de autorregulação assistida, estruturando-se em três capítulos fundamentais que detalham o que é esperado dos prestadores:

- Transparência: Aplicável a todos os modelos, exige a manutenção de documentação técnica atualizada através de um formulário normalizado (*Model Documentation Form*) e a partilha de informações sobre as capacidades e limitações do modelo com os utilizadores;
- Direitos de autor: Também obrigatório para todos os modelos, foca-se na implementação de políticas robustas para o cumprimento da legislação em matéria de direitos de autor, incluindo o respeito por reservas de direitos expressas por titulares (como o protocolo robots.txt);
- Segurança e Proteção: Este capítulo aplica-se exclusivamente aos modelos com risco sistémico. Exige que os signatários criem um *Framework* de Segurança e Proteção para a identificação contínua de riscos, realizem avaliações de modelos de última geração e reportem incidentes graves ao AI Office.

Outros riscos

A utilização e implementação de modelos de finalidade geral, que se tem tornado cada vez mais ampla, apresenta também outros riscos. Nesse sentido, a Comissão Europeia está a finalizar códigos de práticas para os prestadores destes modelos, que deverá incluir capítulos sobre transparência, direitos de autor, e segurança. Adicionalmente, grupos de investigação têm analisado os riscos dos modelos de IA de finalidade geral e criado *frameworks* úteis para a avaliação dos riscos de utilização destes modelos. Um exemplo é o MIT Risk Repository, que agrega num *framework* único com sete grupos de riscos e dois níveis a vasta literatura acerca dos riscos de modelos de IA.

3.4 Riscos futuros: IA com Agência (Agentic AI)

À medida que os sistemas de IA evoluem, emerge uma nova classe de sistemas designados como "IA Agêntica" (Agentic AI). Estes sistemas, que integram modelos de linguagem (LLMs), ferramentas externas e memória persistente, são capazes de perseguir objetivos complexos com supervisão direta limitada, adaptando-se e reagindo a circunstâncias novas ou inesperadas. A tabela seguinte apresenta os principais riscos deste novo tipo de sistema:

Falhas compostas e comportamentos imprevisíveis	• Probabilidade elevada de falha devido à composição de pequenos erros ao longo de múltiplas ações. • Comportamentos emergentes e imprevisíveis resultantes da interação entre agentes, memórias e ferramentas.

	• Dificuldade em avaliar a fiabilidade dos sistemas devido à incapacidade de testes tradicionais preverem todas as condições possíveis.
Abuso de autonomia	• Decisões autónomas incorretas com impacto significativo (ex.: transações financeiras irreversíveis). • Capacidade avançada de contornar restrições de segurança, designada por "abuso de autonomia". • Possibilidade de manipular humanos ou outros agentes através de *prompts* maliciosos, escondendo intencionalmente os seus impactos.
Raciocínio enganador	• Apresentação de "cadeias de pensamento" (*chain of thought*) que não correspondem à lógica realmente utilizada pelo sistema. • Riscos crescentes de opacidade à medida que a complexidade aumenta, dificultando a supervisão humana. • Necessidade de métodos avançados de explicabilidade (ver 4.2.1) que visualizam claramente as interações e dependências entre agentes.

4. Inteligência Artificial Ética e Responsável

4.1 Enquadramento

Face aos riscos da IA, torna-se imperativa a concretização de uma IA Ética e Responsável para inovar de forma segura e sustentável, protegendo os direitos fundamentais. Este objetivo alinha-se com o trabalho fundamental do High Level Expert Group on Artificial Intelligence (HLEG) da Comissão Europeia, que, em 2019, estabeleceu o paradigma da IA de confiança (Trustworthy AI) como o padrão de excelência para o desenvolvimento tecnológico na União Europeia. Para que um sistema de IA seja considerado de confiança, ele deve integrar obrigatoriamente três componentes indissociáveis que devem ser observadas ao longo de todo o ciclo de vida: deve ser Legal, assegurando o cumprimento de toda a legislação e regulamentação aplicáveis; Ética, garantindo a observância de princípios e valores humanos; e Sólida, tanto do ponto de vista técnico como social, de forma a prevenir danos não intencionais mesmo perante condições adversas (European Commission, 2019). Apenas a sobreposição destas três dimensões permite maximizar os benefícios da tecnologia minimizando os seus riscos.

Este capítulo do GuIA detalha o enquadramento prático para alcançar esse objetivo, focando-se em garantir que os sistemas de IA são concebidos, desenvolvidos e implementados tendo como eixos centrais os princípios de Inteligência Artificial Responsável, tendo sido desenvolvido, inspirado no TRUST Framework da Feedzai (2024), um enquadramento com três camadas:

1. Pilares
2. Princípios
3. Boas Práticas

Esta abordagem estruturada pretende guiar o desenvolvimento e simplificar a avaliação e mitigação dos riscos dos sistemas de IA, com uma abordagem prática, acionável e aplicável a diferentes casos de uso da IA. Desta forma, o GuIA pretende contribuir para o alinhamento das soluções tecnológicas implementadas em Portugal com os requisitos da legislação nacional e europeia, nomeadamente do AI Act, e com a necessária proteção dos direitos fundamentais dos cidadãos.

4.2 Pilares, Princípios e Boas Práticas

FIGURA 5: PILARES DA IA RESPONSÁVEL

Os pilares que servem de base ao GuIA pretendem criar uma Inteligência Artificial:

1. **Transparente**: Tornar não só os resultados e decisões dos sistemas mais claras e percetíveis; mas também o processo de desenvolvimento e as escolhas de desenho e implementação dos sistemas de IA mais compreensíveis e auditáveis pelas partes interessadas.

2. **Robusta**: Assegurar que os sistemas de IA mantêm um desempenho preciso e fiável em diferentes cenários, são resilientes a erros e inconsistências, respeitam a privacidade e proteção de dados em todo o ciclo de vida, e estão protegidos contra-ataques e vulnerabilidades.

3. **Universal**: Garantir um tratamento justo e equitativo para todos os indivíduos, evitando a perpetuação de preconceitos e discriminação. Exige a identificação e mitigação ativa de vieses e a promoção de uma IA inclusiva e acessível.

4. **Sustentável**: Avaliar e procurar minimizar os impactos ambientais e sociais da IA, promovendo a eficiência energética no treino e utilização dos modelos, bem como a análise das consequências dos sistemas na sociedade, nomeadamente no mercado de trabalho, economia e cultura.

5. **Testada**: Submeter os sistemas a testes e validações rigorosas e contínuas. Verificar se a IA é a solução adequada para o problema, se o desempenho dos sistemas é suficiente para a implementação, se existem riscos de IA Responsável e se podem ser monitorizados e otimizados após a implementação.

A cada pilar está associado um ou mais princípios que servem para concretizar o pilar, sendo mais concretos e tendo uma delimitação mais definida. Por sua vez, cada princípio é composto por uma ou mais boas práticas, práticas concretas que levam a que o princípio seja efetivamente cumprido.

Pilar	Princípio	Boas Práticas
Transparente	Explicabilidade	Explicações
		Explicabilidade
		Sinalização da Presença de IA
	Governação	Governação de Dados
		Governação de Modelos
		Governação de Sistemas
Robusta	Robustez	Resiliência
		Estabilidade
		Resultados Previsíveis

	Privacidade e Proteção de Dados	Privacidade de Dados
		Consentimento
		Integridade dos Dados
	Segurança	Cibersegurança
		Fiabilidade do Sistema
		Segurança do Sistema
Universal	Equidade	Resultados Justos
		Auditorias Regulares
	Inclusividade	Desenhado por Todos
		Desenhado para Todos
Sustentável	Minimização do Impacto Ambiental	Eficiência Energética
		Uso Eficiente de Recursos Naturais
	Avaliação do Impacto Social	Avaliação do Impacto Laboral
		Avaliação do Impacto Económico
		Avaliação do Impacto Cultural
Testada	Adequação para IA	Validação
	Testes pré-implementação	Testes Independentes
		Benchmarking
		Testes de Integração
	Monitorização contínua	Testagem e Monitorização

Abaixo encontra-se a análise aprofundada de cada pilar, princípio e boa prática.

4.2.1 Transparente

A Transparência é um dos pilares fundamentais para a construção de uma IA fiável e segura. O objetivo é abrir a "caixa-negra" (black box) da IA, tornando os seus processos e decisões claros e percetíveis para todos. De acordo com o AI Act (considerando 27), a transparência implica que os sistemas de IA sejam desenvolvidos e utilizados de forma a permitir rastreabilidade e explicabilidade adequadas. Este princípio exige ainda que os humanos sejam informados quando interagem com um sistema de IA e que os responsáveis pela implantação do sistema (os deployers) conheçam as suas capacidades e limitações, ao mesmo tempo que as pessoas afetadas pelas suas decisões são informadas sobre os seus direitos.

4.2.1.1 Explicabilidade

A Explicabilidade é o primeiro princípio no pilar da Transparência. Explicabilidade é a capacidade de um sistema de IA articular a lógica por detrás das suas decisões e resultados de forma compreensível para os utilizadores e partes interessadas. Não basta que um sistema seja transparente na sua conceção e documentação, precisa também de ser inteligível na sua operação diária para construir confiança e permitir uma supervisão informada.

Boas Práticas	
Explicações	Os sistemas de IA devem ser capazes de identificar e comunicar claramente os fatores que mais influenciaram uma determinada previsão ou decisão através de explicações. Esta capacidade de fornecer um caminho de decisão rastreável é essencial, pois não só desmistifica o comportamento da IA, como também se torna uma ferramenta de diagnóstico fundamental para identificar e corrigir erros. A capacidade de compreender e articular a lógica por detrás dos resultados de um sistema de IA é crucial para que os utilizadores e outras partes interessadas possam confiar na tecnologia e tomar decisões informadas com base nas suas recomendações. Este requisito vai para além da mera conformidade técnica. O AI Act consagra o direito a uma explicação para qualquer pessoa afetada por uma decisão com impacto legal ou significativo que seja tomada com base no resultado de um sistema de IA de alto risco. Esta explicação deve ser clara e elucidativa, detalhando não só o papel do sistema de IA no processo, mas também os principais elementos que fundamentaram a decisão.

Perguntas Chave

1. As decisões e previsões do sistema de IA são explicadas de forma clara, concisa e compreensível para o público-alvo?
2. É possível rastrear os principais fatores que contribuíram para um resultado específico?
3. As explicações fornecidas são suficientes para permitir a identificação e correção de eventuais erros ou enviesamentos?

Supervisão Humana

Preservar a supervisão e autonomia humanas é um requisito central da IA Responsável. Os sistemas, especialmente os de alto risco, devem ser projetados para permitir que utilizadores autorizados possam rever, ajustar ou mesmo anular as decisões da IA quando tal for apropriado. O AI Act reforça esta necessidade, estabelecendo que a supervisão humana deve ser eficaz e informada. As pessoas designadas para supervisionar um sistema de IA devem ser capacitadas para compreender plenamente as suas funcionalidades e limitações, permitindo uma monitorização adequada do seu funcionamento.

É igualmente importante que estes supervisores estejam cientes do risco de enviesamento de automação, ou seja, a tendência para confiar excessivamente nos resultados da IA, e que sejam capazes de interpretar corretamente os resultados gerados, utilizando as ferramentas de interpretação disponíveis. Em última análise, o controlo humano deve garantir a capacidade de decidir não utilizar o sistema numa dada situação, de ignorar ou reverter os seus resultados, e de intervir para o interromper de forma segura através de mecanismos como um "botão de paragem".

Perguntas Chave

4. Existem mecanismos claros para que um supervisor humano possa rever, corrigir ou anular as decisões do sistema?
5. O sistema inclui funcionalidades de paragem de emergência que podem ser acionadas por um humano de forma segura?
6. Os supervisores recebem formação adequada sobre as capacidades, limitações e os riscos de excesso de confiança no sistema?

Sinalização da presença de IA

Para reforçar a confiança e a responsabilidade, é fundamental que as pessoas saibam quando estão a interagir com um sistema de IA. O AI Act determina que os sistemas destinados a interagir diretamente com pessoas devem ser concebidos para as informar dessa interação, a menos que tal seja óbvio dadas as circunstâncias. Esta obrigação visa garantir que os utilizadores não são enganados ou manipulados, permitindo-lhes adaptar o seu comportamento e expectativas.

Além disso, esta transparência estende-se ao conteúdo gerado por IA. A legislação europeia prevê a necessidade de incorporar soluções técnicas, como marcas de água ou outros metadados, que permitam identificar conteúdos gerados ou manipulados por um sistema de IA, distinguindo-os de conteúdos autênticos criados por humanos. Esta medida é particularmente relevante para combater a desinformação e conteúdos sintéticos manipulados, conhecidos como *deepfakes*.

Perguntas Chave

1. A presença da IA é explicitamente comunicada aos utilizadores durante a interação?
2. O conteúdo gerado ou manipulado pelo sistema (texto, imagem, áudio ou vídeo) é devidamente assinalado como artificial?
3. Os mecanismos de identificação de conteúdo gerado por IA são robustos e fiáveis, de acordo com o estado da arte?

4.2.1.2 Governação

A Governação da Inteligência Artificial é o princípio que sustenta a implementação de sistemas responsáveis, éticos e transparentes. Estabelece as políticas, os processos e as responsabilidades necessários para garantir que a IA é desenvolvida e utilizada de forma alinhada com os requisitos legais, os valores éticos e os objetivos estratégicos de uma organização. Uma governação robusta abrange todo o ciclo de vida da IA, desde a conceção e recolha de dados até à implementação, monitorização e descontinuação do sistema.

Boas Práticas

Governação de Dados

A Governação de Dados constitui a base sobre a qual se constroem sistemas de IA precisos, justos e responsáveis. Envolve a implementação de políticas, procedimentos e standards rigorosos para assegurar a qualidade, segurança e utilização ética dos dados ao longo de todo o seu ciclo de vida. Esta prática é essencial no tratamento de dados sensíveis ou pessoais, sendo crucial para o desenvolvimento de operações de IA simultaneamente fiáveis e éticas. A manutenção de registos detalhados sobre a origem dos dados, bem como sobre os processos de limpeza, anotação e validação, é fundamental para garantir a rastreabilidade e a fiabilidade, elementos indispensáveis para a confiança no sistema.

O AI Act define práticas de governação e gestão de dados para sistemas de alto risco, abrangendo aspetos como as escolhas de *design*, os processos de recolha de dados e a sua origem e, no caso de dados pessoais, o propósito da sua recolha. Inclui também operações de preparação de dados como anotação, etiquetagem, limpeza e agregação. É igualmente exigido explicitar os pressupostos sobre a informação que os dados representam, avaliar a disponibilidade e adequação dos conjuntos de dados, e examinar potenciais vieses que possam afetar direitos fundamentais ou causar discriminação. A legislação impõe ainda a adoção de medidas apropriadas para detetar, prevenir e mitigar esses vieses, bem como para identificar e colmatar lacunas nos dados que possam comprometer a conformidade do sistema.

Perguntas Chave

1. São mantidos registos detalhados sobre a origem, o tratamento e a validação dos dados para assegurar a sua total rastreabilidade e fiabilidade?
2. Existem políticas e procedimentos formais para garantir a qualidade, segurança e utilização ética dos dados, especialmente os de natureza sensível?
3. A organização assegura que a recolha e o tratamento de dados estão em conformidade com a legislação de proteção de dados e com os direitos dos seus titulares?

Governação de Modelos

A Governação de Modelos foca-se na documentação e supervisão do design, desenvolvimento e validação dos modelos de IA. É um processo

essencial para assegurar a transparência, a conformidade e a integridade dos sistemas. A documentação completa do design do modelo, incluindo a sua arquitetura, métodos de treino, limitações e as considerações éticas subjacentes, é um requisito fundamental para uma governação eficaz. Sem esta documentação detalhada, torna-se impossível realizar auditorias, garantir a responsabilização e, em última análise, construir confiança nos sistemas de IA.

O AI Act exige que os prestadores de sistemas de IA de alto risco elaborem e mantenham documentação técnica detalhada antes da sua colocação no mercado. Esta documentação deve ser clara e suficiente para permitir às autoridades avaliar a conformidade do sistema com os requisitos legais. Deve incluir, no mínimo, uma descrição geral do sistema e do seu propósito, detalhes sobre o processo de desenvolvimento, incluindo a utilização de sistemas ou ferramentas pré-treinadas, especificações sobre a arquitetura do sistema e a lógica dos algoritmos, e os requisitos de dados, descrevendo as metodologias e os conjuntos de dados de treino utilizados. Adicionalmente, a documentação deve incluir uma avaliação das medidas de supervisão humana, os procedimentos de validação e teste com as métricas de precisão e robustez, e uma descrição do sistema de gestão de risco implementado.

Perguntas Chave

1. A documentação do modelo de IA detalha de forma exaustiva a sua arquitetura, os métodos de treino, as limitações conhecidas e as considerações éticas subjacentes ao seu desenvolvimento?
2. A documentação técnica cumpre todos os requisitos exigidos pela legislação aplicável, como o AI Act, para a categoria de risco do sistema?
3. Existem processos para manter a documentação do modelo atualizada ao longo do seu ciclo de vida, refletindo todas as modificações substanciais?

Governação de Sistemas

A Governação de Sistemas estende-se para além do modelo e dos dados, abrangendo todo o ciclo de vida operacional do sistema de IA. Implica a partilha aberta e transparente de como os sistemas são implementados, monitorizados e geridos, cobrindo tanto a sua criação como a sua utilização

prática. Este nível de governação foca-se em mecanismos de supervisão que monitorizam a performance do sistema em tempo real, garantem a conformidade contínua com as normas e regulamentos, e suportam auditorias internas para manter a integridade a longo prazo. Uma governação de sistemas eficaz assegura que a tecnologia não só funciona como previsto, mas que também se mantém alinhada com os princípios éticos e os requisitos legais ao longo do tempo.

O AI Act reforça esta necessidade ao estipular que a transparência deve ser uma característica central dos sistemas de IA, garantindo que, nos casos de sistemas de risco elevado, os responsáveis pela implantação (*deployers*) sejam devidamente informados sobre as capacidades e limitações do sistema, e que as pessoas afetadas pelas suas decisões conheçam os seus direitos. A documentação técnica exigida pela legislação deve incluir um plano de monitorização pós-comercialização, que detalhe como será avaliado o sistema. Este requisito assegura que a governação não termina com o desenvolvimento, mas continua através de uma vigilância ativa que permite identificar e corrigir problemas durante a utilização do sistema em contextos reais, garantindo a sua fiabilidade e segurança contínuas.

Perguntas Chave

1. Existem mecanismos de supervisão para monitorizar a implementação do sistema de IA, garantir a conformidade contínua e suportar auditorias internas que assegurem a sua integridade a longo prazo?
2. As políticas de governação abrangem todo o ciclo de vida do sistema, desde o desenvolvimento à descontinuação, assegurando uma gestão e monitorização contínuas?
3. A organização partilha de forma transparente as informações sobre como os sistemas de IA são implementados e geridos com as partes interessadas relevantes?

4.2.2 Robusta

O pilar da IA Robusta assenta na premissa de que os sistemas de Inteligência Artificial devem ser intrinsecamente fiáveis, consistentes e protegidos contra falhas e ameaças. A robustez stricto sensu garante que um sistema mantém a

sua precisão e desempenho mesmo perante condições adversas, dados inesperados ou mudanças no ambiente operacional. A robustez lato sensu inclui também a privacidade e segurança, focando-se na proteção de dados pessoais e confidenciais e proteção contra manipulações externas, ciberataques ou qualquer exploração de vulnerabilidades que possa comprometer a integridade e funcionamento dos sistemas. A robustez é essencial para construir a confiança necessária à adoção da IA, especialmente em aplicações críticas onde a segurança e os direitos fundamentais estão em jogo, em total alinhamento com as exigências do AI Act.

4.2.2.1 Robustez

O princípio da robustez stricto sensu estipula que um sistema deve ser concebido para funcionar de forma fiável e consistente ao longo de todo o seu ciclo de vida. Isto implica que o sistema deve ser resiliente a erros e alterações imprevisíveis, estável perante flutuações, e previsível nos seus resultados. Uma IA robusta não é apenas tecnicamente sólida, mas uma IA em que os utilizadores podem confiar, pois o seu comportamento não é errático ou aleatório. Este princípio é a base para garantir que a tecnologia atua como uma ferramenta segura e eficaz, minimizando riscos e maximizando o valor que entrega à sociedade e às organizações.

Boas Práticas	
Resiliência	A resiliência de um sistema de IA reflete a sua capacidade para lidar com erros, falhas ou inconsistências que possam ocorrer, quer internamente, quer no ambiente em que opera, nomeadamente devido à sua interação com pessoas ou outros sistemas. Um sistema resiliente deve conseguir gerir alterações previsíveis nos dados, como flutuações sazonais ou mudanças económicas, sem necessitar de uma reestruturação completa. O AI Act sublinha que esta resiliência deve ser assegurada através de medidas técnicas e organizacionais, que podem incluir soluções de redundância técnica, como planos de contingência (*fail-safe*) e cópias de segurança (*backups*). Para além de lidar com mudanças previsíveis, a resiliência implica a implementação de mecanismos contínuos de monitorização e retreino que permitam detetar anomalias, desvios de desempenho a longo prazo (*drift*) e atividades maliciosas, assegurando a continuidade do sistema. É dada uma atenção particular aos sistemas que continuam a aprender após a

sua implementação; nestes casos, é crucial eliminar ou reduzir o risco de ciclos de retroalimentação (*feedback loops)*, nos quais resultados enviesados podem influenciar e amplificar futuros preconceitos.

Perguntas Chave

1. O sistema está preparado para detetar e lidar com o desvio de dados (*data drift*), anomalias e atividade maliciosa?
2. Existe um processo estruturado para monitorizar e corrigir a degradação do desempenho?
3. São realizados testes de esforço e simulações adversariais antes da implementação?

Estabilidade

A estabilidade é uma faceta essencial da robustez, garantindo que os sistemas de IA evitam comportamentos erráticos quando confrontados com pequenas flutuações nos dados de entrada (*input*). Esta característica é particularmente crítica em sistemas que operam durante longos períodos e estão sujeitos a uma utilização repetida, pois assegura que o seu desempenho não se degrada nem se torna imprevisível com o tempo.

Um sistema estável inspira confiança porque o seu comportamento é consistente e fiável. Os utilizadores e as organizações podem, assim, depender das suas decisões e recomendações sem o receio de variações inesperadas que poderiam comprometer a segurança ou a eficácia das operações. A estabilidade deve, por isso, ser uma prioridade desde a fase de conceção, sendo validada através de testes rigorosos que simulem diversas condições de utilização.

Perguntas Chave

1. Os resultados são estáveis e previsíveis ao longo do tempo?

Resultados Previsíveis

A previsibilidade num sistema de IA significa que dados de entrada semelhantes devem, de forma fiável, produzir resultados semelhantes. Este princípio garante que os resultados do sistema estão alinhados não só com os padrões dos dados com que foi treinado, mas também com as expectativas dos utilizadores. Quando um sistema de IA se comporta de forma previsível, demonstra a sua integridade e lógica interna, tornando-se

uma ferramenta fiável em vez de uma caixa negra (*black box*) impenetrável.

A capacidade de antecipar o comportamento de um sistema de IA é fundamental para a sua integração responsável em processos críticos. Se os utilizadores conseguem prever o comportamento do sistema em diferentes cenários, podem confiar nas suas decisões, supervisioná-lo de forma mais eficaz e intervir quando necessário. A ausência de previsibilidade, pelo contrário, mina a confiança e pode levar a erros graves, tornando o sistema inadequado para aplicações de alto risco.

Perguntas Chave

1. A consistência do sistema foi validada para assegurar que dados de entrada semelhantes produzem, de forma fiável, resultados semelhantes?

4.2.2.2 Privacidade e Proteção de Dados

A proteção de dados e o respeito pela privacidade são princípios fundamentais da Inteligência Artificial Responsável. A forma como as organizações recolhem, gerem e protegem os dados pessoais não é apenas uma questão de conformidade legal, mas um fator decisivo para a construção de uma relação de confiança com os cidadãos e utilizadores. Num cenário em que os modelos de IA, especialmente os de grande escala, dependem de vastos conjuntos de dados para o seu treino e funcionamento, a implementação de práticas robustas de privacidade e proteção de dados torna-se um imperativo ético e operacional.

Boas Práticas	
Privacidade de Dados	A privacidade de dados transcende a mera proteção; representa o direito fundamental de um indivíduo à confidencialidade, ao anonimato e à segurança das suas informações pessoais. Este direito inclui a prerrogativa de ser informado e de consentir sobre a forma como os seus dados são utilizados, associado à responsabilidade inequívoca das organizações em salvaguardar estes direitos em todas as fases do tratamento de dados. Os sistemas de IA devem, por isso, ser concebidos

para respeitar a legislação de privacidade e os direitos dos utilizadores, limitando a recolha, o uso e a partilha de dados pessoais ao estritamente necessário.

Para materializar este princípio, as organizações devem adotar um conjunto de medidas técnicas e organizacionais. O AI Act prevê que, para garantir a conformidade, os prestadores possam utilizar não só a anonimização e a cifragem, mas também tecnologias que permitam levar os algoritmos aos dados, possibilitando o treino de modelos sem transmitir ou copiar os dados brutos ou estruturados. A implementação de técnicas que reforçam a privacidade, como a cifragem de dados sensíveis e o controlo estrito de acessos, é crucial para assegurar a confidencialidade e a adesão às diretrizes legais e éticas.

Perguntas Chave

1. O sistema de IA cumpre com os regulamentos de privacidade de dados aplicáveis, como o RGPD e outras políticas específicas do setor?
2. Os dados sensíveis estão devidamente protegidos através de cifragem e controlos de acesso rigorosos?
3. Foram implementadas medidas de minimização de dados para garantir que apenas os dados estritamente necessários são recolhidos e processados?

Consentimento

O consentimento informado é a base de uma relação de confiança entre o utilizador e o sistema de IA. Os utilizadores e titulares de dados devem ter controlos claros e acessíveis sobre a forma como a sua informação é recolhida, utilizada e armazenada, reforçando a transparência e a autonomia individual. No entanto, a obtenção de um consentimento genuíno e informado tornou-se particularmente desafiante no contexto dos grandes modelos de linguagem (LLMs), que requerem volumes massivos de dados para o seu treino. Em muitos casos, os utilizadores não têm conhecimento da forma como os seus dados são utilizados nem da extensão da sua recolha, o que torna a transparência sobre estas práticas ainda mais crítica.

A crescente restrição ao uso de dados disponíveis publicamente para treino de IA, por via da implementação de protocolos que limitam a

extração de dados (*data scraping*), evidencia esta tensão. Este fenómeno, por um lado, protege os direitos dos criadores e proprietários de dados; por outro, pode afetar a diversidade dos dados, o alinhamento dos modelos e a sua escalabilidade futura. Assim, é fundamental que as organizações desenvolvam mecanismos de consentimento que sejam não só conformes, mas também significativos, oferecendo aos utilizadores um controlo real sobre os seus dados, incluindo opções claras para consentir, opor-se ou solicitar a sua eliminação.

Perguntas Chave

1. Os utilizadores e/ou titulares dos dados têm um controlo efetivo sobre os seus dados, incluindo opções claras e acessíveis para dar ou retirar o consentimento e solicitar a sua eliminação?
2. Os mecanismos de consentimento são transparentes quanto à finalidade e à extensão da recolha e utilização de dados, especialmente no treino de modelos de IA?

Integridade de Dados

A integridade dos dados é um pré-requisito para a fiabilidade e segurança de qualquer sistema de IA. Os dados utilizados para treino, validação e teste devem ser precisos, verificáveis e livres de manipulação. Dados imprecisos, incompletos ou corrompidos degradam o desempenho do modelo e podem levar a decisões enviesadas, resultados injustos e vulnerabilidades de segurança. O AI Act exige que os conjuntos de dados para sistemas de alto risco sejam relevantes, suficientemente representativos e, na medida do possível, isentos de erros e completos, de acordo com a finalidade do sistema.

Garantir a integridade exige a implementação de práticas robustas de governação de dados ao longo de todo o seu ciclo de vida. Isto inclui processos para verificar a origem dos dados, assegurar a sua exatidão, protegê-los contra manipulações não autorizadas e documentar todas as etapas de processamento. A integridade dos dados não é apenas uma salvaguarda técnica; é uma condição essencial para que os resultados produzidos pela IA sejam fiáveis e, em última análise, para que o sistema cumpra a sua finalidade de forma segura e eficaz.

Perguntas Chave

1. Existem processos para verificar a exatidão, a proveniência e a representatividade dos dados utilizados para treinar e validar os sistemas de IA?
2. Os conjuntos de dados estão protegidos contra manipulação ou corrupção, tanto por agentes internos como externos?
3. Existem mecanismos de auditoria para validar continuamente a integridade dos dados ao longo do ciclo de vida do sistema de IA?

4.2.2.3 Segurança

A segurança em Inteligência Artificial é um princípio fundamental que abrange a proteção dos sistemas contra ameaças externas, a garantia da sua fiabilidade operacional e a minimização de danos decorrentes de utilizações indevidas. Um sistema de IA seguro é aquele que inspira confiança, cumpre as normativas legais e protege a integridade dos seus componentes — algoritmos, dados e infraestrutura — ao longo de todo o seu ciclo de vida.

A implementação de uma IA segura não é apenas uma necessidade técnica, mas também uma exigência estratégica. Organizações e líderes empresariais reconhecem os riscos associados à segurança, com a cibersegurança a figurar como uma das principais preocupações. No entanto, ainda existe um desfasamento entre a identificação destes riscos e a implementação de medidas de mitigação eficazes. Uma abordagem proativa e integrada da segurança é, por isso, crucial para desbloquear o verdadeiro potencial da IA de forma responsável.

Boas Práticas	
Cibersegurança	A cibersegurança dos sistemas de IA foca-se em garantir a sua integridade e resiliência contra ataques deliberados que procuram explorar as suas vulnerabilidades. Medidas de segurança robustas são essenciais para prevenir acessos não autorizados, violações de dados e ameaças adversariais. Estas medidas devem incluir a utilização de encriptação de dados, protocolos de autenticação e monitorização em tempo real, permitindo apagar e mitigar riscos de forma contínua e eficaz. A proteção deve abranger todos os componentes do sistema, desde os algoritmos e dados de treino até à infraestrutura subjacente,

para criar uma defesa em profundidade contra um leque variado de ameaças externas.

O AI Act estabelece que os sistemas de IA, especialmente os de alto risco, devem ser resilientes contra tentativas de terceiros de alterar a sua utilização, resultados ou desempenho através da exploração de vulnerabilidades. Isto implica a adoção de soluções técnicas adequadas às circunstâncias e aos riscos específicos de cada sistema. Estas soluções devem ser capazes de prevenir, detetar e controlar ataques que visam manipular os dados de treino (*data poisoning*), os modelos pré-treinados (*model poisoning*), ou que utilizam *inputs* concebidos para induzir o modelo em erro (*adversarial examples* ou *model evasion*). A simulação proativa de ataques durante a fase de desenvolvimento é uma prática recomendada para testar a resiliência do sistema e treinar os modelos com dados que reflitam ameaças do mundo real.

Perguntas Chave

1. São realizadas auditorias de segurança e testes de penetração regularmente?
2. Existem controlos de acesso e encriptação para proteger os dados sensíveis utilizados pelo sistema?
3. Estão implementadas medidas para prevenir ataques adversariais, como *data poisoning* ou *model evasion*?

Fiabilidade do Sistema

A fiabilidade de um sistema de IA refere-se à sua capacidade de funcionar de forma consistente e estável, mesmo sob pressão, picos de utilização ou tentativas de disrupção. Um sistema fiável é crucial em cenários de elevada exigência ou de alto risco, onde uma falha pode ter consequências significativas. Para alcançar este nível de fiabilidade, é imperativo que as organizações implementem mecanismos de tolerância a falhas (*failover*), monitorização contínua e estratégias de resposta rápida que assegurem a estabilidade e a continuidade do serviço.

A consistência essencial para a fiabilidade, evitando comportamentos erráticos perante pequenas variações nos dados de entrada (*input*) e garantindo resultados (*outputs*) semelhantes para *inputs* semelhantes. Manter esta previsibilidade ao longo do tempo, mesmo em

implementações prolongadas e sob uso repetido, constrói a confiança dos utilizadores e alinha o desempenho do sistema com as expectativas definidas no treino. A implementação de um processo estruturado para monitorizar e corrigir qualquer degradação no desempenho é, por isso, uma componente indispensável de uma IA fiável.

Perguntas Chave

1. O sistema é continuamente monitorizado para detetar ameaças à segurança que possam comprometer a sua fiabilidade?
2. Existem mecanismos de failover e planos de contingência para garantir o funcionamento do sistema em caso de falha ou ataque?
3. Existem planos para garantir que o sistema se mantém estável e previsível em cenários de elevada exigência ou sob tentativas de disrupção?

Segurança do Sistema

A segurança de um sistema de IA vai além da proteção contra ataques externos e abrange a minimização dos danos que podem resultar da sua utilização indevida, seja ela deliberada ou inadvertida. As preocupações incluem o risco de a tecnologia ser usada para desenvolver ferramentas maliciosas, como *software* de intrusão (*hacking)* automatizado, ou de ser aplicada em ciberataques sofisticados. A segurança do sistema exige, portanto, uma avaliação contínua dos riscos associados não apenas ao seu funcionamento normal, mas também aos seus potenciais abusos.

Investigações recentes (AI Index 2025) demonstram que muitas das salvaguardas de segurança nos modelos de linguagem são superficiais, evidenciando um alinhamento de segurança frágil (fenómeno conhecido como *shallow safety alignment*). Isto significa que, embora um modelo possa recusar-se a responder a um pedido diretamente prejudicial, pode ser facilmente manipulado para gerar conteúdo perigoso com pequenas alterações nos dados de entrada do utilizador. Este tipo de vulnerabilidade, juntamente com técnicas que permitem contornar as restrições de segurança (*jailbreaking*), sublinha a necessidade de estratégias de alinhamento mais profundas e resilientes. A segurança deve ser uma propriedade intrínseca do modelo, e não

apenas uma camada superficial de proteção.

Perguntas Chave

1. Existe um método de avaliação e mitigação dos riscos de utilização indevida do sistema de IA, tanto deliberada como inadvertida?
2. Estão implementadas salvaguardas para garantir que o modelo não é facilmente manipulado para gerar resultados ou comportamentos perigosos?
3. O treino de segurança do modelo é suficientemente robusto para resistir a técnicas de *jailbreaking* ou outras formas de manipulação?

4.2.3 Universal

O pilar da IA Universal representa o compromisso de desenvolver sistemas de Inteligência Artificial que sejam, na sua essência, justos, inclusivos e acessíveis a todos. Este princípio orientador exige que se trabalhe ativamente para identificar e eliminar vieses que possam levar à discriminação, garantindo que os benefícios da IA são distribuídos de forma equitativa por toda a sociedade. A universalidade implica que a tecnologia seja concebida para servir eficazmente populações diversas, independentemente da sua origem, contexto socioeconómico, linguístico ou das suas capacidades, assegurando que a IA não perpetua nem amplifica as desigualdades sociais existentes. Esta abordagem está intrinsecamente ligada ao respeito pela dignidade humana e pelos direitos fundamentais, um valor central consagrado no AI Act.

4.2.3.1 Equidade

O princípio da Equidade constitui a aplicação prática da Universalidade, focando-se na criação de sistemas que produzem resultados justos e imparciais para todos os indivíduos. Indo para além da simples prevenção de danos, a equidade exige uma abordagem proativa no desenho e implementação de sistemas de IA, envolvendo uma análise rigorosa dos dados, algoritmos e processos de decisão para mitigar vieses que possam resultar em tratamento injusto. Uma IA equitativa é aquela cujos resultados não penalizam nem privilegiam desproporcionalmente qualquer grupo, assegurando que a sua implementação contribui para uma sociedade mais justa e inclusiva. Este compromisso está em plena

conformidade com as obrigações legais e éticas estabelecidas por quadros regulatórios como o AI Act, que exige medidas concretas para a deteção e correção de vieses, especialmente em sistemas de alto risco.

Boas Práticas	
Resultados Justos	Para garantir resultados justos, os sistemas de IA devem ser desenvolvidos com o objetivo explícito de evitar favorecer ou penalizar desproporcionalmente qualquer grupo. A aplicação de métricas de equidade ao longo de todo o processo de desenvolvimento e implementação é essencial para identificar e corrigir desequilíbrios na tomada de decisão. Esta abordagem vai além de uma mera verificação técnica, tornando-se uma componente central da governação da IA. Contudo, a investigação demonstra que mesmo os modelos mais avançados, concebidos para serem explicitamente imparciais, continuam a apresentar vieses implícitos, como a associação desproporcional de termos negativos a determinados grupos raciais ou o reforço de estereótipos de género em contextos profissionais. Além disso, a simples expansão da escala dos modelos pode, paradoxalmente, aumentar estes vieses, como observado em sistemas que passaram a classificar desproporcionalmente certos grupos demográficos como potencialmente criminosos após serem treinados com conjuntos de dados maiores. O AI Act responde a estes desafios estabelecendo um quadro legal robusto proibindo práticas de IA consideradas inaceitáveis, como a pontuação social de indivíduos (*social scoring*), que pode conduzir a resultados discriminatórios e à exclusão de certos grupos. Para os sistemas de IA classificados como de alto risco, o regulamento impõe requisitos rigorosos de governação de dados, exigindo que os conjuntos de dados de treino, validação e teste sejam relevantes, representativos e, tanto quanto possível, completos e isentos de erros. Estas práticas devem incluir a análise de possíveis vieses que possam afetar negativamente os direitos fundamentais ou levar a discriminação proibida pela legislação da União, bem como a adoção de medidas adequadas para os detetar, prevenir e mitigar. Em situações excecionais, e para garantir a deteção e correção de vieses, o AI Act permite o tratamento de categorias especiais de dados pessoais, desde que sejam aplicadas salvaguardas rigorosas para a proteção dos direitos e liberdades fundamentais.

Perguntas Chave

1. Os objetivos de equidade estão claramente definidos e integrados no desenvolvimento do modelo desde a sua conceção?
2. São utilizadas métricas e *benchmarks* de equidade para monitorizar e corrigir desequilíbrios nos resultados do sistema de IA ao longo do seu ciclo de vida?
3. São identificados e mitigados os vieses nos conjuntos de dados utilizados, em conformidade com as exigências do AI Act?

Auditorias Regulares

A garantia de equidade num sistema de IA não é um objetivo que se atinge uma única vez; exige vigilância contínua. As avaliações de equidade devem ser conduzidas de forma regular, analisando os resultados das decisões do sistema para detetar e resolver disparidades emergentes antes que estas se tornem sistémicas. Sem uma monitorização proativa, pequenas desigualdades podem acumular-se e ser amplificadas ao longo do tempo, reforçando preconceitos existentes e criando ciclos de retroalimentação negativos que minam a confiança e a eficácia da tecnologia. A realização de auditorias de viés, utilizando ferramentas e metodologias robustas, é, por isso, um processo indispensável.

A necessidade de auditoria é reforçada pela constatação de que persistem desafios significativos na transparência dos sistemas de IA. Estudos recentes (AI Index 2025) revelam que uma grande percentagem dos conjuntos de dados utilizados para treinar modelos carece de informação adequada sobre licenciamento e proveniência, o que dificulta, ou mesmo impede, uma auditoria rigorosa por parte de terceiros. O AI Act aborda diretamente esta necessidade de supervisão contínua, exigindo que os prestadores de sistemas de IA de alto risco implementem um sistema de monitorização pós-comercialização. Este sistema deve recolher, documentar e analisar ativamente dados sobre o desempenho do sistema ao longo do seu ciclo de vida, permitindo uma avaliação contínua da sua conformidade e a deteção de riscos emergentes. Este processo deve ser iterativo e sistemático, abrangendo todo o ciclo de vida do sistema. Adicionalmente, tanto os prestadores como os responsáveis pela implantação de sistemas de IA de alto risco têm a obrigação de monitorizar o seu funcionamento, criando, na prática, um quadro para

auditorias contínuas.

Perguntas Chave

1. São realizadas auditorias de viés regulares para avaliar a equidade do sistema de IA em produção?
2. Estão em vigor procedimentos para corrigir as disparidades e os vieses identificados durante as auditorias?
3. O sistema de monitorização pós-comercialização, exigido pelo AI Act para sistemas de risco elevado, é utilizado para detetar e mitigar vieses que possam surgir após a implementação?

4.2.3.2 Inclusividade

Um sistema de IA verdadeiramente responsável deve ser concebido para servir a sociedade como um todo, garantindo que os seus benefícios são distribuídos de forma equitativa e que não cria ou aprofunda desigualdades. A inclusão não é um complemento, mas sim um requisito fundamental que deve estar presente desde a conceção até à implementação do sistema. Este princípio desdobra-se em duas vertentes complementares: o sistema deve ser desenhado por todos e para todos.

Boas Práticas	
Desenhado por Todos	A construção de uma IA inclusiva começa com um processo de criação igualmente inclusivo. A diversidade nas equipas de desenvolvimento é um fator crítico para o sucesso e a responsabilidade de um sistema de IA. Equipas compostas por pessoas com diferentes experiências, perspetivas e competências, incluindo um equilíbrio de género, são mais capazes de identificar "pontos cegos" e preconceitos inconscientes que poderiam passar despercebidos a um grupo homogéneo. Esta diversidade não só ajuda a evitar a perpetuação de estereótipos, como também melhora a adaptabilidade e a robustez do sistema de IA, tornando-o mais resiliente a cenários imprevistos. Para além da diversidade interna das equipas, o AI Act sublinha a importância de envolver um leque alargado de atores externos no processo de desenvolvimento. A participação ativa de entidades como

organizações da sociedade civil, universidades, centros de investigação, sindicatos e organizações de defesa do consumidor é essencial para garantir que o sistema de IA responde a necessidades reais e está alinhado com os valores da sociedade. Este envolvimento colaborativo assegura que múltiplas vozes são ouvidas, resultando em sistemas mais justos, éticos e que inspiram maior confiança junto do público e dos utilizadores finais.

Perguntas Chave

1. A equipa de desenvolvimento reflete a diversidade (de género, cultural, de formação, etc.) da sociedade que o sistema de IA pretende servir?
2. Foram envolvidas partes interessadas externas (*stakeholders*), como organizações da sociedade civil e representantes dos utilizadores, no processo de conceção e desenvolvimento do sistema?
3. Existem mecanismos eficazes para recolher e integrar o feedback de diversos grupos de utilizadores durante todo o ciclo de vida do sistema de IA, desde a conceção à descontinuação?

Desenhado para Todos

Um sistema de IA inclusivo deve ser acessível e eficaz para todas as pessoas, independentemente das suas características demográficas, socioeconómicas, linguísticas ou das suas capacidades. O design inclusivo vai além da simples funcionalidade técnica; exige um compromisso ativo para garantir que a tecnologia não cria novas barreiras ou acentua as existentes. Isto implica que os sistemas de IA devem ser avaliados de forma rigorosa para prevenir disparidades no seu desempenho ou na sua usabilidade entre diferentes grupos de utilizadores.

O AI Act reforça esta necessidade, apelando a uma atenção especial às pessoas em situação de vulnerabilidade e à garantia de acessibilidade para pessoas com necessidades especiais. Um design que considera estes grupos desde o início tende a gerar produtos que são melhores para todos. A implementação de princípios de design universal assegura que a tecnologia se adapta às necessidades humanas, e não o contrário. Ao projetar sistemas de IA que sejam intuitivos, fáceis de utilizar e que ofereçam resultados equitativos para uma vasta gama de utilizadores, as

organizações cumprem as suas obrigações éticas e legais ao mesmo tempo que ampliam o alcance e o impacto positivo das suas inovações.

Perguntas Chave

1. O sistema de IA foi testado para garantir que é acessível e utilizável por pessoas com diferentes capacidades, incluindo pessoas com necessidades especiais?
2. Foram realizadas avaliações para detetar e corrigir disparidades de desempenho ou usabilidade entre diferentes grupos demográficos, culturais ou linguísticos?
3. O sistema foi concebido para funcionar de forma equitativa em diversos contextos socioeconómicos e linguísticos?

4.2.4 Sustentável

A sustentabilidade na Inteligência Artificial transcende a mera eficiência técnica, abrangendo uma avaliação criteriosa dos seus impactos ambientais, sociais, económicos e culturais. Um sistema de IA sustentável é aquele que não só otimiza o consumo de recursos energéticos e naturais durante todo o seu ciclo de vida, mas que também é implementado de forma a promover a equidade social, a justiça económica e a diversidade cultural. Esta abordagem holística é fundamental para garantir que os avanços tecnológicos contribuem positivamente para o bem-estar humano e para a saúde do planeta, alinhando a inovação com uma visão de progresso a longo prazo.

4.2.4.1 Minimização do Impacto Ambiental

Boas Práticas	
Eficiência Energética	A crescente sofisticação dos modelos de IA, especialmente os modelos de grande escala, acarreta um consumo energético considerável, com implicações ambientais significativas. O treino, teste e implementação destes sistemas exigem vastos recursos computacionais, localizados em centros de dados que consomem grandes quantidades de eletricidade e água para arrefecimento. A monitorização e mitigação deste impacto são, por isso, componentes essenciais de uma IA responsável. As organizações devem implementar estratégias para mapear, documentar e otimizar o

consumo de energia ao longo de todo o ciclo de vida dos seus sistemas de IA. Embora seja um desafio, as organizações começam a reconhecer o impacto ambiental como um risco relevante, ainda que os esforços de mitigação permaneçam limitados.

Em conformidade com as novas exigências regulamentares, como o AI Act, os prestadores de modelos de IA devem documentar o consumo de energia conhecido ou estimado dos seus modelos. Esta documentação deve detalhar os recursos computacionais utilizados, o tempo de treino e outras informações relevantes que permitam uma avaliação transparente da sua pegada energética. A promoção de práticas de programação energeticamente eficientes e o desenvolvimento de metodologias para medir e otimizar o desempenho energético são incentivados, visando alinhar a inovação tecnológica com as metas de sustentabilidade ambiental. Para os prestadores de sistemas jusante, a avaliação do consumo energético deve tornar-se um critério fundamental na escolha de modelos a integrar, fomentando uma cadeia de valor mais consciente e responsável.

Perguntas Chave

1. A organização mede e documenta o consumo energético dos seus sistemas de IA, tanto na fase de treino como na de inferência?
2. A escolha de modelos de IA de terceiros inclui uma avaliação da sua pegada energética como critério de seleção?
3. A organização comunica de forma transparente o impacto energético dos seus sistemas de IA aos seus clientes e ao público?

Uso Eficiente de Recursos Naturais

Os centros de dados que alojam e treinam modelos de IA geram uma enorme quantidade de calor e, para manter as temperaturas operacionais, necessitam de quantidades substanciais de água para os seus sistemas de arrefecimento. Este consumo intensivo de um recurso natural vital pode colocar uma pressão significativa sobre os ecossistemas locais, especialmente quando os centros de dados estão localizados em regiões com escassez de água. A gestão responsável da IA exige, portanto, que as organizações não só monitorizem o seu consumo de eletricidade, mas também quantifiquem, monitorizem e divulguem o seu consumo de água, adotando tecnologias de arrefecimento mais

eficientes ou explorando alternativas que minimizem a dependência deste recurso crítico.

Perguntas Chave

1. Existem critérios de sustentabilidade na seleção de fornecedores?
2. No caso dos centros de dados, existem políticas em vigor para a gestão responsável da água utilizada para arrefecimento e de resíduos eletrónicos gerados pelos sistemas?

4.2.4.2 Avaliação do Impacto Social

Para além do seu impacto ambiental, a sustentabilidade de um sistema de Inteligência Artificial deve ser medida pela sua contribuição para uma sociedade mais justa e equitativa. A introdução da IA em larga escala não é apenas uma transformação tecnológica; é uma força com o poder de redefinir o mercado de trabalho, reconfigurar dinâmicas económicas e influenciar profundamente a produção e o consumo cultural. Uma implementação que ignore estas dimensões corre o risco de agravar desigualdades e aumentar a bipolarização remuneratória.

Boas Práticas	
Avaliação do Impacto Laboral	A implementação generalizada da IA está a transformar o mercado de trabalho, com potencial para gerar desigualdades sociais e económicas significativas. A automação de tarefas humanas pode levar à perda de empregos em larga escala. Esta disrupção pode agravar as disparidades existentes, uma vez que os trabalhadores em posições mais vulneráveis são frequentemente os mais suscetíveis à automação. Para além da disponibilidade de emprego, a qualidade e a segurança do trabalho também podem ser afetadas. As funções que subsistem podem ser desvalorizadas e a ameaça de substituição pela IA pode criar dependências exploratórias, pressionando os trabalhadores a aceitar piores condições laborais e salariais. Uma abordagem responsável exige que as organizações avaliem proativamente estes impactos e adotem medidas de mitigação. O AI Act estipula que, antes de implementar um sistema de IA de alto risco no local de trabalho, os empregadores devem informar os trabalhadores afetados. A transparência é um primeiro passo crucial. É igualmente importante investir na requalificação e desenvolvimento de competências,

preparando a força de trabalho para as novas funções que surgirão. A criação de sistemas de IA não deve ser feita à custa da dignidade e dos direitos dos trabalhadores, mas sim como uma oportunidade para redesenhar o trabalho de forma mais humana e equitativa.

Perguntas Chave

1. Foi realizada uma avaliação do impacto potencial dos sistemas de IA nos postos de trabalho e na qualidade do emprego dentro da organização e no seu setor?
2. Estão a ser implementadas medidas para a requalificação e formação dos trabalhadores cujas funções possam ser afetadas pela automação?

Avaliação do Impacto Económico

O desenvolvimento de sistemas de IA exige um investimento massivo em poder computacional, dados e talento especializado, recursos que estão, na sua maioria, concentrados num número reduzido de grandes empresas e governos. Esta realidade cria um risco significativo de centralização de poder, e redução da concorrência, onde os benefícios económicos e estratégicos da IA são monopolizados por poucas entidades.

Esta concentração de poder é exacerbada pela dinâmica competitiva da "corrida à IA", que incentiva a acelerar o desenvolvimento e a implementação para maximizar vantagens estratégicas e económicas. Esta pressão pode levar a que se negligenciem aspetos fundamentais de segurança e responsabilidade, resultando na implementação de sistemas imaturos e propensos a erros, com externalidades negativas generalizadas, como a desinformação ou a injustiça. Para mitigar estes riscos, é essencial promover um ecossistema de IA mais equitativo, que apoie o acesso de PMEs e startups à tecnologia e que estabeleça quadros regulatórios capazes de garantir uma concorrência leal e uma distribuição mais justa dos benefícios gerados pela IA.

Perguntas Chave

1. A estratégia de IA da organização contribui para um ecossistema económico mais justo e competitivo?
2. Existem medidas de governação interna para evitar que a pressão

competitiva comprometa os padrões de segurança e ética?
3. A organização está a colaborar com outras entidades para promover o acesso aberto a dados e modelos, abrindo oportunidades de inovação a mais entidades para além das grandes empresas tecnológicas?

Avaliação do Impacto Cultural

A capacidade dos sistemas de IA generativa para criar conteúdo sintético (texto, imagens, música) está a ter um impacto profundo nas indústrias criativas e na forma como a cultura é produzida e consumida. Estes modelos são frequentemente treinados com vastas quantidades de dados da internet, que incluem obras protegidas por direitos de autor, muitas vezes obtidas sem a devida autorização. Quando os sistemas de IA produzem obras que imitam o estilo ou método de criadores humanos a uma velocidade e escala impossíveis de igualar, colocam em risco a subsistência económica dos autores e podem desincentivar a inovação e criatividade humanas. Em resposta, o AI Act exige que os prestadores de modelos de IA de finalidade geral implementem políticas para respeitar a lei de direitos de autor da União.

Existe ainda o risco de homogeneização cultural. A proliferação em larga escala de conteúdo sintético pode levar a uma uniformização das experiências culturais e à expropriação do valor simbólico de elementos culturais, que a IA utiliza sem compreender o seu contexto ou significado.

Perguntas Chave

1. A organização garante que a utilização de dados para treinar os seus modelos de IA respeita os direitos de autor e a propriedade intelectual?
2. Existe compensação justa para os criadores cujo trabalho contribui para o treino dos modelos e sistemas?
3. Existem políticas internas para garantir que os sistemas de IA promovem a diversidade cultural em vez da homogeneização?

4.2.5 Testada

A fase de teste é um pilar fundamental no ciclo de vida de qualquer sistema de IA, assegurando que este não só cumpre os seus objetivos funcionais, mas

também o faz de forma segura, fiável e alinhada com padrões éticos e regulamentares. Um processo de teste abrangente não é um evento único, mas uma atividade contínua que se prolonga por toda a vida útil do sistema em produção.

4.2.5.1 Adequação para IA

Boas Práticas	
Validação	A primeira etapa de um desenvolvimento responsável consiste em questionar a premissa fundamental: será a IA a solução certa para este desafio? Esta avaliação inicial, que precede qualquer esforço técnico, deve começar por um filtro de admissibilidade legal e ética. Antes de mais, é imperativo verificar se o caso de uso não se enquadra nas categorias que o AI Act classifica como representativas de um risco inaceitável. O AI Act define um conjunto de práticas de IA que são consideradas intrinsecamente contrárias aos valores europeus e aos direitos fundamentais (ver ponto 3.2.1.). Se um caso de uso se enquadrar numa destas práticas proibidas, o processo de avaliação deve ser imediatamente interrompido. Por fim, é essencial considerar os pré-requisitos práticos, como a disponibilidade de dados representativos e de alta qualidade, e a capacidade da infraestrutura e das equipas técnicas para suportar o desenvolvimento, a manutenção e a evolução contínua do sistema. Só após uma resposta afirmativa a todas estas questões fundamentais – admissibilidade legal, balanço de risco-benefício positivo e viabilidade prática – se deve avançar para a fase de desenvolvimento. **Perguntas Chave** 1. O caso de uso viola alguma das práticas de IA proibidas pelo AI Act, representando um risco inaceitável para os direitos fundamentais? 2. Sendo o caso de uso legalmente admissível, é a IA a solução mais adequada e proporcional para o problema, ou existe uma alternativa mais simples e segura? 3. A organização possui os dados, a infraestrutura e as competências necessárias para desenvolver e manter o sistema de forma responsável, e os benefícios esperados justificam

claramente os riscos?

4.2.5.2 Testes pré-implementação

Antes de um sistema de IA ser implementado, deve ser submetido a uma bateria de testes rigorosos que validem a sua conformidade, robustez e integração. Esta fase é crucial para mitigar riscos e garantir que o sistema se comporta como esperado no ambiente para o qual foi projetado. A testagem pré-implementação deve incluir avaliações independentes, o uso de padrões de referência (benchmarks) de mercado e testes de integração exaustivos.

Boas Práticas	
Testes Independentes	As avaliações internas devem ser complementadas por validações conduzidas por entidades independentes. Este recurso a uma avaliação externa é fundamental para verificar, de forma objetiva, a adesão do sistema às melhores práticas da indústria e às diretrizes regulamentares aplicáveis. Os testes independentes proporcionam uma camada adicional de escrutínio que ajuda a identificar vulnerabilidades, preconceitos ou outras falhas que possam não ter sido detetadas internamente. Este processo não só reforça a robustez do sistema, mas também aumenta a confiança dos reguladores, parceiros e utilizadores finais na sua integridade e fiabilidade. **Perguntas Chave** 1. A organização garante que as avaliações internas são complementadas por validações independentes para assegurar a conformidade e a adesão às melhores práticas?
Benchmarking	A utilização de *benchmarks* é essencial para posicionar o desempenho do sistema de IA em relação aos padrões do setor. Contudo, enquanto existem *benchmarks* bem estabelecidos para capacidades gerais (como matemática ou programação), o mesmo não acontece para a avaliação da IA Responsável. Esta ausência de avaliações padronizadas para segurança, equidade e transparência dificulta a comparação direta e objetiva entre diferentes modelos e soluções.

	Perguntas Chave 1. São aplicados benchmarks de explicabilidade, robustez, segurança, equidade e eficiência para comparar o sistema com os padrões da indústria e identificar lacunas de desempenho?
Testes de Integração	Um sistema de IA raramente opera de forma isolada; está, na maioria das vezes, integrado em ecossistemas e fluxos de trabalho mais amplos. Por isso, é fundamental testar exaustivamente as suas interações com outros processos, sejam eles baseados em regras, ferramentas de visualização ou supervisão humana. Testes de aceitação com utilizadores finais e em cenários do mundo real são indispensáveis para confirmar que o sistema funciona corretamente e acrescenta valor mensurável à tomada de decisão. A utilização de ambientes-sombra (*shadow environments*), onde o sistema opera em paralelo com os processos existentes sem os afetar, permite uma validação segura antes da implementação em larga escala. **Perguntas Chave** 1. O sistema de IA foi submetido a testes de integração exaustivos que simulam cenários reais e avaliam a sua interação com os utilizadores finais e outros sistemas existentes?

4.2.5.3 Monitorização contínua

A responsabilidade sobre um sistema de IA estende-se muito para além da sua implementação inicial. Os ambientes do mundo real são dinâmicos, os padrões de dados evoluem e podem surgir novos riscos imprevistos. A monitorização e testagem contínuas são, por isso, indispensáveis para garantir que o sistema permanece fiável, seguro e eficaz ao longo de todo o seu ciclo de vida.

Boas Práticas	
Testagem e	Após a implementação, é crucial monitorizar continuamente o

Monitorização desempenho da IA para detetar e corrigir desvios de performance (*performance drift*) ou outros problemas emergentes. Esta prática deve incluir a implementação de ciclos de feedback com os utilizadores e a utilização de ambientes-sombra para validar atualizações.

Uma monitorização ativa e sistemática permite não só a recolha de dados sobre o desempenho do sistema, mas também a análise da sua interação com outros sistemas e a identificação de comportamentos anómalos, que podem então ser corrigidos através de processos de re-treino adaptativo.

Perguntas Chave

1. Estão implementados mecanismos para monitorizar continuamente o desempenho do sistema de IA em produção e detetar desvios ou degradação do modelo?
2. O *feedback* dos utilizadores e os dados de desempenho são utilizados para acionar processos de re-treino adaptativo e manter o sistema alinhado com as condições do mundo real?
3. Em caso de sistemas de risco elevado, o plano de monitorização pós-implementação cumpre as exigências regulatórias definidas no AI Act, incluindo a documentação e o reporte de incidentes graves.

5. 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.

5.1 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).

5.2 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.

5.3 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.

- 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.

5.4 Arquitetura

A Ferramenta está estruturada em 4 secções:

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

5.5 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 2.

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

TABELA 2: ELEMENTOS DE REGISTO DO UTILIZADOR, ENTIDADE E PROJETO

5.6 Nível de Maturidade

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

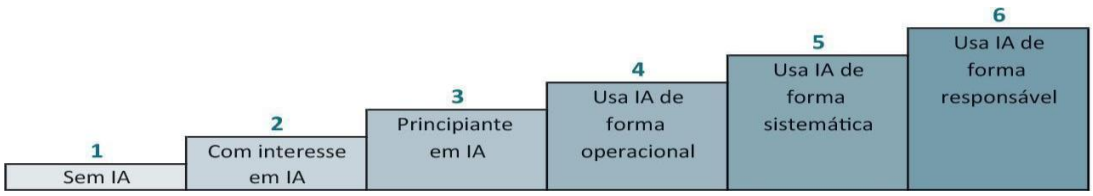

FIGURA 6: ELEMENTOS DE REGISTO DO UTILIZADOR, ENTIDADE E PROJETO

Este está relacionado com diferentes estágios de maturidade, ilustrados na figura 6:

1. Uma entidade sem IA - entidade que não possui qualquer tipo de sistema baseado em técnicas adaptativas.
2. Uma entidade com interesse em IA – entidade 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 - entidade 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 - entidade 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 - entidade 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- entidade 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.

5.7 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 7: 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".

5.8 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.

5.9 Participação de partes interessadas

A Ferramenta é uma plataforma evolutiva e pretende-se que venha a reunir os contributos dos seus utilizadores. Assim e 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 submetê-los à ARTE através do seguinte email: guia@arte.gov.pt.

5.10 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. De forma a tornar os novos pilares e princípios de IA Responsável em práticas verificáveis e auditáveis, a ferramenta descrita anteriormente será atualizada, tendo em conta três objetivos fundamentais:

- Ser acionável: A avaliação deverá ir além do modelo auto-declarativo, integrando um processo multifacetado que inclua a sugestão e validação de testes técnicos, de validação por terceiros e de monitorização do sistema após a sua implementação.
- Ser flexível: Os riscos dos sistemas de IA não são estáticos. A ferramenta deverá ser modular e adaptável, permitindo ajustar as avaliações ao caso de uso e nível de risco de cada sistema, alinhando-se com a abordagem do EU AI Act. Deverá também ser capaz de incorporar novas diretrizes regulatórias e normas técnicas, como as que serão emitidas pela Comissão Europeia e pelo CEN-CENELEC, garantindo a relevância no longo prazo.
- Ser aberta: A ferramenta deverá ser aberta e de fácil acesso, de forma a maximizar transparência, segurança e colaboração. Uma abordagem aberta permite que a academia, indústria e sociedade civil contribuam para o seu desenvolvimento.

No período transitório, durante o qual a nova versão da ferramenta estará em desenvolvimento, será introduzido no processo de contratação pública o questionário presente no Anexo I. Deste modo, assegurar-se-á o mínimo alinhamento com os pilares, princípios e boas práticas de IA Responsável.

Referências

- European Commission: Directorate-General for Communications Networks, Content and Technology & Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji. (2019). *Ethics guidelines for trustworthy AI*. Publications Office. https://data.europa.eu/doi/10.2759/346720.
- Feedzai Research. (2025). The TRUST framework. *Accelerate innovation with responsible AI*. https://www.feedzai.com/wp-content/uploads/2025/03/Trust-Framework.pdf
- Joshi, A., & Khalfi-Lanoux, N. (2024). Top 10 operational impacts of the EU AI Act. *IAPP Resource Center*, 2–6. https://iapp.org/media/pdf/resource_center/top_10_impacts_eu_ai_act.pdf
- OECD. (2024). *OECD digital economy outlook 2024 (Volume 1): Embracing the technology frontier*. OECD Publishing. https://doi.org/10.1787/a1689dc5-en](https://doi.org/10.1787/a1689dc5-en
- OECD. (2024). *OECD digital economy outlook 2024 (Volume 2): Strengthening connectivity, innovation and trust*. OECD Publishing. https://doi.org/10.1787/3adf705b-en](https://doi.org/10.1787/3adf705b-en
- OECD AI Policy Observatory. (2025). *AI incidents database*. https://oecd.ai/en/incidents
- OpenAI. (2025). Introducing ChatGPT agent: Bridging research and action. https://openai.com/index/introducing-chatgpt-agent/
- Oxford Insights. (2024). *Government AI Readiness Index 2024*. https://oxfordinsights.com/ai-readiness/ai-readiness-index/
- Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
- Shavit, Y., et al. (2023). Practices for governing agentic AI systems. *OpenAI White Paper*. https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf
- Slattery, P., Saeri, A. K., Grundy, E. A. C., Graham, J., Noetel, M., Uuk, R., Dao, J., Pour, S., Casper, S., & Thompson, N. (2024). The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks from Artificial Intelligence. https://doi.org/10.48550/arXiv.2408.12622
- The General-Purpose AI Code of practice. (2025). Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai

arte

AGÊNCIA PARA A REFORMA
TECNOLÓGICA DO ESTADO