GUIA PARA A INTELIGÊNCIA ARTIFICIAL

ÉTICA, TRANSPARENTE E RESPONSÁVEL

DEZEMBRO 2020 **VERSÃO BETA**

Projeto GuIA Responsável

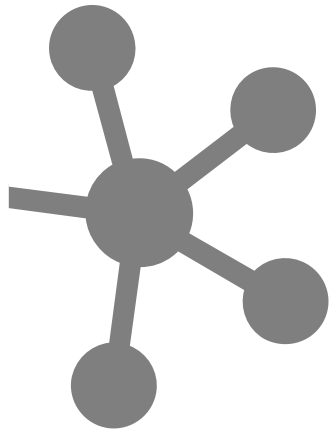
No âmbito da Medida 38 do Programa iSimplex – GuIA Responsável a AMA desenvolveu *guidelines* para o uso responsável da Inteligência Artificial e disponibiliza uma Ferramenta de Avaliação de Risco para aplicação a projetos de Inteligência Artificial, designadamente os que constam do Programa iSimplex 2019.

Estas *guidelines* e a respetiva ferramenta de avaliação de risco constituem um *framework* para a adoção de referências comuns para a implementação de uma Inteligência Artificial ética, transparente e responsável pelo setor público. Mas, acima de tudo, pretende-se que seja uma base para a discussão do tema com a sociedade em geral. Pode encontrar todas as informações em https://tic.gov.pt/pt/web/tic/guia

---

Esta é uma versão preliminar que irá ser disponibilizada a todas as partes interessadas para consulta e reunião de contributos.

ÍNDICE



01 OBJETIVO

I.	II.	III.
Elaborar um Guia que identifique Referências Internacionais, Princípios, Valores e Guidelines para a utilização da Inteligência Artificial na Administração Pública, que possa também ser uma referência no Setor Privado e para a Academia.	Promover e apoiar projetos relacionadas com Ciência dos Dados, Big Data, Machine Learning e Deep Learning. Contribuir para uma melhoria nos projetos associados a Dados e Tecnologias Emergentes, nomeadamente para critérios de avaliação, pareceres prévios e propostas de financiamento.	Desenvolver uma aplicação na web que funcione como Ferramenta de Avaliação, que permita a identificação e a mitigação do Risco.

O PODER DE TRANSFORMAÇÃO DE IA DEVE ESTAR AO SERVIÇO DAS PESSOAS E DO PLANETA.
A IA DESPERTA O MODO COMO AS PESSOAS VÊEM, AGEM E SE ENVOLVEM COM O MEIO ONDE ESTÃO INSERIDOS

VERSÃO DRAFT

Nas cidades, possibilita a análise de tendências de estacionamento, a gestão de energia, a monitorização dos níveis de poluição do ar, a uniformização da recolha de resíduos e a disponibilização de serviços mais diferenciados e ajustados às necessidades de cada cidadão, como nos transportes.

A inovação, a sustentabilidade, a competitividade e investigação científica são facilitadas.

A IA gera efeitos positivos na habitabilidade, nos cuidados de saúde, nos apoios sociais, na criação de emprego, na gestão das obras públicas, na educação e na ação climática.

A integração de informação, tecnologia e inovação consegue melhorar as operações e serviços dentro do setor público e ter infraestruturas de governação mais ágeis e resilientes.

A simplificação da comunicação e a rapidez e melhoria de qualidade dos serviços públicos, aumenta o envolvimento dos cidadãos.

Uma abordagem mais integrada permite a compreensão da comunidade, a avaliação mais acurada das situações e a tomada de decisões ou de respostas mais rápida, eficaz e adaptada, nomeadamente a nível político.

No trabalho verifica-se a redução do tempo de execução de tarefas, a redução de erros humanos, a diminuição dos perigos para o homem, apoio nos trabalhos repetitivos, aumentos de produtividade e eficiência.

IA NO MUNDO

ORGANIZAÇÕES INTERGOVERNAMENTAIS

OCDE

Principles on AI

UNESCO

Elaboration of a Recommendation on the ethics of artificial intelligence

ORGANIZAÇÃO DAS NAÇÕES UNIDAS

Towards an Ethics of Artificial Intelligence

G20

AI Principles

European Commission

Ethics Guidelines for Trustworthy AI
European High Level Expert Group on AI

European Ethical Charter on the use AI of Judicial Systems

White Paper on Artificial Intelligence: Public consultation towards a European approach for excellence and trust

Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment

SETOR PRIVADO

Exemplos

IBM

Everyday Ethics for AI

T · ·

Guidelines for AI

Microsoft

AI Principles

Telefónica

AI Principles

Telia

Guiding Principles on Trust AI Ethics

Google AI

AI at Google Our Principles

ITI

AI Policy Principles

IALATAM

Declaration of the Ethical Principles

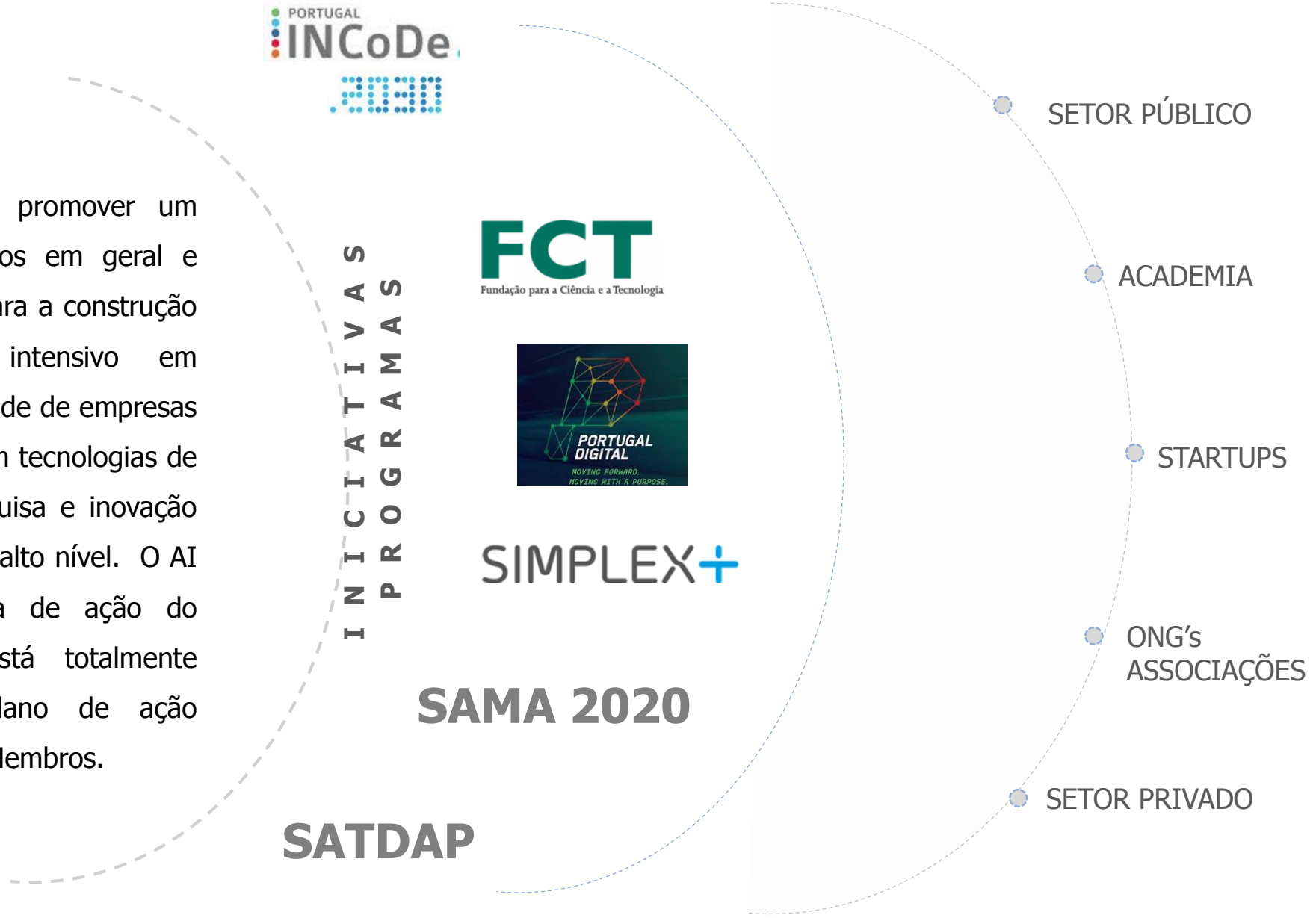
Tencent 腾讯

Six Principles of AI

ERICSSON

Ethics and AI

O objetivo do AI Portugal 2030 é promover um processo coletivo, mobilizando cidadãos em geral e principais stakeholders em particular, para a construção de um mercado de trabalho intensivo em conhecimento, com uma forte comunidade de empresas de vanguarda que produzem e exportam tecnologias de IA, apoiada por comunidades de pesquisa e inovação envolvidas em pesquisas excelentes de alto nível. O AI Portugal 2030 enquadra-se na linha de ação do INCoDe.2030 de investigação e está totalmente alinhada com as diretrizes do plano de ação coordenado da UE e dos seus Estados-Membros.

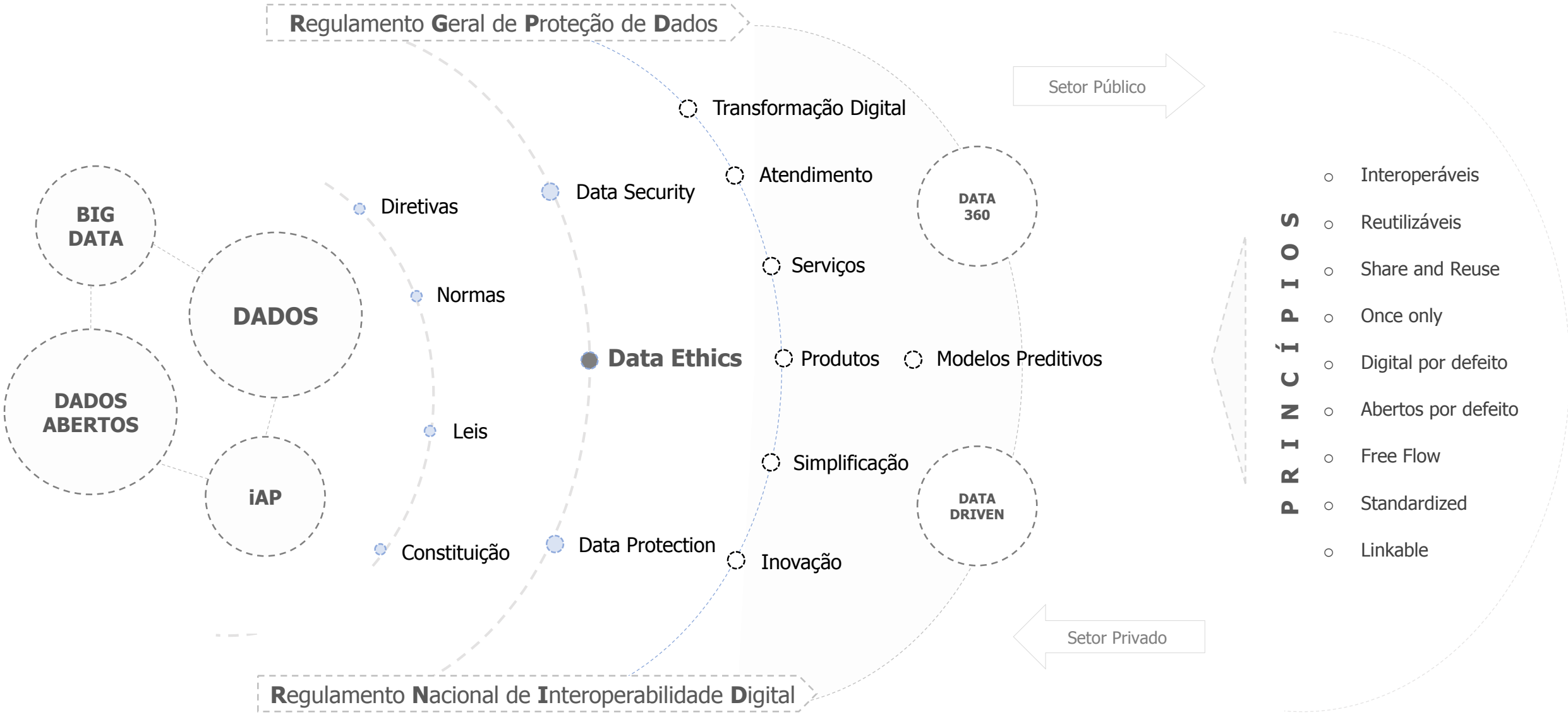


Os stakeholders abrangem todas as organizações e indivíduos envolvidos ou afetados por sistemas de IA, direta ou indiretamente. Os atores de IA acabam por ser um subconjunto dos stakeholders. É esperado um diálogo público moderado pelos governos, bem informado e interativo, incluindo todas as partes interessadas, para melhorar a compreensão da IA, debater oportunidades e desafios relacionados à IA para a economia, a sociedade e o mundo do trabalho, e informar os formuladores de políticas em todos os setores. Promover IA responsável em educação e pesquisa, intercâmbio de conhecimentos e melhores práticas, orientação para conduta empresarial responsável e incentivos para transformar IA responsável em uma vantagem competitiva.

SETOR PÚBLICO	SETOR PRIVADO	ONG's ASSOCIAÇÕES	STARTUPS	ACADEMIA
Infraestruturas de Portugal, IRN Instituto dos Registos e do Notariado, Comboios de Portugal, IMT, IPMA, IMPIC, Associação Nacional Municípios Portugueses, DGADR, Turismo de Portugal, AT Autoridade Tributária e Aduaneira, APA, DGEEC, Segurança Social, ICNF, dgTerritório, IEFP, Hidrográfico, DGRM, IAPMEI, ANACOM, IPN Instituto Pedro Nunes	Uber, REN, CTT, Zurich, Mastercard, NOS, Altice, EDP, Lusíadas, EPAL Grupo Águas de Portugal, Brisa, emel, CUF, SIBS, VISA, Vodafone, Águas de Portugal, ANA Aeroportos de Portugal, Jerónimo Martins, José de Mello Saúde, SONAE, Hospital da Luz	Cáritas Portuguesa, anf, Fundação Francisco Manuel dos Santos, Banco Alimentar contra a Fome, Champalimaud Foundation, AIP Associação Industrial Portuguesa, Fundação Calouste Gulbenkian, anje, Anetie, CIP Confederação Empresarial de Portugal	feedzai, defined crowd, Sword Health, 360imprimir, cheese me, Barkyn, jscrambler, 20tree.ai	NOVA IMS Information Management School, Universidade Católica Portuguesa, iscte Instituto Universitário de Lisboa, INOV inesc, Universidade de Coimbra, ia instituto de astrofísica e ciências do espaço, IST, Universidade de Évora

04 ECOSSISTEMA DE DADOS NA GÉNESE DA IA

VERSÃO DRAFT



5.1 DIMENSÕES

COMO APOIAR A IA QUE DEFENDA A DEMOCRACIA, O ESTADO DE DIREITO E OS DIREITOS FUNDAMENTAIS?

A REFLEXÃO ÉTICA ASSUME UM PAPEL FUNDAMENTAL QUE SE ALICERÇA NOS CONCEITOS DE:

ÉTICA

Os algoritmos oferecem mitigações para tratar vieses éticos?

- O código de ética permitirá a projeção de um conjunto de valores, princípios e diretrizes que acompanhem os desenvolvimentos tecnológicos, bem como, os elementos sociais e políticos associados.
- Um grupo de trabalho de ética e sociologia na IA deve ser estabelecido para investigar coletivamente os impactos éticos, investigar questões, definir diretrizes para as melhores práticas e publicar os conhecimentos adquiridos.

JUSTIÇA E BIAS

Os algoritmos são justos e salvaguardam os utilizadores e beneficiários?

- A abertura à inclusão e à diversidade nos sistemas de IA, aproximam a comunidade através do aumento de confiança nestas tecnologias.
- Um sistema de IA tendencioso pode conduzir ao preconceito e discriminação não intencionais.
- Elementos tendenciosos (BIAS) podem entrar no sistema algorítmico por arquétipos culturais, sociais ou institucionais pré-existentes, ou por limitações técnicas, associadas às escolhas relacionadas com a codificação, recolha, seleção ou utilização dos dados para desenvolver o algoritmo.

RESPONSABILIZAÇÃO

Os algoritmos geram responsabilização, são seguros e passíveis de auditoria?

- Integra os indivíduos responsáveis pelas partes de um sistema de IA, desde a conceção à sua implementação - responsabilidade distribuída .
- Rastreamento completo de procedimentos e resultados de modo transparente, de modo a poderem ser explicados e auditados por terceiros.
- Contabiliza a origem e a utilização de dados, modelos, interfaces de programação de aplicações e outros componentes estruturais desse sistema de IA.
- A responsabilidade por um sistema de IA pode expressar um padrão ético e estar dissociado de consequências legais.

TRANSPARÊNCIA EXPLICABILIDADE

Os algoritmos asseguram a visualização das suas componentes e dos procedimentos aplicados?

- Permite-nos explicar, inspecionar e reproduzir as decisões e a utilização dos dados por esses sistemas.
- Rastreabilidade de desenvolvimento dos processos do sistema de IA
- Comunicação clara das capacidades e nível de precisão do sistema, bem como das suas limitações.
- Explicar como uma decisão foi tomada por um modelo de IA e entender as implicações dos resultados/ impactos decorrentes.
- Consolida a confiança da sociedade em relação à tecnologia.

5.2 PRINCÍPIOS E VALORES

VERSÃO DRAFT

ALINHAMENTO COM OS OBJETIVOS DA SUSTENTABILIDADE SOCIAL, POLITICA, AMBIENTAL, EDUCACIONAL, CIENTIFICA E ECONOMICA

PRINCÍPIOS E VALORES BASILARES PARA UMA IA RESPONSÁVEL E TRANSPARENTE

- Utilização e inovação responsável
- Promoção de um ecossistema digital
- Cooperação entre organismos para uma IA de confiança
- Incorporação de feedback do sector privado, da indústria, de universidades e de organismos da AP
- Promoção da partilha de dados em dados.gov.pt
- Comunicação de possíveis consequências negativas não intencionais
- Partilha de modelos de IA transparentes e explicáveis
- Acessibilidade aos consequentes benefícios
- Promoção de um governo orientado para um serviço centrado no utilizador, na abertura, na colaboração e na acessibilidade
- Benefício e capacitação do maior número de pessoas possível
- Distribuição da prosperidade económica gerada pela IA
- Objeção às tecnologias e sistemas de IA para o fabrico de armas

LEIS, REGULAMENTOS E DIRETRIZES

> Constituição Portuguesa

Dignidade da pessoa humana

Sociedade livre, justa e solidária

> Constituição Europeia

Direitos e liberdades individuais

> Declaração Universal dos Direitos Humanos

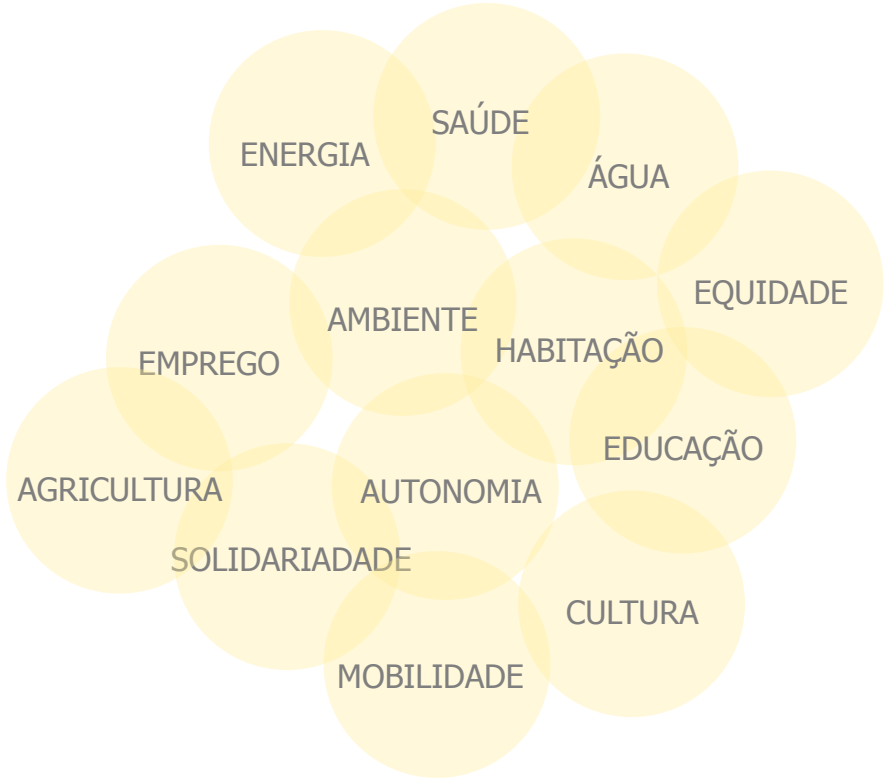
Direito à vida

Direito à segurança

5.3 INCLUSÃO IGUALDADE DESENVOLVIMENTO SUSTENTÁVEL E BEM ESTAR

A IA DEVE SER INCLUSIVA PARA TODOS…

E DEVE ATUAR EM BENEFICIO DAS SEGUINTES ÁREAS



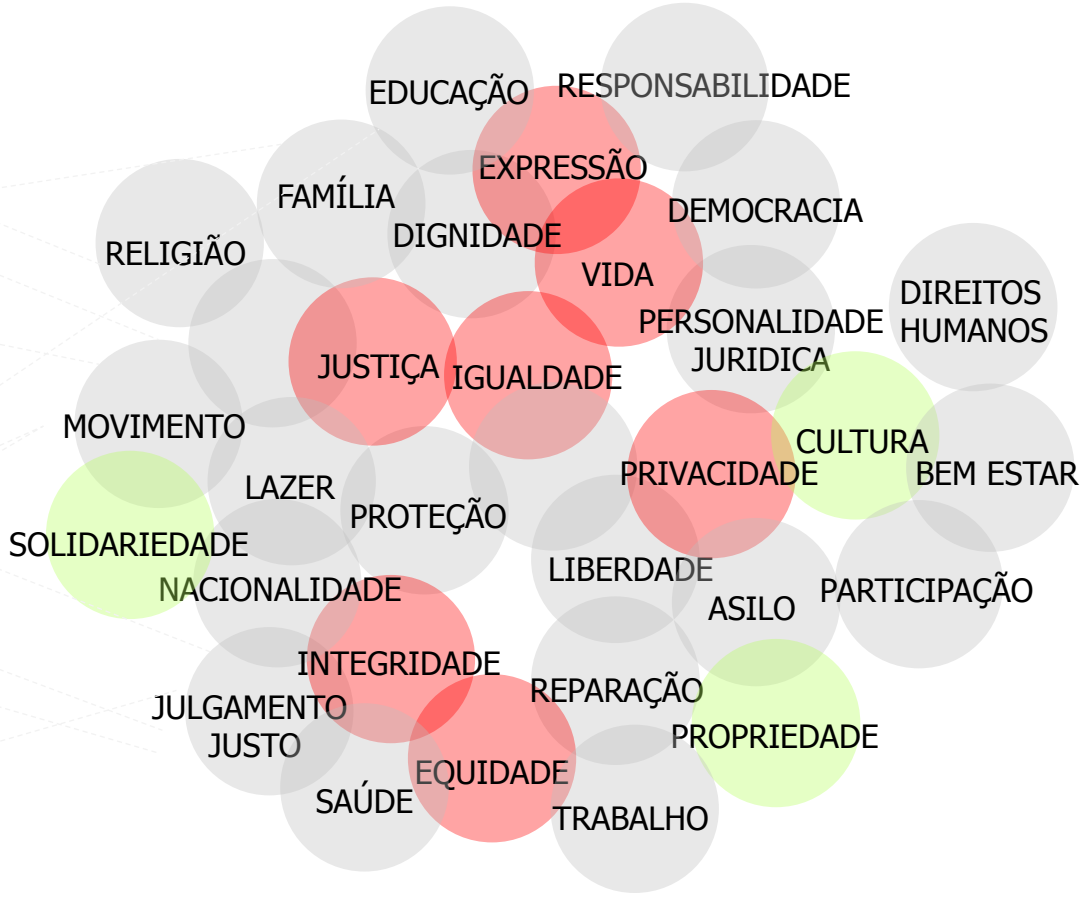
VERSÃO DRAFT

5.3 INCLUSÃO IGUALDADE DESENVOLVIMENTO SUSTENTÁVEL E BEM ESTAR

VERSÃO DRAFT

A IMPORTÂNCIA DE AVALIAR O IMPATO

- Acesso a Empréstimos Bancários
- IA preditiva para o sistema judicial
- Lawyerbots
- Cuidados de Saúde
- Educação
- Chatbots e Atendimento Virtual
- Emprego e Contratação
- Reconhecimento Facial
- Acesso a Habitação Social
- Marketing e Publicidade
- Vigilância
- Tecnologia Militar
- Reconhecimento e transcrição de voz



Ordem de importância da IA:

1º Humanidade > 2º Estado > 3º Organização

5.4 USE CASES

Saúde

- Deteção de padrões microbianos e tumorais em diagnósticos por imagem
- Identificação precoce de pandemias
- Análise de registos médicos para fornecer serviços de saúde mais personalizados, melhores e mais rápidos
- Controlo de qualidade e projeção de novos medicamentos e terapias
- Melhoria dos serviços de agendamento e atendimento
- Redução nos custos de saúde por meio de melhor programação e otimização de ativos de saúde e força de trabalho

Educação

- Tradução de sinais de linguagem gestual para a língua oral
- Construção de pareceres imediatos sobre a escrita dos alunos, permitindo que revisem os seus trabalhos e melhorem rapidamente as suas competências
- Aprendizagem vocacional
- Ensino interativo
- Acessibilidade à educação sem restrições

Mobilidade e Cidades Inteligentes

- Navegação autónoma de carros
- Otimização da experiência dos utilizadores de transportes públicos
- Redução do congestionamento de tráfego por meio de melhores serviços de informação, gerenciamento de semáforos e planeamento de obras viárias
- Controlo do turismo nas cidades
- Gestão de multidões
- Redução do consumo de energia e água

Ambiente e Ação Climática

- Rastreamento e previsão de padrões de poluição do ar, para obter melhores medidas de intervenção na sua qualidade
- Previsão de risco de desastres naturais, criando um sistema de alerta para minimizar o seu impacto
- Monitorização bioacústica e tecnologia móvel para rastrear a saúde das florestas e detetar ameaças
- Medição e modelação de variáveis relacionadas com as alterações climáticas
- Armazenamento de energia de sistemas de energia renovável, por meio de redes inteligentes
- Gestão dos recursos naturais.

Setor Agrícola

- Recolha e processamento de dados climáticos e agrícolas, com a finalidade de melhorar a irrigação de campos
- Aumento da produtividade
- Resolução de desafios associados ao uso inapropriado de pesticidas
- Rastreamento e análise de medidas de controlo de pragas, de modo a ter intervenções mais oportunas e localizadas, para estabilizar a produção agrícola e reduzir o uso de pesticidas

5.4 USE CASES

VERSÃO DRAFT

Justiça

- Recolha e relação de dados relevantes em documentos de casos judiciais, de modo a permitir que advogados pesquisem e defendam casos com maior eficácia

Setor Administrativo

- Promoção da comunicação governamental
- Disponibilização de interfaces de conversação automatizadas com funcionários virtuais para automatizar cenários de atendimento ao cidadão e às empresas
- Reforço da segurança e privacidade nos sistemas informáticos

Setor Social e Solidariedade

- Ajuda de refugiados na procura de emprego, a partir da tradução das suas experiências profissionais
- Determinação do nível de risco de suicídio de jovens LGBTQ, em serviços de ajuda
- Acompanhamento e ajuda à distância de casos de adição
- Identificação, medição e indagação das causas subjacentes da desigualdade
- Construção inteligente capaz de tornar a habitação mais acessível
- Determinação justa de apoios sociais

Cultura

- Combate às notícias falsas;
- Sugestão de atividades culturais

Setor Económico e Financeiro

- Automação de processos transacionais
- Redução do crédito mal parado para cidadãos e empresas
- Deteção de fraude
- Credit scoring
- Processos automatizados de pricing
- Otimização da experiência do usuário, fornecendo sugestões personalizadas
- Antecipação da procura de produtos e gestão de estoque e de entregas
- Automatização de processos na indústria

06 RECOMENDAÇÕES

NO ÂMBITO DO CUMPRIMENTO E PRESERVAÇÃO DE PRINCÍPIOS E VALORES PARA A IA RESPONSÁVEL RECOMENDA-SE:

1 O Controlo Humano

- Controlo humano da tecnologia
- Controlo humano dos algoritmos
- Revisão humana de decisões automatizadas - as pessoas devem ser governadas por pessoas
- Capacidade de reverter decisões automatizadas

2 A Transformação Digital e Tecnológica

- Investimentos em dados e dados abertos
- Esforços para acelerar a digitalização da AP
- Inclusão digital
- Ações da academia, indústria e sociedade civil

3 A Cooperação e Envolvimento

- Empresas de tecnologia de IA
- Fornecedores para a AP
- Organizações de usuários finais do setor privado
- Consultoria
- Especialistas em ética nas tecnologias emergentes e em ciências sociais
- Equipas multidisciplinares com uma variedade de valências
- Apoio de startups

4 A Justiça e Não discriminação

- Prevenção de preconceitos subjacentes aos dados
- Inclusão no planeamento das soluções
- Inclusão no impacto das soluções;
- Dados representativos
- Dados de elevada qualidade
- Reduzir o impato negativo para os funcionários e, quando viável, permitir a sua participação na conceção e implementação desses sistemas

5 A Privacidade

- Controlo de dados do usuário
- Consentimento
- Recomendação e informação de leis de proteção de dados
- Capacidade de restringir o processamento
- Direito à retificação
- Direito de apagar registo

06 RECOMENDAÇÕES

NO ÂMBITO DO CUMPRIMENTO E PRESERVAÇÃO DE PRINCÍPIOS E VALORES PARA A IA RESPONSÁVEL RECOMENDA-SE:

6 A Promoção de Valores Humanos

- Focado no benefício da sociedade
- Refletir os Valores e a Ética do Setor Público, bem como as obrigações internacionais e direitos humanos
- Acesso à tecnologia
- Acesso à informação
- Replicabilidade por indivíduos nas mesmas circunstâncias

7 A Responsabilidade Profissional

- Colaboração entre atores e stakeholders
- Design responsável
- Consideração de efeitos a longo prazo
- Integridade e excelência científica
- Precisão por meio de análises aprofundadas em todas as etapas
- Qualificar e certificar fornecedores

8 A Responsabilização

- Recomendação para novos regulamentos
- Avaliação com impactos mensuráveis
- Requisitos para avaliação e auditoria
- Verificabilidade e replicabilidade
- Responsabilidade legal
- Capacidade de intervir
- Responsabilidade ambiental
- Criação de órgão de monitorização
- Correção de decisões automatizadas
- Explicação satisfatória e auditável da ocorrência de erros

9 A Segurança e Proteção

- Mecanismos de confiabilidade
- Mecanismos de previsibilidade
- Geração de alarmística
- O Governo deve colaborar estreitamente com os técnicos e investigadores para investigar, prevenir e mitigar os potenciais usos maliciosos de IA

10 A Transparência e Explicabilidade

- Análise de dados recorrendo a código aberto
- Notificação quando um sistema de IA toma uma decisão sobre um indivíduo ou um grupo de indivíduos
- Requisito de relatórios regulares ao longo de todo o ciclo de vida da solução
- Direito à informação por parte dos utilizadores e ou beneficiários
- Contratação aberta de tecnologia para o Governo
- Acesso à explicação das decisões tomadas.
- Comunicação à comunidade dos efeitos da implementação de IA

06 RECOMENDAÇÕES

NUMA PERSPETIVA GERAL DE CONCEÇÃO DE SOLUÇÕES DE IA, É RECOMENDADO ÀS PARTES INTERESSADAS:

Comités

- Criação de um Comité Ético e de um Comité de Experts, que inclua profissionais das áreas em que são utilizadas tecnologias de IA (por exemplo, um painel de médicos e um painel de juízes)

Desenho de um roadmap

- Planeamento e design do processo de recolha e processamento de dados e construção de modelos associados
- Verificação e validação dos dados
- Desenvolvimento de testes
- Operacionalização e monitorização
- Garantir a acessibilidade desde o início

Planeamento de um Projeto

Responder às questões "Como?":

- Inovar de forma responsável
- Promover um ecossistema digital
- A cooperação entre organismos pode promover a confiança na IA
- Incorporar feedback do sector privado, indústria, academia, organismos da AP e promover a partilha de dados em dados.gov.pt
- Identificar consequências negativas e implicações éticas
- Partilhar modelos de IA transparentes e explicáveis
- Gerar benefícios e identificá-los
- Ajudar o Governo a ser mais eficiente e providenciar melhores serviço
- identificar o que os fornecedores com experiência comprovada técnica e ética

Algoritmos gerados

- Pressupostos em que se baseiam
- Atualização sistemática a que deverão estar sujeitos
- Avaliação de replicabilidade
- Grau de supervisão humana sobre as decisões
- Capacidade para prever eventos raros
- Previsão e mensuração do impacto social e económico
- Certificações necessárias
- Lista de indicadores de fiabilidade
- Identificação de métricas para avaliar o treino e monitorização
- Garantia de padrões de excelência científica
- Grau de abertura na perspetiva de serem auditados e serem detetados eventuais erros

07 FERRAMENTA DE AVALIAÇÃO DE RISCO

VERSÃO DRAFT

DIMENSÕES

Relatório de Avaliação e Recomendações

RESPONSABILIZAÇÃO | TRANSPARÊNCIA | EXPLICABILIDADE | JUSTIÇA | ÉTICA

Qualquer nível de maturidade

by design — Qualquer etapa do ciclo do projeto — by evolution

BENEFICIOS

- Eliminar o efeito black-box
- Como incorporar a IA nas Entidades
- Reduzir Bias
- Proteger pessoas vulneráveis
- Não discriminação
- Interdisciplinaridade
- Identificar riscos e impactos
- Monitorizar resultados
- Melhorar de modo contínuo o desempenho dos sistemas
- Melhorar a política e os mecanismos de mitigação
- Processos de inspeção/auditoria dos sistemas mais eficazes
- Segurança, qualidade e proteção dos sistemas mais eficientes
- Compreender os resultados no contexto nacional e setorial
- Sustentabilidade ambiental, social e económica mais efetiva.

	VERSÃO	ESTADO	DATA	SUMÁRIO
	COMPLETA 1.0	CONSULTA	31/12/2020	DRAFT INICIAL
	RESUMIDA 1.0	CONSULTA	31/12/2020	DRAFT INICIAL

Para mais informações ou para fazer chegar contributos e sugestões envie e-mail para: guia@ama.pt