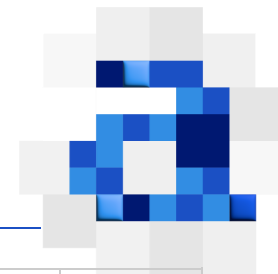
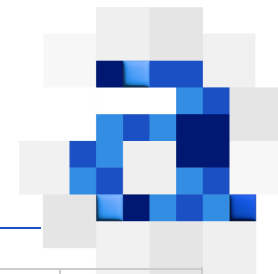


ANEXO I | Questionário GuIA

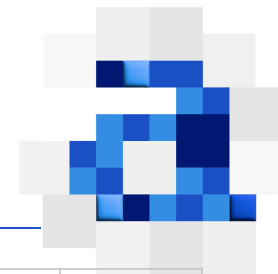
Transparente	Explicabilidade	Explicações	As decisões e previsões do sistema de IA são explicadas de forma clara, concisa e compreensível para o público-alvo?	☐
			É possível rastrear os principais fatores que contribuíram para um resultado específico?	☐
			As explicações fornecidas são suficientes para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



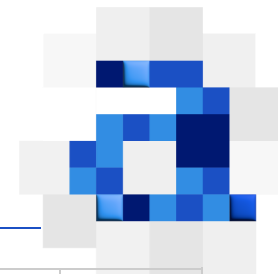
Robusta	Robustez		Existem políticas e procedimentos formais para garantir a qualidade, segurança e utilização ética dos dados, especialmente os de natureza sensível?	☐
			A organização assegura que a recolha e o tratamento de dados estão em conformidade com a legislação de proteção de dados e com os direitos dos seus titulares?	☐
		Governança de Modelos	A documentação do modelo de IA detalha de forma exaustiva a sua arquitetura, os métodos de treino, as limitações conhecidas e as considerações éticas subjacentes ao seu desenvolvimento?	☐
			A documentação técnica cumpre todos os requisitos exigidos pela legislação aplicável, como o EU AI Act, para a categoria de risco do sistema?	☐
			Existem processos para manter a documentação do modelo atualizada ao longo do seu ciclo de vida, refletindo todas as modificações substanciais?	☐
		Governança de Sistemas	Existem mecanismos de supervisão para monitorizar a implementação do sistema de IA, garantir a conformidade contínua e suportar auditorias internas que assegurem a sua integridade a longo prazo?	☐
			As políticas de governança abrangem todo o ciclo de vida do sistema, desde o desenvolvimento à descontinuação, assegurando uma gestão e monitorização contínuas?	☐
			A organização partilha de forma transparente as informações sobre como os sistemas de IA são implementados e geridos com as partes interessadas relevantes?	☐
	Robustez	Resiliência	O sistema está preparado para detetar e lidar com o desvio de dados (<i>data drift</i>), anomalias e atividade maliciosa?	☐
			Existe um processo estruturado para monitorizar e corrigir a degradação do desempenho?	☐
			São realizados testes de esforço e simulações “adversariais” antes da implementação?	☐



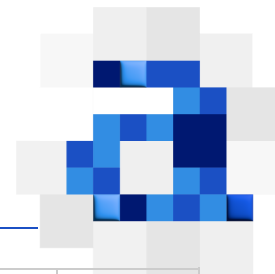
		Estabilidade	Os resultados são estáveis e previsíveis ao longo do tempo?	☐
		Resultados Previsíveis	A consistência do sistema foi validada para assegurar que dados de entrada semelhantes produzem, de forma fiável, resultados semelhantes?	☐
	Privacidade e Proteção de Dados	Privacidade de Dados	O sistema de IA cumpre com os regulamentos de privacidade de dados aplicáveis, como o RGPD e outras políticas específicas do setor?	☐
			Os dados sensíveis estão devidamente protegidos através de cifragem e controlos de acesso rigorosos?	☐
			Foram implementadas medidas de minimização de dados para garantir que apenas os dados estritamente necessários são recolhidos e processados?	☐
		Consentimento	Os utilizadores e/ou titulares dos dados têm um controlo efetivo sobre os seus dados, incluindo opções claras e acessíveis para dar ou retirar o consentimento e solicitar a sua eliminação?	☐
			Os mecanismos de consentimento são transparentes quanto à finalidade e à extensão da recolha e utilização de dados, especialmente no treino de modelos de IA?	☐
		Integridade dos Dados	Existem processos para verificar a exatidão, a proveniência e a representatividade dos dados utilizados para treinar e validar os sistemas de IA?	☐
			Os conjuntos de dados estão protegidos contra manipulação ou corrupção, tanto por agentes internos como externos?	☐
			Existem mecanismos de auditoria para validar continuamente a integridade dos dados ao longo do ciclo de vida do sistema de IA?	☐
	Segurança	Cibersegurança	São realizadas auditorias de segurança e testes de penetração regularmente?	☐



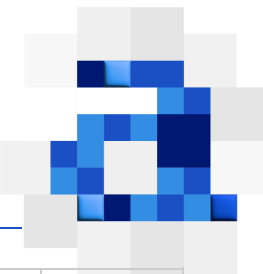
Universal			Existem controlos de acesso e encriptação para proteger os dados sensíveis utilizados pelo sistema?	☐
			Estão implementadas medidas para prevenir ataques “adversariais”, como <i>data poisoning</i> ou <i>model evasion</i> ?	☐
		Fiabilidade do Sistema	O sistema é continuamente monitorizado para detetar ameaças à segurança que possam comprometer a sua fiabilidade?	☐
			Existem mecanismos de <i>failover</i> e planos de contingência para garantir o funcionamento do sistema em caso de falha ou ataque?	☐
			Existem planos para garantir que o sistema se mantém estável e previsível em cenários de elevada exigência ou sob tentativas de disrupção?	☐
		Segurança do Sistema	Existe um método de avaliação e mitigação os riscos de utilização indevida do sistema de IA, tanto deliberada como inadvertida?	☐
			Estão implementadas salvaguardas para garantir que o modelo não é facilmente manipulado para gerar resultados ou comportamentos perigosos?	☐
			O treino de segurança do modelo é suficientemente robusto para resistir a técnicas de <i>jailbreaking</i> ou outras formas de manipulação?	☐
		Universal	Equidade	Resultados Justos
São utilizadas métricas e <i>benchmarks</i> de equidade para monitorizar e corrigir desequilíbrios nos resultados do sistema de IA ao longo do seu ciclo de vida?	☐			
São identificados e mitigados os vieses nos conjuntos de dados utilizados, em conformidade com as exigências do EU AI Act?	☐			



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



	Uso Eficiente de Recursos Naturais		A organização comunica de forma transparente o impacto energético dos seus sistemas de IA aos seus clientes e ao público?	☐
			Existem critérios de sustentabilidade na seleção de fornecedores de <i>hardware</i> e <i>software</i> ?	☐
			Existem políticas em vigor para a gestão responsável da água utilizada para arrefecer e de resíduos eletrónicos gerados pelos sistemas?	☐
	Avaliação do Impacto Social	Avaliação do Impacto Laboral	Foi realizada uma avaliação do impacto potencial dos sistemas de IA nos postos de trabalho e na qualidade do emprego dentro da organização e no seu setor?	☐
			Estão a ser implementadas medidas para a requalificação e formação dos trabalhadores cujas funções possam ser afetadas pela automação?	☐
		Avaliação do Impacto Económico	A estratégia de IA da organização contribui para um ecossistema económico mais justo e competitivo?	☐
			Existem medidas de governação interna para evitar que a pressão competitiva comprometa os padrões de segurança e ética?	☐
			A organização está a colaborar com outras entidades para promover o acesso aberto a dados e modelos, abrindo oportunidades de inovação a mais entidades para além das grandes empresas tecnológicas?	☐
		Avaliação do Impacto Cultural	A organização garante que a utilização de dados para treinar os seus modelos de IA respeita os direitos de autor e a propriedade intelectual?	☐
			Existe compensação justa os criadores cujo trabalho contribui para o treino dos modelos e sistemas?	☐
			Existem políticas internas para garantir que os sistemas de IA promovem a diversidade cultural em vez da homogeneização?	☐



Testada	Adequação para IA	Validação	O caso de uso viola alguma das práticas de IA proibidas pelo EU AI Act, representando um risco inaceitável para os direitos fundamentais?	☐
			Sendo o caso de uso legalmente admissível, é a IA a solução mais adequada e proporcional para o problema, ou existe uma alternativa mais simples e segura?	☐
			A organização possui os dados, a infraestrutura e as competências necessárias para desenvolver e manter o sistema de forma responsável, e os benefícios esperados justificam claramente os riscos?	☐
	Testes pré-implementação	Testes Independentes	A organização garante que as avaliações internas são complementadas por validações independentes para assegurar a conformidade e a adesão às melhores práticas?	☐
		Benchmarking	São aplicados <i>benchmarks</i> de explicabilidade, robustez, segurança, equidade e eficiência para comparar o sistema com os padrões da indústria e identificar lacunas de desempenho?	☐
		Testes de Integração	O sistema de IA foi submetido a testes de integração exaustivos que simulam cenários reais e avaliam a sua interação com os utilizadores finais e outros sistemas existentes?	☐
	Monitorização contínua	Testagem e Monitorização	Estão implementados mecanismos para monitorizar continuamente o desempenho do sistema de IA em produção e detetar desvios ou degradação do modelo?	☐
			O feedback dos utilizadores e os dados de desempenho são utilizados para acionar processos de re-treino adaptativo e manter o sistema alinhado com as condições do mundo real?	☐
			Em caso de sistemas de risco elevado, o plano de monitorização pós-implementação cumpre as exigências regulatórias definidas no EU AI Act, incluindo a documentação e o reporte de incidentes graves?	☐