

Regular a Inteligência Artificial nos Cuidados de Saúde Públicos para Prevenir o Viés de Gênero e Racial

**Um Modelo Regulatório para a IA Biomédica
Alinhado com a Carta dos Direitos Fundamentais**

Documento de política



Co-funded by
the European Union



Introdução

A inteligência artificial (IA) está cada vez mais integrada nos sistemas de saúde em toda a Europa, apoiando o diagnóstico, a tomada de decisões clínicas, o planeamento de tratamentos e a previsão de risco dos doentes. Estas tecnologias oferecem oportunidades consideráveis para melhorar a eficiência e os resultados dos cuidados de saúde. Simultaneamente, evidências crescentes demonstram que **os sistemas de IA podem reproduzir ou amplificar as desigualdades existentes nos cuidados de saúde**.¹

Vários mecanismos podem conduzir a resultados enviesados na tomada de decisões apoiada por IA nos cuidados de saúde. Estes incluem desequilíbrios nos dados de treino, erros de medição em variáveis clínicas, indicadores proxy inadequados, e diferenças entre os ambientes de desenvolvimento e os contextos reais de implementação.² O viés pode também surgir quando os sistemas de IA são introduzidos nos fluxos de trabalho clínicos sem formação adequada para os profissionais de saúde ou sem mecanismos substanciais de supervisão humana e de responsabilização para monitorizar o seu desempenho em diferentes populações de doentes.³

Além disso, a investigação empírica tem demonstrado que **estes vieses podem ter consequências graves em contextos clínicos**. Estudos têm reportado disparidades persistentes no diagnóstico e tratamento de condições importantes. As mulheres e os doentes de minorias raciais/étnicas com doença cardiovascular recebem frequentemente um diagnóstico e tratamento mais tardios ou menos intensivos em comparação com homens e doentes de raça branca,⁴ e as diferenças raciais e étnicas afetam os cuidados e os resultados na diabetes,⁵ enquanto condições de saúde mental como a depressão são subdiagnosticadas e subtratadas tanto em mulheres como em

¹ Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y., *Inteligência artificial nos cuidados de saúde: passado, presente e futuro*, *Stroke and Vascular Neurology* 2(4): 230–243, 2017.

² Vokinger, K. N., Feuerriegel, S., & Kesselheim, A. S., *Mitigar o viés na aprendizagem automática em medicina*, *Communications Medicine* 1, 25, 2021.

³ Panesar, S., Kliot, M., Parrish, R., Fernandez-Miranda, J., Cagle, Y., & Britz, G.W., *Promessas e perigos da inteligência artificial em neurocirurgia*, *Neurosurgery* 87(1): 33–44, 2020.

⁴ Balla, S., Gomez, S.E., & Rodriguez, F., *Desigualdades nos cuidados e nos resultados cardiovasculares em mulheres de minorias raciais/étnicas*, *Current Treatment Options in Cardiovascular Medicine* 22(12):75, 2020.

⁵ Young, C. & Myers, A. K., *Desigualdades raciais e étnicas nos cuidados clínicos e na gestão da diabetes: uma revisão narrativa*, *Endocrine Practice* 29(4): 295–300, 2023.

populações minoritárias.⁶ Um exemplo bem documentado é um estudo exploratório no *Journal of Medical Internet Research* que identificou disparidades de desempenho baseadas no sexo em algoritmos de aprendizagem automática utilizados para a previsão de doença cardíaca, constatando, em particular, que as doentes do sexo feminino estavam consistentemente sub-representadas nos conjuntos de dados de treino e que os modelos apresentavam taxas mais elevadas de falsos negativos para as mulheres, demonstrando como dados enviesados podem contribuir para um **desempenho diagnóstico desigual**.⁷ Alargando esta preocupação à equidade racial, outro estudo verificou que modelos de IA treinados para prever a gravidade da depressão a partir de linguagem em redes sociais tiveram um bom desempenho para participantes de raça branca, mas um desempenho fraco para participantes negros, revelando disparidades de desempenho baseadas na raça que refletem diferenças nos padrões de expressão não captadas pelos modelos.⁸

Para enfrentar estes riscos de viés, são necessários **modelos regulatórios que traduzam princípios éticos e legais internacionais e europeus em mecanismos de governação concretos** que as instituições de saúde possam implementar na prática. Na União Europeia (UE), o panorama regulatório evoluiu rapidamente com a adoção do Regulamento de Inteligência Artificial (AI Act), que estabelece obrigações para os sistemas de IA de risco elevado e para os responsáveis pela sua implantação, como os hospitais.⁹ Contudo, a aplicação destes requisitos legais como procedimentos práticos no seio das instituições de saúde continua a ser um desafio significativo.

Este documento de política propõe **um modelo regulatório para a governação da IA biomédica que se alinha com o acervo da UE**, incluindo a Carta dos Direitos Fundamentais da UE, e que foi concebido para os sistemas públicos de saúde e para as organizações da sociedade civil (OSC). O modelo proposto está ancorado em instrumentos regulatórios europeus, incluindo o Regulamento de IA (AI Act), o Regulamento dos Dispositivos Médicos (MDR), o Regulamento dos Dispositivos Médicos para Diagnóstico In Vitro (IVDR), o Regulamento Geral sobre a Proteção de Dados (RGPD), e o emergente Espaço Europeu de Dados de Saúde (EHDS). Traduz princípios de

⁶ Sorkin, D.H., Ngo-Metzger, Q., Billimek, J., August, K.J., Greenfield, S., & Kaplan, S.H., *Subdiagnóstico e subtratamento da depressão em doentes com diabetes tipo 2 de diversas origens raciais/étnicas*, *Diabetes Care* 34(3): 598–600, 2011.

⁷ Straw, I., Rees, G., & Nachev, P., *Disparidades de desempenho baseadas no sexo em algoritmos de aprendizagem automática para a previsão de doença cardíaca: Estudo Exploratório*, *Journal of Medical Internet Research* 26: e46936, 2024.

⁸ Rai, S., Stade, E. C., Giorgi, S., Francisco, A., Ungar, L. H., Curtis, B., & Guntuku, S. C., *Marcadores linguísticos-chave da depressão nas redes sociais dependem da raça*, *Proceedings of the National Academy of Sciences* 121(14): e2319837121, 2024.

⁹ Regulation (EU) 2024/1689 of the European Parliament and of the Council, *Criação de regras harmonizadas em matéria de inteligência artificial (AI Act)*, PE/24/2024/REV/1, Official Journal of the European Union, 13 June 2024.

direitos fundamentais, em particular a dignidade humana, a igualdade, a não discriminação e o direito aos cuidados de saúde, em salvaguardas práticas aplicáveis a todo o ciclo de vida dos sistemas de IA. As referências a exemplos selecionados de jurisdições fora da UE são incluídas apenas quando fornecem ensinamentos transferíveis e destinam-se a complementar o quadro jurídico e baseado em direitos europeu que sustenta o modelo.

Adicionalmente, o modelo regulatório baseia-se em abordagens regulatórias já estabelecidas noutros setores de risco elevado, incluindo a gestão da segurança na aviação, a governação do risco de modelos no setor financeiro, a farmacovigilância na regulação dos medicamentos, e as avaliações de impacto algorítmico na administração pública. Estas abordagens fornecem ensinamentos valiosos para a construção de quadros de governação capazes de responder aos complexos riscos associados ao ciclo de vida da IA biomédica, e foram aqui adaptadas para incorporar explicitamente considerações de equidade de género e racial ao longo de todo o ciclo de vida da IA.

O modelo fornece às instituições de saúde e às OSC **um quadro de governação prático que abrange a aquisição, a validação, a implantação, a monitorização e a responsabilização** relativamente aos sistemas de IA biomédica. Ao incorporar proteções de direitos fundamentais em processos operacionais, este quadro visa reduzir os riscos de viés de género e racial na tomada de decisões em saúde apoiada por IA e contribuir para resultados de saúde mais equitativos em toda a UE. O documento oferece ainda **um conjunto de políticas, ferramentas e artefactos práticos que os hospitais e as OSC podem adaptar a diferentes contextos nacionais**, de forma a implementar o modelo de forma consistente em contextos reais.

O Quadro Jurídico Europeu para a IA nos Cuidados de Saúde

A UE estabeleceu um ambiente regulatório multifacetado para a IA biomédica, composto por quatro camadas jurídicas complementares: o Regulamento de IA (AI Act), o direito dos dispositivos médicos e a orientação relacionada, a proteção de dados, e a governação dos dados de saúde.

O AI Act estabelece um quadro regulatório baseado no risco para os sistemas de IA. Muitas aplicações na área da saúde, incluindo ferramentas de apoio à decisão clínica e algoritmos de diagnóstico, integram a categoria dos sistemas de IA de risco elevado, os quais devem cumprir requisitos rigorosos relativos à gestão do risco, à governação de dados, à transparência, à supervisão humana, e à exatidão. Os requisitos aplicáveis incluem também a identificação e a mitigação dos riscos de resultados discriminatórios através de uma análise desagregada das populações, tratando o sexo e o género, a origem étnica e outras categorias sociais como variáveis transversais. **As instituições de saúde que implantam sistemas de IA têm também obrigações específicas nos termos**

Financiado pela União Europeia. Os pontos de vista e as opiniões expressas são no entanto apenas os do(s) autor(es) e não refletem necessariamente os da União Europeia ou da Agência de Execução Europeia do Trabalho, da Educação e da Cultura (EACEA). Nem a União Europeia nem a autoridade que concede a subvenção podem ser consideradas responsáveis por eles.

Código do projeto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

do AI Act. Estas incluem assegurar que os sistemas são utilizados em conformidade com as instruções do fornecedor, monitorizar o seu desempenho, manter registos operacionais, e comunicar incidentes graves às autoridades competentes, nomeadamente quando esses incidentes possam afetar de forma desproporcionada diferentes populações de doentes.

No que respeita à aplicação do AI Act, as práticas proibidas e algumas regras de supervisão e governação já se aplicam desde 2025.¹⁰ O próximo marco importante é agosto de 2026, altura em que a maioria das obrigações previstas no regulamento se torna aplicável (incluindo os sistemas de risco elevado do Anexo III e os deveres de transparência do Art. 50.º). A IA de risco elevado incorporada em produtos regulamentados (o que inclui muitos sistemas de IA em dispositivos médicos) só se torna aplicável a partir de agosto de 2027.

Para os hospitais, as disposições do AI Act mais relevantes na prática são:

- **Requisitos para sistemas de risco elevado.** Os fornecedores devem implementar um sistema de gestão do risco ao longo do ciclo de vida (Art. 9.º) e práticas de governação de dados que incluam: requisitos explícitos de deteção e mitigação de viés (Art. 10.º), incluindo a avaliação da representatividade dos dados e de potenciais disparidades de desempenho entre populações através de análise desagregada por sexo e género e por origem racial ou étnica; salvaguardas de transparência e disponibilização de informação aos responsáveis pela implantação (Art. 13.º); supervisão humana concebida para prevenir o viés de automação e permitir procedimentos de anulação ou interrupção (Art. 14.º); e controlos de exatidão, robustez e cibersegurança, com atenção explícita a ciclos de retroação e a ataques específicos a sistemas de IA (Art. 15.º), os quais podem, de outro modo, agravar as desigualdades existentes se não forem devidamente abordados.
- **Obrigações dos responsáveis pela implantação.** Os hospitais e outros responsáveis pela implantação devem utilizar os sistemas de IA de risco elevado em conformidade com as instruções do fornecedor (Art. 26.º); atribuir uma supervisão humana competente (Art. 26.º); garantir que os dados de entrada são representativos e completos quando o responsável pela implantação controla esses dados (Art. 26.º), considerando também (sempre que viável) a diversidade demográfica das populações clínicas; monitorizar o funcionamento do sistema (Art. 26.º); conservar os registos gerados automaticamente durante, pelo menos, seis meses (Art. 26.º); e comunicar, em tempo útil, os incidentes graves

¹⁰ European Commission, *Calendário de implementação do AI Act da UE*, Brussels: European Commission, 2024.

aos fornecedores e às autoridades competentes (Arts. 26.º e 73.º), particularmente quando exista risco de resultados clínicos desiguais.

- **Avaliação de Impacto sobre os Direitos Fundamentais (FRIA).** Antes de implantar determinados sistemas de IA de risco elevado, os organismos regidos pelo direito público (e as entidades privadas que prestam serviços públicos) devem realizar uma FRIA, descrevendo os processos, os grupos afetados, os riscos, as medidas de supervisão humana e os mecanismos de mitigação ou de reclamação, devendo posteriormente notificar a autoridade de fiscalização do mercado (Art. 28.º). Este requisito é diretamente relevante para os hospitais públicos em muitos Estados-Membros, na medida em que fornece um mecanismo para identificar e abordar, antes da implantação, os riscos de discriminação e de impactos desiguais, avaliados através de análise desagregada por sexo e género e por raça/etnicidade.

Para além do AI Act, **os sistemas de IA utilizados em contextos médicos estão também sujeitos ao direito dos dispositivos médicos e à orientação relacionada**, incluindo o *Regulamento dos Dispositivos Médicos (MDR)*¹¹ e o *Regulamento dos Dispositivos Médicos para Diagnóstico In Vitro (IVDR)*,¹² que regulam a segurança e o desempenho das tecnologias médicas. Estes regulamentos exigem avaliação clínica e vigilância pós-comercialização para o software de dispositivos médicos, incluindo ferramentas baseadas em IA, com atenção ao desempenho equitativo entre populações de doentes desagregadas, sempre que os dados o permitam. Adicionalmente, a *Orientação sobre a Qualificação e Classificação de Software ao abrigo do MDR/IVDR*, do Grupo de Coordenação de Dispositivos Médicos (MDCG), fornece diretrizes sobre como qualificar e classificar software nos termos do MDR e do IVDR.¹³ Esta sublinha que a classificação depende da finalidade pretendida, e não da localização do software (por exemplo, baseado na cloud ou integrado), e inclui exemplos retirados de sistemas de informação hospitalar e de software de apoio à decisão clínica, destacando a necessidade de considerar impactos diferenciados em diversas populações de doentes. A *Orientação sobre a Avaliação Clínica e de Desempenho de Software de Dispositivos Médicos* do MDCG fornece apoio adicional, reforçando a expectativa de que a evidência deve ser proporcional ao papel clínico do software, incluindo evidência de que o desempenho da

¹¹ European Union, *Regulamento (UE) 2017/745 relativo aos Dispositivos Médicos (MDR)*, Official Journal of the European Union, L117, Brussels: EU, 2017.

¹² European Union, *Regulamento (UE) 2017/746 relativo aos Dispositivos Médicos de Diagnóstico In Vitro (IVDR)*, Official Journal of the European Union, L117, Brussels: EU, 2017.

¹³ Medical Device Coordination Group, *Orientação sobre a Qualificação e a Classificação de Software no âmbito do Regulamento (UE) 2017/745 (MDR) e do Regulamento (UE) 2017/746 (IVDR)*, Brussels: MDCG, 2019.

IA é consistente entre populações diferenciadas por sexo e género e por origem racial ou étnica.¹⁴

A relação entre o AI Act e o direito dos dispositivos médicos existente é clarificada na *Orientação sobre a Interação entre o AI Act e o MDR/IVDR*.¹⁵ O documento explica que o AI Act se aplica em conjunto com estes dois regulamentos. Na prática, isto significa que **muitos sistemas de IA médicos têm de cumprir ambos os quadros simultaneamente**. A orientação esclarece também quando a IA médica se qualifica como de risco elevado nos termos do AI Act, o que normalmente ocorre quando o próprio sistema de IA constitui um dispositivo médico (ou uma componente de segurança de um dispositivo médico) e está, por isso, sujeito a uma avaliação de conformidade por terceiros realizada por um organismo notificado. Para a aquisição hospitalar, isto tem implicações práticas. Os hospitais devem exigir que os fornecedores demonstrem o estatuto regulatório dos seus sistemas, incluindo a classificação MDR/IVDR, o envolvimento de um organismo notificado quando aplicável, e a conformidade com os requisitos de risco elevado do AI Act, bem como a documentação do desempenho por grupos populacionais desagregados e das medidas de mitigação do viés, quando relevante. Os contratos de aquisição devem também exigir que os fornecedores disponibilizem a documentação, a informação de transparência e as medidas de supervisão necessárias aos responsáveis pela implantação nos termos do AI Act, incluindo disposições que garantam resultados equitativos para todos os doentes.

Vários **outros instrumentos da UE são igualmente relevantes para a governação prática da IA na saúde**. O *Regulamento Geral sobre a Proteção de Dados (RGPD)* mantém-se como uma restrição fundamental para a IA biomédica, dado que os dados de saúde constituem dados pessoais de categoria especial, e por isso as implantações hospitalares requerem frequentemente *Avaliações de Impacto sobre a Proteção de Dados (DPIA)*. O AI Act (Arts. 26.º, n.º 9, e 27.º, n.º 4) associa explicitamente a prática dos responsáveis pela implantação às obrigações de DPIA (utilização da informação do fornecedor para apoiar a DPIA) e posiciona a FRIA como complementar à DPIA, podendo esta incluir a consideração de impactos diferenciados entre populações diferenciadas por sexo e género e por origem racial/étnica. Além disso, o *Espaço Europeu de Dados de Saúde (EHDS)* visa facilitar o acesso seguro aos dados de saúde para a prestação de cuidados e para a investigação, mantendo simultaneamente salvaguardas robustas de proteção de dados.¹⁶ Espera-se que o EHDS desempenhe um papel fundamental na criação de conjuntos de dados (datasets) mais representativos para o desenvolvimento da IA

¹⁴ Medical Device Coordination Group, *Orientação sobre a Avaliação Clínica (MDR) e a Avaliação de Desempenho (IVDR) de Software em Dispositivos Médicos*, Brussels: MDCG, 2020.

¹⁵ Medical Device Coordination Group & European AI Board, *Orientação sobre a Interação entre o MDR/IVDR e o AI Act*, Brussels: European Commission, 2025.

¹⁶ European Union, *Regulamento relativo ao Espaço Europeu de Dados de Saúde (EHDS)*, Brussels: European Union, 2024.

biomédica, apoiando a avaliação centrada na equidade e reduzindo o risco de sub-representação de populações de doentes com determinadas características demográficas.¹⁷

Por último, no que respeita à cibersegurança, a Agência da União Europeia para a Cibersegurança produziu orientações sobre práticas de cibersegurança para a IA. Estas defendem que as normas organizacionais já existentes (por exemplo, a gestão da segurança e da qualidade ao abrigo da ISO) podem ser adaptadas a contextos de IA.¹⁸ A *Orientação sobre Cibersegurança para Dispositivos Médicos* do MDCG descreve a forma como os fabricantes devem cumprir os requisitos de cibersegurança do MDR e do IVDR através da gestão do risco e de expectativas mínimas de cibersegurança.¹⁹ É importante notar que uma governação sólida da cibersegurança sustenta uma implantação equitativa da IA: proteger a integridade dos dados, o acesso e a fiabilidade do sistema ajuda a prevenir a amplificação não intencional das desigualdades de género e raciais/étnicas.

Em conjunto, estes instrumentos jurídicos criam um ambiente regulatório complexo no qual os sistemas de IA na saúde devem operar. O modelo regulatório de IA para hospitais públicos europeus aqui proposto foi concebido explicitamente em torno destas quatro camadas jurídicas fundamentais que regem a prática da IA na saúde. Em conformidade, o modelo prevê mecanismos para monitorizar e mitigar impactos desiguais entre populações diferenciadas por sexo e género e por origem racial ou étnica ao longo de todo o ciclo de vida da IA.

Quadros de Governação Europeus e Internacionais Aplicáveis à IA Responsável nos Cuidados de Saúde

É importante notar que os quadros regulatórios e de normalização europeus discutidos anteriormente constituem o fundamento jurídico primário para a governação da IA biomédica no seio da União Europeia. Para além do ambiente regulatório da UE para a IA biomédica, vários instrumentos internacionais não vinculativos fornecem orientação para a gestão dos riscos da IA. É de notar que os **quadros de governação adicionais aqui apresentados são incluídos para demonstrar técnicas de implementação que podem ser transferidas para o contexto da UE, permanecendo, no entanto, plenamente**

¹⁷ European Commission, *Espaço Europeu de Dados de Saúde (EHDS)*, Brussels: European Commission, 2025.

¹⁸ European Union Agency for Cybersecurity (ENISA), *Cibersegurança da IA e Normalização*, Athens: ENISA, 2023.

¹⁹ Medical Device Coordination Group, *MDCG 2019-16 Rev. 1 – Orientação sobre Cibersegurança para Dispositivos Médicos*, Brussels: European Commission, 2020.

subordinados aos requisitos jurídicos europeus e aos princípios dos direitos fundamentais. São utilizados de forma seletiva como analogias de outras jurisdições onde foram criados, na prática, mecanismos de governação do ciclo de vida comparáveis, particularmente em áreas como a gestão do risco de modelos, a auditoria algorítmica e a avaliação de impacto institucional.

Por exemplo, o *Quadro de Gestão do Risco de IA* (AI RMF) do National Institute of Standards and Technology (NIST) oferece uma abordagem abrangente para a gestão dos riscos da IA, através de uma abordagem ao ciclo de vida organizada em torno da governação, do mapeamento do risco, da medição do risco e da gestão do risco.²⁰ O quadro reconhece explicitamente a equidade e o combate ao viés prejudicial como dimensões centrais da IA de confiança. Adicionalmente, como parte do mapeamento e da medição do risco, encoraja as organizações a identificar e a avaliar impactos diferenciados entre as populações de doentes, considerando também as diferenças de sexo e género e raciais/étnicas. O objetivo é incorporar a confiança e os controlos de risco ao longo da conceção, da implantação e da monitorização, combinando a avaliação técnica com a responsabilização organizacional. Para os hospitais, a força do AI RMF reside na sua capacidade de ser traduzido em práticas concretas, tais como **funções claramente definidas para a supervisão da equidade, a avaliação metódica do desempenho em populações desagregadas, e estratégias de mitigação documentadas**, mantendo simultaneamente uma perspetiva mais ampla baseada em direitos.

No que respeita aos sistemas de gestão, a norma *ISO/IEC 42001* da International Organization for Standardization estabelece requisitos para um sistema organizacional de gestão da IA, centrando-se em políticas, objetivos, processos e melhoria contínua para uma IA responsável.²¹ Embora não seja prescritiva quanto a danos específicos, apoia a integração de princípios de equidade e não discriminação nas políticas organizacionais e nos controlos de risco, permitindo que as instituições abordem potenciais vieses de género e raciais de forma coordenada. De forma complementar, a norma *ISO/IEC 23894* fornece orientação sobre a integração da gestão do risco no desenvolvimento e na implantação da IA, incluindo a identificação e o tratamento de riscos relacionados com impactos não intencionais ou desproporcionados em diferentes grupos populacionais.²² Especificamente para a IA clínica, a norma *ISO 14971* define princípios de gestão do risco amplamente utilizados na regulação de dispositivos médicos, os quais descrevem a identificação sistemática de perigos, a avaliação do risco,

²⁰ National Institute of Standards and Technology, *Quadro de Gestão do Risco em Inteligência Artificial (AI RMF 1.0)*, Gaithersburg, MD: NIST, 2023.

²¹ International Organization for Standardization, *ISO/IEC 42001:2023 – Sistema de Gestão da Inteligência Artificial*, Geneva: ISO, 2023.

²² International Organization for Standardization, *ISO/IEC 23894:2023 – Gestão do Risco em Inteligência Artificial*, Geneva: ISO, 2023.

o controlo do risco e a monitorização ao longo do ciclo de vida.²³ Estes processos podem ser alargados para considerar o desempenho ou os resultados desiguais entre populações desagregadas como parte da identificação de danos. As normas descritas são valiosas na medida em que sugerem **uma estrutura de sistema de gestão que os hospitais podem adotar**. Fornecem também uma linguagem comum através da qual os requisitos relacionados com a equidade, como a avaliação de populações desagregadas e a mitigação do viés, podem ser especificados na aquisição e na contratação com fornecedores.

Os quadros de governação ética fornecem igualmente orientação normativa importante. As *Orientações Éticas para uma IA de Confiança* da Comissão Europeia constituem um instrumento de política útil, ainda que não vinculativo.²⁴ Definem sete requisitos centrais para sistemas de IA responsáveis, incluindo a diversidade, a não discriminação e a equidade, os quais apelam explicitamente à prevenção do viés e à consideração de impactos entre populações com diferentes características demográficas, como o sexo e o género e a origem racial/étnica. A *Lista de Avaliação para uma Inteligência Artificial de Confiança (ALTAI)*, uma ferramenta voluntária de autoavaliação desenvolvida pelo Grupo de Peritos de Alto Nível sobre Inteligência Artificial da Comissão Europeia, operacionaliza essas orientações sob a forma de uma lista de autoavaliação.²⁵ Num modelo regulatório para a saúde, o valor da ALTAI reside no seu **formato orientado por perguntas, que pode ser adaptado em listas de verificação práticas para a aquisição e a implantação**, incentivando as partes interessadas a considerar questões como a representatividade dos dados de treino, o risco de resultados clínicos diferenciados, e a adequação das salvaguardas contra efeitos discriminatórios, incorporando assim considerações de equidade na tomada de decisões do quotidiano por clínicos, especialistas de TI e gestores hospitalares.

Numa perspetiva de direitos humanos, a Convenção-Quadro do Conselho da Europa sobre IA constitui um instrumento jurídico relevante, explicitamente ancorado nos princípios da dignidade humana, da transparência, da responsabilização, da igualdade e não discriminação, e da privacidade, fornecendo assim uma base jurídica para abordar efeitos discriminatórios, incluindo os relacionados com o viés de género e racial em sistemas de IA.²⁶ De igual modo, os *Princípios sobre Inteligência Artificial* da Organização

²³ International Organization for Standardization, *ISO 14971:2019 – Dispositivos Médicos: Aplicação da Gestão do Risco a Dispositivos Médicos*, Geneva: ISO, 2019.

²⁴ European Commission, *Orientações Éticas para uma IA de Confiança*, Brussels: European Commission, 2019.

²⁵ European Commission, *Lista de Avaliação para uma Inteligência Artificial de Confiança (ALTAI) – Autoavaliação*, Brussels: European Commission, 2020.

²⁶ Council of Europe, *Convenção-Quadro sobre Inteligência Artificial (CAI)*, Strasbourg: Council of Europe, 2023.

para a Cooperação e Desenvolvimento Económico (OCDE)²⁷ e a *Recomendação sobre a Ética da IA* da UNESCO²⁸ oferecem orientação normativa internacional que incentiva a identificação e a mitigação de impactos injustos ou diferenciados entre populações desagregadas, reforçando a expectativa de que os sistemas de IA não devem reproduzir nem amplificar desigualdades estruturais. No domínio da governação ética específica à saúde, a Organização Mundial da Saúde articulou seis princípios centrais para a IA na saúde: autonomia, bem-estar, explicabilidade, responsabilidade, equidade e capacidade de resposta.²⁹ O princípio da **equidade exige explicitamente atenção às diferenças no acesso, no desempenho e nos resultados entre diversas populações de doentes, considerando também as diferenças de sexo e género e raciais ou étnicas.**

Estes quadros de governação internacionais correspondem diretamente aos requisitos alinhados com a Carta em matéria de dignidade, igualdade, não discriminação e direito à saúde, partilhando uma ênfase comum na governação ao longo do ciclo de vida — princípios que podem ser diretamente traduzidos em controlos práticos para as instituições de saúde.

Lições de Modelos Regulatórios de Outros Setores

Setores de risco elevado, como a aviação, a banca e o setor farmacêutico, desenvolveram há muito sistemas de governação sofisticados para a gestão de riscos tecnológicos. Estes setores oferecem abordagens regulatórias valiosas que podem ser adaptadas à governação da IA na saúde e à gestão dos desafios relacionados com o viés de género e racial.

Embora o viés de género e racial não constitua, em sentido estrito, um risco tecnológico, manifesta-se como um risco sociotécnico quando os dados, os algoritmos e as práticas organizacionais interagem com as desigualdades estruturais existentes nos cuidados de saúde. Neste contexto, o termo "risco" refere-se à manifestação do viés, ao nível do sistema, nos resultados, nas decisões e nos desfechos clínicos, e não à sua origem. Os próprios cuidados de saúde já fornecem evidência extensa destes efeitos, incluindo disparidades documentadas no diagnóstico, no tratamento e na previsão de risco entre populações diferenciadas por sexo e género e por raça/etnicidade. Os exemplos de outros setores complementam, assim, a evidência proveniente dos cuidados de saúde, ilustrando mecanismos de governação já consolidados para a gestão de sistemas

²⁷ Organisation for Economic Co-operation and Development, *Princípios da OCDE sobre Inteligência Artificial*, Paris: OECD, 2019.

²⁸ UNESCO, *Recomendação sobre a Ética da Inteligência Artificial*, Paris: UNESCO, 2021.

²⁹ World Health Organization, *Ética e Governação da Inteligência Artificial para a Saúde*, Geneva: WHO, 2021.

complexos e de elevado impacto, em conjunto com os resultados empíricos da área da saúde que demonstram a necessidade destes controlos regulatórios.

Um exemplo particularmente relevante é o *Sistema de Gestão da Segurança (SMS)* utilizado na aviação. A regulamentação internacional da aviação civil trata a segurança não como um exercício único de certificação, mas como um processo contínuo, transversal a toda a organização, que permite a tomada de decisões com base na identificação de riscos, na gestão do risco, na garantia da segurança (monitorização e auditoria), na gestão da mudança, e na melhoria contínua.³⁰ Aplicar esta lógica à IA na saúde significaria **estabelecer um "Sistema de Gestão da Segurança da IA" que incorpore explicitamente considerações de equidade**, incluindo: (1) um canal de comunicação de riscos e quase-incidentes para os clínicos, que permita captar potenciais incidentes de viés de género ou racial; (2) um registo formal de risco para cada sistema de IA, documentando as diferenças de desempenho entre populações de doentes desagregadas; (3) gatilhos de controlo de alterações que avaliem o impacto de atualizações do modelo, de desvios nos dados, ou de alterações no fluxo de trabalho, sobre os resultados de equidade; e (4) revisões periódicas de garantia da segurança, com medidas corretivas que abordem tanto os riscos técnicos como as desigualdades de género ou raciais identificadas. A ênfase explícita do AI Act (Arts. 9.º e 72.º) na gestão do risco ao longo do ciclo de vida e na monitorização pós-comercialização está em estreita consonância com esta lógica do SMS.

Uma segunda abordagem útil provém da regulação financeira, onde ferramentas quantitativas complexas são tratadas como "modelos" que requerem um desenvolvimento robusto, uma validação independente e supervisão de governação. Na supervisão bancária, foram implementados quadros de governação do risco comparáveis em jurisdições como os Estados Unidos (por exemplo, a *Orientação de Supervisão sobre a Gestão do Risco de Modelos* da Reserva Federal dos EUA), os quais constituem um exemplo de como a validação, a supervisão e a revisão independente de modelos podem ser implementadas.³¹ Abordagens semelhantes poderiam ser aplicadas aos sistemas de IA biomédica, de forma a garantir que **as ferramentas de apoio à decisão clínica são validadas de forma independente (interna ou externamente) antes de serem introduzidas** em contextos de cuidados de saúde, com uma **avaliação explícita do desempenho entre populações diferenciadas por sexo e género, e por características raciais ou étnicas**. Os artefactos mínimos de validação devem documentar a finalidade pretendida do sistema, os seus limites operacionais, os modos de falha conhecidos, e o desempenho em populações desagregadas, incluindo

³⁰ International Civil Aviation Organization, *Manual de Gestão da Segurança (Doc 9859)*, 4th ed., Montréal: ICAO, 2018.

³¹ Board of Governors of the Federal Reserve System and Office of the Comptroller of the Currency, *Orientação de Supervisão sobre a Gestão do Risco de Modelos (SR 11-7)*, Washington, D.C., 4 April 2011.

quaisquer disparidades ou medidas de mitigação aplicadas. A revalidação deve ser desencadeada sempre que o sistema seja implantado num novo local, aplicado a uma população de doentes diferente, ligado a novos equipamentos ou protocolos, atualizado pelo fornecedor, ou apresente sinais de degradação do desempenho. Estas práticas estão em conformidade com o AI Act (Arts. 13.º, 14.º e 60.º), que exige testes, transparência para os responsáveis pela implantação, e supervisão humana concebida para detetar e prevenir o viés de automação, garantindo que os cuidados apoiados por IA são simultaneamente seguros e equitativos entre diversas populações.

No que diz respeito à governação pós-implantação, a regulação farmacêutica fornece um modelo robusto: na farmacovigilância, as reações adversas e os sinais de segurança são sistematicamente recolhidos e analisados para detetar riscos emergentes. Em particular, o *Módulo VI das Boas Práticas de Farmacovigilância (GVP)* da Agência Europeia de Medicamentos estabelece expectativas detalhadas para a recolha, gestão e submissão de relatos de suspeitas de reações adversas, refletindo uma abordagem de ecossistema à comunicação e ao acompanhamento de incidentes.³² Um **sistema comparável de "vigilância da IA"** poderia permitir aos hospitais comunicar e analisar **incidentes** relacionados com decisões clínicas apoiadas por IA, com particular atenção aos resultados de equidade. Os elementos centrais incluiriam: (1) comunicação sistemática que capte quaisquer desigualdades observadas nos cuidados ou nos resultados, em função de diferenças de sexo e género ou raciais/étnicas; (2) documentação obrigatória de incidentes graves em que o viés possa ter contribuído para um tratamento diferenciado; (3) obrigações de notificar o fornecedor e as autoridades competentes sobre preocupações relacionadas com a equidade (notificando as autoridades caso o viés conduza a danos graves, como erros de diagnóstico ou tratamento desigual, mantendo as desigualdades de natureza rotineira no âmbito da monitorização interna da qualidade); e (4) funções de deteção de tendências e sinais que identifiquem padrões emergentes de viés entre diferentes departamentos, unidades de saúde, ou a rede de cuidados de saúde em sentido mais amplo. Esta monitorização permitiria medidas de mitigação proativas e uma melhoria contínua, garantindo que os cuidados apoiados por IA são não só seguros, mas também equitativos.

Por último, ferramentas de governação do setor público, como as avaliações de impacto algorítmico e as auditorias de viés, oferecem métodos práticos para avaliar o impacto social dos sistemas de tomada de decisão automatizados, incluindo potenciais disparidades em função de diferenças de sexo e género ou raciais e étnicas. Estas ferramentas traduzem princípios éticos e normas de direitos humanos em formatos acionáveis que as organizações podem utilizar para antecipar e mitigar desigualdades antes da implantação. Por exemplo, a *Avaliação de Impacto Algorítmico* federal do

³² European Medicines Agency, *GVP Módulo VI: Recolha e Submissão de Suspeitas de Reações Adversas*, London: EMA, 2017

Canadá classifica os riscos dos sistemas de IA e associa-os a medidas de mitigação.³³ Esta ferramenta enquadra explicitamente a implantação como um problema de governação orientado pelo impacto. Adicionalmente, a abordagem da cidade de Nova Iorque às ferramentas automatizadas de decisão em matéria de emprego exige auditorias de viés e obrigações de divulgação ao público.³⁴ De forma correspondente, os hospitais e as OSC que trabalham na área da saúde poderiam implementar **um pacote de "Avaliação de Impacto e Auditoria da IA" que combine um questionário estruturado centrado na equidade com auditorias de equidade independentes e periódicas**, produzindo um resumo publicável do desempenho entre populações desagregadas e dos esforços de mitigação. A implementação de um pacote deste tipo contribuiria para promover a transparência, a responsabilização pública e a confiança dos doentes nos cuidados apoiados por IA.

A inclusão de modelos regulatórios de setores como a segurança da aviação, a gestão do risco de modelos financeiros, a farmacovigilância e a governação algorítmica destina-se a constituir uma transferência deliberada de princípios de governação já estabelecidos. Estes domínios foram selecionados por implementarem, na prática, abordagens formalizadas para a deteção, a monitorização e a mitigação de riscos em sistemas complexos, nos quais podem surgir impactos diferenciados sobre os indivíduos. Vários já incorporam, em termos operacionais, práticas de equidade, incluindo a estratificação demográfica do risco na farmacovigilância, a validação de modelos e os controlos de viés na supervisão financeira, e a comunicação sistemática de eventos adversos que afetam populações específicas. A sua relevância reside tanto na semelhança conceptual como na utilização comprovada de mecanismos de governação ao longo do ciclo de vida, de comunicação obrigatória, e de supervisão independente, que podem ser diretamente adaptados para abordar os riscos de equidade de género e racial nos sistemas de IA biomédica.

Modelo para um modelo regulatório de IA alinhado com a Carta, destinado a hospitais públicos e OSC

Esta secção apresenta uma arquitetura de modelo regulatório **ancorada nos princípios da Carta da UE, expressa, contudo, sob a forma de procedimentos práticos que os hospitais e as OSC podem implementar** na prática. A Carta fornece o fundamento normativo para a governação das tecnologias de saúde na Europa, particularmente através dos princípios da dignidade humana, da igualdade, da não discriminação, e do

³³ Treasury Board of Canada Secretariat, *Avaliação de Impacto Algorítmico*, Ottawa: Government of Canada, 2026.

³⁴ New York City Department of Consumer and Worker Protection, *Ferramentas Automatizadas de Decisão em Matéria de Emprego*, New York: City of New York, 2023.

direito aos cuidados de saúde.³⁵ Estes direitos são especialmente relevantes quando sistemas automatizados influenciam decisões clínicas que afetam os cuidados prestados aos doentes, incluindo potenciais lacunas entre populações diferenciadas por sexo e género e por origem racial/étnica.

Ao nível mais fundamental, o Art. 1.º da Carta, que estabelece que a dignidade humana é inviolável, sustenta requisitos de governação que garantem que as ferramentas de IA são utilizadas de forma a preservar o julgamento humano e a respeitar a autonomia do doente. Em termos práticos, este princípio exige **salvaguardas contra a confiança excessiva em recomendações automatizadas, mecanismos para contestar decisões assistidas por IA, e estruturas de supervisão clínica que previnam o viés de automação** - a tendência dos decisores humanos para confiarem excessivamente nos resultados algorítmicos, o que pode amplificar as desigualdades caso as recomendações da IA desfavoreçam sistematicamente determinadas populações desagregadas.

De forma semelhante, os Arts. 20.º e 21.º, que consagram a igualdade perante a lei e proíbem a discriminação, fornecem o fundamento jurídico para obrigações explícitas de equidade e de mitigação do viés na IA biomédica. Os hospitais que implantam sistemas de IA devem, por isso, implementar **processos para monitorizar o desempenho do sistema entre diversas populações de doentes**, incluindo o sexo e o género, a raça ou etnicidade, e combinações interseccionais destas características. Sempre que sejam detetadas desigualdades no desempenho ou nos resultados, devem ser implementadas medidas corretivas através do ajustamento do modelo, de alterações no fluxo de trabalho clínico, ou de percursos de decisão alternativos, garantindo que a implantação da IA não consolida desigualdades estruturais.

Adicionalmente, o Art. 35.º da Carta garante o direito a um elevado nível de proteção da saúde humana, sustentando assim ainda mais o princípio de que os sistemas de IA só podem ser implantados quando contribuam de forma demonstrável para cuidados de saúde seguros, eficazes e equitativos. As tecnologias de IA não devem introduzir novas barreiras ao acesso aos cuidados, nem agravar as disparidades existentes que afetam populações de doentes historicamente sub-servidas ou de outra forma desfavorecidas, devendo o seu impacto sobre essas populações ser avaliado de forma contínua.

Para garantir estes direitos na prática, o modelo regulatório proposto neste documento de política adota uma **abordagem de governação ao longo do ciclo de vida, que estrutura a supervisão da IA em torno de uma série de pontos de decisão obrigatórios, ou "gates"** (portões/etapas). Estes gates refletem a lógica regulatória integrada no AI Act da UE e na legislação de saúde relacionada, fornecendo simultaneamente aos hospitais um percurso de governação claro e auditável. O modelo parte do princípio de que o hospital é, regra geral, um responsável pela implantação (nos termos do AI Act) e

³⁵ European Union, *Carta dos Direitos Fundamentais da União Europeia*, 2012/C 326/02, Strasbourg: EU.

um prestador de cuidados de saúde nos termos do direito nacional da saúde, mas contempla também o caso em que o hospital se torna fornecedor/fabricante (como no exemplo da IA desenvolvida internamente).

Os hospitais são os principais responsáveis pela implantação e pelo controlo da execução dos sistemas de IA, enquanto as OSC atuam como organismos independentes de supervisão, auditoria e monitorização baseada em direitos, sem controlo direto sobre as decisões diárias relativas à implantação clínica.

A implementação deste modelo regulatório segue um princípio de proporcionalidade, reconhecendo que os **hospitais e as OSC podem ter níveis diferentes de capacidade técnica, organizacional e analítica entre Estados-Membros e sistemas de saúde**. Os requisitos mínimos obrigatórios aplicam-se de forma uniforme, de modo a garantir um nível de base de segurança, transparência e salvaguardas de equidade. Adicionalmente, práticas mais avançadas, tais como uma análise mais aprofundada de populações desagregadas, auditorias externas, e ferramentas sofisticadas de monitorização do viés, podem ser implementadas de forma progressiva, em função da capacidade institucional. O anexo técnico segue igualmente esta abordagem, distinguindo entre requisitos mínimos e requisitos recomendados.

1. **Etapa de classificação e definição de âmbito.** Antes da aquisição ou do desenvolvimento, as instituições têm de determinar o estatuto regulatório de uma determinada ferramenta de IA. Isto inclui avaliar se o sistema se qualifica como software de dispositivo médico nos termos do MDR ou do IVDR, se constitui um sistema de IA de risco elevado nos termos do AI Act, ou se se enquadra numa categoria de risco mais baixo que ainda assim exige supervisão de governação. Nesta fase inicial, os hospitais podem sinalizar ferramentas cuja utilização pretendida possa ter implicações de equidade acrescidas ou afetar diversas populações de doentes, assinalando que estas ferramentas exigirão uma avaliação mais detalhada do viés de género e racial em fases posteriores. A orientação produzida pelo MDCG sobre a qualificação e a classificação de software fornece exemplos práticos que podem ajudar a evitar a classificação regulatória incorreta e a preparar o terreno para as avaliações de viés e de equidade.³⁶ Os hospitais conduzem a classificação e a determinação regulatória. As OSC fornecem contributos consultivos independentes sobre os riscos relacionados com a equidade, mas não têm poder de decisão nesta etapa.

³⁶ Medical Device Coordination Group, *Orientação sobre a Qualificação e a Classificação de Software no âmbito do Regulamento (UE) 2017/745 (MDR) e do Regulamento (UE) 2017/746 (IVDR)*, Brussels: MDCG, 2019.

2. **Etapa de pré-aquisição e contratação.** Os hospitais exercem frequentemente a sua maior influência sobre a governação da IA durante a fase de aquisição. Nesta etapa, as instituições devem garantir que os fornecedores se comprometem contratualmente a abordar os riscos de equidade, incluindo o viés de género e racial, ao longo de todo o ciclo de vida da IA. Os fornecedores devem ser obrigados a fornecer documentação que demonstre a conformidade regulatória, incluindo evidência de avaliações de conformidade quando relevante, resultados da validação de desempenho, e informação necessária para que os responsáveis pela implantação cumpram as suas obrigações nos termos do AI Act. Os contratos de aquisição devem atribuir explicitamente as responsabilidades entre o fornecedor e o responsável pela implantação, incluindo requisitos relativos à transparência, à documentação, ao registo, à supervisão humana, e ao apoio pós-implantação. O objetivo nesta etapa é incorporar a consciência do viés e a responsabilização nas obrigações contratuais, criando uma base para uma implantação segura e equitativa. Os hospitais são responsáveis pelas decisões de aquisição e pelos acordos contratuais. As OSC podem rever a documentação de aquisição quanto a considerações de equidade e de direitos, e emitir recomendações não vinculativas.
3. **Etapa de validação.** A terceira etapa é a etapa de validação, que garante que as ferramentas de IA são avaliadas de forma independente antes de entrarem nos fluxos de trabalho clínicos. A validação deve avaliar tanto o desempenho global como os resultados entre populações de doentes desagregadas, com particular atenção ao sexo e género e à origem racial/étnica. Este passo é fundamental porque os sistemas de IA biomédica podem apresentar um desempenho diferenciado entre populações quando os dados de treino não são representativos ou quando são ignoradas diferenças contextuais entre os ambientes de desenvolvimento e de implantação.³⁷ A avaliação deve também documentar quaisquer medidas de mitigação aplicadas e destacar os riscos de equidade remanescentes, fornecendo aos hospitais evidência para tomarem decisões informadas sobre a implantação. Os hospitais são responsáveis por realizar ou encomendar a validação e por interpretar os resultados para as decisões de implantação. As OSC fornecem uma revisão independente do desempenho desagregado e dos riscos de viés, sem controlo sobre os resultados da validação.
4. **Etapa de implantação.** Uma vez que um sistema tenha passado a validação, a quarta etapa é a etapa de implantação. Os hospitais precisam de garantir que as condições de funcionamento sustentam uma utilização segura, equitativa e responsável. Isto inclui a atribuição de pessoal de supervisão devidamente formado, a implementação de mecanismos que permitam aos clínicos anular ou

³⁷ Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H., *Garantir a Justiça na Aprendizagem Automática para Promover a Equidade em Saúde*, *Annals of Internal Medicine*, 169(12), 866-872, 2018.

desativar as recomendações algorítmicas, e a garantia de que os utilizadores compreendem a finalidade pretendida do sistema, as suas limitações, e quaisquer considerações relativas ao desempenho desagregado. Os clínicos devem ser informados sobre potenciais variações injustas nos resultados quando são consideradas diferenças de sexo e género ou raciais/étnicas, e os doentes devem ser informados quando ferramentas de decisão apoiadas por IA são utilizadas nos seus cuidados. Os hospitais mantêm a responsabilidade plena pelas decisões de implantação, pela integração clínica, e pela supervisão da implementação. As OSC não têm qualquer papel na implantação, mas podem rever a documentação relativa à preparação para a implantação numa perspetiva de equidade e de direitos.

5. **Etapas de monitorização, auditoria e responsabilização.** Esta etapa final rege a fase pós-implantação do ciclo de vida da IA. Os sistemas de IA devem ser sujeitos a monitorização contínua quanto ao desempenho global, às disparidades entre populações desagregadas, e a incidentes de segurança ou de equidade. Os hospitais devem manter registos do funcionamento do sistema, implementar procedimentos normalizados de comunicação de incidentes que captem potenciais vieses de género ou raça, e reavaliar periodicamente o desempenho do sistema através de exercícios de revalidação. Quando surjam desigualdades ou riscos de viés que não possam ser eficazmente mitigados, o sistema deve ser modificado, suspenso ou desativado. As conclusões devem alimentar tanto a governação interna como, quando aplicável, a aprendizagem ao nível da rede mais ampla, de forma a prevenir a recorrência do viés. Os hospitais são responsáveis pela monitorização contínua, pelo registo, pela comunicação de incidentes, e pela implementação de medidas de mitigação. As OSC realizam auditorias independentes, avaliam os riscos sistémicos de equidade entre diferentes unidades, e apoiam o encaminhamento de disparidades persistentes, sem controlo operacional sobre as modificações do sistema.

Em conjunto, estas etapas do ciclo de vida transformam os princípios de direitos fundamentais em processos de governação concretos, capazes de abordar os riscos de viés de género e racial nos sistemas de IA clínicos.

Figura 1: Percurso de Governação da IA Alinhado com a Carta, criado pelo projeto AEQUITAS

Percurso de Governação da IA Alinhado com a Carta, criado pelo projeto AEQUITAS PARA IA BIOMÉDICA NOS CUIDADOS DE SAÚDE PÚBLICOS



Políticas práticas, ferramentas operacionais e artefactos de governação para profissionais de hospitais e OSC

Financiado pela União Europeia. Os pontos de vista e as opiniões expressas são no entanto apenas os do(s) autor(es) e não refletem necessariamente os da União Europeia ou da Agência de Execução Europeia do Trabalho, da Educação e da Cultura (EACEA). Nem a União Europeia nem a autoridade que concede a subvenção podem ser consideradas responsáveis por eles.

Código do projeto: 101215009 – AEQUITAS – CERV-2024-CHAR-LITI

Para que funcione eficazmente em contextos reais de cuidados de saúde, o modelo regulatório proposto neste documento de política prevê a criação de instrumentos práticos que possam ser reutilizados entre instituições e adaptados a diferentes contextos nacionais. Sublinha-se a necessidade de procedimentos, documentação e modelos normalizados que permitam tanto aos hospitais como às OSC administrar os sistemas de IA de forma consistente. O conjunto de instrumentos a seguir apresentado está alinhado com os requisitos do AI Act, com a orientação relativa ao MDR e ao IVDR, e com práticas de governação de outros setores.

Fase de aquisição e contratação

- **Pacote de aquisição e contratação.** A comunidade de aquisição de IA da UE publicou cláusulas contratuais-tipo (versões para sistemas de risco elevado e não elevado) e atualizou-as de acordo com o AI Act, encorajando explicitamente as autoridades públicas a utilizá-las e fornecendo comentários para a sua implementação.³⁸ No caso da IA na saúde, e com vista a avaliar e mitigar os riscos de viés de género e racial, estas cláusulas podem ser adaptadas num anexo específico para cada hospital, enquanto artefacto autónomo, que exija, no mínimo:
 - (1) comunicação normalizada da composição dos dados de treino, incluindo a representação por sexo e género e por origem racial/étnica, para apoiar a avaliação da representatividade e do risco de viés;
 - (2) divulgação da proveniência dos dados, incluindo se os conjuntos de dados são proprietários ou não proprietários, as suas condições de acessibilidade, e as implicações para a transparência, a validação independente, e a avaliação do viés de género e racial;
 - (3) divulgação do desempenho entre populações desagregadas e da evidência de validação, incluindo resultados estratificados por sexo e género e por origem racial ou étnica, documentação de quaisquer disparidades conhecidas, medidas de mitigação do viés aplicadas durante o desenvolvimento, e testes de validação de suporte (incluindo considerações específicas ao local relevantes para a equidade e a justiça);
 - (4) requisitos de registo e auditabilidade, garantindo que o sistema é capaz de gerar registos de utilização que possam apoiar auditorias independentes relativas a viés ou a impactos diferenciados;
 - (5) obrigações de atualização e de controlo de alterações, exigindo que o fornecedor informe o hospital sobre atualizações do modelo, reciclagem do treino, ou alterações no fluxo de trabalho, e que documente os seus

³⁸ European Commission, *Cláusulas Contratuais-Tipo de IA da UE Atualizadas Já Disponíveis*, Brussels: European Commission, 2025.

potenciais efeitos no desempenho relativamente a populações diferenciadas;

- (6) compromissos de resposta a incidentes relacionados com o viés, obrigando o fornecedor a comunicar erros ou desigualdades identificadas que afetem populações de doentes diferenciadas por sexo e género ou por origem racial/étnica, e a delinear medidas de correção;
- (7) atribuição clara de responsabilidades, especificando as obrigações do fornecedor e do responsável pela implantação relativamente ao cumprimento dos requisitos do AI Act e à gestão dos riscos relacionados com a equidade e o viés.

Etapas de validação e pré-implantação

- **Política de governação de dados de saúde e de representatividade.** Para mitigar o viés de género e racial, todos os hospitais e OSC que implantam IA devem adotar uma política formal que garanta que os dados utilizados para treino, ajuste fino, ou fins de implementação são representativos, completos, e precisos para as populações servidas. Isto inclui avaliar a composição demográfica, documentar as populações de doentes sub-representadas, e definir medidas de mitigação quando necessário. A política deve abranger tanto os conjuntos de dados desenvolvidos internamente como os obtidos externamente, incluindo dados proprietários e de acesso livre, e exigir a verificação do desempenho desagregado e de quaisquer estratégias de mitigação do viés aplicadas pelos fornecedores. Esta governação está em conformidade com o AI Act (Art. 10.º), que exige a análise e a mitigação dos riscos de viés, e reflete as melhores práticas na literatura sobre a mitigação do viés em IA na saúde, abrangendo todo o ciclo de vida, desde a recolha de dados até à monitorização pós-implantação.³⁹ Esta política não constitui um artefacto em si, mas sim uma camada de política que orienta outros artefactos. Funciona como um princípio pré-implantação e alimenta diretamente a monitorização, a auditoria e a responsabilização contínuas.
- **Base de referência de cibersegurança e modelo de ameaças específico à IA.** Os sistemas de IA hospitalares podem ser vulneráveis não só a ataques convencionais, como o envenenamento de dados, os exemplos adversariais, ou a evasão do modelo, mas também a manipulações ou erros que agravam o viés de género ou racial, por exemplo, afetando de forma desproporcionada estratos demográficos sub-representados nos dados de treino ou de implementação. O AI Act (Art. 15.º) exige resiliência contra alterações não autorizadas dos resultados ou do desempenho, sublinhando a necessidade de salvaguardas

³⁹ Char, D.S., Shah, N.H., & Magnus, D., *Implementar a Aprendizagem Automática nos Cuidados de Saúde — Abordar os Desafios Éticos*, *New England Journal of Medicine* 378: 981-983 (2018).

específicas à IA. Os hospitais devem, por isso, estabelecer uma base de referência de cibersegurança e um modelo de ameaças específico à IA que considere explicitamente os riscos relacionados com o viés, a par das ameaças técnicas. Este documento vivo funciona tanto como política como ferramenta prática, servindo de lista de verificação pré-implantação para garantir a implementação de controlos de cibersegurança alinhados com o MDR e o IVDR, com a governação organizacional da segurança da informação (por exemplo, a ISO/IEC 27001), e com a modelação de ameaças e os planos de teste sensíveis ao viés.⁴⁰ Apoiar também a monitorização contínua pós-implantação, de forma a detetar riscos técnicos e relacionados com o viés emergentes, e a responder de forma proativa. Ao implementar esta política, os hospitais podem reforçar a resiliência do sistema, salvaguardar a equidade na tomada de decisões apoiada por IA, e manter a conformidade com o AI Act (Arts. 9.º a 15.º) e com os requisitos do MDR/IVDR.

- **Um "dossier do sistema de IA" por ferramenta.** Este artefacto é um documento normalizado, mantido para cada sistema de IA antes da sua implantação numa organização de saúde. O dossier deve resumir a finalidade pretendida do sistema, a integração no fluxo de trabalho clínico, os resultados da validação, e os indicadores de desempenho, com particular atenção ao desempenho entre populações de doentes diferenciadas por sexo e género ou por origem racial/étnica, e a quaisquer desigualdades conhecidas ou medidas de mitigação. Deve documentar os mecanismos de supervisão humana, esclarecer quais as decisões que permanecem exclusivamente sob o controlo do clínico, e assinalar quaisquer contraindicações relevantes para a segurança do doente e para a equidade. Ao fornecer uma visão formalizada e consciente do viés sobre o sistema de IA, o dossier apoia a transparência, a governação interna, e a supervisão regulatória, garantindo que as decisões pré-implantação têm explicitamente em conta os riscos de resultados desiguais.
- **Um "pacote combinado de FRIA e DPIA".** Os hospitais públicos que implantam sistemas de IA de risco elevado são cada vez mais obrigados a avaliar os riscos para os direitos fundamentais, bem como os riscos para a proteção dos dados pessoais. A integração destas avaliações num único modelo, que satisfaça as disposições do AI Act (Art. 27.º), do RGPD, e da DPIA, pode reduzir a carga administrativa, garantindo simultaneamente que as questões centrais sobre as populações afetadas, os danos potenciais, as estratégias de mitigação, e os mecanismos de reclamação são sistematicamente abordadas, com particular

⁴⁰ Medical Device Coordination Group, *MDCG 2019-16 Rev. 1 - Orientação sobre Cibersegurança para Dispositivos Médicos*, Brussels: European Commission, 2020.

atenção às potenciais desigualdades que afetam populações de doentes diferenciadas por sexo e género e por raça/etnicidade.

O "pacote combinado de FRIA e DPIA" funciona principalmente como um artefacto de auditoria ex ante e periodicamente atualizado, situado na fase de pré-implantação e no âmbito da camada de auditoria de equidade e de direitos. Estas avaliações constituem avaliações organizadas e documentadas dos riscos para os direitos fundamentais e para a proteção de dados, funcionando como pontos de controlo de governação, e não como ferramentas de monitorização contínua. Embora devam ser atualizadas caso os riscos ou as condições do sistema se alterem, a sua função central é fornecer uma avaliação sistemática e revisável dos impactos potenciais antes e durante a implantação, apoiando a conformidade tanto com o AI Act como com o RGPD, e ajudando a identificar e a mitigar os vieses relacionados com o género ou com a raça.

Para reforçar a vertente dos direitos humanos, deve prever-se uma consulta significativa com as OSC na conceção das avaliações de impacto, de forma a incorporar a sua experiência em matéria de direitos. Por exemplo, a Rede Europeia de Instituições Nacionais de Direitos Humanos (ENNHRI) insta a Comissão Europeia a complementar os modelos com orientação e consulta pública, de forma a garantir FRIA eficazes que considerem a equidade nas dimensões de género e racial/étnica.⁴¹

- **Uma "política de controlo de alterações".** Esta política rege as atualizações aos sistemas de IA, incluindo o retreino do modelo, as atualizações de software, ou as modificações nos fluxos de trabalho clínicos. As organizações devem definir claramente o que constitui uma "alteração material" que exija revalidação ou aprovação, garantindo que as atualizações não introduzem riscos imprevistos para os cuidados prestados aos doentes. Em particular, as alterações devem ser avaliadas quanto a potenciais impactos na equidade de género e racial, incluindo alterações no desempenho desagregado ou resultados diferenciados entre populações com características demográficas distintas. Os hospitais devem também garantir que os clínicos e o restante pessoal relevante são informados sobre modificações que possam afetar os resultados dos cuidados prestados a diversas populações de doentes, permitindo uma supervisão e medidas de mitigação adequadas. Ao incorporar considerações de equidade na gestão de

⁴¹ European Network of National Human Rights Institutions (ENNHRI), *Declaração sobre a Garantia de Avaliações de Impacto sobre os Direitos Fundamentais (FRIA) Eficazes nos termos do AI Act da UE*, Brussels: ENNHRI, 2025.

alterações, os hospitais mantêm tanto a segurança como a justiça da tomada de decisões apoiada por IA na prática clínica.

- **Um "sistema de gestão da qualidade" (QMS) para responsáveis pela implantação que são também fornecedores.** Outra dimensão importante do modelo diz respeito à IA desenvolvida internamente por hospitais ou instituições de investigação. As organizações públicas de saúde desenvolvem ou adaptam cada vez mais as suas próprias ferramentas de IA, particularmente em centros médicos académicos. Nesses casos, a instituição pode efetivamente atuar como fabricante ou fornecedor, o que introduz responsabilidades regulatórias adicionais. A orientação europeia sobre dispositivos médicos desenvolvidos internamente esclarece que as instituições de saúde devem operar sistemas de gestão da qualidade e manter documentação técnica que demonstre que as ferramentas desenvolvidas internamente respondem a necessidades específicas dos doentes que não podem ser satisfeitas através de produtos comerciais.⁴² O QMS deve garantir que os resultados da IA são avaliados quanto a um desempenho diferenciado entre populações de doentes diferenciadas por sexo e género e por origem racial/étnica, que as estratégias de mitigação são documentadas, e que os mecanismos de supervisão humana preservam explicitamente a segurança do doente e a equidade. Mesmo quando se aplicam isenções regulatórias a dispositivos desenvolvidos internamente, as obrigações decorrentes dos direitos fundamentais e dos quadros mais amplos de governação da IA permanecem plenamente aplicáveis, fazendo do QMS uma ferramenta crítica para incorporar a equidade na conceção, na validação, e na implantação dos sistemas de IA biomédica.

Etapas de monitorização, auditoria e responsabilização

- **Um "relatório de desempenho desagregado por populações e de equidade".** Dado que as variações no desempenho algorítmico podem contribuir para resultados de saúde desiguais, os hospitais não devem apenas exigir que os fornecedores forneçam informação relevante antes da aquisição, mas devem também produzir análises internas que mostrem como os sistemas se comportam entre doentes com diferentes características demográficas. Estas análises devem centrar-se no desempenho estratificado por sexo e género e por diferenças raciais ou étnicas relevantes, sempre que legal e exequível, bem como na justificação de quaisquer lacunas de desempenho inevitáveis. Quando existam lacunas de desempenho, as organizações devem documentar as

⁴² Medical Device Coordination Group, *MDCG 2023-1 – Orientação sobre a Isenção para Instituições de Saúde (MDR/IVDR)*, Brussels: European Commission, 2023.

estratégias de mitigação e monitorizar a eficácia dessas medidas ao longo do tempo. Esta abordagem reflete expectativas emergentes na regulação europeia da IA, segundo as quais as organizações devem avaliar e mitigar ativamente os riscos de viés nos conjuntos de dados e nos resultados dos modelos. O AI Act (Art. 10.º) exige explicitamente a análise e a mitigação dos riscos de viés nos conjuntos de dados, e permite o tratamento limitado de dados de categoria especial para efeitos de deteção do viés, sob salvaguardas estritas.

O "relatório de desempenho desagregado por populações e de equidade" funciona principalmente como uma ferramenta de monitorização contínua no âmbito do quadro de governação. Exige a recolha contínua de dados, o acompanhamento regular do desempenho do sistema entre os grupos demográficos relevantes, e a atualização contínua das medidas de mitigação sempre que sejam identificados resultados diferenciados. Isto permite que as organizações de saúde detetem e abordem os riscos de viés ao longo do tempo. Simultaneamente, as principais conclusões deste relatório, tais como as desigualdades identificadas e a eficácia das estratégias de mitigação, devem ser sintetizadas e incluídas na "auditoria de equidade e de direitos", garantindo transparência e responsabilização pública.

- **Um regime de auditoria de duas camadas, composto por uma auditoria técnica e uma auditoria de equidade e de direitos.** Um regime de auditoria viável para os hospitais deve equilibrar o rigor com a praticabilidade, e ser compreensível tanto para o pessoal como para o público. Neste contexto, uma auditoria é uma revisão formalizada e periódica da conformidade, do desempenho, e dos impactos sobre os direitos de um sistema de IA, concebida para complementar a monitorização contínua, e não para a substituir.

Auditoria técnica: esta camada constitui uma revisão operacional confidencial e abrange verificações detalhadas que permanecem internas, incluindo testes de segurança, avaliações de robustez, validação do desempenho clínico, e conformidade com a governação de dados. Deve ser dada particular atenção ao desempenho entre diferenças de sexo e género e raciais ou étnicas, avaliando se os resultados do modelo desfavorecem sistematicamente alguma população. Estas atividades devem estar alinhadas com as disposições relevantes do AI Act (Arts. 9.º a 15.º), do MDR, e do IVDR, quando aplicável.

Auditoria de equidade e de direitos: esta segunda camada centra-se na transparência, na responsabilização, e na mitigação do viés. Deve documentar as desigualdades nos resultados, as populações incluídas nos testes, as limitações conhecidas, e os mecanismos de reclamação, com uma análise explícita dos

impactos de género e raciais. Deve também ser recolhida evidência de que a supervisão humana é significativa e capaz de corrigir resultados enviesados. Seguindo a abordagem de "auditoria de viés e divulgação" utilizada na regulação do emprego,⁴³ um resumo dirigido ao público em geral pode comunicar as principais conclusões, as estratégias de mitigação, e as melhorias implementadas, sem revelar informação proprietária ou confidencial. O "pacote combinado de FRIA e DPIA" situa-se também no âmbito da camada de auditoria de equidade e de direitos, reforçando a avaliação rigorosa dos direitos fundamentais, da proteção de dados, e da equidade demográfica.

- **Um "registo de vigilância da IA".** A monitorização pós-implantação é também apoiada através da criação de um registo de vigilância da IA, inspirado nos sistemas de farmacovigilância utilizados na regulação dos medicamentos.⁴⁴ Os profissionais de saúde devem poder comunicar incidentes, quase-incidentes, ou resultados inesperados relacionados com decisões apoiadas por IA, com campos explícitos para captar quaisquer disparidades de sexo e género ou raciais/étnicas observadas. Os relatos devem ser sistematicamente analisados para identificar padrões emergentes, incluindo impactos diferenciados em grupos demográficos específicos, permitindo às organizações detetar sinais de segurança, preocupações de equidade, ou tendências de viés. Quando necessário, os incidentes, em particular os que indiquem um tratamento desigual ou risco para grupos protegidos, devem ser comunicados prontamente aos fornecedores do sistema, aos comités de governação hospitalar relevantes, e às autoridades reguladoras. O registo deve também alimentar as auditorias periódicas de equidade e de direitos, apoiando a melhoria contínua e a mitigação do viés na implantação clínica de IA.
- **Colaboração e partilha de incidentes ao nível da rede.** A rede europeia de OSC e hospitais públicos desenvolvida no âmbito do quadro AEQUITAS funciona como uma camada de governação partilhada, fornecendo um vocabulário comum e normas de comunicação para incidentes e conclusões relacionadas com o viés de género e racial em sistemas de IA. Esta rede apoia a aprendizagem entre diferentes locais e a disseminação de boas práticas para a deteção, a mitigação, e a correção das desigualdades, de forma semelhante às abordagens utilizadas em setores de elevada segurança e na farmacovigilância. Ao associar dados sobre incidentes de viés, desempenho desagregado, e estratégias de mitigação, a rede

⁴³ New York City Department of Consumer and Worker Protection, *Ferramentas Automatizadas de Decisão em Matéria de Emprego (AEDT)*, New York: City of New York, 2025.

⁴⁴ European Medicines Agency, *GVP Módulo VI – Recolha e Submissão de Suspeitas de Reações Adversas (Rev 2)*, London: EMA, 2017.

permite que os hospitais e as OSC identifiquem padrões sistémicos de desigualdade e coordenem intervenções específicas. À medida que o EHDS facilita tanto a utilização primária (prestação de cuidados) como a utilização secundária (investigação e política) num quadro regulado, a rede pode alinhar as suas práticas de partilha de dados e de aprendizagem com os calendários e os atos de execução do EHDS, à medida que estes evoluem, apoiando uma implantação da IA mais equitativa e transparente entre instituições.

- **Módulo de formação para clínicos e peritos das OSC.** Para mitigar o viés de automação e as desigualdades na IA clínica, o modelo regulatório AEQUITAS inclui formação obrigatória, concebida para dotar os profissionais de saúde das competências necessárias para compreender as limitações dos sistemas de IA, interpretar os resultados de forma crítica, e anular ou interromper os sistemas quando os resultados possam produzir desfechos enviesados ou desiguais. A formação dá ênfase à consciencialização sobre os riscos de viés de género e racial, aos princípios de direitos fundamentais, e a considerações éticas, incluindo a equidade, a não discriminação, e a inclusão dos doentes.

Este módulo pode ser visto como um artefacto facilitador, apoiando ativamente a utilização do modelo regulatório, ao garantir que o pessoal responsável pela supervisão da IA pode implementar de forma competente os procedimentos de monitorização, validação e responsabilização, com uma perspetiva de equidade. A formação deve incluir exercícios baseados em casos práticos sobre a deteção e a mitigação de disparidades de sexo e género e raciais/étnicas.

Conclusão

A IA oferece oportunidades significativas para melhorar os resultados em saúde, mas, sem uma governação cuidadosa, pode reproduzir ou agravar as desigualdades de sexo e género e raciais ou étnicas, bem como outras formas de viés sistémico. O modelo regulatório apresentado neste documento de política traduz princípios de direitos fundamentais em salvaguardas concretas ao longo do ciclo de vida da IA, desde a aquisição e a validação até à implantação, à monitorização e à responsabilização, com atenção explícita à deteção e à mitigação de lacunas nos resultados entre diferentes populações de doentes. Ao implementar uma arquitetura de governação da IA estruturada em torno de um conjunto claro de etapas do ciclo de vida, políticas, ferramentas práticas e artefactos, os hospitais e as OSC podem avaliar, mitigar, e documentar sistematicamente os riscos, incluindo os impactos desiguais sobre populações de doentes marginalizadas. Ao integrar lições de outros setores, requisitos jurídicos da UE, e quadros de governação internacionais, o modelo regulatório apoia os

profissionais de saúde na administração responsável dos sistemas de IA, garantindo cuidados de saúde equitativos, seguros e transparentes para todos os doentes, independentemente do seu sexo e género, origem racial/étnica, ou outras características.

Financiado pela União Europeia. Os pontos de vista e as opiniões expressas são no entanto apenas os do(s) autor(es) e não refletem necessariamente os da União Europeia ou da Agência de Execução Europeia do Trabalho, da Educação e da Cultura (EACEA). Nem a União Europeia nem a autoridade que concede a subvenção podem ser consideradas responsáveis por eles.
Código do projeto: 101215009 – AEQUITAS – CERV-2024-CHAR-LITI