

GuIA

Guia para a
Inteligência Artificial



Versão do documento: v.1.2.

Data: junho de 2025



SIGLAS E ACRÓNIMOS

AI-HLEG	<i>High-Level Expert Group on Artificial Intelligence</i>
AP	Administração Pública
AMA	Agência para a Modernização Administrativa
CE	Comissão Europeia
DL	<i>Deep Learning</i>
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 dos Bancos Centrais das 19 maiores economias do mundo e UE
IA	Inteligência Artificial
iAP	Interoperabilidade na Administração Pública
IoT	<i>Internet of Things</i>
ITS	<i>Intelligent Tutoring Systems</i>
I&D	Investigação e Desenvolvimento
LGBTQ+	Lésbicas, Gays, Bissexuais, Travestis, Transexuais, Transgéneros, Queer e outras entidades de género
LLM	<i>Large Language Model</i>
ML	<i>Machine Learning</i>
NLP	<i>Natural Language Processing</i>
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.....	6
1. Enquadramento	8
1.1. Âmbito e Estrutura do Guia	8
1.2. Glossário	9
1.3. O que é a IA?	10
1.4. Como funciona a IA?.....	11
1.5. IA na sociedade e as suas implicações.....	11
1.6. IA no mundo.....	15
1.7. Regulamento Europeu AI Act	17
1.8. Tendências Digitais Globais (OCDE) - Tendências digitais globais desafios emergentes.....	18
2. IA em Portugal	18
2.1. Estratégia Digital Nacional (EDN).....	19
2.2. Dimensões Estratégicas da EDN.....	20
2.3. Plano de Ação 2025–2026.....	20
2.4. Interoperabilidade e Dados Abertos.....	20
2.5. Ecossistemas e Atores	21
2.6. Ecossistema de dados na Génese da IA na AP	21
2.7. Inovação em Portugal.....	31
2.8. Impacto da IA no mercado de trabalho português.....	32
3. Riscos da Inteligência Artificial.....	33
3.1. Enquadramento	33
3.2. Níveis de risco no EU AI Act.....	33
3.2.1. Risco Inaceitável.....	34
3.2.2. Risco Elevado	35
3.2.3. Risco Limitado (Riscos de Transparência)	35
3.2.4. Risco Mínimo ou Nulo.....	35
3.3. Modelos de IA de finalidade geral	35
3.4. Riscos futuros: IA Agêntica (Agentic AI)	36



4. Inteligência Artificial Ética e Responsável	36
4.1. Enquadramento	36
4.2. Pilares.....	37
4.2.1 transparente.....	38
4.2.1.1. Explicabilidade	38
4.2.1.2. Governança.....	39
4.2.2. Robusta	41
4.2.2.1. Robustez	41
4.2.2.2.Privacidade e Proteção de Dados.....	42
4.2.2.3. Segurança	44
4.2.3. Universal	46
4.2.3.1. Equidade	46
4.2.3.2. Inclusividade	47
4.2.4. Sustentável.....	48
4.2.4.1. Minimização do Impacto Ambiental.....	49
4.2.4.2. Avaliação do Impacto Social	50
4.2.5. Testada	51
4.2.5.1. Adequação para IA.....	51
4.2.5.2. Testes pré-implementação	52
4.2.5.3. Monitorização contínua.....	53
5. Ferramenta de Avaliação de Riscos	55
5.1. Objetivos	55
5.2. Destinatários	55
5.3. Benefícios	56
5.4. Arquitetura	56
5.5. Utilização.....	57
5.6. Nível de Maturidade.....	57
5.7. Recomendações.....	58
5.8. Relatório de avaliação	59
5.9. Participação de partes interessadas	59
5.10. Programa de avaliação plurianual.....	59

ANEXO I | Questionário Guia..... Erro! Marcador não definido.



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 Modernização Administrativa (AMA) vem uma vez mais assumir o seu papel enquanto instituição pública responsável pela promoção e desenvolvimento da modernização administrativa em Portugal propondo, através do projeto IA Responsável, a exploração de caminhos para se conseguir uma IA que considere estas questões. 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 Pública (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 de 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 análise 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.

As referências bibliográficas que fundamentaram o desenvolvimento deste projeto, incluem as de autoria de Organizações Intergovernamentais, da Comissão Europeia (CE), do Setor Privado, de *Developers* e de Consultoras especializadas.



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.

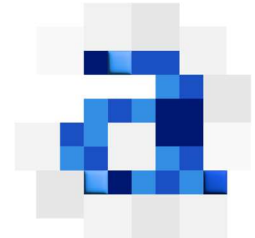
ALGORITMO: sequência lógica e finita de instruções que visam atingir um determinado propósito.

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



- Controlo de qualidade de alimentos e de medicamentos.

No Setor da **EDUCAÇÃO:**

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

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

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

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

- Rastreamento e previsão de padrões de poluição do ar por meio de sensores, para conseguir melhores medidas de intervenção na qualidade do ar;
- Identificação de pontos de vulnerabilidade nas reservas naturais, protegendo os 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;
- Detecção de fraude;
- *Credit scoring*;
- Algoritmos para *trading*;
- Processos automatizados de *pricing*;
- Controlo de práticas abusivas para o consumidor;
- Impedimento da prosperidade de monopólios, por exemplo, através da monitorização de transações;
- Otimização da experiência do utilizador, fornecendo sugestões personalizadas, navegação baseada em preferências e pesquisa de produtos suportada em imagens, para o setor de vendas a retalho. Ainda nesse setor, antecipação da procura pelos clientes, gestão de existências e de entregas;
- Automatização de processos na indústria, com impactos na engenharia, na cadeia de fornecimento, na gestão de produto, nos custos de produção, na manutenção, na garantia de qualidade e na logística e armazenamento.

1.6. IA no mundo

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

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

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*, a UE foi desenvolvida regulamentação que promove um 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) - Tendências digitais globais 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 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 monitorização 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](https://portal.dados.gov.pt) são instrumentos centrais para garantir a articulação ente 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).

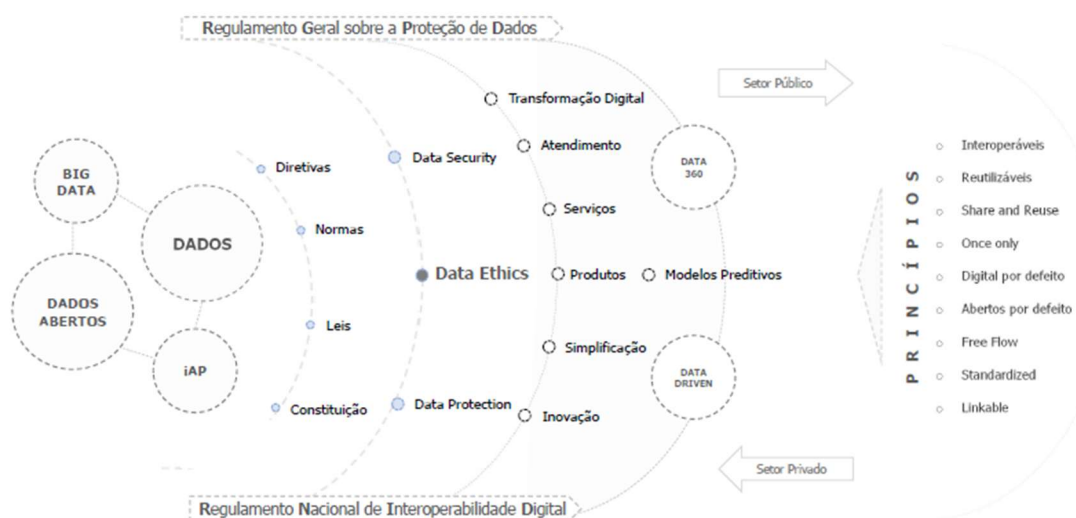


FIGURA 1: ECOSSISTEMA DE DADOS NA ADMINISTRAÇÃO PÚBLICA

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](https://iap.gov.pt), 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 promovam 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:

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



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

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

Referem-se ainda as seguintes normas e plano estratégico:

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

DATA 360

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



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

BIG DATA

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

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

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

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

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



FIGURA 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 perspectiva, 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 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 sociais;
- Apoiar decisões baseadas em dados;
- Promover sustentabilidade e inclusão.

Exemplos de inovação no sector público em Portugal nos últimos 2 anos:

- **Fábricas de IA da EuroHPC:** Portugal integra esta rede europeia de centros de excelência em IA, com foco em soluções industriais avançadas. A iniciativa é liderada pela Fundação para a Ciência e a Tecnologia (FCT), promovendo inovação em setores estratégicos;
- **LUMIS – Lisbon Unit for Learning and Intelligent Systems:** O Instituto Superior Técnico (IST) lidera este centro de investigação que desenvolve sistemas inteligentes com foco em aprendizagem automática e IA ética, reforçando a posição de Portugal na liderança europeia em IA;
- **HumanE-AI-Net:** O INESC TEC participa nesta rede europeia que trabalha no desenvolvimento de uma IA mais humanizada, ética e centrada no utilizador, com impacto direto em políticas públicas e serviços digitais inclusivos;
- **iCare4NextG:** Projeto que aplica IA em cuidados de saúde preventivos e assistência domiciliária, promovendo a personalização e eficiência dos serviços de saúde, especialmente para populações envelhecidas;
- **Bio-Waste2Carbon:** Iniciativa que transforma resíduos florestais em energia sustentável com apoio de IA, contribuindo para a transição energética e a economia circular.

Estes projetos demonstram como Portugal está a aplicar IA de forma estratégica e responsável, alinhando inovação tecnológica com sustentabilidade e bem-estar social.

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.

De acordo com o estudo “**A Inteligência Artificial está a gerar valor em todos os setores em Portugal**” da Microsoft, embora 54% das organizações do setor público em Portugal já tenham implementado soluções de Inteligência Artificial, apenas 5% afirmam estar a obter um valor significativo no serviço prestado aos cidadãos com essa adoção.

Isto demonstra que, apesar do progresso na implementação, a maturidade e o impacto efetivo da IA ainda são limitados, sendo necessário reforçar a capacidade de escalar projetos e gerar valor mensurável. O setor da saúde continua a destacar-se como um dos mais avançados na aplicação de IA, especialmente em áreas como diagnóstico assistido, gestão hospitalar e monitorização de dados clínicos.

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 EU AI Act, um risco é definido como a combinação da probabilidade de ocorrência de um dano com a severidade desse mesmo dano. 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.

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.

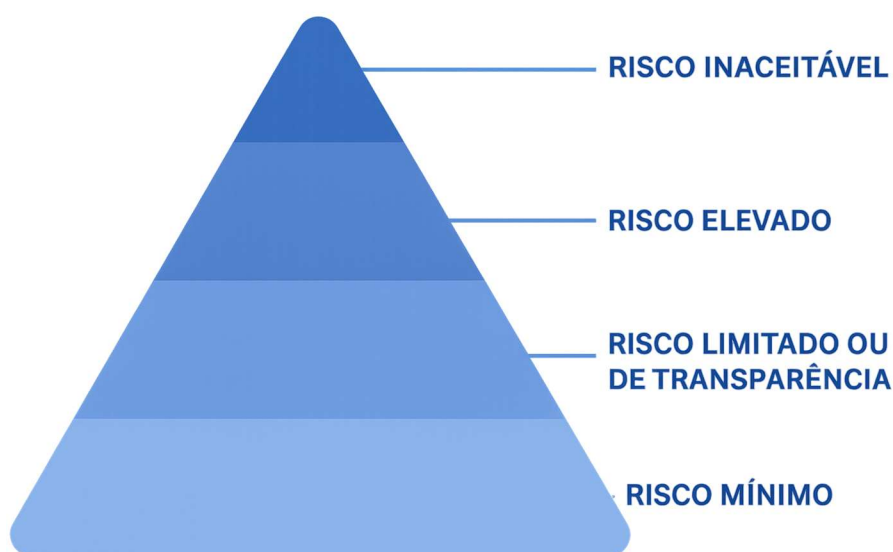
3.2. Níveis de risco no EU AI Act

O Regulamento Europeu da Inteligência Artificial (EU AI Act) estabelece um quadro legal harmonizado para garantir que os sistemas de IA desenvolvidos e utilizados na União Europeia sejam centrados no ser humano, fiáveis e respeitadores dos direitos fundamentais. Para isso, adota uma



abordagem baseada no risco, definindo-o como a combinação entre a probabilidade de ocorrência de danos e a sua gravidade. Esta definição abrange tanto danos materiais como imateriais, e considera o ciclo de vida completo do sistema, desde a instalação até à manutenção.

Com base nesta abordagem, o EU AI Act classifica os sistemas de IA em quatro níveis de risco: risco inaceitável (proibido), risco elevado (sujeito a requisitos rigorosos), risco limitado (com obrigações de transparência) e risco mínimo ou nulo (sem regulação específica). Esta estratificação permite adaptar as exigências legais à gravidade dos riscos associados, promovendo uma utilização segura e ética da IA em toda a União Europeia.



3.2.1. Risco Inaceitável

O EU AI Act proíbe um conjunto de práticas de Inteligência Artificial consideradas de risco inaceitável, por representarem uma ameaça direta à segurança, dignidade e direitos fundamentais dos cidadãos. Estas práticas incluem o uso de técnicas subliminares ou manipuladoras que distorcem o comportamento humano, a exploração de vulnerabilidades de pessoas em situação de fragilidade (como crianças ou pessoas com deficiência), e a classificação social (*social scoring*), que avalia indivíduos com base em comportamentos ou características pessoais, resultando em tratamentos discriminatórios.

Outras práticas proibidas incluem a avaliação preditiva de risco criminal baseada apenas em perfis, a criação de bases de dados de reconhecimento facial por recolha aleatória de imagens, o reconhecimento de emoções em contextos laborais ou educativos (exceto por razões médicas ou de segurança), a categorização biométrica sensível (como inferência de raça ou orientação sexual), e a identificação biométrica remota em tempo real em espaços públicos, salvo exceções muito restritas e autorizadas. Estas proibições refletem o compromisso da União Europeia com uma IA ética, segura e centrada no ser humano.



3.2.2. Risco Elevado

O EU AI Act proíbe práticas de Inteligência Artificial classificadas como de risco inaceitável, por representarem ameaças diretas à segurança, dignidade e direitos fundamentais. Entre estas estão técnicas manipuladoras, exploração de vulnerabilidades, classificação social, avaliação preditiva de crimes com base em perfis, e reconhecimento de emoções em contextos sensíveis.

Também são proibidas a recolha aleatória de imagens para bases de dados de reconhecimento facial, a categorização biométrica sensível e a identificação biométrica remota em tempo real, salvo exceções rigorosamente controladas. Estas proibições refletem o compromisso da UE com uma IA ética, segura e centrada nas pessoas.

3.2.3. Risco Limitado (Riscos de Transparência)

A categoria de risco limitado, definida pelo Artigo 50.º do EU AI Act, aplica-se a sistemas de IA cujo principal risco está na falta de transparência, e não em danos diretos à segurança ou aos direitos fundamentais. Nestes casos, o regulamento impõe obrigações específicas para garantir que os utilizadores estão cientes da presença de IA e podem tomar decisões informadas.

Entre os exemplos abrangidos estão: sistemas que interagem diretamente com pessoas (como *chatbots*), que devem informar claramente que são máquinas; sistemas que geram conteúdos sintéticos (áudio, imagem, vídeo ou texto), que devem ser identificáveis como artificiais, especialmente em casos de *deepfakes*; e sistemas de reconhecimento de emoções ou categorização biométrica, que exigem aviso prévio às pessoas expostas.

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 EU 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 EU AI Act introduz uma abordagem específica para os modelos de IA de finalidade geral, como os modelos generativos ou fundacionais, distinguindo-os dos sistemas de IA tradicionais. Enquanto os sistemas são aplicações finais com as quais os utilizadores interagem, os modelos são os motores subjacentes, treinados com grandes volumes de dados para executar múltiplas tarefas. Estes modelos



são concebidos para serem reutilizados em diferentes aplicações e não são regulados da mesma forma que os sistemas, excetuando quando são comercializados.

Dentro desta categoria, o regulamento identifica os modelos com **riscos sistêmicos**, ou seja, com potencial de causar impactos negativos em larga escala na saúde pública, segurança ou direitos fundamentais. Estes modelos são sujeitos a obrigações mais rigorosas, especialmente quando o seu treino ultrapassa 10^{25} FLOPs. A Comissão Europeia pode ainda classificá-los com base em critérios qualitativos. Para mitigar outros riscos associados, estão a ser desenvolvidos códigos de conduta sobre transparência, direitos de autor e segurança, complementados por frameworks como o MIT Risk Repository, que sistematizam os principais riscos destes modelos.

3.4. Riscos futuros: IA Agêntica (*Agentic AI*)

A "IA Agêntica" representa uma nova geração de sistemas de inteligência artificial que combinam modelos de linguagem, ferramentas externas e memória persistente, permitindo-lhes agir com autonomia e adaptar-se a contextos imprevistos. Estes sistemas são capazes de perseguir objetivos complexos com supervisão humana limitada, o que levanta novos desafios em termos de segurança, previsibilidade e controlo.

Entre os principais riscos estão: **falhas compostas**, resultantes da acumulação de pequenos erros que geram comportamentos imprevisíveis; **abuso de autonomia**, em que os sistemas tomam decisões incorretas ou contornam restrições de segurança; e **raciocínio enganador**, quando os sistemas apresentam justificações que não correspondem à lógica real usada, dificultando a supervisão e exigindo métodos avançados de explicabilidade.

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 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 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 rigorosos requisitos da legislação nacional e europeia, nomeadamente do EU AI Act, e com a necessária proteção dos direitos fundamentais dos cidadãos.

4.2. Pilares



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

1. **Transparente:** Tornar, por um lado, os resultados e decisões dos sistemas mais claras e perceptíveis. Por outro lado, tornar 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.



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 perceptíveis para todos. De acordo com o EU 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.

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.

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.

Perguntas Chave



1. Existem mecanismos claros para que um supervisor humano possa rever, corrigir ou anular as decisões do sistema?
2. O sistema inclui funcionalidades de paragem de emergência que podem ser acionadas por um humano de forma segura?

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 EU 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.

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?

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.



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?

Governança de Modelos

A Governança 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.

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 EU AI Act, para a categoria de risco do sistema?

Governança de Sistemas

A Governança 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.

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?

4.2.2. Robusta

O pilar da IA Robusta defende que os sistemas de Inteligência Artificial devem ser fiáveis, consistentes e resistentes a falhas e ameaças. A robustez *stricto sensu* assegura que o desempenho e a precisão do sistema se mantêm mesmo em condições adversas. Já a robustez *lato sensu* abrange aspetos como a privacidade e segurança, protegendo dados pessoais e prevenindo ciberataques ou manipulações externas. Esta robustez é fundamental para garantir a confiança na IA, especialmente em contextos críticos, e está em conformidade com o EU 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.

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?



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.

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.

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.

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 EU 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.

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.

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.

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 interrupção. 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.

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?

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.



Perguntas Chave

1. Existe um método de avaliação e mitigação os 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?

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.

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. Ir 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.

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.

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?

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.

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?

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.

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?

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.

Perguntas Chave

1. O sistema de IA foi testado para garantir que é acessível e utilizável por pessoas com diferentes capacidades, incluindo pessoas com deficiência?
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, contribuindo para a emissão de gases com efeito de estufa. 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.

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 de *hardware* e *software*? Existem critérios de sustentabilidade na seleção de fornecedores de *hardware* e *software*?



2. Existem políticas em vigor para a gestão responsável da água utilizada para arrefecer 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 centralizar o poder.

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 tornar-se mais monótonas e a ameaça de substituição pela IA pode criar dependências exploratórias, pressionando os trabalhadores a aceitar piores condições laborais e salariais.

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.



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?

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 artistas e podem desincentivar a inovação e criatividade humanas.

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 os criadores cujo trabalho contribui para o treino dos modelos e sistemas?

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 EU AI Act classifica como representativas de um risco inaceitável.

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 EU 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?

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.

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.

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 Monitorização

Após a implementação, é crucial monitorizar continuamente o 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.

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?



5. Ferramenta de Avaliação de Riscos

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:

- Conjunto de perguntas do tipo binário, *likert* ou escolha múltipla, validado anualmente;
- Ponderações atribuídas pela AMA e pelo utilizador a cada uma das cinco dimensões, validadas anualmente;



- Pontuação de avaliação;
- E matriz de recomendações associada ao nível de maturidade em que se encontra a entidade.

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 explícitas na figura 3.

Todas as perguntas são de resposta obrigatória.

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

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

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

FIGURA 3: ELEMENTOS DE REGISTO DO UTILIZADOR, ENTIDADE E PROJETO.

5.6. Nível de Maturidade

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

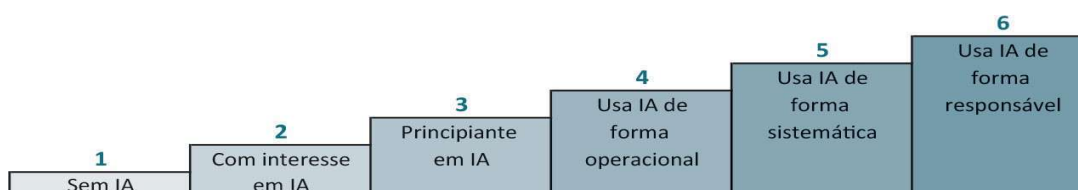


FIGURA 4: NÍVEIS DE MATURIDADE / ESTÁGIOS DE MATURIDADE



Este está relacionado com diferentes estágios de maturidade, ilustrados na figura 4:

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

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 planejar	Ótimo para a fase em que se encontra
Usa IA de forma operacional	Recomendamos planejar 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 5: 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.

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 AMA através do seguinte email: guia@ama.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:

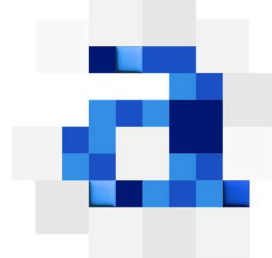
- Ser acionável: A avaliação deverá ir além do modelo auto-declarativo, integrando um processo multifacetado que inclua a sugestão e validação de testes técnicos, de validação por terceiros e de monitorização do sistema após a sua implementação.
- Ser flexível: Os riscos dos sistemas de IA não são estáticos. 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



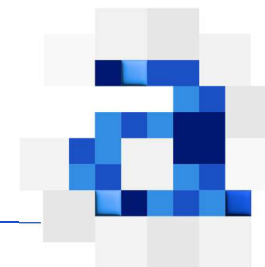
emitidas pela Comissão Europeia e pelo CEN-CENELEC, garantindo a relevância no longo prazo.

- 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.

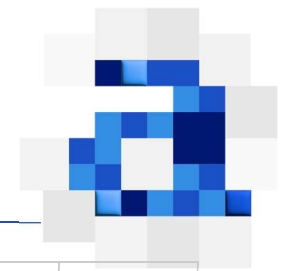
No período transitório, durante o qual a nova versão da ferramenta estará em desenvolvimento, será introduzido no processo de contratação pública o questionário presente no Anexo I. Deste modo, assegurar-se-á o mínimo alinhamento com os pilares, princípios e boas práticas de IA Responsável.



ANEXO I | Questionário GuIA

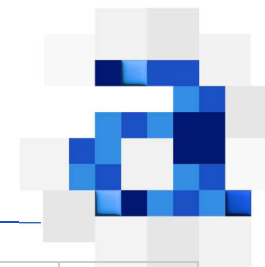


Transparente	Explicabilidade	Explicações	As decisões e previsões do sistema de IA são explicadas de forma clara, concisa e compreensível para o público-alvo?	☐
			É possível rastrear os principais fatores que contribuíram para um resultado específico?	☐
			As explicações fornecidas são suficientes para permitir a identificação e correção de eventuais erros ou enviesamentos?	☐
		Supervisão Humana	Existem mecanismos claros para que um supervisor humano possa rever, corrigir ou anular as decisões do sistema?	☐
			O sistema inclui funcionalidades de paragem de emergência que podem ser acionadas por um humano de forma segura?	☐
			Os supervisores recebem formação adequada sobre as capacidades, limitações e os riscos de excesso de confiança no sistema?	☐
		Sinalização da Presença de IA	A presença da IA é explicitamente comunicada aos utilizadores durante a interação?	☐
			O conteúdo gerado ou manipulado pelo sistema (texto, imagem, áudio ou vídeo) é devidamente assinalado como artificial?	☐
			Os mecanismos de identificação de conteúdo gerado por IA são robustos e fiáveis, de acordo com o estado da arte?	☐
	Governança	Governança de Dados	São mantidos registos detalhados sobre a origem, o tratamento e a validação dos dados para assegurar a sua total rastreabilidade e fiabilidade?	☐
			Existem políticas e procedimentos formais para garantir a qualidade, segurança e utilização ética dos dados, especialmente os de natureza sensível?	☐



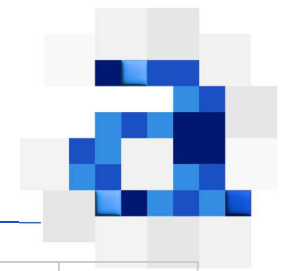
GUIA PARA A INTELIGÊNCIA ARTIFICIAL

			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?	☐
		Governança de Modelos	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?	☐
			A documentação técnica cumpre todos os requisitos exigidos pela legislação aplicável, como o EU AI Act, para a categoria de risco do sistema?	☐
			Existem processos para manter a documentação do modelo atualizada ao longo do seu ciclo de vida, refletindo todas as modificações substanciais?	☐
		Governança de Sistemas	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?	☐
			As políticas de governança abrangem todo o ciclo de vida do sistema, desde o desenvolvimento à descontinuação, assegurando uma gestão e monitorização contínuas?	☐
			A organização partilha de forma transparente as informações sobre como os sistemas de IA são implementados e geridos com as partes interessadas relevantes?	☐
Robusta	Robustez	Resiliência	O sistema está preparado para detetar e lidar com o desvio de dados (<i>data drift</i>), anomalias e atividade maliciosa?	☐
			Existe um processo estruturado para monitorizar e corrigir a degradação do desempenho?	☐
			São realizados testes de esforço e simulações “adversariais” antes da implementação?	☐
		Estabilidade	Os resultados são estáveis e previsíveis ao longo do tempo?	☐
		Resultados Previsíveis	A consistência do sistema foi validada para assegurar que dados de entrada semelhantes produzem, de forma fiável, resultados semelhantes?	☐



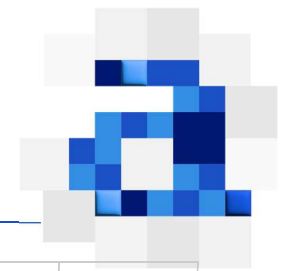
GUIA PARA A INTELIGÊNCIA ARTIFICIAL

	Privacidade e Proteção de Dados	Privacidade de Dados	O sistema de IA cumpre com os regulamentos de privacidade de dados aplicáveis, como o RGPD e outras políticas específicas do setor?	☐
			Os dados sensíveis estão devidamente protegidos através de cifragem e controlos de acesso rigorosos?	☐
			Foram implementadas medidas de minimização de dados para garantir que apenas os dados estritamente necessários são recolhidos e processados?	☐
		Consentimento	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?	☐
			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 dos Dados	Existem processos para verificar a exatidão, a proveniência e a representatividade dos dados utilizados para treinar e validar os sistemas de IA?	☐
	Os conjuntos de dados estão protegidos contra manipulação ou corrupção, tanto por agentes internos como externos?		☐	
	Existem mecanismos de auditoria para validar continuamente a integridade dos dados ao longo do ciclo de vida do sistema de IA?		☐	
	Segurança	Cibersegurança	São realizadas auditorias de segurança e testes de penetração regularmente?	☐
Existem controlos de acesso e encriptação para proteger os dados sensíveis utilizados pelo sistema?			☐	
Estão implementadas medidas para prevenir ataques “adversariais”, como <i>data poisoning</i> ou <i>modelo evasion</i> ?			☐	



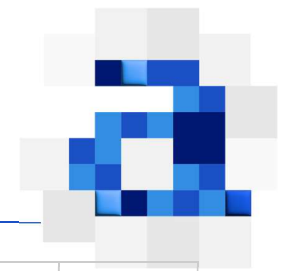
GUIA PARA A INTELIGÊNCIA ARTIFICIAL

Universal	Equidade	Fiabilidade do Sistema	O sistema é continuamente monitorizado para detetar ameaças à segurança que possam comprometer a sua fiabilidade?	☐
			Existem mecanismos de <i>failover</i> e planos de contingência para garantir o funcionamento do sistema em caso de falha ou ataque?	☐
			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	Existe um método de avaliação e mitigação os riscos de utilização indevida do sistema de IA, tanto deliberada como inadvertida?	☐
			Estão implementadas salvaguardas para garantir que o modelo não é facilmente manipulado para gerar resultados ou comportamentos perigosos?	☐
			O treino de segurança do modelo é suficientemente robusto para resistir a técnicas de <i>jailbreaking</i> ou outras formas de manipulação?	☐
Universal	Equidade	Resultados Justos	Os objetivos de equidade estão claramente definidos e integrados no desenvolvimento do modelo desde a sua conceção?	☐
			São utilizadas métricas e <i>benchmarks</i> de equidade para monitorizar e corrigir desequilíbrios nos resultados do sistema de IA ao longo do seu ciclo de vida?	☐
			São identificados e mitigados os vieses nos conjuntos de dados utilizados, em conformidade com as exigências do EU AI Act?	☐
		Auditorias Regulares	São realizadas auditorias de viés regulares para avaliar a equidade do sistema de IA em produção?	☐
			Estão em vigor procedimentos para corrigir as disparidades e os vieses identificados durante as auditorias?	☐



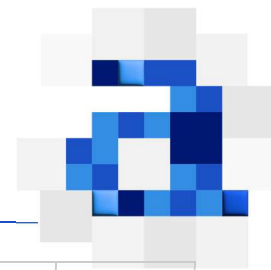
GUIA PARA A INTELIGÊNCIA ARTIFICIAL

	Inclusividade		O sistema de monitorização pós-comercialização, exigido pelo EU AI Act para sistemas de risco elevado, é utilizado para detetar e mitigar vieses que possam surgir após a implementação?	☐
		Desenhado por Todos	A equipa de desenvolvimento reflete a diversidade (de género, cultural, de formação, etc.) da sociedade que o sistema de IA pretende servir?	☐
			Foram envolvidas partes interessadas externas (<i>stakeholders</i>), como organizações da sociedade civil e representantes dos utilizadores, no processo de conceção e desenvolvimento do sistema?	☐
			Existem mecanismos para recolher e integrar o feedback de diversos grupos durante todo o ciclo de vida da IA?	☐
		Desenhado para Todos	O sistema de IA foi testado para garantir que é acessível e utilizável por pessoas com diferentes capacidades, incluindo pessoas com deficiência?	☐
			Foram realizadas avaliações para detetar e corrigir disparidades de desempenho ou usabilidade entre diferentes grupos demográficos, culturais ou linguísticos?	☐
			O sistema foi concebido para funcionar de forma equitativa em diversos contextos socioeconómicos e linguísticos?	☐
Sustentável	Minimização do Impacto Ambiental	Eficiência Energética	A organização mede e documenta o consumo energético dos seus sistemas de IA, tanto na fase de treino como na de inferência?	☐
			A escolha de modelos de IA de terceiros inclui uma avaliação da sua pegada energética como critério de seleção?	☐
			A organização comunica de forma transparente o impacto energético dos seus sistemas de IA aos seus clientes e ao público?	☐
			Existem critérios de sustentabilidade na seleção de fornecedores de <i>hardware</i> e <i>software</i> ?	☐



GUIA PARA A INTELIGÊNCIA ARTIFICIAL

	Avaliação do Impacto Social	Uso Eficiente de Recursos Naturais	Existem políticas em vigor para a gestão responsável da água utilizada para arrefecer e de resíduos eletrónicos gerados pelos sistemas?	☐
		Avaliação do Impacto Laboral	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?	☐
			Estão a ser implementadas medidas para a requalificação e formação dos trabalhadores cujas funções possam ser afetadas pela automação?	☐
		Avaliação do Impacto Económico	A estratégia de IA da organização contribui para um ecossistema económico mais justo e competitivo?	☐
			Existem medidas de governação interna para evitar que a pressão competitiva comprometa os padrões de segurança e ética?	☐
			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 organização garante que a utilização de dados para treinar os seus modelos de IA respeita os direitos de autor e a propriedade intelectual?	☐
			Existe compensação justa os criadores cujo trabalho contribui para o treino dos modelos e sistemas?	☐
			Existem políticas internas para garantir que os sistemas de IA promovem a diversidade cultural em vez da homogeneização?	☐
		Validação	O caso de uso viola alguma das práticas de IA proibidas pelo EU AI Act, representando um risco inaceitável para os direitos fundamentais?	☐
Testada	Adequação para IA		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?	☐



	Testes pré-implementação		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?	☐
		Testes Independentes	A organização garante que as avaliações internas são complementadas por validações independentes para assegurar a conformidade e a adesão às melhores práticas?	☐
		Benchmarking	São aplicados <i>benchmarks</i> de explicabilidade, robustez, segurança, equidade e eficiência para comparar o sistema com os padrões da indústria e identificar lacunas de desempenho?	☐
		Testes de Integração	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?	☐
	Monitorização contínua	Testagem e Monitorização	Estão implementados mecanismos para monitorizar continuamente o desempenho do sistema de IA em produção e detetar desvios ou degradação do modelo?	☐
			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?	☐
			Em caso de sistemas de risco elevado, o plano de monitorização pós-implementação cumpre as exigências regulatórias definidas no EU AI Act, incluindo a documentação e o reporte de incidentes graves?	☐

ama