

Guia

Guia para a Inteligência Artificial

GUIA PARA UMA INTELIGÊNCIA ARTIFICIAL
ÉTICA, TRANSPARENTE E RESPONSÁVEL
NA ADMINISTRAÇÃO PÚBLICA

DIMENSÕES DE AVALIAÇÃO



Este documento resume conteúdos que integram o GuIA para a Inteligência Artificial Ética, Transparente e Responsável na Administração Pública, elaborado pela AMA e visa constituir uma leitura mais simples e focada nas

Dimensões da Avaliação 2/3

É essencial que exista uma concordância entre a aplicação de sistemas de IA para assistência da sociedade e delegação de decisões, e os **valores** que definem uma sociedade.

Para conseguir uma IA Responsável, é preciso garantir que a utilização desta tecnologia:

melhore e defenda uma **cultura democrática**, o **Estado de direito** e os **direitos fundamentais** de uma sociedade.

A reflexão para uma Inteligência Artificial ética assume assim um papel fundamental que se deve alicerçar nos conceitos de: **Responsabilização**, **Transparência**, **Explicabilidade**, **Justiça** e **Ética**.



Estes conceitos materializam-se quando sistemas baseados em técnicas adaptativas, i.e. sistemas inteligentes:

- Cumprem a execução de responsabilidades e possibilidade plena de auditoria/inspeção (**Responsabilização**);
- Asseguram a visualização das suas componentes e dos procedimentos aplicados (**Transparência**);
- Explicam o funcionamento e as implicações decorrentes das funções computacionais (**Explicabilidade**);
- Incorporam garantias e salvaguardas aos utilizadores e beneficiários, sendo justos e não discriminativos (**Justiça**); e
- Asseguram mecanismos efetivos de mitigação de vieses inesperados, diminuindo desta forma possíveis riscos éticos (**Ética**).

Um sistema com **IA Responsável** garante que utilizadores, empresas e governos:

- Não incorram ou sofram consequências associadas a vieses produzidos por sistemas inteligentes, que de alguma forma instanciem comportamentos diferentes dos planeados;
- Respeitem os valores morais da sociedade que os alberga.



RESPONSABILIZAÇÃO

Uma das dimensões da IA que gera maior controvérsia é a responsabilização.

Os maiores obstáculos surgem quando se procura responder a questões como:

- Quem está no controlo?
- Como sabemos se um sistema de IA cumpre o propósito que realmente procuramos?

- Os dados são interpretados com precisão?
- Como podemos codificar um sistema de IA para cumprir obrigações legais como a Declaração Universal de Direitos Humanos ou a Lei de Privacidade?
- Quem ou o que é responsável pelos efeitos gerados da utilização dessa tecnologia, sejam benéficos ou perversos?
- O que será uma IA responsável?

A responsabilização de um sistema de IA inclui todos os **indivíduos responsáveis** por:

- Cada uma das partes num fluxo de trabalho complexo dum sistema com IA, **desde a conceção à sua implementação**;
- **Monitorização** da solução;
- **Revisões** ao sistema na sua fase evolutiva;
- **Suporte** ao utilizador/beneficiário; e
- **Supervisão e gestão** de riscos éticos, antes e após estar em operação.

A garantia de uma IA ética carece assim de **mecanismos que assegurem a responsabilização** pela criação, desenvolvimento e/ou utilização de sistemas de IA, e **minimizem os efeitos da parcialidade e opacidade** inerentes às redes de autonomia artificiais.

Tais mecanismos estão intimamente relacionados com:

para poderem
ser explicados
e auditados por
terceiros

a gestão de risco e
o rastreamento completo de procedimentos e resultados

Isto significa:

- Ter em conta a origem e a utilização de **dados**;
- **Modelos**;
- **Interfaces de programação de aplicações**; e
- Outros **componentes estruturais** desse sistema com IA.

Uma das ferramentas à tomada de decisões de IA Responsável é a definição de limites e regulamentos, por legisladores e autoridades, para os temas de recolha e utilização de dados e de certificados elegíveis para a produção de sistemas que utilizam ia.



TRANSPARÊNCIA E EXPLICABILIDADE

A transparência e a explicabilidade são duas **dimensões complementares**, uma vez que a primeira consolida o cumprimento da segunda.

A **transparência** num sistema com IA é garantida pela possibilidade de **interpretação** desse mesmo sistema, isto é, pela **clareza** e **inteligibilidade** do seu conteúdo.

A criação de algoritmos e sistemas de IA transparentes permite-nos:

explicar,
inspecionar e
reproduzir

os processos técnicos implícitos,
o raciocínio que fundamentou as decisões e
e a utilização dos dados por esses sistemas

Aumenta-se a transparência a partir de:

- **Documentação** com informação relativa: à recolha, ao armazenamento e à utilização de dados.

- **Comunicação clara:** das capacidades, do nível de precisão e das limitações do sistema de IA.

- **Expressa autorização:** para a recolha, o armazenamento e a utilização dos dados pessoais ou sensíveis.

- Disponibilização da documentação em **repositórios de livre acesso**, quer internamente, quer para os utilizadores/beneficiários da solução que utiliza IA.

Rastreabilidade do ciclo de procedimentos de desenvolvimento do sistema de IA

A partir da transparência de informação inerente aos sistemas de IA, **mecanismos de explicabilidade** permitem:

- confirmar que o sistema funcionou como esperado;
- prever o que é possível desencadear a partir de diferentes ações computacionais; e
- Reduzir a resistência à utilização e aceitação da IA, ao consolidar a confiança da sociedade em relação à tecnologia.

No final, mais que **conhecer** questões técnicas específicas, como os algoritmos, trata-se de ser capaz de **explicar** como uma decisão foi tomada por um modelo de IA e **entender** as implicações dos resultados/impactos decorrentes, esperados ou não.

Um sistema com ia transparente e passível de explicação torna os valores de base da entidade que o desenvolveu explícitos e conduz à responsabilização na tomada de decisões.



JUSTIÇA

A justiça é um conceito que se refere ao estado ideal de ausência de preconceito ou favoritismo direcionado a um indivíduo ou grupo, com base nas suas características herdadas ou adquiridas.

Este engloba os ideais de **equidade, imparcialidade, igualitarismo** e **não discriminação**.

É necessário que o sistema com IA, em particular o algoritmo, garanta:

- Os direitos individuais (p.e., liberdade de expressão e proteção dos dados);

- Os direitos coletivos (p.e., direito à saúde, à educação, à cultura, à vacinação e à justiça);
- E o cumprimento de normativos.

A abertura à inclusão e à diversidade nos sistemas com IA, aproximam a comunidade através do **aumento de confiança** nestas tecnologias.

É intuito de uma IA ética **dirigir-se a todos os utilizadores**, permitindo que todos possam utilizar os seus produtos e serviços.

Viés

A introdução ou criação de **viés** em sistemas de IA está intimamente ligada ao princípio de **justiça**.

Elementos tendenciosos podem entrar no sistema algorítmico de IA por:

A perpetuação inconsciente destas tendências pode conduzir a **resultados injustos, ao preconceito e à discriminação não intencional**.

Inclusão inadvertida de histórias tendenciosas, incompletas e imprecisas;
Arquétipos culturais, sociais ou institucionais pré-existentes;
Limitações técnicas; e
Modelos de má governança.

Estes podem ser consequência do **algoritmo** desenhado ou das escolhas relacionadas com a codificação, recolha, seleção ou utilização dos **dados** para alimentar o algoritmo.

No processo de construção de um sistema com IA, algumas questões podem ser formuladas com a finalidade de **identificar elementos tendenciosos**, entre as quais:

- Consideraram-se a diversidade e representatividade dos utilizadores finais e / ou sujeitos nos dados?
- Testaram-se grupos-alvo específicos ou casos de utilização problemáticos?
- Utilizaram-se ferramentas técnicas disponíveis publicamente, de última geração, para melhorar a compreensão dos dados, modelo e desempenho?
- Implementaram-se metodologias de testagem e monitorização para todo o processo do sistema com IA?

É essencial o desenvolvimento de procedimentos padronizados que verifiquem preconceitos ou outros objetos de discriminação adaptados ao número crescente de dados disponíveis e da celeridade desse mesmo crescimento.

ÉTICA



Embora com evidente potencial de orientar serviços e aumentar a eficiência e eficácia das instituições governamentais, a **consciencialização de riscos** inerentes à implementação de Inteligência Artificial é perentória. Tal é indissociável de uma **aplicação responsável e ética**, ponderada em função do contexto e abordada metodicamente.

A confiança em sistemas com IA aumenta com:

- Uma abordagem previsível e transparente;
- A compreensão da sua complexidade; e
- O objetivo primordial de criar benefício, aumentar o poder individual e a proteção do indivíduo e dos bens comuns à sociedade.

Uma **estrutura legal** robusta não é suficiente per se e deve ser acompanhada de um **código de ética**.

Este permitirá a projeção de um **conjunto de valores, princípios e diretrizes** que acompanhem os desenvolvimentos tecnológicos, bem como, os elementos sociais e políticos associados.

A sua aplicação deve estar assegurada em **todas as fases do ciclo** de um sistema com IA e deve ser cumprida por todos os indivíduos associados.

Os princípios éticos da IA estão em consonância com os princípios éticos para qualquer iniciativa de dados. Releva-se: **o respeito pelo homem, o respeito pelos direitos humanos, a participação e a responsabilidade pelas escolhas e decisões**.

O ciclo de valor dos dados abrange as etapas de recolha e criação, de armazenamento, proteção e processamento, de partilha, organização e publicação, e de utilização e reutilização. O sucesso da aplicação da IA resulta do sucesso de todas as etapas.

