GuIA para uma Inteligência Artificial Ética, Transparente e Responsável na AP

GuiA

Guia para a Inteligência Artificial

01 Inteligência Artificial na Administração Pública

Responsável

01 Benefícios da Inteligência Artificial

- Ganhos de eficiência e ganhos de produtividade, com a respetiva redução de custos para a AP
- Tempo poupado para os cidadãos, entidades públicas e empresas
- Rastreamento e previsão de mudanças em tempo real, que permitem respostas mais adequadas e atempadas, com correspondente agilidade nos processos. Ex. Pandemia
- *Melhores e mais inovadores serviços públicos*
- Respostas mais rápidas e adequadas a contextos de crise
- Desenvolvimento económico e social
- Antecipação e resposta a desafios (na componente de Gestão e Planeamento)

01 ÂMBITO

iSIMPLEX

No âmbito da Medida 38 do Programa iSimplex – GuIA Responsável, a AMA desenvolveu um guia com orientações para o uso responsável da Inteligência Artificial que constitui uma referência para a implementação, pelo setor público, de uma Inteligência Artificial ética, transparente e responsável.

Foi também disponibilizada uma Ferramenta de Avaliação de Risco para aplicação em todos os projetos de Inteligência Artificial na Administração Pública.

Embora dirigidos especificamente aos projetos de Inteligência Artificial na AP, estes instrumentos podem ser utilizados noutros contextos, nomeadamente no Setor Privado e Academia.

GUIA COMPLETO VERSÃO INTEGRAL

SUMÁRIO EXECUTIVO

VERSÕES CURTAS

FOLHETO

FERRAMENTA AVALIAÇÃO do RISCO

01 BENCHMARK

Intergovernmental Organizations

OCDE

UNESCO

UNITED NATIONS

G20

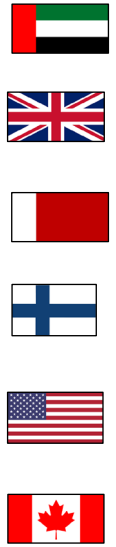
European Commission

- High-level expert group on AI
- European Ethical Charter on the use of AI in judicial systems
- White Paper on AI
- Assessment List for Trustworthy AI

Private Sector

- Microsoft
- Telefónica
- Google AI
- Deloitte.
- IBM
- McKinsey & Company

National level (examples)



01 GuIA

Descreve o contexto nacional e mundial da IA e identifica (algumas) das mudanças emergentes desta tecnologia.

Apresenta as **dimensões**, valores e princípios para uma IA Responsável e ética.

Sugere um framework conceptual para a implementação de projetos de IA Responsáveis na AP (e outros setores).

Explicação da Ferramenta de Avaliação.

01 Princípios, valores e Recomendações

Garantir a segurança, a privacidade e a proteção dos dados

Inclusão, igualdade, liberdade de expressão, desenvolvimento sustentável e bem-estar

As decisões dos sistemas estarem alinhadas com os humanos na ética e na moral

Respeito e Equidade

Qualidade e representatividade dos Dados

Eliminar os preconceitos que podem estar subjacentes nos próprios Dados

Favorecer a habitação, cultura, educação e saúde; etc.

Controlo humano da tecnologia e algoritmos

Revisão humana de decisões automatizadas. As pessoas devem ser governadas por pessoas

Capacidade de reverter decisões automatizadas

Promover o acesso à tecnologia e à informação para todos. Direito à informação

01 Princípios, valores e Recomendações (Algoritmos)

Os pressupostos em que são baseados

A atualização sistemática a que devem estar sujeitos

A avaliação da replicabilidade de soluções desenvolvidas

O grau de supervisão humana sobre as decisões

A capacidade para prever eventos raros

A previsão e mensuração do impacto social resultante da sua aplicação

As certificações necessárias

Os indicadores de fiabilidade

A identificação de métricas para avaliar o treino e monitorização

O grau de abertura na perspetiva de serem auditados e serem detetados eventuais erros

01 DIMENSÕES

RESPONSABILIZAÇÃO

Estão identificados todos os responsáveis?

ÉTICA

Os algoritmos e os dados são isentos de vieses éticos?

EXPLICABILIDADE

Os algoritmos podem ser compreendidos?

JUSTIÇA

Os algoritmos e os dados são justos e não discriminam?

TRANSPARÊNCIA

Os algoritmos são disponibilizados?

Framework conceptual estruturado em 5 dimensões

01 GuIA

Sumário Executivo

1. Enquadramento
2. IA em Portugal
3. IA ética, responsável e transparente
4. IA maliciosa
5. Recomendações
6. Barreiras e desafios
7. Ferramenta de avaliação de risco
8. Fontes Complementares
9. Anexos

01 GuIA

Revisão e contributos de especialistas (exteriores ao proj.).

Apresentação pública.

Inquérito aberto (13 perguntas).

Período de consulta pública (1 mês).

Incorporação de sugestões e correções.

02 Ferramenta de Avaliação

02 Ferramenta de Avaliação

- Analisar a suscetibilidade de projetos de IA (algoritmos), relativamente às cinco dimensões subjacentes a uma IA Responsável (alinhamento com o GuIA)

- Enriquecer a avaliação no âmbito do Parecer Prévio para projetos de IA e DS (sensibilização)
  - Apresentação obrigatória do relatório para autoavaliação e informação (anexo)

- Estabelecer referenciais de avaliação para projetos de IA
- Comparar os resultados obtidos e definir referência avaliativa nacional e setorial
- Fomentar a partilha de boas práticas, para soluções de IA/DS que tenham em vista o bem público

02 Ferramenta de Avaliação

CMD/CC

REGISTO

RESPONSABILIZAÇÃO

16Q

Permitir a afetação de responsabilidades e possibilidade plena de auditoria/inspeção

ÉTICA

24Q

Assegurar mecanismos efetivos de mitigação de enviesamentos inesperados mitigando, riscos éticos

EXPLICABILIDADE

11Q

Explicar o funcionamento e as implicações decorrentes das funções computacionais

JUSTIÇA

11Q

Incorporar garantias e salvaguardas aos utilizadores e beneficiários

TRANSPARÊNCIA

17Q

Assegurar o acesso às suas componentes e aos procedimentos aplicados

SCORE

AVALIAÇÃO LONGITUDINAL

AVALIAÇÃO DIMENSIONAL

REFERENCIAS

RELATÓRIO PDF

RESPONSABILIZAÇÃO

- Existe um responsável pela proteção de dados privados ou sensíveis utilizados pelo algoritmo?
- Existe um responsável pela gestão (identificação e mitigação) de potenciais erros ou enviesamentos resultantes do algoritmo?
- O algoritmo tem supervisão humana ao longo do seu ciclo de vida, especialmente na geração das saídas?
- O algoritmo utilizado no projeto é alvo de revisão com periodicidade adequada, incluindo atualizações decorrentes da evolução legislativa?

02 Ferramenta de Avaliação

ÉTICA

- O sistema utiliza funções computacionais para controlo de enviesamentos éticos associados à discriminação de pessoas/ grupos vulneráveis?
- O sistema possui meios para a submissão de reclamações associadas ao seu funcionamento?
- O sistema possui procedimentos para a avaliação externa regular do seu funcionamento, por exemplo por parte de um comité de ética independente?

EXPLICABILIDADE

- Existe explicação sobre a funcionamento do algoritmo?
- Existe explicação clara sobre as saídas produzidas pelo algoritmo?
- Existe explicação clara sobre os impactos da aplicação do algoritmo?

02 Ferramenta de Avaliação

JUSTIÇA

- Na fase de planeamento do projeto a equipa definiu valores e princípios de justiça com vista a evitar preconceitos e exclusão?
- O algoritmo é robusto e permite nova aprendizagem/utilização de novos dados/tratamento de situações não previstas em tempo real?
- O algoritmo tira partido de ausências ou lacunas normativas ao longo do ciclo de vida do projeto?

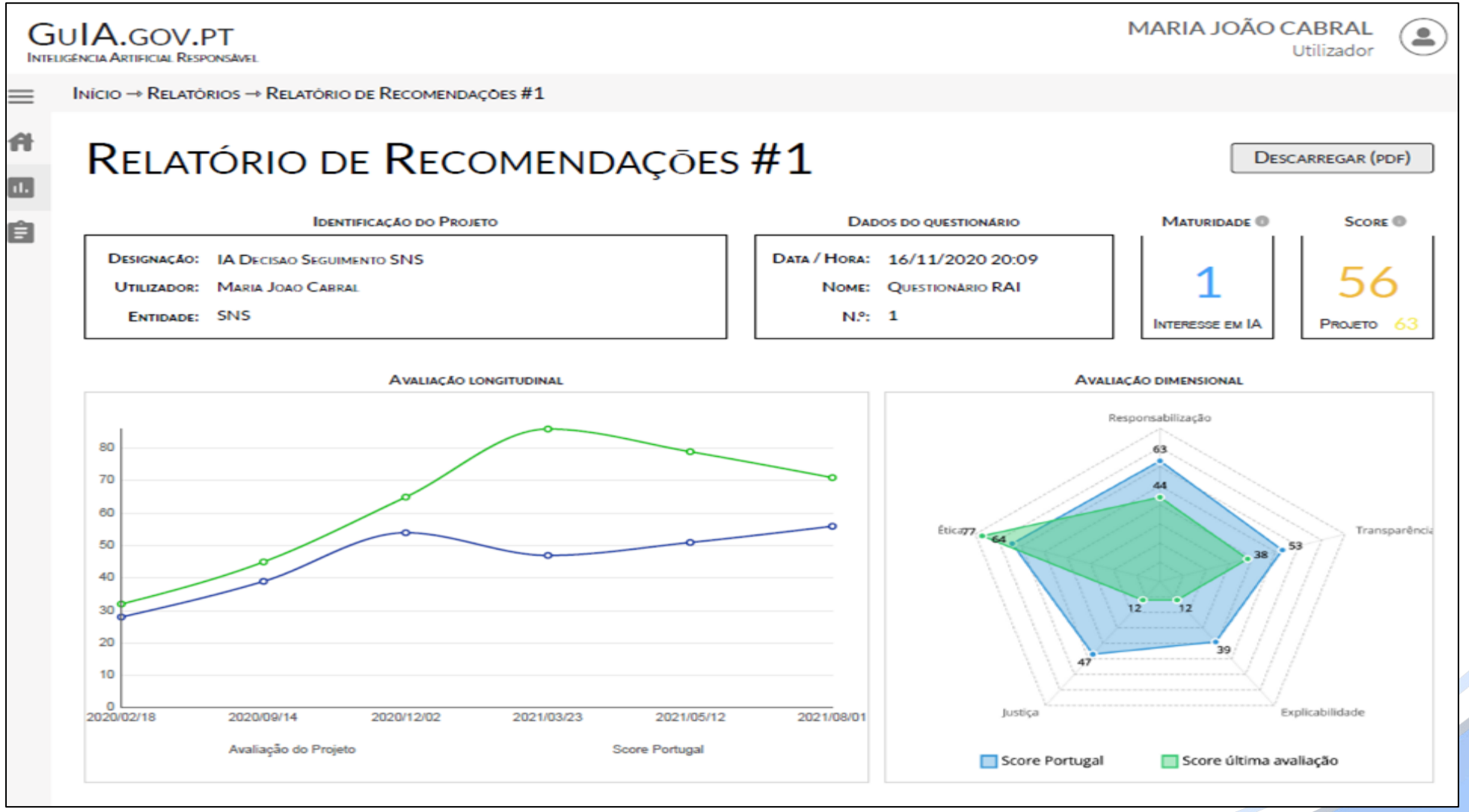
02 Ferramenta de Avaliação

TRANSPARÊNCIA

- O utilizador tem conhecimento dos dados que são utilizados no Sistema?
- O sistema disponibiliza, de forma visível e acessível, as opções de autorização para processamento dos dados pessoais do utilizador?
- A documentação associada ao projeto (p.e. descrição, códigos, catálogos de dados, entre outros) está disponível para consulta em plataformas colaborativas (e.g., Github)?

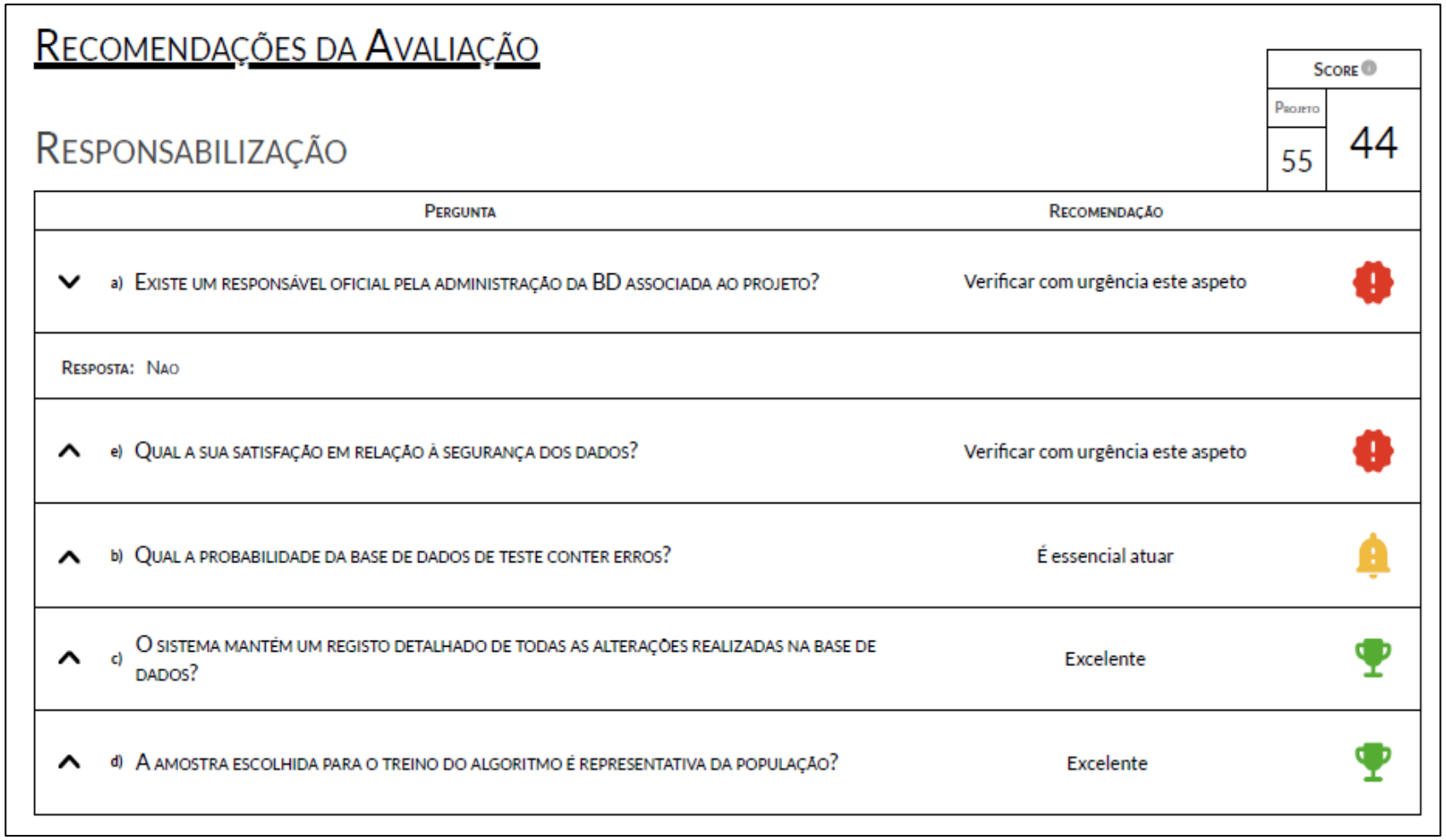
02 Ferramenta de Avaliação

Output



02 Ferramenta de Avaliação

Output



Recomendações da Avaliação

Responsabilização

Score	
Projeto	55
	44

Pergunta	Recomendação	
a) Existe um responsável oficial pela administração da BD associada ao projeto?	Verificar com urgência este aspeto	
Resposta: Nao		
e) Qual a sua satisfação em relação à segurança dos dados?	Verificar com urgência este aspeto	
b) Qual a probabilidade da base de dados de teste conter erros?	É essencial atuar	
c) O sistema mantém um registo detalhado de todas as alterações realizadas na base de dados?	Excelente	
d) A amostra escolhida para o treino do algoritmo é representativa da população?	Excelente	

GuIA Responsável

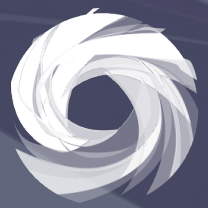
https://tic.gov.pt/areas-tematicas/inteligencia-artificial

Ferramenta de Avaliação

http://toolia.tic.gov.pt/

OBRIGADO

António Soares
antonio.soares@ama.gov.pt
TicAPP - Equipa de Dados e Interoperabilidade



tic.gov.pt