ESCOLA
SUPERIOR
DE TECNOLOGIA
E GESTÃO
POLITÉCNICO
DO PORTO

MESTRADO
Praticas Jurídico-Digitais

Implicações Jurídicas do Regulamento da Inteligência Artificial para o Comércio Digital na União Europeia: Uma Perspetiva Informática

Mariana Isabel Serra Massa

10/**2025**

ESCOLA
SUPERIOR
DE TECNOLOGIA
E GESTÃO
POLITÉCNICO
DO PORTO

MESTRADO
Práticas Jurídico-Digitais

Implicações Jurídicas do Regulamento da Inteligência Artificial para o Comércio Digital na União Europeia: Uma Perspetiva Informática

Mariana Isabel Serra Massa
8190268

Orientadores

Doutora Patrícia dos Anjos Azevedo
Doutor Ricardo Jorge da Silva Santos

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção o do grau de Mestre em Práticas Jurídico-Digitais pela Escola Superior de Tecnologia e Gestão do Instituto Politécnico do Porto.

10/**2025**

Declaração de Integridade

Eu, Mariana Isabel Serra Massa, estudante nº 8190268, do Mestrado de Práticas Jurídico-Digitais da Escola Superior de Tecnologia e Gestão do Instituto Politécnico do Porto, declaro que não fiz plágio nem auto-plágio, pelo que o trabalho intitulado "Implicações jurídicas do Regulamento da Inteligência Artificial para o Comércio Digital na União Europeia: Uma Perspetiva Informática" é original e da minha autoria, não tendo sido usado previamente para qualquer outro fim.

Mais declaro que todas as fontes usadas estão citadas, no texto e na bibliografia final, segundo as regras de referenciação adotadas na instituição.

"O país onde o comércio é mais livre será sempre o mais rico e próspero, guardadas as proporções".

Voltaire

Agradecimentos

Concluída mais uma etapa do meu percurso académico, não poderia deixar de expressar a minha sincera gratidão a todos aqueles que, de diferentes formas, contribuíram para que fosse possível alcançar este objetivo.

Primeiramente, agradeço aos meus pais, ao meu irmão e ao meu namorado, por todo o amor e apoio incondicional, não apenas nesta etapa da minha vida, mas em todos os momentos em que precisei de confiança, incentivo e carinho.

Aos meus orientadores, Doutora Patrícia Anjos Azevedo e Doutor Ricardo Jorge da Silva Santos, agradeço pelos valiosos ensinamentos, pela constante disponibilidade e pela dedicação demonstrada ao longo da elaboração deste projeto.

Por fim, e de forma igualmente importante, agradeço às minhas amigas e a todos os colegas de mestrado por todas as palavras de incentivo e pelo apoio continuo.

A todos, o meu muito obrigada!

Resumo

O Regulamento 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024 (*i.e.* Regulamento da Inteligência Artificial) entrou em vigor a 1 de agosto de 2024, prevendo-se a sua aplicação integral a partir de 2 de agosto de 2026. No entanto, certas disposições específicas como as proibições e as obrigações relativas à literacia em IA tornaram-se aplicáveis mais cedo, isto é, a 2 de fevereiro de 2025.

O aludido Regulamento da Inteligência Artificial da União Europeia vem estabelecer um regime jurídico harmonizado para o desenvolvimento, comercialização e utilização de sistemas de IA no mercado interno europeu.

A principal finalidade do Regulamento é a de garantir que a utilização da inteligência artificial respeita os direitos fundamentais, a segurança e os valores da União, promovendo simultaneamente a inovação tecnológica responsável e os perigos inerentes, designadamente a opacidade dos algoritmos, a discriminação automatizada e o impacto sobre os direitos dos consumidores e trabalhadores. O Regulamento tem ainda na sua génese a ambição estratégica da UE de estabelecer um enquadramento normativo que reforce a confiança digital e a soberania tecnológica europeia.

Palavras-chave inteligência artificial; regulamento, União Europeia; informática; algoritmos

Abstract

Regulation 2024/1689 of the European Parliament and of the Council of 13 June 2024 (i.e. Artificial Intelligence Act) entered into force on 1 August 2024 and is expected to apply in full of 2 August 2026. However, certain specific provisions such as the prohibitions and obligations on AI literacy became applicable earlier, on 2 February 2025. The European Union Artificial Intelligence Regulation establishes a harmonised legal framework for the development, commercialisation and use of AI systems in the European internal market.

The main purpose of the Regulation is to ensure that the use of artificial intelligence respects the fundamental rights, safety and values of the Union, while promoting responsible technological innovation and the inherent dangers, including the opacity of algorithms, automated discrimination and the impact on consumer and workers' rights. It also stems from the EU's strategic ambition to establish a regulatory framework that strengthens digital trust and consolidates European technological sovereignty.

Keywords *artificial intelligence; regulation, European Union; computer science; algorithms*

Todas as referências a preceitos legislativos
sem indicação do texto legal
devem entender-se feitas para o Regulamento 2024/1689 do Parlamento Europeu e
do Conselho, de 13 de junho de 2024,
que cria regras harmonizadas em matéria de inteligência artificial.

Índice

Índice de Figuras

Siglas e Abreviaturas

AIA	*AI Act*
AIMS	*AI Management System*
AI RMF	*Artificial Intelligence Risk Management Framework*
al.	alínea
ALTAI	*Assessment List for Trustworthy Artificial Intelligence*
art.	artigo
cit.	citado
cfr.	confrontar
disp.	disponível em
ed.	edição
e.g.	por exemplo
etc	entre outros
GPAI	*Global Partnership on Artifical Intelligence*
HLEG	*High-Level Expert Group*
IA	Inteligência Artificial
id.	idem
i.e.	isto é
IMDA	*Infocomm Media Devepment*
LIME	*Local Interpretable Model-agnostic Explanations*
ML	*Machine Learning*
n.º	número
NIST	*National Institute of Standards and Technology*
p.	página
pp.	páginas
PDCA	*Plan, Do, Check, Act*
PDPC	*Personal Data Protection Commission*
RGPD	Regulamento Geral sobre a Proteção de Dados
SHAP	*Shapley Additive exPlanation*
ss.	seguintes
UE	União Europeia
vol.	Volume

Introdução

De acordo com HAENLEIN e KAPLAN, a inteligência artificial consiste na capacidade de um sistema interpretar dados externos corretamente, aprender com eles e utilizar essa aprendizagem para atingir objetivos e tarefas específicas através de uma adaptação flexível[1]. Por seu turno, FABIO MORANDÍN-AHUERMA define a IA como sendo a capacidade de uma máquina ou sistema informático simular e executar tarefas que normalmente requerem inteligência humana, como o raciocínio lógico, a aprendizagem e a resolução de problemas[2].

Não obstante a maioria das pessoas não ter presente uma definição formalística de inteligência artificial, a verdade é que quase ninguém hoje fica indiferente à mesma. De facto, a rápida evolução da inteligência artificial tem transformado profundamente o modo como indivíduos, empresas e instituições interagem no espaço digital, especialmente no domínio do comércio eletrónico.

No entanto, esta transformação levanta questões significativas do ponto de vista jurídico, ético e técnico, pelo que, em resposta a estes desafios, a União Europeia aprovou, em 2024, o Regulamento 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024[3], que cria regras harmonizadas em matéria de inteligência artificial e que altera os Regulamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e as Diretivas 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (*i.e.* Regulamento da Inteligência Artificial), sendo este o primeiro instrumento legislativo de grande escala destinado a disciplinar o uso da IA com base numa abordagem centrada no risco e na proteção dos direitos fundamentais.

Conforme nos refere GAON, por estarmos perante uma novidade (*i.e.* pelo menos na escala atual) e com impactos brutais no comércio e no dia-a-dia das pessoas, a IA carece de regulamentação ajustada[4].

O Regulamento da IA surge, assim, no contexto de uma estratégia europeia mais ampla para a transição digital e para a consolidação da soberania tecnológica da União, articulando

[1] Cfr. HAENLEIN, M. e KAPLAN, A. (2019) – *A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence*. California Management Review, 61, pp. 14 a 15.

[2] Cfr. MORANDÍN-AHUERMA, Fabio (2022) – *What is Artificial Intelligence?* International Journal of Research Publication and Reviews, p. 29.

[3] Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32024R1689.

[4] Cfr. GAON, A. (2021) – *Reframing the artificial intelligence concept.* The Future of Copyright in the Age of Artificial Intelligence.

preocupações relacionadas com a segurança, a transparência, a não discriminação, a responsabilidade e a explicabilidade dos sistemas automatizados. Ora, tendo sido inspirado por princípios éticos delineados em documentos como as Orientações éticas para uma IA de confiança (2019)[5] e pelo Livro Branco sobre a IA (2020)[6], o regulamento consubstancia uma tentativa clara de equilibrar a promoção da inovação com a salvaguarda do interesse público e da confiança digital.

No comércio digital europeu, onde a IA é frequentemente empregue em processos de recomendação, personalização de conteúdos, gestão automatizada de contratos e decisões algorítmicas com impacto direto sobre consumidores e fornecedores, é fundamental que existam normas claras e transparentes que a regulem.

Isto posto, no presente estudo propomo-nos analisar, numa perspetiva informática e jurídica, as principais implicações da entrada em vigor do Regulamento da IA para o comércio digital na União Europeia. Neste sentido, serão abordadas as exigências técnicas e legais impostas pelo novo regulamento, os seus impactos práticos sobre os agentes económicos, bem como os desafios que a sua implementação poderá representar em termos de conformidade, interoperabilidade e fiscalização. O nosso estudo parte da premissa de que uma compreensão integrada entre o Direito e as Ciências da Computação é indispensável para uma aplicação eficaz e justa das novas normas no ecossistema digital europeu.

[5] Cfr. https://op.europa.eu/pt/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1.

[6] *Id.*

1 Evolução do Regulamento e enquadramento histórico

Segundo STUART RUSSELL e PETER NORVIG, a inteligência artificial é a área da ciência e da engenharia que se dedica a criar agentes racionais, isto é, qualquer entidade que consegue perceber o seu ambiente e agir de forma a maximizar as probabilidades de alcançar os seus objetivos[7]. Por seu turno, WOLFHART TOTSCHNIG defende que, embora não exista uma definição única para o conceito de inteligência artificial, podemos dizer que, na prática, trata-se de um sistema que integra componentes ciberfísicos e *software* de forma autónoma e auto-organizada[8].

Por outras palavras, a inteligência artificial é capaz de funcionar de forma independente, sem necessidade de intervenção humana contínua.

Destarte, estamos perante uma verdadeira revolução na forma como interagimos com as máquinas. É certo que a inteligência artificial não surgiu agora, mas a mesma nunca tinha estado ao nosso alcance de forma tão massiva e com tantos impactos no nosso dia-a-dia.

Neste sentido, o Regulamento de Inteligência Artificial da União Europeia emergiu num contexto político e jurídico de crescente preocupação com o impacto das novas tecnologias no tecido social, económico e jurídico da Europa.

A União Europeia é uma entidade supranacional que, ao longo dos anos, tem procurado afirmar-se como uma potência normativa nos mais diversos campos, não sendo o domínio digital exceção. Assim, a UE pretendeu adotar uma abordagem regulatória que visa equilibrar a promoção da inovação com a salvaguarda dos direitos fundamentais. Este regulamento que surge como o primeiro do seu género a nível mundial, constitui uma peça central da estratégia digital europeia, tendo por objetivo criar um quadro jurídico harmonizado que permita um desenvolvimento e utilização responsáveis e seguros da Inteligência Artificial no mercado interno da União Europeia.

O contexto político que conduz à elaboração deste regulamento é marcado por uma visão estratégica que procura promover uma IA centrada no ser humano[9].

---

[7] Cfr. RUSSELL, Stuart e NORVIG, Peter (2022) – *Inteligência artificial: um enfoque moderno*. Rio de Janeiro: Elsevier, p. 5.

[8] Cfr. TOTSCHNIG, Wolfhart (2020) – *Fully Autonomous AI. Science and Engineering Ethics*, 26, pp. 2473 a 2485.

[9] Neste sentido, *vide* a Comunicação da Comissão ao Parlamento Europeu, ao Conselho, ao Comité Económico e Social Europeu e ao Comité das Regiões, 8.4.2019 COM(2019) 168 final, Disp. in https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:52019DC0168&from=EN.

É, aliás, curioso que o tema da inteligência artificial só tenha tido uma verdadeira explosão a partir do ano de 2023. Porém, a verdade é que o seu desenvolvimento e história são bem mais longínquos.

Neste sentido, ao longo do tempo, a União Europeia tem procurado prevenir-se para os seus riscos e, ao mesmo tempo, beneficiar das suas virtudes.

Assim, logo em 2018, a Comissão Europeia publicou a Comunicação "Inteligência Artificial para a Europa"[10], comunicação essa que lançou as bases para uma estratégia europeia neste domínio e no mesmo ano surgiu também o Plano Coordenado sobre a Inteligência Artificial.

Em 2019, o Grupo de Peritos de Alto Nível em Inteligência Artificial (*i.e. HLEG*) publicou as Diretrizes Éticas para uma IA de Confiança, com vista a garantir que a IA respeita os direitos fundamentais e valores europeus[11].

Um ano depois, em 2020, lançado o Livro Branco sobre a Inteligência Artificial, o qual cria uma abordagem europeia para a excelência e a confiança, tendo proposto medidas regulatórias e políticas para equilibrar inovação e proteção dos direitos dos cidadãos[12].

Em 2021 foi adotada a Comunicação "Promover uma abordagem europeia para a Inteligência Artificial" (*i.e.* COM/2021/205), que acompanhou a proposta de regulamento e a atualização do Plano Coordenado[13].

No ano de 2022, a Comissão apresentou propostas relativas à responsabilidade civil aplicável à IA[14].

Já em 2024 foi então aprovado o Regulamento (UE) 2024/1689 relativo à Inteligência Artificial (*i.e. AI Act*), o primeiro regulamento abrangente do mundo sobre IA, aplicável de forma faseada em função do risco da tecnologia[15].

Desta forma, verificamos que o contexto político foi marcado por uma visão estratégica, que visa garantir uma IA segura, ética, transparente e respeitadora dos direitos fundamentais, sendo que esta visão surge num momento em que a inteligência artificial se destaca como

---

[10] Cfr. https://digital-strategy.ec.europa.eu/en/library/communication-artificial-intelligence-europe.

[11] Cfr. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[12] Cfr. https://commission.europa.eu/document/download/d2ec4039-c5be-423a-81ef-b9e44e79825b_pt?filename=commission white-paper-artificial-intelligence-feb2020_pt.pdf

[13] Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A52021DC0205.

[14] Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:52022PC0496&from=EN.

[15] Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A32024R1689.

um dos motores principais da inovação tecnológica e transformação social, trazendo benefícios económicos e sociais brutais, mas também comportando riscos para a segurança, privacidade, a igualdade de tratamento e outros direitos fundamentais consagrados na Carta dos Direitos Fundamentais da União Europeia[16].

Ao criar um quadro jurídico harmonizado para todos os Estados-Membros, a UE procura estabelecer regras claras e uniformes para o desenvolvimento, comercialização e utilização dos sistemas de IA, com especial ênfase na proteção dos cidadãos contra usos abusivos e nocivos desta tecnologia.

[16] Cfr. *https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:12016P/TXT*.

2 O Regulamento 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024, cria regras harmonizadas em matéria de IA

2.1 Objetivos do Regulamento

Os objetivos do chamado *AI Act* da União Europeia vêm plasmados, desde logo, no parágrafo primeiro do Regulamento 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024.

Ora, de acordo com o referido parágrafo, verificamos que o regulamento tem por finalidade estabelecer um regime jurídico uniforme no espaço da União Europeia relativamente aos sistemas de inteligência artificial, de modo a reforçar o bom funcionamento do mercado interno e a evitar disparidades regulatórias entre os diferentes Estados-Membros[17].

Neste sentido, o seu foco central do aludido regulamento reside em assegurar que o desenvolvimento, a comercialização, a entrada em funcionamento e a utilização dos sistemas de IA ocorram em conformidade com os valores fundamentais da União, promovendo uma inteligência artificial de confiança e orientada para o ser humano. Ao mesmo tempo, o regulamento visa garantir a proteção da saúde, da segurança e dos direitos fundamentais consagrados na Carta dos Direitos Fundamentais da União Europeia, salvaguardando bens jurídicos como a democracia, o Estado de direito e a preservação do ambiente, bem como prevenindo consequências nocivas associadas ao uso da tecnologia.

Paralelamente, o regulamento visa fomentar a inovação e consolidar a livre circulação de bens e serviços relacionados com a inteligência artificial no mercado interno, proibindo os Estados-Membros de imporem restrições adicionais à sua introdução ou utilização, salvo quando previstas pelo próprio quadro normativo europeu.

A acrescer ao exposto, o parágrafo 6.º vem ainda salientar que, dada a relevância e o impacto da inteligência artificial na sociedade, é essencial que o seu desenvolvimento e regulamentação respeitem os valores da União e os direitos fundamentais previstos nos Tratados e na Carta, assegurando que se trate de uma tecnologia centrada no ser humano e orientada para a promoção do bem-estar coletivo[18].

Para tal, de acordo com o parágrafo 8.º, a UE considera que é necessário criar um quadro jurídico europeu que harmonize as regras aplicáveis à inteligência artificial, assegurando tanto

[17] Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689.

[18] Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689.

o seu desenvolvimento, utilização e integração no mercado interno, como a proteção de interesses públicos fundamentais, entre os quais se destacam a saúde, a segurança, os direitos fundamentais, a democracia, o Estado de direito e o ambiente. Para tal, impõe-se regular a colocação no mercado e a utilização de determinados sistemas de IA, garantindo simultaneamente a livre circulação de produtos e serviços e um elevado nível de confiança.

Desta forma, o *AI Act* deve apoiar a inovação, em particular das pequenas e médias empresas e das *startups*, promovendo um ecossistema europeu alinhado com os valores da União e capaz de posicionar a Europa na liderança global de uma inteligência artificial ética, segura, centrada no ser humano e promotora de bem-estar coletivo.

Neste sentido, o art.º 1, n.º 1 do Regulamento concretiza o objetivo central do mesmo é reforçar o mercado interno por via da promoção de uma inteligência artificial fiável e centrada no ser humano, garantindo a proteção da saúde, da segurança, dos direitos fundamentais, da democracia, do Estado de direito e do ambiente, ao mesmo tempo que se fomentam a inovação e a prevenção de impactos negativos decorrentes da utilização dos sistemas de IA.

A acrescer ao exposto, o n.º 2 do mesmo artigo salienta que o regulamento prevê a aplicação de regras harmonizadas para a colocação no mercado, a entrada em funcionamento e a utilização de sistemas de inteligência artificial na União (*e.g.* alínea a), determina igualmente a proibição de certas práticas de IA consideradas incompatíveis com os valores da União (*e.g.* alínea b) e estabelece requisitos específicos para sistemas de alto risco, bem como obrigações para os respetivos operadores (*e.g.* alínea c). Por outro lado, o mesmo define regras de transparência aplicáveis a determinados sistemas de IA (*e.g.* alínea d), consagra também normas harmonizadas para a colocação no mercado de modelos de finalidade geral (*e.g.* alínea e), inclui disposições sobre o acompanhamento e a fiscalização do mercado, a governação e a aplicação da lei (*e.g.* alínea f) e prevê medidas de apoio à inovação, com especial foco nas pequenas e médias empresas, incluindo aquelas em fase inicial.

2.2 Âmbito de aplicação do Regulamento

2.2.1 Casos de aplicação positiva

Com vista a concretizar os objetivos atrás enunciados, o parágrafo 9.º do Regulamento salienta que o mesmo vem estabelecer regras harmonizadas sobre a colocação no mercado, a colocação em serviço e a utilização de sistemas de inteligência artificial de alto risco, em articulação com o regime jurídico europeu existente.

Ora, ainda de acordo com o referido parágrafo, estas regras aplicam-se transversalmente a todos os setores, sem prejuízo da legislação já em vigor da União, designadamente em matéria de proteção de dados, defesa do consumidor, segurança no trabalho, relações laborais e outros direitos fundamentais, que permanecem intocados e plenamente aplicáveis. Assim, garantem-se os direitos e vias de recurso previstos na ordem jurídica da União, incluindo a indemnização por eventuais danos, bem como a proteção de matérias próprias da legislação laboral nacional, desde que conforme ao direito da União. Ademais, o regulamento reforça a eficácia desses direitos ao introduzir obrigações específicas de transparência, documentação e manutenção de registos aplicáveis a todos os intervenientes na cadeia de valor da IA, assegurando simultaneamente que normas nacionais com objetivos legítimos de interesse público, como a proteção dos trabalhadores ou dos menores, continuem a ser respeitadas[19].

Concretizando o referido âmbito de aplicação, o art.º 2.º, n.º 1 do Regulamento dispõe que o mesmo se aplica aos prestadores que disponibilizem no mercado ou em serviço sistemas de inteligência artificial, bem como modelos de finalidade geral, no território da União, independentemente de estarem estabelecidos na União ou num país terceiro (*e.g.* alínea a). Abrange igualmente os responsáveis pela implantação de sistemas de IA que tenham estabelecimento ou estejam localizados na União (*e.g.* alínea b), assim como aqueles situados fora da União, quando os resultados dos sistemas por si operados sejam utilizados no espaço europeu (*e.g.* alínea c).

Inclui ainda os importadores e distribuidores de sistemas de IA (*e.g.* alínea d), os fabricantes que coloquem no mercado ou em serviço sistemas integrados nos seus produtos sob a sua própria marca (*e.g.* alínea e), bem como os mandatários que atuem em nome de prestadores não estabelecidos na União (*e.g.* alínea f). Finalmente, o regulamento é aplicável às pessoas localizadas na União que sejam afetadas pela utilização de sistemas de IA (*e.g.* alínea g).

2.2.2 Situações de exclusão de aplicação

Em sentido inverso ao atrás referido, o art.º 2.º é bastante exaustivo no que concerne aos casos de exclusão de aplicação do presente regulamento.

[19] Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689.

Assim, o regulamento exclui do seu âmbito de aplicação os domínios não abrangidos pelo direito da União e todas as matérias relativas à segurança nacional, não afetando as competências dos Estados-Membros nessa área. Por outro lado, o mesmo não se aplica a sistemas de IA utilizados exclusivamente para fins militares, de defesa ou de segurança nacional, independentemente da entidade envolvida, incluindo situações em que os resultados sejam utilizados na União com essas finalidades (*e.g.* n.º 3). O regulamento também não se aplica a autoridades públicas de países terceiros ou a organizações internacionais quando utilizem sistemas de IA em contextos de cooperação policial ou judiciária internacional com a União, desde que existam garantias adequadas de proteção dos direitos fundamentais (*e.g.* n.º 4)[20].

Não obstante a aprovação do regulamento, mantém-se intocada a responsabilidade dos prestadores de serviços intermediários prevista no Regulamento (UE) 2022/2065[21] (*e.g.* n.º 5).

Ademais, o regulamento não abrange sistemas ou modelos de IA desenvolvidos exclusivamente para fins de investigação e desenvolvimento científicos, bem como as atividades de investigação e teste anteriores à entrada no mercado ou colocação em serviço, salvo quando envolvam testagem em condições reais (*e.g.* n.ºs 6 e 8). Isto é, sem prejuízo das regras da União já existentes em matéria de proteção dos consumidores e de segurança dos produtos (*e.g.* n.º 9), não se aplicando aos usos pessoais e não profissionais de sistemas de IA por pessoas singulares (*e.g.* n.º 10). Importa sublinhar que o regulamento não impede a adoção, pelos Estados-Membros ou pela União, de normas mais favoráveis de proteção dos trabalhadores relativamente à utilização da IA em contexto laboral, incluindo através de convenções coletivas (*e.g.* n.º 11)[22].

Por fim, o regulamento também não se aplica a sistemas de IA disponibilizados sob licenças gratuitas e de código aberto, salvo quando sejam sistemas de alto risco ou abrangidos por disposições específicas dos artigos 5.º ou 50.º do mesmo regulamento (*e.g.* n.º 12).

[20] Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689.

[21] O chamado Regulamento dos Serviços Digitais, Regulamento (UE) 2022/2065 do Parlamento Europeu e do Conselho, de 19 de outubro de 2022, Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32022R2065.

[22] Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689.

2.3 Esclarecimento conceptual

No que concerne aos conceitos inerentes à aplicação do Regulamento 2024/1689 do Parlamento Europeu e do Conselho de 13 de junho de 2024, o artigo 3.º é extremamente exaustivo.

Assim, para efeitos do regulamento, e salientando-se aqui as definições mais relevantes para o presente estudo, verificamos que, nos termos do disposto no art.º 3.º, entende-se por (1) sistema de IA qualquer tecnologia autónoma e adaptável que, a partir de dados, gera previsões, recomendações ou decisões com impacto em ambientes físicos ou virtuais e por (2) risco, a combinação da probabilidade de ocorrência de um dano com a gravidade desse dano.

Já o conceito de (3) prestador é definido como sendo a pessoa ou entidade que desenvolve ou manda desenvolver e coloca no mercado um sistema ou modelo de IA sob o seu nome ou marca.

Por outro lado, (9) colocação no mercado é a primeira disponibilização de um sistema ou modelo de IA na União, (10) disponibilização no mercado é o fornecimento comercial, pago ou gratuito, no território da União e (11) colocação em serviço é a primeira utilização do sistema na União segundo a finalidade prevista.

Por outro lado, (27) norma harmonizada é uma norma europeia definida no Regulamento (UE) 1025/2012[23] e (28) especificação comum é um conjunto de especificações técnicas previsto no mesmo regulamento para cumprir requisitos jurídicos.

Em termos mais técnico-informáticos, (29) dados de treino são os utilizados para ajustar parâmetros de um sistema de IA; (30) dados de validação servem para avaliar e corrigir o processo de aprendizagem do sistema treinado; (31) conjunto de dados de validação é um subconjunto de treino usado especificamente para essa avaliação; (32) dados de teste permitem confirmar o desempenho do sistema antes da sua colocação no mercado; (33) dados de entrada são os fornecidos ou obtidos pelo sistema para gerar resultados; (34) dados biométricos são dados pessoais sobre características físicas, fisiológicas ou comportamentais tratados por meios técnicos; (35) identificação biométrica é o reconhecimento automatizado de identidade comparando dados com uma base de referência; (36) verificação biométrica é a autenticação "um para um" de uma identidade com base em dados previamente fornecidos;

[23] Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32012R1025.

(37) categorias especiais de dados pessoais são os dados sensíveis definidos na legislação europeia, como saúde ou opiniões políticas; (38) dados operacionais sensíveis são informações ligadas à prevenção e repressão de crimes cuja divulgação comprometa processos; (39) sistema de reconhecimento de emoções identifica ou infere emoções através de dados biométricos; (40) sistema de categorização biométrica afeta pessoas a categorias com base em dados biométricos, salvo uso técnico acessório; (41) sistema de identificação biométrica à distância identifica pessoas sem participação ativa, através de comparação com bases de dados.

O (42) sistema de identificação biométrica à distância em tempo real é um sistema que realiza essa identificação sem atraso significativo, incluindo ligeiro atraso; (43) sistema de identificação biométrica à distância em diferido é o que não funciona em tempo real; e (44) espaço acessível ao público é qualquer local físico, público ou privado, de acesso a número indeterminado de pessoas, ainda que condicionado.

A acrescer ao exposto, (49) incidente grave é qualquer falha de um sistema de IA que cause morte, graves danos à saúde, perturbações críticas, violação de direitos fundamentais ou danos sérios a bens ou ao ambiente; (50) dados pessoais são os definidos no artigo 4.º do Regulamento (UE) 2016/679[24]; (51) dados não pessoais são todos os que não se enquadram nessa definição. Já (52) definição de perfis é a prevista no artigo 4.º do mesmo regulamento. Por outro lado, (53) plano de testagem em condições reais é o documento que descreve objetivos, métodos, âmbito e controlo desses testes; (54) plano do ambiente de testagem é o documento acordado entre prestador e autoridade que define condições, objetivos e requisitos do ambiente de ensaio; (55) ambiente de testagem da regulamentação da IA é um espaço controlado, supervisionado por autoridade, que permite desenvolver e validar sistemas de IA, inclusive em condições reais, por tempo limitado.

Ademais, (57) testagem em condições reais é a utilização temporária de um sistema de IA em contexto real, para verificar a sua conformidade, sem se considerar como colocação no

[24] Artigo 4.º

Definições

"Para efeitos do presente regulamento, entende-se por:

1) «Dados pessoais», informação relativa a uma pessoa singular identificada ou identificável («titular dos dados»); é considerada identificável uma pessoa singular que possa ser identificada, direta ou indiretamente, em especial por referência a um identificador, como por exemplo um nome, um número de identificação, dados de localização, identificadores por via eletrónica ou a um ou mais elementos específicos da identidade física, fisiológica, genética, mental, económica, cultural ou social dessa pessoa singular", Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679.

mercado ou em serviço; (58) participante é a pessoa que integra essa testagem; (59) consentimento informado é a declaração livre e consciente do participante de aceitar participar após conhecer todos os aspetos relevantes; (60) falsificações profundas são conteúdos multimédia manipulados por IA para parecerem autênticos; (61) infração generalizada é uma violação do direito da União que afeta coletivamente pessoas em pelo menos dois ou três Estados-Membros; (62) infraestrutura crítica é a definida na Diretiva (UE) 2022/2557; (63) modelo de IA de finalidade geral é aquele com amplitude e adaptabilidade significativas, capaz de executar diversas tarefas e ser integrado em múltiplos sistemas, salvo os usados apenas para investigação ou protótipos; (64) capacidades de elevado impacto são as equivalentes ou superiores às dos modelos mais avançados; (65) risco sistémico é o risco decorrente dessas capacidades de elevado impacto que pode afetar de forma significativa a saúde, a segurança, os direitos fundamentais ou a sociedade.

Por último, de acordo com o Regulamento, (66) sistema de IA de finalidade geral é aquele baseado num modelo de finalidade geral e apto a servir propósitos múltiplos de forma autónoma ou integrada; (67) operação de vírgula flutuante é qualquer operação matemática com números em vírgula flutuante representados digitalmente; e (68) prestador a jusante é quem coloca no mercado um sistema de IA, integrando nele um modelo próprio ou obtido de outra entidade.

2.4 Obrigações para operadores de sistemas de IA

De acordo com o parágrafo 8.º do art. 3.º um "operador" é um prestador, fabricante de produtos que é responsável pela implantação, mandatário, importador ou distribuidor e, por seu turno, "sistema de IA" é um sistema que é baseado em máquinas, o qual é concebido para operar com níveis de autonomia variáveis, o qual pode apresentar capacidade de adaptação após a implantação e que, para objetivos explícitos ou implícitos, e com base nos dados de entrada que recebe, infere a forma de gerar resultados, tais como previsões, conteúdos, recomendações ou decisões que poderão influenciar ambientes físicos ou virtuais[25].

Ora, neste sentido, verificamos que o Regulamento estabelece diversas obrigações para os operadores.

[25] Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689.

Desde logo, o art. 9.º estabelece que, relativamente aos sistemas de IA classificados como de alto risco, o processo deve incluir, em primeiro lugar, a identificação e análise dos riscos conhecidos ou previsíveis que possam afetar a saúde, a segurança ou os direitos fundamentais, sempre que o sistema seja utilizado de acordo com a sua finalidade (*e.g.* art. 9.º, n.º 2, alínea a). Em seguida, exige-se a estimativa e avaliação dos riscos que possam surgir também em cenários de utilização indevida, desde que esta seja razoavelmente previsível (*e.g.* art. 9.º, n.º 2, alínea b). Além disso, o sistema deve contemplar a consideração de outros riscos detetados com base em dados recolhidos através do acompanhamento pós-comercialização do sistema, conforme previsto no art.º 72.º (*e.g.* art. 9.º, n.º 2, alínea c).

Por fim, é imperativo que sejam adotadas medidas de gestão de riscos específicas e adequadas para dar resposta aos riscos identificados nas etapas anteriores (*e.g.* art. 9.º, n.º 2, alínea d).

Ademais, – e, aliás, conforme já vimos – o conceito de operador abrange diferentes categorias de profissionais, sendo que os seus deveres são distribuídos pelos diferentes operadores que intervêm no ciclo de vida dos sistemas de IA. Ora vejamos:

Em primeiro lugar, temos o prestador.

Podemos afirmar que o prestador é o principal responsável pela conceção e pelo "estado de conformidade" do sistema, pelo que, deve construir e documentar o sistema em conformidade com os requisitos técnicos e organizacionais plasmados nos arts. 9.º a 15.º. Além disso, o mesmo tem de implantar um sistema de gestão da qualidade (*e.g.* art. 17.º), manter um plano de vigilância pós-comercialização (*e.g.* art. 18.º), reportar incidentes graves às autoridades competentes e cooperar com estas, facultando documentação e registos (*e.g.* art. 20.º).

A conformidade formal inclui ainda a marcação CE e a declaração de conformidade da União (*e.g.* art. 47.º e 48.º), bem como o registo na base de dados da UE antes da colocação no mercado (*e.g.* art.º 49.º).

Por seu turno, o mandatário (representante autorizado) atua como rosto do prestador dentro da União quando este não se encontra estabelecido no território europeu.

Nos termos do art. 22.º, o mandatário deve conservar a documentação técnica, disponibilizá-la às autoridades competentes e servir de interlocutor em todas as matérias de conformidade.

Já os importadores e distribuidores funcionam como pontos de controlo no mercado.

Desta forma, o art. 23.º obriga o importador a verificar, antes da colocação no mercado, a presença da marcação CE, da declaração de conformidade e das instruções, devendo ainda cooperar com autoridades e adotar medidas corretivas quando necessário e, por outro lado, o art.º 24.º impõe ao distribuidor deveres semelhantes de verificação e de diligência, incluindo a obrigação de suspender a disponibilização e notificar autoridades caso existam riscos ou dúvidas de conformidade.

Quem implanta ou utiliza o sistema tem as suas responsabilidades operacionais plasmadas no art. 26.º, o qual estabelece que o mesmo deve ser usado segundo as instruções do prestador, assegurando a qualidade dos dados de entrada sob o seu controlo, implementando medidas técnicas e organizacionais adequadas (*i.e.* incluindo supervisão humana), conservar registos e comunicar incidentes tanto ao prestador como às autoridades.

Por sua vez, os responsáveis pela implantação que integrem ou coloquem no mercado sistemas de IA sob o seu nome assumem as obrigações correspondentes a prestadores.

Quanto a este aspeto, o art. 27.º define que os mesmos passam a responder pela conformidade técnica, documentação, marcação CE e reporte, como se fossem prestadores originais.

Por último, mas não menos importante, verificamos que o Regulamento prevê uma redistribuição de responsabilidades ao longo da cadeia de valor.

Neste sentido, o art. 25.º estipula que quem altere substancialmente um sistema de IA passa a assumir obrigações próprias de prestador, e que todos os operadores têm deveres de partilha de informação entre si, de forma a garantir a conformidade em todas as fases do ciclo de vida do sistema.

2.5 Categorias de sistemas de IA sujeitas a regras específicas

A matéria da categoria de sistemas de IA e os graus de risco daria, por si só, para um trabalho individual. Porém, não obstante não ser este o foco principal do presente estudo, tentaremos – ainda assim – apresentar uma breve visão sobre este aspeto.

Ora, o Regulamento (UE) 2024/1689 organiza os sistemas de inteligência artificial em categorias distintas, sujeitas a regimes específicos em função do seu nível de risco.

No patamar mais forte estão os sistemas de "risco inaceitável"[26], expressamente proibidos pelo art. 5.º.

Incluem-se no risco inaceitável as técnicas subliminares ou manipulativas que alterem de forma significativa o comportamento das pessoas, a exploração de vulnerabilidades individuais ou sociais (*e.g.* idade ou deficiência), os sistemas de pontuação social que resultem em discriminações, a avaliação de risco criminal baseada em perfis, a criação de bases de dados faciais por recolha indiscriminada de imagens, o reconhecimento remoto de emoções em contextos laborais ou educativos (*i.e.* com exceções médicas ou de segurança), a categorização biométrica em função de características sensíveis e a identificação biométrica remota em tempo real em espaços públicos, salvo exceções muito restritas ligadas à aplicação da lei.

A segunda categoria "hierárquica" é a dos sistemas de risco elevado, definida no art. 6.º e no Anexo III.

Como nos salientam STETTINGER e WEISSENSTEINER, é necessário a aplicação de uma metodologia adequada a este tipo de categoria, por forma a ser possível garantir a fiabilidade dos sistemas de IA, sem, no entanto, comprometer os direitos dos cidadãos[27].

De acordo com o n.º 1 do art. 6.º um sistema de inteligência artificial é considerado de risco elevado sempre que se verifiquem cumulativamente duas condições.

Em primeiro lugar, quando se destina a ser utilizado como componente de segurança de um produto abrangido pela legislação de harmonização da União enumerada no anexo I, ou quando ele próprio constitui esse tipo de produto.

Em segundo lugar, quando esse produto, ou o próprio sistema de IA enquanto produto autónomo, esteja sujeito a uma avaliação de conformidade por entidades terceiras antes de ser colocado no mercado ou em serviço, nos termos da mesma legislação referida no anexo I.

Numa categoria intermédia encontram-se os sistemas de risco limitado, que não exigem o cumprimento das obrigações pesadas aplicáveis aos de alto risco, mas estão sujeitos a requisitos de transparência. O Regulamento, mais concretamente no seu art. 50.º[28], impõe

---

[26] Cfr. SHEIKIN, A. (2024) – *Prohibited Artificial Intelligence Practices In The Legislation Of The European Union.* LEGAL ORDER: History, Theory, Practice.

[27] Cfr. STETTINGER, G., WEISSENSTEINER, P. (2024) – *Trustworthiness Assurance Assessment for High-Risk AI-Based Systems. IEEE Access, 12*, pp. 22718-22745.

[28] Cfr. *https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689.*

que os utilizadores sejam informados de que estão a interagir com um sistema de IA, que sejam assinalados os conteúdos gerados por estes sistemas e que exista rotulagem clara em casos como *chatbots* ou sistemas de geração de imagem, vídeo ou texto.

Por sua vez, os sistemas de risco mínimo – como filtros de *spam* ou videojogos que utilizam IA – não são objeto de obrigações específicas e ficam isentos de requisitos adicionais.

Por último, cumpre deixar ainda uma breve nota para o facto de o Regulamento prever uma categoria autónoma para os modelos de IA de finalidade geral (*i.e. GPAI*), regulados nos arts. 53.º a 56.º.

Estes modelos, utilizados em múltiplas aplicações, têm de cumprir obrigações próprias de transparência, incluindo a disponibilização de documentação técnica, a divulgação de informações sobre os conteúdos utilizados no treino e o respeito por normas de direitos de autor. Quando se trate de modelos de finalidade geral suscetíveis de gerar riscos sistémicos, aplicam-se requisitos acrescidos, como a realização de testes antagónicos, a avaliação do impacto dos riscos, a adoção de medidas de mitigação e o reporte de incidentes relevantes.

Face ao exposto, verificamos que o Regulamento que aqui nos encontramos a estudar constrói um quadro escalonado de obrigações, a saber: proibição absoluta para riscos inaceitáveis, forte regulação para riscos elevados, transparência para riscos limitados, ausência de imposições para riscos mínimos; e um regime específico para os modelos de finalidade geral.

2.6 Mecanismos de fiscalização

Com vista ao cabal funcionamento e aplicação do *AI Act*, o Regulamento veio prever diversos mecanismos de fiscalização e também de sanções.

Ora, desde logo, cumpre referir o art. 3.º, parágrafo 26 do Regulamento 2024/1689 do Parlamento Europeu e do Conselho de 13 de junho de 2024, que vem definir "Autoridade de fiscalização do mercado" como a autoridade nacional que realiza as atividades e toma as medidas previstas no Regulamento (UE) 2019/1020.

Por outro lado, temos o parágrafo 48 do mesmo artigo, o qual enuncia que "Autoridade nacional competente" consiste numa autoridade notificadora ou uma autoridade de fiscalização do mercado; e, por outro lado, no que diz respeito aos sistemas de IA colocados em serviço ou utilizados pelas instituições, agências, serviços e organismos da União, quaisquer referências às autoridades nacionais competentes ou às autoridades de fiscalização do mercado no

presente regulamento devem ser entendidas como referências à Autoridade Europeia para a Proteção de Dados[29].

A acrescer ao exposto, o art. 45.º, n.º 1, alínea d) do Regulamento estabelece, de forma expressa, que quando forem para tal notificados, os organismos devem comunicar à autoridade notificadora pedidos de informação que tenham recebido das autoridades de fiscalização do mercado sobre as atividades de avaliação da conformidade. Dentro deste ponto, mas entrando em concreto no tema das autoridades de fiscalização, cumpre salientar o art. 70.º, n.º 1 do Regulamento, o qual nos refere que cada Estado-Membro cria ou designa pelo menos uma autoridade notificadora e pelo menos uma autoridade de fiscalização do mercado para efeitos do presente regulamento como autoridades nacionais competentes.

No que concerne à fiscalização do mercado e controlo dos sistemas de IA presentes no mercado da União, o art. 74.º, n.º 2 dispõe que as autoridades de fiscalização do mercado têm de comunicar anualmente à Comissão Europeia e às autoridades nacionais de concorrência as informações relevantes recolhidas nas suas atividades, onde se incluem as práticas proibidas detetadas e as medidas adotadas. Ainda no mesmo artigo, mas no seu n.º 11, temos que as autoridades de fiscalização do mercado e a Comissão podem propor e realizar, em conjunto ou separadamente, atividades e investigações destinadas a promover a conformidade, detetar incumprimentos e emitir orientações sobre sistemas de IA de alto risco, com apoio de coordenação do Serviço para a IA[30].

A acrescer ao exposto, cremos que é ainda pertinente fazer referência ao art. 75.º, n.º 1 e 2 do Regulamento, o qual estabelece que o serviço para a IA supervisiona a conformidade de sistemas de IA baseados em modelos de finalidade geral quando desenvolvidos pelo mesmo prestador, dispondo dos poderes de uma autoridade de fiscalização do mercado, e coopera com estas autoridades sempre que existam indícios de incumprimento em sistemas de alto risco.

Ainda no que respeita aos poderes, o n.º 1 do art. 76.º do Regulamento é claro ao dispor que as autoridades de fiscalização do mercado devem ter competências e poderes para assegurar que a testagem em condições reais está em conformidade com o regulamento.

[29] Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689.

[30] Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689.

2.7 Sanções passíveis de aplicação

O objetivo da previsão do Regulamento passa pelo seu cumprimento, a verdade é que, bem sabemos que existirão sempre incumprimentos.

Neste sentido, entra em primeiro lugar o papel da fiscalização, por forma a apreciar se o Regulamento se encontra a ser cumprido. Caso se verifique a existência de incumprimento, então deverão ser aplicadas sanções ao prevaricador.

Neste seguimento, o art. 99.º, n.º 1 do Regulamento prevê que cabe aos Estados Membros a definição de um regime de sanções e medidas de execução para infrações ao regulamento, garantindo que as mesmas sejam eficazes, proporcionadas e dissuasivas, tendo em conta as orientações da Comissão e a viabilidade económica das PME e *startups*.

Sem prejuízo desta competência reservada aos Estados-Membros, o n.º 3 e 4 do mesmo artigo vêm estabelecer que o incumprimento da proibição das práticas de IA previstas no art.º 5.º é sujeito a coimas muito elevadas, as quais podem alcançar 35 milhões de euros ou, no caso de o infrator ser uma empresa, até 7% do volume de negócios anual mundial do exercício anterior, aplicando-se sempre o valor mais elevado. Já as violações de outras disposições do regulamento, não abrangidas por aquela proibição, mas relacionadas com operadores e organismos notificados, são sancionadas com coimas que podem ir até 15 milhões de euros ou até 3% do volume de negócios anual mundial, novamente prevalecendo o montante mais elevado. Nesta categoria incluem-se, entre outras, as obrigações dos prestadores, dos mandatários, dos importadores, dos distribuidores, dos responsáveis pela implantação, bem como as exigências impostas aos organismos notificados e as regras de transparência que vinculam tanto os prestadores como os responsáveis pela implantação[31].

Ora, o regime sancionatório, assim desenhado, distingue com clareza as infrações mais graves, ligadas à violação das proibições centrais, das restantes obrigações de conformidade, garantindo uma resposta diferenciada mas sempre proporcional e dissuasora.

A acrescer ao exposto, um aspeto que julgamos ser interessante salientar prende-se com a aplicação do regime sancionatório aos próprios organismos europeus.

Nesta linha, o art.º 100.º, n.º 1 do Regulamento vem dispor que a Autoridade Europeia para a Proteção de Dados tem competência para aplicar coimas às instituições, órgãos e organismos da União Europeia abrangidos pelo regulamento. No momento de decidir se deve

[31] Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689.

impor uma coima e de fixar o respetivo valor, é necessário avaliar as circunstâncias concretas de cada caso, sendo que, para esse efeito, são considerados diversos fatores, entre os quais a natureza, a gravidade e a duração da infração, bem como as suas consequências, nomeadamente em função da finalidade do sistema de inteligência artificial envolvido, do número de pessoas atingidas e da dimensão dos danos sofridos, importando ainda ponderar o grau de responsabilidade da entidade em causa, à luz das medidas técnicas e organizacionais que tenha adotado, e as iniciativas que esta tenha desenvolvido para reduzir os prejuízos das pessoas afetadas[32].

[32] Cfr. https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689.

3 A implementação de sistemas de IA em conformidade com o Regulamento – desafios e oportunidades para a área da informática

3.1 Barreiras técnicas para atendimento dos requisitos legais

De acordo com SABRINA KUTSCHER, com o *AI Act* da União Europeia (*i.e. AIA*), a União apresentou uma nova resposta à evolução digital, desta vez incidindo diretamente sobre os sistemas de inteligência artificial, considerando a referida autora que esta regulação poderá vir a definir o rumo da forma como as autoridades lidam com a crescente utilização da IA, dado que vem estabelecer um conjunto de regras harmonizadas para a inteligência artificial[33].

Assim, conforme tivemos oportunidade de analisar, o Regulamento Europeu relativo à inteligência artificial, sendo plenamente inovador, traz diversos aspetos positivos na regulamentação da IA.

No entanto, como qualquer normativo e, em especial, como um normativo aplicável a uma verdadeira "revolução", o mesmo traz desafios e dificuldades.

Ora, como salientam JÚLIO MARQUES e PABLO DE ABREU VIEIRA, desde logo, a principal barreira técnica na implementação de sistemas de IA em conformidade com o Regulamento AI decorre da complexidade para garantir transparência e explicabilidade dos modelos, sendo a necessidade de explicação das decisões automatizadas um requisito legal que desafia diretamente as engenharias de IA atualmente disponíveis[34].

Tal significa que existe uma grande dificuldade em operacionalizar princípios jurídicos, como a transparência, a explicabilidade e a robustez, em sistemas baseados em modelos de elevada complexidade, como os de *machine learning* profundo. Ademais, a exigência de documentação exaustiva, rastreabilidade dos dados e mecanismos de auditoria contínua implica o desenvolvimento de novas metodologias técnicas, ainda em fase embrionária, capazes de traduzir obrigações legais em soluções concretas de engenharia.

Acresce a necessidade de conciliar estas exigências com a eficiência computacional e a inovação, o que muitas vezes gera tensões entre conformidade regulatória e viabilidade tecnológica.

---

[33] Cfr. KUTSCHER, Sabrina (2025) – *The EU AI Act: Law of Unintended Consequences?* Technology and Regulation, p. 317.

[34] Cfr. MARQUES, Júlio e VIEIRA, Pablo de Abreu (2023) – *Explorando a Explicabilidade da Inteligência Artificial – Técnicas para Compreender e Interpretar Modelos de Aprendizado de Máquina,* pp. 2 a 15.

Noutro âmbito, PEDRO SANTIAGO RIBEIRO e CATARINA CAMACHO CORREIA enunciam a questão da literacia informático-legal como uma dificuldade acrescida, considerando a previsão do art. 4.º do Regulamento, que determina uma obrigação fundamental para os prestadores e responsáveis pela implantação de sistemas de IA de garantirem a literacia dos seus colaboradores[35].

Além do exposto, a diversidade das infraestruturas tecnológicas e a falta de padronização entre diferentes Estados-Membros dificultam a interoperabilidade dos sistemas de IA, necessária para assegurar a conformidade uniforme e integrada com múltiplos requisitos legais, o que representa um desafio técnico e administrativo significativo. Ademais, com o ciclo acelerado de evolução tecnológica na área da IA e com a introdução frequente de modelos de finalidade geral, torna-se fundamental que os sistemas de avaliação e certificação acompanhem o ritmo, fator que representa complexidade para a informática em termos de gestão da evolução tecnológica e conformidade legal.

3.2 Desafios na integração de segurança, privacidade e ética no desenvolvimento de IA

Dado o seu carácter de novidade, de desconhecido e de sofisticação, a integração de segurança, privacidade e ética no desenvolvimento da inteligência artificial coloca desafios de elevada complexidade, tanto do ponto de vista técnico como normativo.

Um dos grandes desafios, de acordo com MARIO GIULIO BERTORELLI e ROBERT PRAAS, prende-se com a capacidade efetiva de implementação.

Os referidos autores salientam que o Gabinete para a IA da União Europeia, encarregado de supervisionar a aplicação do Regulamento, dispõe de uma equipa reduzida e de um orçamento que representa apenas metade dos recursos atribuídos ao Instituto de Segurança de IA do Reino Unido[36].

No que concerne à privacidade e ética, KAMILA e JASROTIA salientam que, além dos desafios técnicos, existem implicações significativas para os direitos das pessoas, como o uso indiscriminado de dados pessoais e a transparência nas decisões tomadas por sistemas

[35] Cfr. RIBEIRO, Pedro Santiago e CORREIA, Catarina Camacho (2025) – *Regulamento de IA: desafios e oportunidades das novas normas já em vigor*, Disp. *in* https://www.pwc.pt/pt/sala-imprensa/artigos-opiniao/2025/regulamento-inteligencia-artificial.html.

[36] Cfr. BERTORELLI, Mario Giulio e PRAAS, Robert (2025) – *O Regulamento da IA e o desafio de lidar com uma tecnologia em rápida evolução*, Disp. *in* https://voltportugal.org/noticias/o-regulamento-da-ia-e-o-desafio-de-lidar-com-uma-tecnologia-em-rapida-evolucao.

automatizados, sendo por isso necessário equilibrar a inovação com a proteção dos direitos fundamentais, garantindo que a IA seja desenvolvida e utilizada de forma responsável e ética[37]. Já de acordo com BELK, um dos maiores desafios reporta-se à complexidade de integrar segurança tecnológica, respeito pela privacidade e princípios éticos durante o ciclo de desenvolvimento da IA, o que sucede devido às características intrínsecas da IA, como a opacidade, complexidade algorítmica, dependência intensiva de dados e autonomia decisória, aspetos que aumentam o potencial de riscos para direitos fundamentais, como dignidade, privacidade e não discriminação[38].

Com a crescente sofisticação dos modelos de IA torna-se necessário o estabelecimento de arquiteturas resilientes a ataques e mecanismos de prevenção de usos indevidos, sem que isso comprometa a transparência e auditabilidade das decisões algorítmicas[39].

Já no plano ético, SAURA, RIBEIRO-SORIANO e PALACIOS-MARQUÉS consideram que a principal dificuldade reside em traduzir princípios abstratos, como justiça, não discriminação e responsabilidade, em padrões técnicos verificáveis, evitando tanto o "*ethics washing*" como a estagnação normativa, pelo que, o verdadeiro desafio consiste, assim, em conceber estruturas regulatórias e metodológicas capazes de garantir simultaneamente robustez técnica, respeito pelos direitos fundamentais e legitimidade social no ciclo de vida da IA[40].

Por outro lado, temos também a questão da segurança, em concreto, da defesa nacional. Quanto a este aspeto, MARIO GIULIO BERTORELLI e ROBERT PRAAS salientam a imprecisão de algumas definições cruciais, de entre as quais chamam à atenção para o uso do termo "segurança nacional", sendo que, atualmente, o *AI Act* não se aplica aos sistemas de IA utilizados pelos Estados-Membros para fins de "segurança nacional", o que pode causar entropia interpretativa[41].

---

[37] Cfr. KAMILA, M. e JASROTIA, S. (2023) – *Ethical issues in the development of artificial intelligence: recognizing the risks.* International Journal of Ethics and Systems.

[38] Cfr. BELK, R. (2020) – *Ethical issues in service robotics and artificial intelligence.* The Service Industries Journal, 41, pp. 860 - 876.

[39] Cfr. KINGSTON, J. (2017) – *Using artificial intelligence to support compliance with the general data protection regulation.* Artificial Intelligence and Law, 25, pp. 429 - 443.

[40] Cfr. SAURA, J., RIBEIRO-SORIANO, D. e PALACIOS-MARQUÉS, D. (2022) – *Assessing behavioral data science privacy issues in government artificial intelligence deployment.* Gov. Inf. Q., 39, 101679.

[41] Cfr. BERTORELLI, Mario Giulio e PRAAS, Robert (2025) – *O Regulamento da IA e o desafio de lidar com uma tecnologia em rápida evolução*, Disp. *in* https://voltportugal.org/noticias/o-regulamento-da-ia-e-o-desafio-de-lidar-com-uma-tecnologia-em-rapida-evolucao.

3.3 Oportunidades de inovação tecnológica dentro do respeito pelo Regulamento

Uma das grandes críticas que é feita às instituições europeias passa pelo seu alargado acervo legislativo.

De facto, este aspeto tem levado alguns críticos, como ANTÓNIO BRANCO, a afirmarem que "Os Estados Unidos vão continuar a inovar, a China vai continuar a copiar e a Europa vai continuar a regular"[42].

Não obstante entendermos a crítica, cremos que o Regulamento Europeu da Inteligência Artificial não deve ser visto unicamente como um conjunto de limites impostos ao setor tecnológico, mas antes como um quadro jurídico que abre um leque de oportunidades para inovação responsável e competitiva.

Conforme nos refere GABRIEL OSÓRIO DE BARROS, no "Destaca-se a abordagem baseada no risco adotada pelo *AI Act*, que diferencia os sistemas de IA com base no seu potencial risco, impondo requisitos mais rigorosos para aqueles considerados de alto risco, ao mesmo tempo que procura promover a inovação tecnológica. Esta abordagem equilibra a necessidade de proteger os cidadãos e a sociedade com o imperativo de apoiar o progresso tecnológico e a competitividade da UE"[43]. Ora, conforme facilmente se conclui, a classificação dos sistemas de IA por níveis de risco obriga os criadores a repensar a arquitetura tecnológica, incentivando soluções mais robustas, transparentes e seguras desde a sua génese.

Cremos que esta abordagem não só promove confiança social e adesão de mercado, como estimula investimento em áreas como mecanismos de monitorização contínua e auditoria independente, que tendem a elevar os padrões de qualidade do ecossistema europeu.

Por outro lado, podemos ainda afirmar que a conformidade com o *AI Act* pode ainda impulsionar inovação no domínio da interoperabilidade, ao exigir documentação clara, gestão de dados rigorosa e mecanismos de responsabilização verificáveis.

Com um enquadramento global, temos a criação de um vasto espaço para o desenvolvimento de ferramentas que simplifiquem a condução de avaliações de impacto, monitorizem em tempo real parâmetros éticos e de segurança, e potenciem a *compliance* como um serviço competitivo, pelo que, para as empresas tecnológicas, estar em conformidade deixa de ser

[42] Cfr. https://cnnportugal.iol.pt/eua/china/investigador-portugues-sobre-inteligencia-artificial-os-eua-vao-continuar-a-inovar-a-china-vai-continuar-a-copiar-e-a-europa-vai-continuar-a-regular/20231209/65747668d34e65afa2f874dd.

[43] Cfr. BARROS, Gabriel Osório (2024) – *Regulamentação da Inteligência Artificial na União Europeia: Uma análise do AI Act*, Gabinete de Estratégia e Estudos do Ministério da Economia, p. 19.

um mero encargo e converte-se em sinal distintivo de fiabilidade, criando novas cadeias de valor baseadas na confiança e não apenas na eficiência técnica.

Ademais, GABRIEL OSÓRIO DE BARROS salienta "ainda que o *AI Act* impõe obrigações regulatórias principalmente para dois tipos de operadores: os fornecedores de sistemas de IA e os utilizadores desses sistemas [pelo que], é essencial que as empresas estejam preparadas para cumprir as novas regras"[44]. Ora, seguindo esta interpretação, podemos afirmar que outro vetor de inovação consiste no incentivo à investigação em técnicas de mitigação de riscos, como a redução de enviesamentos algorítmicos, a adoção de *federated learning*[45] para proteger a privacidade individual e o desenvolvimento de metodologias de teste mais avançadas para ambientes críticos, como saúde, transportes e infraestrutura pública. O *AI Act*, ao reconhecer setores de alto risco, obriga a que estes se tornem laboratórios de excelência regulatória e tecnológica, pelo que, ao terem que o cumprir, as empresas europeias estarão dentro de um verdadeiro polo de inovação ética e segura.

[44] *Id.*

[45] Para mais desenvolvimentos sobre este tema, vide WEN, J., ZHANG, Z., LAN, Y., CUI, Z., CAI, J., & ZHANG, W. (2022) – *A survey on federated learning: challenges and applications. International Journal of Machine Learning and Cybernetics*, 14, pp. 513 - 535.

4 *Framework* proposta para auditoria e *compliance* de IA no *e-commerce*

De acordo com MANNING, o *compliance* corresponde ao conjunto de medidas e procedimentos destinados a assegurar que uma empresa atue em conformidade com a legislação, regulamentos, normas aplicáveis e princípios éticos, tanto internos como externos[46].

Neste sentido, o *compliance* aplicado a sistemas de inteligência artificial regulados no comércio digital pode ser compreendido como o conjunto de práticas, normas e mecanismos de governança destinados a assegurar que o desenvolvimento, a implementação e a utilização de tais sistemas estejam em conformidade com requisitos legais, regulamentares, éticos e técnicos, pelo que, se trata necessariamente de um processo dinâmico de alinhamento contínuo entre a inovação tecnológica e o quadro normativo vigente, em especial quando os sistemas de IA são utilizados em contextos de elevado impacto socioeconómico, como a contratação eletrónica, a gestão de plataformas digitais ou a personalização de serviços no comércio *online*.

Um exemplo que podemos referir relativamente à aplicação do *compliance* em sistemas de IA regulados reporta-se ao setor bancário. Como nos refere PEDRO MAIA, "O *compliance* é comummente identificado como uma das áreas da atividade bancária que mais apetência mostra para o uso de inteligência artificial, sendo que os benefícios, neste plano, são evidentes: a inteligência artificial oferece a possibilidade de analisar, filtrar, *etc.*, o universo total das operações – independentemente do seu montante, do lugar em que sejam ordenadas, da jurisdição a que pertençam os beneficiários, da hora e do dia de semana em que ocorram, *etc.* –, em tempo real – determinando, por exemplo, o bloqueio de uma operação de pagamento com um cartão de crédito –, considerando um acervo de informação (*e.g. big data*) inacessível ao conhecimento humano[47].

Ademais, a inteligência artificial tem emergido como uma ferramenta transformadora no comércio digital, permitindo que as empresas analisem e compreendam o comportamento do consumidor como nunca computacionalmente. Com o *output* da análise e tratamento de dados de tais sistemas informáticos, *insights* comerciais valiosos são obtidos. Os dados têm as mais diversas fontes como redes sociais, *websites*, aplicações móveis, *e-mails*, *etc.*, que,

[46] Cfr. MANNING, L. (2020) – *Moving from a compliance-based to an integrity-based organizational climate in the food supply chain*. Comprehensive reviews in food science and food safety, 19 3, pp. 995-1017

[47] Cfr. MAIA, Pedro (2021) – *Compliance bancário na era da inteligência artificial – uma breve introdução*, Revista Julgar, n.º 45, pp.187 e 188.

razoavelmente, apontam para padrões de compra, preferências dos consumidores, entre outros"[48].

Ora, neste sentido, não há dúvidas de que o *compliance* em sistemas de IA no comércio digital é fundamental.

Desde logo, podemos afirmar que o *compliance* em sistemas de IA no comércio digital protege os consumidores contra abusos e discriminações ocultas em algoritmos de recomendação, precificação ou *scoring* de crédito, reduz riscos jurídicos e reputacionais através de auditorias e monitorização; promove competitividade sustentável, uma vez que a conformidade ética e regulatória é valorizada pelo mercado e investidores; incentiva inovação responsável, orientando o desenvolvimento de IA por princípios éticos e legais; e reforça a governação interna e a *accountability*, garantindo a rastreabilidade das decisões algorítmicas e procede a uma documentação clara do ciclo de vida dos sistemas.

4.1 Arquitetura técnica de sistemas *compliance* de IA

Como verificámos anteriormente, uma *framework* para auditoria e *compliance* de IA consiste no estabelecimento de um conjunto de processos, critérios e ferramentas para garantir que os sistemas sejam seguros, éticos e cumpram a lei.

Conforme nos refere WOLTERS KLUWER, o estabelecimento de uma *framework* eficaz deve estabelecer claramente a *governance* da IA, definindo responsabilidades, papéis e processos de supervisão para garantir o alinhamento com políticas internas e externas, sendo que a *governance* deve incluir mecanismos de controlo internos robustos para mitigar riscos, assegurar a conformidade legal e fomentar a cultura de ética e responsabilidade tecnológica[49].

Por outro lado, de acordo com CRUZ, BERTOLLO e CAMARGO, para o estabelecimento de uma *framework* eficaz é necessário proceder à identificação, avaliação e gestão contínua dos riscos associados aos sistemas de IA, incluindo riscos de segurança, privacidade, vieses

[48] Cfr. MASSA, Mariana, FLORES, Cláudio e SANTOS, Ricardo (2025) – *Inteligência Artificial (IA) no Comércio Digital: Oportunidades e Desafios*, Centro de Inovação e Investigação em Ciências Empresariais e Sistemas de Informação, Escola Superior de Tecnologia e Gestão / Politécnico do Porto, p. 2.

[49] Cfr. WOLTERS KLUWER (2024) – *Solucionando o enigma: Aplicando a IA Generativa em Atividades de Auditoria Interna*, Internal Audit Foundation, p. 6.

discriminatórios, e falhas de funcionamento[50]. Quer isto dizer que a *framework* deverá contemplar mecanismos para detetar e corrigir desvios, realizando avaliações periódicas, com vista a assegurar a integridade dos sistemas. Ademais, a *framework* deve prever processos para auditoria interna e externa permanente, permitindo identificar anomalias, fraudes ou violações rapidamente e deverá assegurar o cumprimento de padrões éticos como o respeito pela autonomia humana, prevenção de danos, equidade e inclusão, justiça e não discriminação. Os referidos princípios devem orientar a conceção e utilização responsável dos sistemas de IA, prevenindo impactos sociais negativos e fortalecendo a confiança pública.

Para uma compreensão mais precisa do alcance e relevância das *frameworks* em matéria de inteligência artificial, é fundamental apresentar exemplos de aplicação. Nesse contexto, destacam-se iniciativas como a *framework* do *National Institute of Standards and Technology* (*NIST*), voltado para a identificação, avaliação e mitigação de riscos em sistemas de IA; a *framework ALTAI* da União Europeia, que enfatiza princípios éticos e confiabilidade; e a *AI Verify Testing Framework*, que apoia a validação e o monitoramento da transparência e robustez das tecnologias de IA.

4.1.1 *NIST AI Risk Management Framework* (*i.e. AI RMF*)

De acordo com CAMERON F. CARRY, a criação da *Artificial Intelligence Risk Management Framework* foi determinada pela *National Artificial Intelligence Initiative Act*, integrada na lei de autorização da defesa nacional de 2020. O documento segue o modelo de outros quadros de gestão de risco do *NIST*, como a *Cybersecurity Framework* (2014) e a *Privacy Framework* (2020) e, tal como estes, resultou de um processo participativo, com várias versões preliminares submetidas a consulta pública, oficinas de trabalho e outros mecanismos de envolvimento da sociedade. O resultado é um instrumento concebido como "documento vivo": voluntário, respeitador de direitos, transversal a todos os setores e independente do caso de uso, podendo ser aplicado em organizações de qualquer dimensão. À semelhança dos quadros anteriores, estrutura-se em "funções nucleares", subcategorias e perfis de implementação[51].

[50] Cfr. CRUZ, R., BERTOLLO, D. e CAMARGO, M. (2020) – *O Impacto da Inteligência Artificial na Auditoria: Uma Revisão Bibliográfica*. XX Mostra de Iniciação Científica. UCS-PPGA, p. 15.

[51] Cfr. CARRY, Cameron F. (2023) – *Commentary NIST's AI Risk Management Framework plants a flag in the AI*, Cfr. https://www.brookings.edu/articles/nists-ai-risk-management-framework-plants-a-flag-in-the-ai-debate.

Em concreto, SCOTT D. ROSE afirma que a referida *framework* detalha processos para mapear riscos, definir métricas técnicas, como precisão e *fairness*, criar planos de resposta e fortalecer controles internos e mecanismos de *accountability*[52].

Numa outra perspetiva, DON MCCLEAN salienta que a *framework NIST RMF* é especialmente recomendada para empresas com forte exposição regulatória, como plataformas financeiras ou de comércio digital, dando suporte à observabilidade de modelos, *logs* auditáveis e resposta a incidentes[53].

O núcleo da *AI RMF* constitui-se por quatro principais funções que norteiam as atividades de governança e mitigação de riscos de inteligência artificial.

Figura 1 - Esquema das quatro funções principais da gestão de riscos em IA, fonte: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf, p. 25

A framework dita que as funções são organizadas em atividades de gestão de risco de IA ao mais alto nível. A governação é concebida de forma transversal, garantindo que a informação é disseminada e integrada ao longo das restantes três funções. Em detalhe, as quatro funções subdividem-se em categorias, e estas em subcategorias permitindo assim que a *AI RMF* abranja ações e *outputs* que favorecem o diálogo, a compreensão e a implementação de práticas eficazes para gerir os riscos de IA, bem como a promoção do desenvolvimento de sistemas de IA fiáveis[54].

[52] Cfr. ROSE, Scott D. (2021) – *Planning for a Zero Trust Architecture*, p. 16.

[53] Cfr. MACLEAN, Don (2017) – *The NIST Risk Management Framework: Problems and recommendations.* Cyber Security: A Peer-Reviewed Journal.

[54] Cfr. National Institute of Standards and Technology. (2023) - *Artificial Intelligence Risk Management Framework (AI RMF) 1.0* (NIST AI 100-1). U.S. Department of Commerce.

Figura 2 - Tabela das categorias e subcategorias de cada função e o seu objetivo, fonte: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

Função	Categorias	Objetivos
GOVERN	1. Governança 2. Políticas 3. Cultura de risco 4. Supervisão 5. Responsabilidade 6. Documentação	i. Estabelecer e manter uma cultura organizacional que prioriza a gestão de riscos em IA. ii. Envolver políticas, responsabilidades e processos de supervisão.
MAP	1. Definição do contexto 2. Âmbito do sistema 3. Identificação de riscos 4. Envolvimento das partes interessadas	i. Identificar contextos, riscos e objetivos do sistema de IA. ii. Analisar *stakeholders*, potenciais impactos e propósitos do sistema.
MEASURE	1. Métricas 2. Validação 3. Testes 4. Avaliação de dados 5. Modelos 6. Quantificação de incertezas	i. Avaliar e quantificar riscos, desempenho e impactos do sistema de IA. ii. Empregar métricas, testes e auditorias para medir a confiança, equidade, segurança, entre outros fatores.
MANAGE	1. Resposta ao risco 2. Monitorização 3. Gestão de incidentes 4. Melhoria contínua	i. Implementar ações para priorizar, responder e monitorar riscos ao longo do ciclo de vida da IA. ii. Incluir planos de mitigação, correção e melhoria contínua.

4.1.1.1 *NIST AI RMF vs. AI Act*

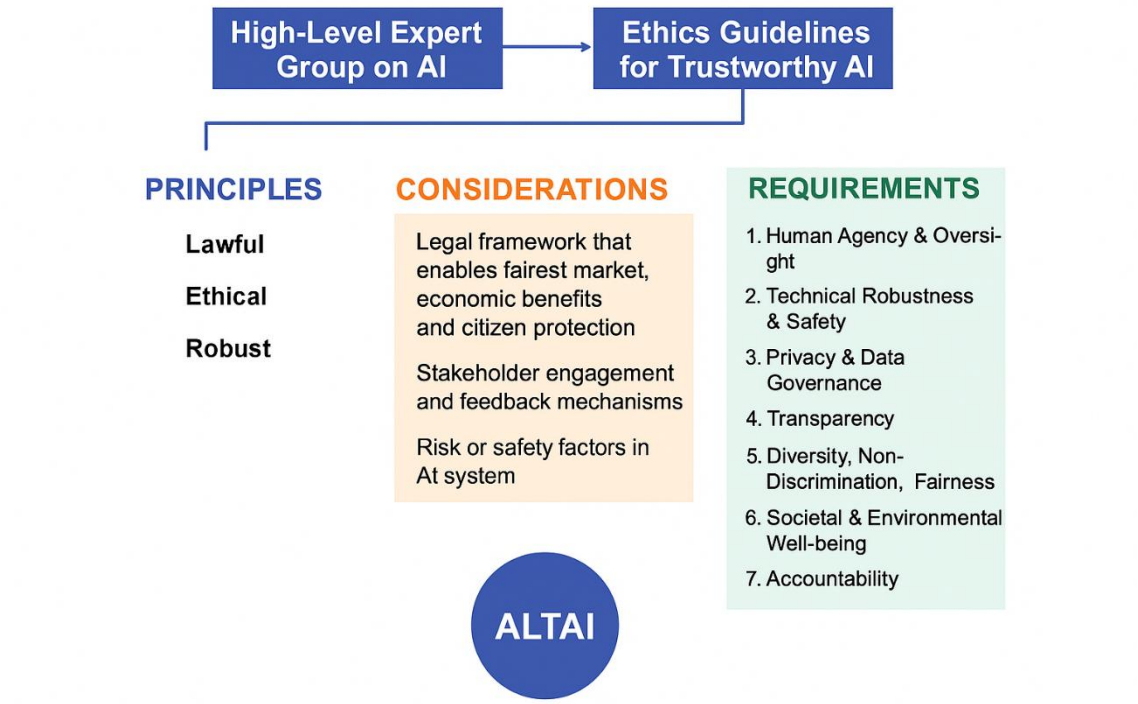
Embora partilhem consideráveis particularidades, o *AI Act* (*i.e.* UE) e a *NIST AI RMF* (*i.e.* EUA) desencontram-se no alcance e na aplicação. O *AI Act*, juridicamente vinculativo, aplica-se a qualquer utilizador de sistemas de IA que opere na UE, classificando as soluções por níveis de risco e impondo requisitos rigorosos aos sistemas de alto risco. Em caso de incumprimento, poderão ser aplicadas coimas até 35 milhões de euros ou 7% do volume de negócios global. Por outro lado, a *NIST AI RMF* é um quadro voluntário de gestão de risco, focado na mitigação de riscos e com princípios e práticas de autorregulação que promovem confiança e transparência. Embora não imponha sanções legais, a não conformidade pode trazer riscos reputacionais e até operacionais.

4.1.2 *Assessment List for Trustworthy Artificial Intelligence (i.e. ALTAI)*

A 17 de julho de 2020, o *High-Level Expert Group on Artificial Intelligence (i.e. AI HLEG)*, convocado pela Comissão Europeia, apresentou a versão final da *Assessment List for Trustworthy Artificial Intelligence (i.e. ALTAI)*, desenvolvida após um processo-teste que contou com a participação de 350 *stakeholders*. A solução, com os seus princípios e considerações,

compromete-se com a operacionalização dos sete princípios fundamentais para a construção de uma inteligência artificial, definidos nas anteriores *Ethics Guidelines for Trustworthy AI* (2019).

Figura 3 - Estrutura conceptual das *Ethics Guidelines for Trustworthy AI*, fonte: https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2023.1020592/full



Segundo HERRON, a legalidade exige a adoção de um quadro jurídico adequado e impõe à indústria a responsabilidade de participar ativamente no debate de políticas públicas, procurando um enquadramento regulatório que assegure um funcionamento mais justo do mercado, garantindo simultaneamente que os benefícios económicos se distribuem de forma equitativa e que é conferida a máxima proteção aos cidadãos[55]. A conceção deste primeiro domínio constitui a principal atividade com a qual a Comissão Europeia está atualmente envolvida no momento da redação deste documento. A dimensão ética requer o envolvimento contínuo das partes interessadas e a criação de mecanismos de *feedback* que permitam recolher e incorporar as suas perspectivas, assegurando que o desenvolvimento e aplicação de IA permanecem alinhados com os valores sociais e humanos, conforme referem RADCLYFFE e NODELL[56]. Por sua vez, a robustez implica a identificação e mitigação dos

[55] Cfr. HERRON, M. (2020) - *A Civil Liability Regime for AI?* Lexology.

[56] Cfr. RADCLYFFE, C., and NODDEL. R. (2020) - *Ethical by design: Measuring and managing digital ethics in the enterprise.*

fatores de risco e de segurança relevantes durante a conceção, desenvolvimento e implementação dos sistemas de IA promovendo a adoção e conformidade com normas técnicas regulatórias que garantam o seu desempenho fiável e seguro, como defendem HAMON *et al.*[57].

Ao definir os três pilares desta forma, o *HLEG* reconheceu três diferentes domínios de governação que requerem consideração e exigem uma gestão atenta com as melhores práticas para ser passível de se alcançar confiança no mecanismo de IA.

4.1.2.1. *Ethics Guidelines for Trustworthy AI*

Figura 4 - Diretrizes de Avaliação Ética da Inteligência Artificial segundo o *HLEG-AI*, fonte: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Guideline	Descrição
Human Agency & Oversight	Testar a organização quanto à maturidade com que o *comiteé* considera ser o papel do *human-in-command, human-in-the-loop, or human-on-the-loop* no contexto do sistema de IA. O *HLEG-AI* defende que os sistemas de IA devem acentuar o florescimento da democracia e apoiar a agência, conduzindo a uma sociedade mais equitativa.
Technical Robustness & Safety	O princípio da precaução está em mente quando o foco muda para o requisito de que sistemas de IA sejam desenvolvidos com uma abordagem preventiva do risco. É sugerida uma reflexão deliberada dos danos razoavelmente prováveis e que estes sejam minimizados através do processo de design. Esta é a secção com mais perguntas, o que reflete a urgência em torno do risco e da segurança que é sentida pelo *HLEG-AI*.
Privacy & Data Governance	Obrigações devem ser estipuladas ao desenvolvedor de sistemas para garantir que a privacidade seja projetada para o sistema desde o momento de design, ao mesmo tempo que garante que altos padrões de governança de dados estejam em vigor.
Transparency	As demandas relativas à transparência agrupam-se em três distintos temas: os controlos em torno da proveniência dos dados e modelos num sistema de IA; a medida em que as decisões alcançadas pelo sistema de IA são explicadas e a compreensão dos utilizadores avaliada; a qualidade da divulgação e comunicação aos utilizadores da existência e funcionamento do sistema de IA.
Diversity, Non-discrimination & Fairness	Na consideração da diversidade, não discriminação e equidade, o *HLEG-AI* está profundamente ciente dos potenciais danos que podem ser causados pela coleta de dados numa sociedade desigual e o processamento destes por equipes não diversas. A probabilidade deste conjunto de eventualidades resultar no agravamento das desigualdades existentes, a menos que receba especial atenção, seria, de outro modo, demasiado elevada.
Societal and Environmental Well-being	A preocupação de que uma implementação inadequada da tecnologia possa ter potencial para perturbar o tecido social é confirmada nas questões desta seção. Reconhece-se que há um equilíbrio a encontrar entre beneficiar todos os seres humanos no presente e as gerações futuras. Mais uma vez, o florescimento da democracia tem de ser visto como preocupação, e especialmente a pluralidade de valores e opções de vida dos indivíduos.
Accountability	As ansiedades comuns perante um sistema de IA são suficientes para causar danos em casos em que seja perdido o controlo dos modelos ou este descontrole não seja percetível pelas equipas devido à complexidade do sistema.

De acordo com JOBIN *et al.*, têm sido registadas várias críticas às abordagens assentes apenas em princípios de ética e de governança da IA. Estes autores destacam que, embora

[57] Cfr. Hamon, R., Junklewitz, H., and Sanchez, I. (2020) - *Robustness and Explainability of Artificial Intelligence. Technical report*, Publications Office of the European Union, Luxembourg.

muitas organizações adotem tais *guidelines* como um primeiro passo relevante, acabam frequentemente por considerar que esses princípios são suficientes por si só, negligenciando a necessidade de os operacionalizar através de mecanismos técnicos, normativos e verificáveis[58].

4.1.3 *AI Verify Testing Framework*

A *AI Verify Testing Framework* constitui-se como um instrumento de governação concebido para apoiar organizações na avaliação da implementação responsável de sistemas de inteligência artificial. Desenvolvido pela Infocomm *Media Development Authority (i.e. IMDA)* e pela *Personal Data Protection Commission (PDPC)* [59], foi oficialmente lançada em 2022, tornado *open-source* em 2023 e, posteriormente, ampliada em 2025 para incluir orientações específicas sobre os riscos emergentes da inteligência artificial generativa. Essa atualização assegura a sua harmonização com referenciais internacionais como a *NIST AI Risk Management Framework (i.e.* EUA), a ISO/IEC 42001:2023, o *AI Act (i.e.* União Europeia) e o *Hiroshima Process Code of Conduct*.

A metodologia assenta em quatro eixos principais: princípios orientadores, resultados desejados, processos de verificação e evidências documentais e é destinado a desenvolvedores, equipas internas de conformidade e auditores externos. Em contributo para o reforço da confiança dos *stakeholders*, a *framework* requer *reports* de conformidade, autoavaliações sistemáticas e documentação estruturada de práticas responsáveis de IA.

4.2 *Phase-based audit process*

O Processo de Auditoria Faseado aplicado à gestão de IA, em conformidade com a ISO/IEC 42001:2023 e a *NIST AI RMF*, surge como delineador e avaliador da conformidade e maturidade de sistemas de inteligência artificial. O processo, faseado em quatro momentos interdependentes, visa assegurar que *o AI Management System (i.e. AIMS)* é auditado de forma progressiva, desde a preparação até a melhoria contínua.

[58] Cfr. Jobin, A., Ienca, M., & Vayena, E. (2019) - The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399.

[59] Cfr. ALLEN, Jason Grant; LOO, Jane; LUNA CAMPOVERDE, José Luis (2025) - *Governing Intelligence: Singapore's Evolving AI Governance Framework. Cambridge Forum on AI: Law and Governance*, Cambridge University Press, 2025, p. 9–10.

Enquanto a *NIST AI RMF* aborda a gestão de risco e confiança através dos domínios *GOVERN, MAP, MEASURE* e *MANAGE*[60], a norma ISO tenciona representar a estrutura do sistema de gestão de IA, com políticas, responsabilidades, ciclo *PDCA*, entre outros[61]. Com base nestes referenciais, é possível enunciar que o Phase-based Audit Process integra, assim, ambos os modelos, garantindo uma abordagem harmonizada entre gestão (*e.g.* ISO/IEC 42001), risco (*e.g. NIST AI RMF*) e conformidade regulatória (*e.g. AI Act*).

Figura 5 - Plano de Auditoria de IA com Alinhamento a ISO/IEC 42001, *NIST AI RMF* e *AI Act*

Fase	Semanas	Objetivos	Relação com a ISO/IEC 42001	Relação com *NIST AI RMF*	Relação com *o AI Act*
1. *Planning & Scoping*	1-2	i. Definir escopo, objetivos e critérios de auditoria. ii. Identificar riscos, partes interessadas e contexto operacional do sistema de IA.	Cláusulas 4-6 Contexto da organização, liderança, planeamento	*MAP* Definição de contexto, riscos e objetivos	i. **Art. 9.º** Sistema de gestão de risco ii. **Art. 10.º (1-3)** Governança e qualidade de dados iii. **Art. 17.º** Gestão de ciclo de vida e design responsável
2. *Document Review & Control Assessment*	3-6	Avaliar políticas, procedimentos, registos e evidências da implementação do AIMS e dos controlos de risco.	Cláusulas 7-8 Suporte, operação e controlo documentado	*GOVERN* Políticas, papéis, responsabilidades, transparência	i. **Art. 11.º** Documentação técnica ii. **Art. 12.º** Registos automático iii. **Art. 14.º** Supervisão humana iv. **Art. 18.º** Conformidade técnica
3. *Implementation & Verification*	7-8	i. Avaliar a conformidade real do sistema de IA, testando medições, controlos e processos. ii. Verificar eficácia de mitigação de riscos.	Cláusula 9 Avaliação de desempenho e verificação	*MEASURE/MANAGE* Medição, validação e resposta a riscos	i. **Art. 15.º** Robustez, precisão e segurança ii. **Art. 16.º** Obrigações do fornecedor iii. **Art. 20.º** Avaliação de conformidade iv. **Art. 21.º** Procedimentos internos de controlo
4. *Reporting & Continuous Improvement*	9-10	i. Relatar resultados, não conformidades e ações corretivas. ii. Garantir melhoria contínua do sistema de gestão de IA.	Cláusula 10 Melhoria	*MANAGE* Gestão, mitigação e melhoria contínua.	i. **Art. 19.º** Atualização pós-comercialização e monitorização

[60] https://www.nist.gov/itl/ai-risk-management-framework

[61] https://www.iso.org/standard/81230.html

					ii. Art. 23.º Correção e mitigação de não conformidades iii. Art. 72.º–74.º Supervisão e sanções

4.3 Ferramentas e tecnologias

Canalizando conformidades com os requisitos estabelecidos pelo *AI Act* no contexto dos sistemas de inteligência artificial, é fundamental e imprescindível a integração de um conjunto de ferramentas técnicas que suportem os demais processos de monitorização, validação, explicabilidade e auditoria. Primeiramente e antes das referências às ferramentas mais aplicadas nos casos de uso, apresentamos, através da seguinte tabela, a analogia entre os requisitos legais do *AI Act* e possíveis ferramentas e tecnologias.

Figura 6 - Ferramentas e Técnicas para Conformidade com o AI Act

Requisito Legal do AI Act	Implementação Técnica	Ferramenta/Tecnologia
Gestão de Risco (e.g. art. 9º)	i. Inventário de modelos ii. Classificação por risco iii. Planos de mitigação documentados iv. Simulação de cenários adversos	*E.g. NIST AI RMF, ISO/IEC 42001:2023, AI Verify*
Dados de Treino e de Teste (e.g. art. 10º)	i. *Datasheets for Datasets* ii. Análise de viés pré-treinamento iii. Versionamento de *datasets* iv. Validação continua de qualidade	*E.g. DVC, Great Expectations, Aequitas, Datasheets for Datasets*
Documentação Técnica (e.g. art. 11º)	i. Relatórios automatizados de treino ii. *Logs* de decisões algorítmicas iii. *Model cards* para cada modelo	*E.g. MLfow, Weigths & Biases, Model Cards Toolkit*
Registo e *Logging* (e.g. art. 12º)	i. *Logging* de inputs e outputs ii. *Logging* de decisões contestadas iii. Retenção segura de *logs* para auditoria	*E.g. MLflow, Apache Kafka, Elastic Stack*
Transparência (e.g. art. 13º)	i. Interfaces de explicabilidade para utilizadores ii. *APIs* de autoria iii. Relatórios de funcionamento acessíveis	*E.g. LIME, SHAP, Human-in-the-Loop APIs, Active Learning frameworks*
Supervisão Humana (e.g. art. 14º)	i. Painéis de controlo para intervenção	*E.g. AI Verify, Human-in-the-Loop APIs, Active Learning frameworks*
Robustez, Precisão e Cibersegurança (e.g. art. 15º)	i. Testes de *adversarial robustness* ii. Validação em múltiplos cenários iii. Monitorização de *drift* e anomalias	*E.g. IBM Adversarial, Robustness Toolbox, TensorFlow Privacy, Alibi Detect*
Proteção de Dados (e.g. AI Act art. 23º, RGPD)	i. Pseudonimização/anonimização ii. Controlo granular de consentimento iii. Monitorização de acessos a dados pessoais	*E.g. GDPR Toolbox, PrivBayes, OpenDP (i.e.Differential Privacy)*

4.3.1 Microsoft *Fairlearn*

O *Fairlearn* é um *toolkit open-source* desenvolvido pela *Microsoft*, cuja finalidade é apoiar organizações na identificação e mitigação de potenciais riscos de injustiça e discriminação algorítmica (*e.g.* arts. 9.º e 10.º do *AI Act*) em sistemas de inteligência artificial. A ferramenta assenta em dois componentes principais: *mitigation,* que fornece algoritmos destinados a reduzir enviesamentos (*i.e. biases*) durante o processo de treino ou utilização de modelos preditivos; e *assessment,* que dispõe de métricas e relatórios de avaliação de equidade, permitindo comparar o desempenho do modelo entre diferentes grupos (e.g. género, idade, origem étnica ou outra variável sensível). Deste modo, a ferramenta não apenas mede disparidades, mas também sugere estratégias para as mitigar[62].

Um exemplo concreto da aplicação do *Fairlearn* no comércio digital verifica-se nos sistemas de recomendação e personalização de ofertas, utilizados para orientar as escolhas dos consumidores. Nestes contextos, existe o risco de surgirem enviesamentos algorítmicos, suscetíveis de gerar tratamentos desiguais, como a atribuição de promoções mais vantajosas ou de recomendações de produtos premium a determinados perfis, em função do género ou da condição socioeconómica. O *Fairlearn* permite identificar essas desigualdades e aplicar técnicas de mitigação, ajustando o modelo de modo a promover equidade nas interações comerciais, sem comprometer significativamente a performance global do sistema.

4.3.2 IBM *Fairness* 360

O *Fairness* 360 (*i.e. AIF360*) é um *toolkit open-source* desenvolvido pela *IBM*, concebido para a deteção, avaliação e mitigação de *bias* algorítmicos em sistemas de inteligência artificial, conduzindo assim as diretrizes de alto nível, como a equidade, a transparência e a responsabilidade, ao longo de todo o ciclo de vida dos modelos de IA. Disponibilizando um conjunto alargado de métricas de equidade e algoritmos de mitigação, aplicáveis nas fases de *pre-processing, in-processing* e *post-processing, e*stas funcionalidades permitem comparar o desempenho de modelos entre diferentes grupos demográficos, como género, idade ou condição socioeconómica, identificar potenciais desigualdades e implementar medidas corretivas que reduzam práticas discriminatórias não intencionais.

[62] https://fairlearn.org/

No contexto do comércio digital, o *AIF360* pode ser utilizado para monitorizar sistemas de recomendação e personalização de ofertas, assegurando que os algoritmos não favorecem injustamente determinados grupos de consumidores. Por exemplo, pode identificar situações em que promoções ou produtos de maior valor são sistematicamente apresentados apenas a perfis masculinos ou de maior rendimento, propondo ajustes que garantam um tratamento equitativo entre utilizadores[63].

4.3.2 - LIME/SHAP

LIME (i.e. Local Interpretable Model-agnostic Explanations) e SHAP (i.e. Shapley Additive exPlanations) constituem duas abordagens amplamente usadas em algoritmos de IA para o desenvolvimento de modelos mais interpretáveis, especialmente em contextos regulatórios e éticos.

O *LIME*, técnica destinada a fornecimento de explicações locais para previsões individuais de modelos complexos, funciona apresentando modelos interpretáveis, mais simples, geralmente em regressão linear, que aproxima o comportamento do modelo complexo na vizinhança de instâncias específicas. A abordagem, especializada na execução de tarefas de classificação de texto, imagens e séries temporais, permite compreender decisões individuais de forma intuitiva. Contudo, são frequentes as explicações geradas pelo *LIME* que apresentam instabilidade, devido à aleatoriedade envolvida na criação das perturbações dos dados utilizadas para a análise[64].

O SHAP (*i.e. Shapley Additive exPlanations*), por sua vez, atribui valores de importância aos *inputs*, com *reports* de impacto de cada característica na previsão do modelo. A ferramenta tem forte base na teoria dos jogos cooperativos em valores de *Shapley* para distribuir as variáveis de forma justa. Tem a sua aplicação em modelos de árvore de decisão, redes neurais e modelos lineares, oferecendo explicações consistentes e fundamentadas teoricamente[65].

---

[63] https://research.ibm.com/blog/ai-fairness-360

[64] Cfr. ZAFAR, Muhammad Rizwan; KHAN, Nasir M. (2019) - *DLIME: A Deterministic Local Interpretable Model-Agnostic Explanations Approach for Computer-Aided Diagnosis Systems*. arXiv preprint p. 1.

[65] Cfr. SALIH, Ahmed M.; RAISI-ESTABRAGH, Zahra; BOSCOLO GALAZZO, Ilaria; RADEVA, Petia; PETERSEN, Steffen E.; LEKADIR, Karim; MENEGÁZ, Gloria (2025) - *A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. Advanced Intelligent Systems*, 7(1), DOI: 10.1002/aisy.202400304.

No *e-commerce*, técnicas de explicabilidade como *SHAP* e *LIME* permitem compreender de que forma variáveis associadas aos clientes, (*e.g.* comportamento de navegação e dados demográficos) influenciam o desempenho de campanhas publicitárias. Partindo do exemplo que é identificado que um determinado grupo demográfico responde melhor a um tipo específico de promoção, é despoletado o direcionamento dos esforços e recursos de forma mais eficaz, sem comprometer a performance global do sistema[66].

4.3.3 *Databricks MLflow*

MLflow é uma plataforma *open-source* da *Databricks*, conjeturada para a facilitação da gestão completa do ciclo de vida de projetos de *machine learning*, promovendo maior reprodutibilidade e transparência — e aspetos essenciais no contexto da governação ética e responsável da inteligência artificial[67].

Esta organiza-se em quatro componentes principais: *MLflow Tracking*, que regista e compara métricas, parâmetros e resultados de diferentes execuções experimentais; o *MLflow Projects*, que padroniza a estrutura dos projetos de *ML*, facilitando a portabilidade entre equipas e ambientes; o *MLflow Models*, que gere o ciclo de vida dos modelos, desde a fase de treino até à sua implementação; e o *MLflow Registry*, um repositório central que permite o controlo de versões e a gestão colaborativa de modelos[68].

Um exemplo prático de aplicação do *MLflow* pode ser observado em plataformas de comércio eletrónico, onde modelos de recomendação de produtos são constantemente atualizados. Através do *MLflow*, as equipas de dados podem registar as diferentes versões dos modelos testados, acompanhar métricas de desempenho (*e.g.* precisão ou taxa de conversão) e selecionar, de forma transparente, o modelo mais adequado para implementação. Este processo contribui para uma gestão mais eficiente e auditável dos sistemas de IA, assegurando a sua fiabilidade e conformidade com princípios de responsabilidade algorítmica.

4.3.4 *Great Expectations*

O *Great Expectations* é uma ferramenta *open-source*, amplamente reconhecida pela sua eficácia na validação, documentação e monitorização da qualidade de dados, ao longo de

[66] https://paanalytics.net/blog/interpretacao-de-shap-e-lime-em-machine-learning/

[67] https://mlflow.org/

[68] https://www.mlflow.org/docs/latest/ml/model-registry

todo o ciclo de vida de sistemas de IA e *ML*. Assegura assim que os dados utilizados em processos analíticos e preditivos são confiáveis, consistentes e conformes com requisitos predefinidos, mitigando riscos associados a erros, enviesamentos e violações de conformidade.

A ferramenta baseia-se na criação de expectativas, isto é, regras ou testes automatizados que descrevem o comportamento esperado dos dados. Estas expectativas devem ser aplicadas durante a recolha, transformação e carregamento de dados *(e.g. ETL pipelines)*, bem como nas fases de treino e validação de modelos, garantindo integridade e rastreabilidade em todo o processo.

No contexto de governação de IA e conformidade com o *AI Act*, o *Great Expectations* contribui para a transparência e responsabilização, assegurando que os dados utilizados para o treino de modelos cumprem critérios de qualidade e não introduzem enviesamentos estruturais. Esta prática é particularmente relevante em setores como o comércio eletrónico, onde pequenas inconsistências nos dados de clientes ou transações podem comprometer recomendações, previsões de procura ou decisões automatizadas[69]

Um exemplo prático de aplicação ocorre na validação de dados de clientes e produtos em plataformas de *e-commerce*, antes de alimentar modelos de recomendação. O *Great Expectations* verifica, de forma automática, a consistência de campos como categorias, preços e histórico de compras, assegurando que o modelo é treinado com informação coerente e atualizada, reduzindo assim erros e promovendo maior fiabilidade das previsões.

4.4 Aplicação no contexto do *e-commerce* europeu

Queremos, então, propor uma possível arquitetura de *compliance* e governação algorítmica aplicável a sistemas de recomendação para *e-commerce* no contexto da UE (figura 7), concebida no alinhamento com os referenciais a *NIST AI Risk Management Framework*, a *Assessment List for Trustworthy AI (UE)* e a *AI Verify*. O objetivo desta proposta parte por assegurar que os mecanismos de recomendação utilizados em ambientes comerciais digitais operem de forma transparente, justa e auditável, respeitando os princípios de confiabilidade, accountability e explicabilidade exigidos pelo *AI Act*.

A *benchmark* adota uma abordagem modular, estruturada em quatro camadas: *Data Governance*, *Model Monitoring*, *Explainability* e *Audit Trail*. Estas abrangem todo o ciclo de vida

[69] https://greatexpectations.io/

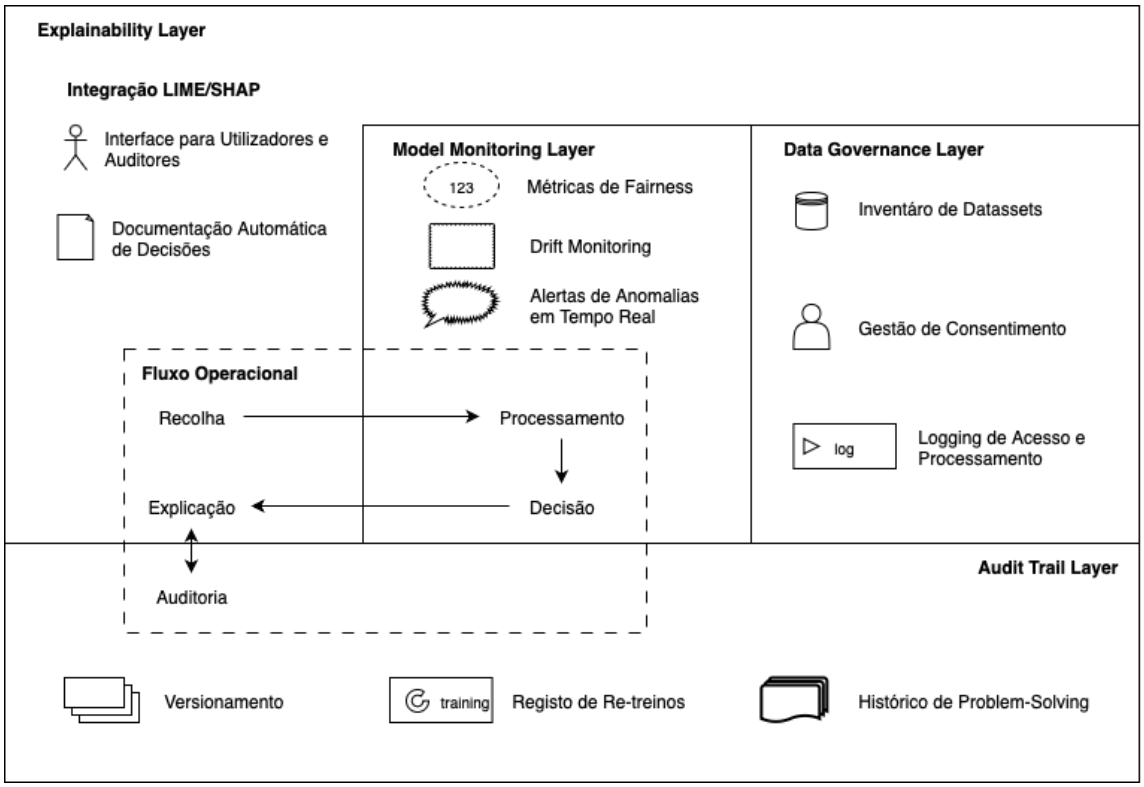
dos dados e modelos. Estas camadas, em permanente comunicação, permitem a rastreabilidade de decisões algorítmicas, o controlo de riscos de enviesamento, a documentação contínua das operações e a verificação independente por auditores.

No contexto do *e-commerce europeu*, a proposta vai para além da conformidade regulatória, com o reforço da confiança do consumidor, a mitigação de riscos reputacionais e a promoção de decisões automatizadas alinhadas com princípios éticos e legais. Assim, o fluxo operacional segue a cadeia:

$$\text{Recolha} \rightarrow \text{Processamento} \rightarrow \text{Decisão} \rightarrow \text{Explicação} \rightarrow \text{Auditoria}$$

Não obstante, querem-se controlos contínuos em cada camada, assegurando que decisões algorítmicas permanecem rastreáveis, explicáveis e auditáveis em tempo real.

Figura 7 – Proposta arquitetura de um sistema de *compliance* de IA no contexto do *e-commerce* europeu



4.4.2 *Data Governance Layer*

Fundamentando os princípios de transparência e rastreabilidade da *NIST AI RMF* e *AL-TAI*, a relativa camada é a base de controlo do ciclo de vida dos dados que, para conformidade com o *GDPR* e *AI Act* (*e.g.* art. 10.º), deve incluir: um catálogo de *datasets*, isto é, um inventário automatizado que identifica a origem, formato, propósito e sensibilidade dos dados; uma gestão de consentimento, por exemplo a partir do motor de *compliance* para recolha e revogação dinâmica de consentimentos; *logging* de acesso e processamento com registos imutáveis para auditorias.

4.4.2 *Model Monitoring Layer*

Para assegurar a *robustness* e a *accountability*, segundo o *AI Verify* e as funções do *NIST RMF* (*MEASURE*, *MANAGE*), é expectável que a referente camada supervisione o desempenho e a equidade do sistema de recomendação, a partir de: métricas de *fairness* como a *Disparate Impact Ratio* e a *Equal Opportunity Difference* que medem desigualdades em recomendações; monitorização de data-drift através do *Kolmogorov–Smirnov test* e do *Population Stability Index*, que avaliam a deriva de dados e modelos; alertas em tempo real com *pipelines* de monitorização contínua com *Prometheus* e *Grafana* para sinalizar anomalias.

4.4.3 *Explainability Layer*

Com vista na redução da assimetria da informação, no reforço da confiança e na garantia de uma transparência algorítmica, querem-se modelos interpretáveis locais com integração de *SHAP* e *LIME* para explicações de cada recomendação; uma *interface* para auditores e utilizadores com *dashboards* e justificações em linguagem natural com métricas de impacto; documentação automática capaz de gerar relatórios técnicos conforme *AI Act* art. 13º e o quarto princípio da *ALTAI* (*Transparency*).

4.4.4 *Audit Trail Layer*

Por fim e forma a cumprir os requisitos de *accountability* e governance previstos na AI *Verify Framework*, a auditoria da arquitetura da figura 7 necessitará rastreabilidade total do histórico do sistema; versionamento via *MLflow*, isto é, registo de modelos, hiperparâmetros e datasets associados; registo de re-treinos com histórico temporal de atualizações de modelos com metadados de validação; gestão de contestações a partir de, por exemplo, um repositório de revisões manuais, decisões revertidas e justificações arquivadas.

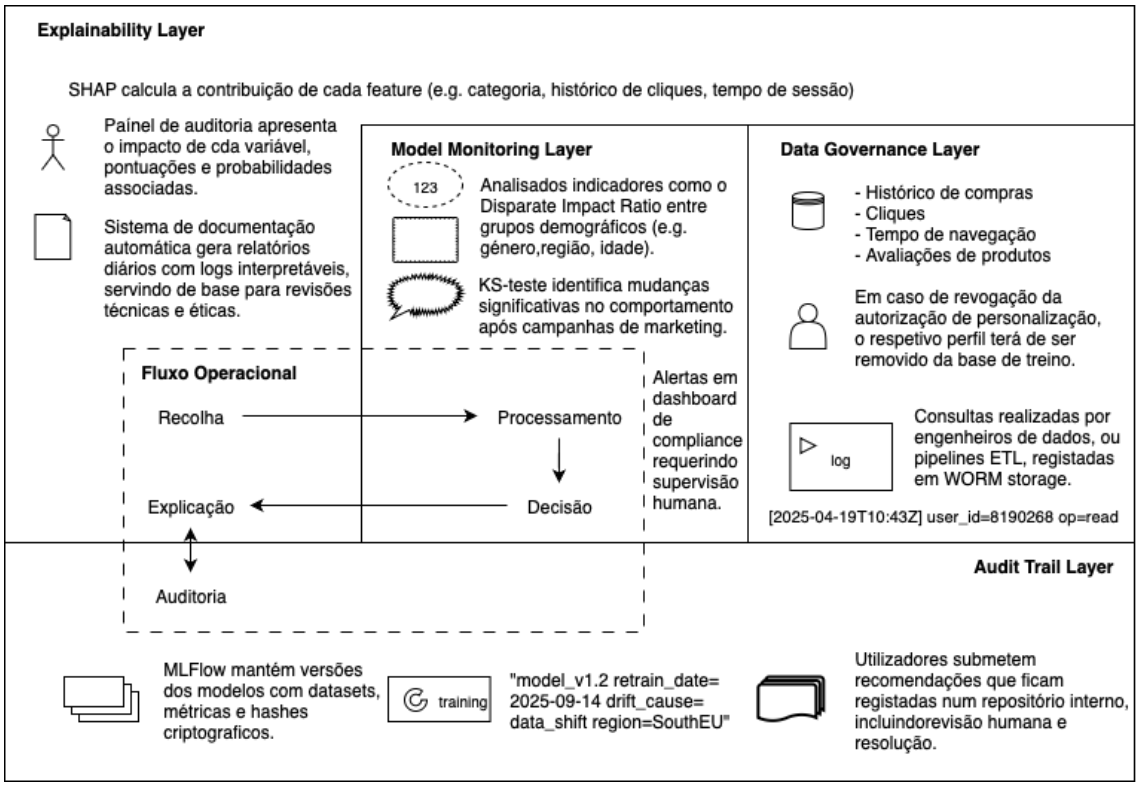
4.5 Caso de estudo simulado

A *Euro eShop* é uma empresa fictícia de *e-commerce* que opera em vários países da União Europeia. A companhia utiliza algoritmos de recomendação para personalizar produtos, preços e campanhas de *marketing*. Com a entrada em vigor do *AI Act*, a empresa precisa de demonstrar governança, transparência e auditabilidade nos seus sistemas de IA.

A loja eletrónica centraliza os seus dados através de num catálogo automatizado de *datasets*, enriquecido com metadados sobre origem, consentimento e finalidade de uso. O modelo de recomendação é sujeito a monotorização contínua, com foco em *fairness*, *performance* e *drift detection*. Cada recomendação apresentada ao utilizador é acompanhada por uma explicação inteligível, como "Este produto foi sugerido porque utilizou produtos similares na categoria 'eletrónica' e avaliou positivamente marcas relacionadas."

A camada de *Audit Trail da figura 7 e 8*, fronteira de todas as restantes camadas, garante a rastreabilidade e integridade histórica de todos os modelos e decisões.

Figura 8– Proposta de arquitetura aplicada ao *e-commerce* europeu



4.5.5 Validação do *framework*

O *framework* da figura surge em conformidade com Art. 9º *(i.e.* Gestão de Riscos) e Art. 12º *(e.g. logging) do AI Act*, com processos de auditoria independente que certificam transparência e ausência de discriminação, bem como a publicação de um relatório de IA responsável trimestral, aumentando confiança do consumidor e valor reputacional.

A arquitetura assegura assim que as decisões algorítmicas são rastreáveis, verificáveis e eticamente alinhadas com os princípios europeus de transparência, justiça e confiabilidade.

5 Diretrizes técnicas para desenvolvimento responsável de IA

As diretrizes apresentadas neste capítulo visam estabelecer uma ponte entre o enquadramento jurídico do *AI Act* e a sua aplicação prática no domínio do comércio digital e de outros contextos de utilização da inteligência artificial. Ao articular requisitos normativos com soluções técnicas e operacionais pretende-se facilitar a implementação responsável e eficaz de sistemas de IA em ambientes empresariais.

Importa salientar que os parâmetros e métricas apresentados possuem carácter indicativo e adaptativo, devendo ser ajustados em função do nível de risco associado, da natureza dos dados tratados e das orientações futuras emanadas pela Comissão Europeia ou pelas autoridades nacionais competentes

5.1 Fundamentos ético-regulatórios; transparência, responsabilidade e explicabilidade

O desenvolvimento e a utilização de sistemas de inteligência artificial suscitam desafios significativos, particularmente no que respeita à transparência, à responsabilidade e à explicabilidade das decisões automatizadas.

A conformidade com o quadro normativo vigente, bem como a promoção de práticas que reforcem a confiança pública, exigem a implementação de medidas estruturadas e consistentes.

Desde logo, é necessário proceder à criação de um registo exaustivo de todo o ciclo de vida do sistema de IA, para assegurar transparência e rastreabilidade.

Conforme salientam SCHOR e BLACKWELL, os referidos registos deverão incluir a definição de objetivos do sistema, descrição dos dados utilizados para treino, metodologias de validação, métricas de desempenho, modelação das decisões e versões dos algoritmos, sendo que este registo serve não apenas para fins de auditoria, mas também para fornecer uma base sólida em caso de litígio ou investigação regulatória[70].

Por outro lado, é necessário proceder à clarificação dos critérios decisórios.

[70] Cfr. SCHOR, B. e BLACKWELL, A. (2024) - *Meaningful Transparency for Clinicians: Operationalising HCXAI Research with Gynaecologists*. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency.*

De acordo com HENN *et al.*, os sistemas de IA devem ser desenvolvidos de forma que os critérios que fundamentam as suas decisões possam ser compreendidos por utilizadores, reguladores e outras partes interessadas, designadamente pacientes médicos[71].

Ainda nesta linha, torna-se necessário que seja clarificado o sistema de gestão e assunção de responsabilidades.

Como salienta ENGSTROM, para mitigar riscos legais e reputacionais, é imperativo definir claramente os papéis e responsabilidades de todas as entidades envolvidas no ciclo de vida da IA, incluindo desenvolvedores, operadores e decisores finais[72]. Desta forma, torna-se necessária a implementação de políticas internas de gestão, que integram a realização de auditorias periódicas e a aplicação de mecanismos de supervisão, o que contribuiu para a *accountability* do sistema, garantindo que as ações ou falhas possam ser assumidas e corrigidas de forma inequívoca.

Assim, a transparência, a responsabilidade e a explicabilidade constituem pilares essenciais para o desenvolvimento de sistemas de IA.

Desta forma, a manutenção de um registo de documentação clara sobre o ciclo de vida dos modelos, mecanismos de divulgação informativa, com atribuição inequívoca de papéis e deveres na sua aplicação, respeito pelas estruturas de *governance* algorítmica com rastreabilidade e contratos que prevejam obrigações específicas, bem como a adoção de modelos interpretáveis ou de ferramentas que viabilizem explicações compreensíveis e úteis e que assegurem a possibilidade de contestação das decisões automatizadas, são essenciais para garantir que a opacidade tecnológica não comprometa os direitos fundamentais nem a tutela jurisdicional efetiva.

5.2 Segurança da informação e proteção de dados pessoais

A acrescer ao exposto no ponto anterior, cumpre ainda salientar que a utilização de sistemas de inteligência artificial implica a recolha, processamento e armazenamento de grandes volumes de dados, frequentemente sensíveis e identificáveis. Destarte, é necessário que

[71] Cfr. HENN, J., VANDEMEULEBROUCKE, T., HATTERSCHEIDT, S., DOHMEN, J., KALFF, J., WYNSBERGHE, A. e MATTHAEI, H. (2025) – *German surgeons' perspective on the application of artificial intelligence in clinical decision-making.* International Journal of Computer Assisted Radiology and Surgery, 20, pp. 825 - 835.

[72] Cfr. ENGSTROM, D. (2020) – *Algorithmic Accountability in the Administrative State.* Yale Journal on Regulation, 37, p. 1.

os sistemas estejam em conformidade com normas de proteção de dados, como o Regulamento Geral sobre a Proteção de Dados (*i.e. RGPD*), sendo ainda essencial a implementação de medidas de segurança da informação, para salvaguardar os direitos fundamentais e preservar a integridade do sistema.

Sobre este aspeto, KINGSTON salienta que os mecanismos de proteção de dados devem harmonizar-se com regimes jurídicos já consolidados, como o Regulamento Geral de Proteção de Dados, mas também antecipar riscos emergentes, como a inferência de informação sensível a partir de padrões aparentemente anónimos[73]. A acrescer ao exposto, a recolha de dados deve ser limitada ao estritamente necessário para atingir os objetivos do sistema de IA, pelo que, a definição clara do propósito do tratamento e a documentação deste processo auxiliam na garantia de que os dados não sejam utilizados de forma abusiva ou desviada.

Como salientam GERKE, MINSSEN e COHEN, este princípio de minimização reduz riscos de exposição indevida e facilita o cumprimento do *RGPD*[74].

Por outro lado, AUGUSTO *et al.* salientam que, sempre que possível, os dados pessoais devem ser pseudonimizados ou anonimizados antes de serem utilizados em processos de treino ou validação de modelos[75].

Neste sentido, a anonimização fortalece a proteção da privacidade, enquanto a pseudonimização permite manutenção de utilidade analítica, mitigando riscos de reidentificação e legalidade.

Ainda nesta linha, quando estamos perante dados que revelam origem racial, opinião política, convicções religiosas ou dados biométricos, devem ser aplicados cuidados adicionais. A sua utilização deve ser justificada legalmente, sujeita a consentimento esclarecido[76] e ainda acompanhada de medidas de mitigação de risco de discriminação ou prejuízo aos indivíduos.

---

[73] Cfr. KINGSTON, J. (2017) – *Using artificial intelligence to support compliance with the general data protection regulation.* Artificial Intelligence and Law, 25, pp. 429 - 443.

[74] Cfr. GERKE, S., MINSSEN, T., & COHEN, G. (2020) – *Ethical and legal challenges of artificial intelligence-driven healthcare.* Artificial Intelligence in Healthcare, p. 295 - 336.

[75] Cfr. AUGUSTO, C., OLIVERO, M., MORÁN, J., MORALES, L., RIVA, C., AROBA, J. e TUYA, J. (2020) – *Test-Driven Anonymization in Health Data: A Case Study on Assistive Reproduction.* 2020 IEEE International Conference On Artificial Intelligence Testing (AITest), pp. 81-82.

[76] Para maior aprofundamento sobre este conceito, *vide* NOGAROLI, Rafaella (2023) – *Responsabilidade civil médica na inteligência artificial: culpa médica e deveres de conduta no século XXI*, Universidade Federal do Paraná, Curitiba, p. 39.

5.3 Guidelines técnico-operacionais

O desenvolvimento responsável de sistemas de inteligência artificial, no contexto do comércio digital, exige a conciliação entre princípios jurídicos e práticas verificáveis. Nesta linha, diversos organismos internacionais e empresas de referência têm proposto modelos de governança e padrões operacionais que complementam o enquadramento normativo europeu. Entre as principais referências destacam-se os *Princípios de IA Responsável da Google*[77] e o *Microsoft Responsible AI Standard*[78], ambos orientados pela transparência, equidade, segurança, privacidade e responsabilidade.

5.3.1 Documentação de dados de treino

O artigo 10.º do *AI Act* estabelece que os sistemas de alto risco devem ser desenvolvidos com *datasets* adequados, relevantes e representativos. Do ponto de vista técnico, este requisito traduz-se na necessidade de documentar exaustivamente os dados utilizados em todas as fases do ciclo de vida da IA, assegurando qualidade, integridade e rastreabilidade.

Com isto em mente, instruímos, sugerindo, a cada atualização significativa do *dataset*, as seguintes técnicas para a implementação responsável de sistemas de compliance de IA.

Ficha técnica do *datasheet*, conforme modelo proposto por Gebru *et al.*[79], contemplando:

i. Motivação: finalidade e contexto da recolha
ii. Composição: tipologia de dados e número de instâncias
iii. Processo de recolha: fontes e metodologia adotada
iv. Pré-processamento: transformações, limpezas ou normalizações aplicadas
v. Distribuição: análise de eventuais desequilíbrios ou lacunas de representatividade

Análise de viés pré-treino, com base nas orientações pré-dispostas[80]:

i. Cálculo de métricas de equidade (*e.g. Disparate Impact Ratio, Equal Opportunity Difference*)
ii. Documentação de sub-representações

[77] https://ai.google/responsibility/principles/

[78] https://www.microsoft.com/en-us/ai/responsible-ai

[79] Cfr. Gebru, T., et al. (2018) - *Datasheets for Datasets*. arXiv preprint arXiv:1803.09010.

[80] Cfr. National Institute of Standards and Technology. (2023) - *Artificial Intelligence Risk Management Framework (AI RMF) 1.0* (NIST AI 100-1). U.S. Department of Commerce, p. 20-24.

iii. Avaliação de variáveis sensíveis ou correlatadas (*e.g. proxy variables*)

Versionamento e integridade dos dados, em conformidade com as práticas de *DataOps*[81]:

i. Utilização de repositórios controlados (*e.g. Git LFS, DVC*)
ii. Manutenção de histórico de alterações (*e.g. changelogs*)
iii. Verificação de integridade através de *hashes* criptografados

5.4 Casos de uso específicos no *e-commerce*

A aplicação prática das diretrizes do regulamento europeu no comércio digital exige atenção na classificação de risco prevista no artigo 6.º, que considera sistemas de IA de alto risco aqueles que afetam significativamente a saúde, segurança ou direitos fundamentais dos indivíduos. Sendo os principais contextos funcionais:

i. Sistemas de recomendação
ii. Modelos de preços dinâmicos
iii. Mecanismos de deteção de fraude
iv. Chatbots de apoio ao cliente

Em cada caso, a conformidade deve ser assegurada por mecanismos de auditoria técnica, métricas de fairness, documentação transparente e supervisão humana contínua.

5.4.1 Sistemas de recomendação

Em sistemas que analisam padrões de consumo e o comportamento de utilizadores para personalizar sugestões de produtos, o risco jurídico situa-se principalmente na eventual discriminação algorítmica e na opacidade das decisões automatizadas. Para mitigar tais riscos, devem ser implementadas medidas que assegurem a rastreabilidade dos dados de entrada e das decisões, a validação periódica de eventuais vieses nos resultados e a disponibilização de explicações interpretáveis, recorrendo às técnicas anteriormente mencionadas.

5.4.2 Sistema de preços dinâmicos

Sistemas de preços dinâmicos utilizam modelos de IA para ajustar valores de produtos em tempo real, com base em variáveis como a procura, o histórico de compras e o perfil

[81] Cfr. IBM Developer. (2020) – *An introduction to the DataOps discipline*. IBM. https://developer.ibm.com/articles/an-introduction-to-the-dataops-discipline/

do utilizador. De acordo com o artigo 5.º e anexo III do *AI Act*, este tipo de sistema pode, em casos, ser classificado como de alto risco, devido ao seu potencial de gerar discriminação económica e de produzir impactos significativos sobre os consumidores.

Do ponto de vista técnico, recomenda-se a validação periódica da equidade e do impacto dos modelos, a definição de limites automáticos de variação de preços, a disponibilização de explicações acessíveis sobre as alterações aplicadas e a implementação de um *dashboard* de monitorização para as equipas de compliance. Ferramentas como *Fairlearn*, *AI Fairness 360*, *MLflow* e o *Explainable AI SDK* podem ser utilizadas para apoiar estas práticas de verificação, auditoria e transparência.

5.4.3 Deteção de fraude

Os sistemas de deteção de fraude baseiam-se em técnicas de *ML* para identificar padrões anómalos em transações financeiras, *outputs* estes que identificam comportamentos suspeitos de forma precoce, no entanto devem cumprir com os afirmados requisitos de robustez técnica, proteção de dados pessoais e justificação transparente das decisões automatizadas.

A norma *ISO/IEC 23894:2023* recomenda a utilização de mecanismos de *anomaly detection* combinados com revisões humanas, de modo a equilibrar a eficiência algorítmica com a supervisão responsável. Para garantir a conformidade e a fiabilidade, é essencial implementar processos de anonimização dos dados de treino, assegurar a verificação sistemática de falsos positivos e falsos negativos e manter um registo auditável de todas as decisões de bloqueio ou sinalização de transações, permitindo rastreabilidade e prestação de contas (*i.e. accountability*).

5.4.4 *Chatbots* de apoio ao cliente

Os *chatbots* constituem sistemas de interação automatizada com utilizadores, predominantes na prestação de apoio, mais concretamente na resposta a pedidos de informação ou na realização de operações básicas em ambientes digitais. De acordo com o artigo 52.º do *AI Act*, é obrigatório fornecer ao utilizador informação clara de que está a interagir com um sistema de inteligência artificial, evitando confusão ou engano quanto à natureza da entidade com quem comunica. Além disso, devem ser mantidos registos auditáveis de *logs* e respostas, assegurada a supervisão humana em interações sensíveis — como as que envolvem decisões pessoais, financeiras ou de saúde — e aplicados filtros de conteúdo e mecanismos de

verificação da consistência informacional, de modo a prevenir respostas incorretas, enviesadas ou potencialmente prejudiciais.

5.5 Checklist de implementação prática

Com base no *EU AI Act Compliance Checker*[82], a verificação de conformidade de sistemas de inteligência artificial pode ser estruturada em quatro fases chave: pré-desenvolvimento, desenvolvimento, pré-implantação e pós-implantação. Seguindo os marcos abordagem, mapeamos os requisitos legais e técnicos do *AI Act* de forma sistemática.

Figura 9 - *Checklist* de implementação pratica com base no *EU AI Act Compliance Checker*

Fases	*Checklists*
Pré-desenvolvimento	✓ Avaliar necessidade de IA, garantindo que a sua utilização é justificada e proporcional ao objetivo. ✓ Classificar risco, determinando se o sistema é de alto risco ou de outra categoria legal. ✓ Avaliar impacto sempre que envolva dados pessoais ou decisões automatizadas. ✓ Definir métricas de equidade e robustez técnica. ✓ Aprovação ética e registo de governança do sistema.
Desenvolvimento	✓ Documentar *datsets*, com *datasheets* que descrevem origem, limitações e autorizações de uso. ✓ Versionar código e modelos, com registo de alterações e documentação técnica. ✓ Testar viéses integradas em *pipelines* CI/CD. Integrar explicabilidade e interpretações claras das decisões algorítmicas. ✓ Registar logs de treino e inferir a garantia da rastreabilidade e auditoria.
Pré-implantação	✓ Auditar internamente conforme análises técnicas, ética e jurísdição). ✓ Testar adversariais em ambiente de validação (*staging*). ✓ Documentar assegurando transparência perante auditorias. ✓ Criar interfaces de transparência e supervisão humana, especialmente para sistemas de alto risco.
Pós-implantação	✓ Monitorizar o desempenho contínuo e detetar desvios emergentes. ✓ Recolher *feedback* e tratar contestações. ✓ Ser auditado periodicamente. ✓ Atualizar documentação e *datasets*.

5.6 Métricas de *compliance*

As métricas de *compliance* surgem como indicadores quantitativos e qualitativos na avaliação se um sistema de IA cumpre com os requisitos legais, éticos e técnicos aplicáveis. Conhecidas pela conversão de princípios abstratos como justiça, transparência, robustez e responsabilidade em medidas objetivas e verificáveis, permitem a conformidade demonstrada de forma mensurável. Vão, ainda, no alinhamento com o nível de risco e a natureza do sistema. De acordo com a *NIST AI RMF* (2023) e a norma *ISO/IEC 42001:2023*, a definição e

[82] https://artificialintelligenceact.eu/assessment/eu-ai-act-compliance-checker/

monitorização destas métricas são essenciais para garantir confiança, rastreabilidade e melhoria contínua dos sistemas de IA. É imperativo que a seleção das métricas seja proporcional ao nível de risco e à natureza do sistema.

As métricas prioritárias incluem:

i. Métricas de equidade (*e.g. fairness*), (*e.g. Equalized Odds, Disparate Impact Ratio)*
ii. Métricas de transparência e explicabilidade (*e.g.* tempo médio de explicação, nível de complexidade do modelo)
iii. Métricas de robustez (*e.g.* resistência a dados adversariais, taxa de erro sob perturbação controlada)
iv. Métricas de responsabilidade (*i.e.* accountability; *e.g.* tempo de resposta a incidentes, número de revisões humanas efetuadas)

A documentação de *compliance* deve justificar a escolha de cada métrica, descrevendo as razões da seleção, os critérios de medição e as fontes normativas de referência. Só assim, o processo assegura rastreabilidade, coerência metodológica e transparência, reforçando a confiança dos utilizadores e facilitando auditorias internas e externas.

5.7 Processos de monitorização contínua

Em conformidade com os artigos 9.º, 13.º e 14.º do *AI Act*, a monitorização contínua configura-se como um componente fundamental da gestão de risco e da responsabilização no desenvolvimento e operação de sistemas de inteligência artificial. Para assegurar a conformidade com os princípios regulatórios e técnicos, o processo deve contemplar mecanismos automáticos de deteção de *drift* de dados e desempenho; revisões periódicas conduzidas por equipas de auditoria interna; canais de *feedback* e contestação de decisões automatizas; *immutable logging* de operações relevantes; atualização programada de modelos e *datasets*, em função das alterações verificadas no contexto operativo e nos requisitos legais ou comerciais.

Queremos salientar a importância da integração de ferramentas *de MLOps* para a esperada operacionalização eficiente do ciclo de vida da IA e da reprodutividade e governança técnica em conformidade com os padrões exigidos.

6 Conclusões

De acordo com STUART RUSSELL e PETER NORVIG, a inteligência artificial é a área que procura desenvolver agentes capazes de perceber o meio envolvente e agir de forma racional para atingir os seus objetivos[83].

Por seu turno, WOLFHART TOTSCHNIG sublinha que, apesar de não existir uma definição universal do termo, a IA pode ser entendida, na prática, como um sistema que combina componentes ciberfísicos e *software*, funcionando de forma autónoma e auto-organizada[84].

Considerando estes aspetos, temos a emergência do Regulamento Europeu aplicável à inteligência artificial (*i.e. AI Act*).

Segundo SABRINA KUTSCHER, o *AI Act* da União Europeia consubstancia uma resposta à evolução digital, ao focar-se diretamente nos sistemas de inteligência artificial, servindo como guia orientador da forma como as autoridades enfrentam o uso crescente da IA, ao criar um quadro de regras harmonizadas para a sua aplicação[85]. De facto, o Regulamento (UE) 2024/1689 representa um verdadeiro marco na consolidação de uma política europeia coerente em matéria de inteligência artificial, com impacto direto e significativo no domínio do comércio digital. A sua génese encontra raízes na estratégia da União Europeia de afirmar-se como líder global na regulação da tecnologia, em equilíbrio entre a promoção da inovação e a salvaguarda de valores fundamentais, como a proteção dos direitos fundamentais, a segurança jurídica e a confiança dos consumidores.

Cremos que a análise realizada ao longo do presente estudo permitiu evidenciar que o Regulamento representa um marco jurídico e também tecnológico de extrema importância para o comércio digital na União Europeia, porquanto, ao estabelecer regras harmonizadas para a utilização de sistemas de inteligência artificial.

Em concreto, o legislador europeu procurou conciliar a necessidade de fomentar a inovação com a exigência de salvaguardar direitos fundamentais, segurança, transparência e confiança nos mercados digitais. Na linha daquele que costuma ser o apanágio do legislador europeu, este regulamento concretiza uma abordagem preventiva e proporcional. Consideramos que a análise desenvolvida ao longo deste trabalho demonstra que o Regulamento não

[83] Cfr. RUSSELL, Stuart e NORVIG, Peter (2022) – *Inteligência artificial: um enfoque moderno*. Rio de Janeiro: Elsevier, p. 5.

[84] Cfr. TOTSCHNIG, Wolfhart (2020) – *Fully Autonomous AI. Science and Engineering Ethics*, 26, pp. 2473 a 2485.

[85] Cfr. KUTSCHER, Sabrina (2025) – *The EU AI Act: Law of Unintended Consequences?*, Technology and Regulation, p. 317.

se limita a uma abordagem normativa clássica, mas introduz um modelo de governação tecnológica orientado pelo risco, estabelecendo obrigações diferenciadas para operadores de sistemas de IA. Esta lógica assegura proporcionalidade e, simultaneamente, cria um ambiente regulatório capaz de reforçar a credibilidade das soluções digitais aplicadas ao comércio.

Esta opção reforça a coerência com princípios da ordem jurídica europeia, designadamente a proteção da dignidade humana, da privacidade e da igualdade, constituindo igualmente um instrumento de harmonização legal dentro do mercado interno.

Em termos estritamente informáticos, cremos que o regulamento gera simultaneamente desafios técnicos e oportunidades.

Os requisitos relativos à explicabilidade, robustez, proteção de dados e supervisão humana colocam uma pressão acrescida sobre os criadores e operadores de IA, impondo-lhes uma adaptação estrutural dos processos de conceção e auditoria tecnológica, bem como no que concerne à integração de mecanismos de transparência, explicabilidade e mitigação de riscos éticos nos sistemas de IA.

Contudo, esses mesmos requisitos podem fortalecer dinâmicas de inovação responsável, aumentando a credibilidade das soluções digitais perante consumidores e parceiros comerciais.

Neste sentido, a construção de *frameworks* de auditoria e *compliance* – conforme analisada no presente estudo – apresenta-se com extrema importância para viabilizar a operacionalização das normas europeias em contexto empresarial. O recurso a modelos já consagrados internacionalmente, como a *NIST Framework*, ou a *ALTAI*, que permitem articular boas práticas com exigências regulatórias.

No entanto, sublinha-se a necessidade de se desenvolver metodologias adaptadas à realidade europeia e ao setor específico do comércio digital.

Importa, por fim, destacar que a eficácia do regime dependerá não apenas da previsão legal apresenta pelas normas, mas também da capacidade de implementar mecanismos técnicos de conformidade contínua e da formação de uma cultura organizacional orientada para a ética digital.

Desta forma, o sucesso do Regulamento dependerá de uma interação dinâmica entre o Direito e a informática, em que o cumprimento legal, a inovação tecnológica e a proteção dos direitos fundamentais se assumem como pilares indissociáveis para o futuro do comércio digital europeu.

Bibliografia

- ALLEN, J. G., LOO, J., & LUNA CAMPOVERDE, J. L. (2025). Governing intelligence: Singapore's evolving AI governance framework. Cambridge Forum on AI: Law and Governance, Cambridge University Press, 9–10.
- ALVES, Ana Paula (2021) – *Responsabilidade Civil do Estado Português por conduta de robô cirúrgico integrado no Serviço Nacional de Saúde.* Lex Medicinae. Revista Portuguesa de Direito da Saúde, 18(36)...
- AUGUSTO, C., OLIVERO, M., MORÁN, J., MORALES, L., RIVA, C., AROBA, J. e TUYA, J. (2020) – *Test-Driven Anonymization in Health Data: A Case Study on Assistive Reproduction.* 2020 IEEE International Conference On Artificial Intelligence Testing (AITest).
- AZEVEDO, F. (2019) – *O Consentimento Informado Silenciado na Esfera da TeleMedicina***.** Cadernos de Lex Medicinae, 1(4).
- BARROS, Gabriel Osório (2024) – *Regulamentação da Inteligência Artificial na União Europeia: Uma análise do AI Act,* Gabinete de Estratégia e Estudos do Ministério da Economia.
- BELK, R. (2020) – *Ethical issues in service robotics and artificial intelligence*. The Service Industries Journal, 41.
- BERTORELLI, Mario Giulio e PRAAS, Robert (2025) – *O Regulamento da IA e o desafio de lidar com uma tecnologia em rápida evolução***,** Disp. *in* https://voltportugal.org/noticias/o-regulamento-da-ia-e-o-desafio-de-lidar-com-uma-tecnologia-em-rapida-evolucao**.**
- CARRY, Cameron F. (2023) – *Commentary NIST's AI Risk Management Framework plants a flag in the AI*, Cfr. https://www.brookings.edu/articles/nists-ai-risk-management-framework-plants-a-flag-in-the-ai-debate.
- COMUNICAÇÃO DA COMISSÃO AO PARLAMENTO EUROPEU, AO CONSELHO, AO COMITÉ ECONÓMICO E SOCIAL EUROPEU E AO COMITÉ DAS REGIÕES, 8.4.2019 COM(2019) 168 final, Disp. in https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:52019DC0168&from=EN.
- CRUZ, R., BERTOLLO, D. e CAMARGO, M. (2020) – *O Impacto da Inteligência Artificial na Auditoria: Uma Revisão Bibliográfica. XX Mostra de Iniciação Científica.* UCS-PPGA.
- DANTAS, E., NOGAROLI, R. (2020). *Consentimento informado do paciente frente às novas tecnologias da saúde (teleMedicina, cirurgia robótica e inteligência artificial).* Lex Medicinae. Revista Portuguesa de Direito da Saúde, 17(33), 25-63.
- DASGUPTA, P., JONES, A., GILL, I. (2004). *Robotic urological surgery: a perspective*. BJU International, 95(1).
- DEVELOPER, IBM. (2020) - *An introduction to the DataOps discipline.* IBM.
- HAMON, R., JUNKLEWITZ, H., & SANCHEZ, I. (2020) - *Robustness and explainability of artificial intelligence***.** Technical Report, Publications Office of the European Union, Luxembourg.

- ENGSTROM, D. (2020) – *Algorithmic Accountability in the Administrative State*. Yale Journal on Regulation, 37.
- GAON, A. (2021) – *Reframing the artificial intelligence concept*. The Future of Copyright in the Age of Artificial Intelligence.
- GEBRU, T., et al. (2018) - *Datasheets for datasets*. arXiv preprint arXiv:1803.09010.
- GERKE, S., MINSSEN, T., & COHEN, G. (2020) – *Ethical and legal challenges of artificial intelligence-driven healthcare*. Artificial Intelligence in Healthcare.
- HAENLEIN, M. e KAPLAN, A. (2019) – *A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence*. California Management Review, 61.
- HERRON, M. (2020) - *A civil liability regime for AI? Lexology.*
- HENN, J., VANDEMEULEBROUCKE, T., HATTERSCHEIDT, S., DOHMEN, J., KALFF, J., WYNSBERGHE, A. e MATTHAEI, H. (2025) – *German surgeons' perspective on the application of artificial intelligence in clinical decision-making.* International Journal of Computer Assisted Radiology and Surgery, 20.
- IQBAL, U., CELI, L., & LI, Y. (2020) – *How Can Artificial Intelligence Make Medicine More Preemptive?* Journal of Medical Internet Research.
- JOBIN, A., IENCA, M., & VAYENA, E. (2019) - *The global landscape of AI ethics guidelines.* Nature Machine Intelligence, 1(9), 389–399.
- KAMILA, M. e JASROTIA, S. (2023) – *Ethical issues in the development of artificial intelligence:* KAMILA, M. e JASROTIA, S. (2023) – *Ethical issues in the development of artificial intelligence: recognizing the risks.* International Journal of Ethics and Systems.
- KINGSTON, J. (2017) – *Using artificial intelligence to support compliance with the general data protection regulation.* Artificial Intelligence and Law, 25.
- KIRILLOVA, Elena, BLINKOV, Oleg, OGNEVA, Natalija, VRAZHNOV, AAeksey, & SERGEEVA, Natal'ja (2020) – *Artificial Intelligence as a New Category of Civil Law.* Journal of Advanced Research in Law and Economics, 11.
- KUTSCHER, Sabrina (2025) – *The EU AI Act: Law of Unintended Consequences?,* Technology and Regulation.
- MACLEAN, Don (2017) – *The NIST Risk Management Framework: Problems and recommendations.* Cyber Security: A Peer-Reviewed Journal.
- MAIA, Pedro (2021) – *Compliance bancário na era da inteligência artificial — uma breve introdução,* Revista Julgar, n.º 45.
- MANNING, L. (2020) – *Moving from a compliance-based to an integrity-based organizational climate in the food supply chain.* Comprehensive reviews in food science and food safety, 19 3
- MARQUES, Júlio e VIEIRA, Pablo de Abreu (2023) – *Explorando a Explicabilidade da Inteligência Artificial - Técnicas para Compreender e Interpretar Modelos de Aprendizado de Máquina.*

- MASSA, Mariana, FLORES, Cláudio e SANTOS, Ricardo (2025) – *Inteligência Artificial (IA) no Comércio Digital: Oportunidades e Desafios*, Centro de Inovação e Investigação em Ciências Empresariais e Sistemas de Informação, Escola Superior de Tecnologia e Gestão / Politécnico do Porto.
- MOINGEON, P., KUENEMANN, M., & GUEDJ, M. (2021) – *Artificial Intelligence-enhanced drug design and development: toward a computational precision medicine.* Drug discovery today.
- MORANDÍN-AHUERMA, Fabio (2022) – *What is Artificial Intelligence?.* International Journal of Research Publication and Reviews.
- NELSON, Scott, WALSH, Colin, OLSEN, Casey, MCLAUGHLIN, Andre, LEGRAND, Joseph, SCHUTZ, Nick, & LASKO, Thomas (2020) – *Demystifying artificial intelligence in pharmacy.American journal of health-system pharmacy*: AJHP: official journal of the American Society of Health-System Pharmacists.
- NOGAROLI, Rafaella (2023) – *Responsabilidade civil médica na inteligência artificial: culpa médica e deveres de conduta no século XXI*, Universidade Federal do Paraná, Curitiba.
- NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. (2023) - *Artificial Intelligence Risk Management Framework (AI RMF) 1.0 (NIST AI 100-1).* U.S. Department of Commerce.
- OCDE (2024) – *OECD AI Principles overview*, OECD.AI Policy Observatory, Cfr. https://oecd.ai/en/ai-principles.
- RIBEIRO, Pedro Santiago e CORREIA, Catarina Camacho (2025) – *Regulamento de IA: desafios e oportunidades das novas normas já em vigor*, Disp. *in* https://www.pwc.pt/pt/sala-imprensa/artigos-opiniao/2025/regulamento-inteligencia-artificial.html.
- RADCLYFFE, C., & NODDEL, R. (2020) - *Ethical by design: Measuring and managing digital ethics in the enterprise.*
- ROSE, Scott D. (2021) – *Planning for a Zero Trust Architecture.*
- RUSSELL, Stuart e NORVIG, Peter (2022) – *Inteligência artificial: um enfoque moderno.* Rio de Janeiro: Elsevier.
- SÆTRA, Henrik Skaug (2020) – *A shallow defence of a technocracy of artificial intelligence: Examining the political harms of algorithmic governance in the domain of government.* Technology in Society, 62.
- SAURA, J., RIBEIRO-SORIANO, D. e PALACIOS-MARQUÉS, D. (2022) – *Assessing behavioral data science privacy issues in government artificial intelligence deployment.* Gov. Inf. Q., 39, 101679.
- SCHOR, B. e BLACKWELL, A. (2024). Meaningful Transparency for Clinicians: Operationalising HCXAI Research with Gynaecologists**.** *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency.*
- SHEIKIN, A. (2024) – *Prohibited Artificial Intelligence Practices In The Legislation Of The European Union.* LEGAL ORDER: History, Theory, Practice.

- STETTINGER, G., WEISSENSTEINER, P. (2024) – *Trustworthiness Assurance Assessment for High-Risk AI-Based Systems*. IEEE Access, 12, pp. 22718–22745.
- TOTSCHNIG, Wolfhart (2020) – *Fully Autonomous AI. Science and Engineering Ethics*, 26.
- WEN, J., ZHANG, Z., LAN, Y., CUI, Z., CAI, J., & ZHANG, W. (2022) – *A survey on federated learning: challenges and applications. International Journal of Machine Learning and Cybernetics*, 14.
- WOLTERS KLUWER (2024) – *Solucionando o enigma: Aplicando a IA Generativa em Atividades de Auditoria Interna*, Internal Audit Foundation.
- YU, C., LIN, Y., Lin, C., LIN, S., WU, J., & CHANG, S. (2020) – *Development of an Online Health Care Assessment for Preventive Medicine: A Machine Learning Approach.* Journal of Medical Internet Research.
- ZAFAR, M. R., & KHAN, N. M. (2019) - *DLIME: A deterministic local interpretable model-agnostic explanations approach for computer-aided diagnosis systems. arXiv preprint, 1.*
- SALIH, A. M., RAISI-ESTABRAGH, Z., BOSCOLO GALAZZO, I., RADEVA, P., PETERSEN, S. E., LEKADIR, K., & MENEGÁZ, G. (2025) - *A perspective on explainable artificial intelligence methods: SHAP and LIME*. Advanced Intelligent Systems, 7(1). DOI: 10.1002/aisy.202400304.

Webgrafia

- https://ai.google/responsibility/principles/
- https://artificialintelligenceact.eu/assessment/eu-ai-act-compliance-checker/
- https://commission.europa.eu/document/download/d2ec4039-c5be-423a-81ef-b9e44e79825b_pt?filename=commission-white-paper-artificial-intelligence-feb2020_pt.pdf
- https://cnnportugal.iol.pt/eua/china/investigador-portugues-sobre-inteligencia-artificial-os-eua-vao-continuar-a-inovar-a-china-vai-continuar-a-copiar-e-a-europa-vai-continuar-a-regular/20231209/65747668d34e65afa2f874dd
- https://digital-strategy.ec.europa.eu/en/library/communication-artificial-intelligence-europe
- https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
- https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:12016P/TXT
- https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32012R1025
- https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679
- hthttps://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32022R2065
- https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CLEX:52022PC0496&from=EN
- https://eur-lex.europa.eu/legal-content/PT/TXT/uri=CELEX%3A32024R1689
- https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32024R1689
- https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689
- https://eur-lex.europa.eu/legal-content/PT/TXT/ uri=CELEX%3A52021DC0205
- https://fairlearn.org/
- https://greatexpectations.io/
- https://mlflow.org/
- https://op.europa.eu/pt/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1
- https://paanalytics.net/blog/interpretacao-de-shap-e-lime-em-machine-learning
  https://www.nist.gov/itl/ai-risk-management-framework
- https://research.ibm.com/blog/ai-fairness-360

- https://www.deloitte.com/br/pt/services/consulting-risk/analysis/desafios-governanca-uso-inteligencia-artificial-pelas-empresas.html
- https://www.microsoft.com/en-us/ai/responsible-ai
- https://www.mlflow.org/docs/latest/ml/model-registry
- https://www.iso.org/standard/81230.html
- https://www.nist.gov/itl/ai-risk-management-framework
- https://www.prnewswire.com/news-releases/deloitte-introduces-trustworthy-ai-framework-to-guide-organizations-in-ethical-application-of-technology-in-the-age-of-with-301118495.html