



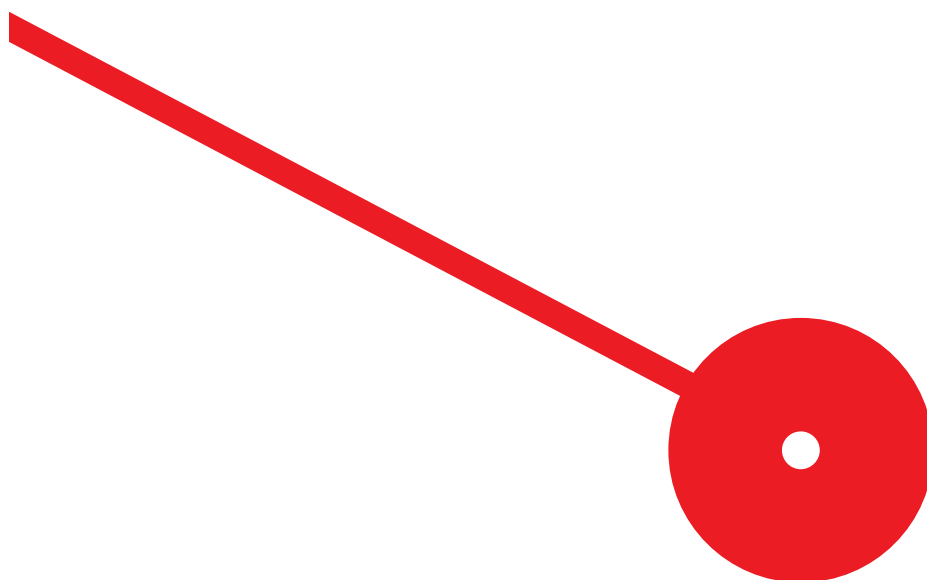
Ética e Regulação da Inteligência Artificial no contexto das Instituições Públicas Portuguesas

Ana Filipa da Silva Martins

12/2024

Versão final

Esta versão contém as críticas e sugestões dos elementos do júri

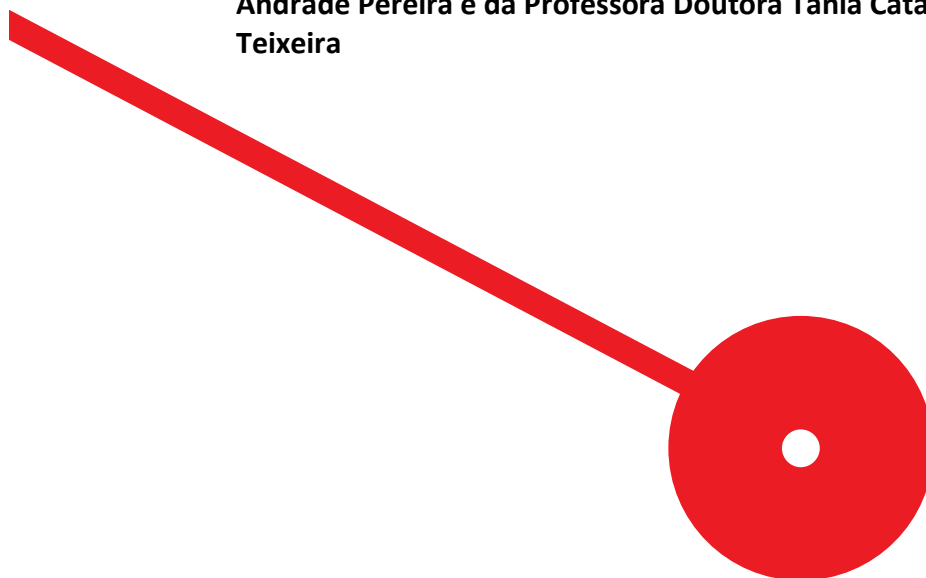




Ética e Regulação da Inteligência Artificial no contexto das Instituições Públicas Portuguesas

Ana Filipa da Silva Martins

Dissertação de Mestrado apresentado ao Instituto Superior de Contabilidade e Administração do Porto para a obtenção do grau de Mestre em Gestão das Organizações - Ramo Gestão Pública, sob orientação do Professor Doutor Adalmiro Álvaro Malheiro de Castro Andrade Pereira e da Professora Doutora Tânia Catarina Vilaça Teixeira



Agradecimentos

Gostaria de manifestar a minha profunda gratidão aos meus orientadores, Adalmiro Álvaro Malheiro de Castro Andrade Pereira e Tânia Catarina Vilaça Teixeira, pela paciência e apoio constante durante todo o processo de elaboração desta dissertação. As suas experiências e empenho foram cruciais para a realização deste trabalho, e estou extremamente agradecida pelas valiosas orientações e sugestões que me ofereceram.

Agradeço também à minha família, que esteve sempre presente, oferecendo-me amor, compreensão e estímulo nos momentos mais difíceis. Sem o suporte e a motivação que recebi da vossa parte, este percurso teria sido totalmente diferente.

A todos que, de alguma forma, contribuíram para a concretização deste trabalho, o meu sincero obrigado.

Resumo:

A crescente adoção da Inteligência Artificial (IA) nas instituições públicas representa uma mudança significativa na forma como os serviços governamentais são prestados e administrados. Com o potencial de aumentar a eficiência operacional, aprimorar a tomada de decisões e personalizar a interação com os cidadãos, a IA emerge como uma ferramenta promissora para enfrentar os desafios atuais das administrações públicas. Contudo, esta revolução tecnológica traz consigo uma série de complexidades. Questões éticas, preocupações relacionadas com a privacidade e a necessidade de uma regulação apropriada são assuntos que requerem uma reflexão cuidadosa e uma abordagem responsável.

Palavras chave: Inteligência artificial, Ética, Regulação, Instituições Públicas

Abstract:

The growing adoption of Artificial Intelligence (AI) in public institutions represents a significant change in how government services are delivered and managed. With the potential to enhance operational efficiency, improve decision-making, and personalize interactions with citizens, AI emerges as a promising tool to tackle the current challenges faced by public administrations. However, this technological revolution brings with it a range of complexities. Ethical issues, concerns related to privacy, and the need for appropriate regulation are topics that require careful consideration and a responsible approach.

Key words: Artificial Intelligence, Ethics, Regulation, Public Institutions

Índice

Capítulo I – Introdução	1
1.1 A Inteligência Artificial nas Instituições Públicas	1
1.2 Justificação da escolha do tema	1
1.3 Objetivos	2
1.4 Estrutura da dissertação	2
Capítulo II – Conceitos fundamentais e o papel da ética.....	4
2.1 Definição de Inteligência Artificial	4
2.1.1 Tipos de Inteligência Artificial.....	4
2.2 Aplicações da IA nas instituições públicas	6
2.2.1 Benefícios da implementação da IA	6
2.2.2 Riscos da implementação da IA	7
2.3 O papel da ética	9
2.3.1 Princípios éticos para o desenvolvimento da IA	10
2.3.2 Transparência e Responsabilidade.....	12
2.3.3 Questões de justiça e não discriminação.....	14
2.4 Ética na tecnologia: um panorama global.....	16
2.4.1 Panorama atual da IA nas instituições públicas em Portugal.....	18
Capítulo III – O guia da AMA e regulação da IA na Administração Pública portuguesa.....	19
3.1 Análise do Guia da AMA	19
3.1.1 Princípios Fundamentais	19
3.1.2 Recomendações práticas	20
3.1.3 Recomendações e orientações	21
3.1.4 Desafios e oportunidades da aplicação do guia.....	22
3.2 Exemplos de implementação da IA em instituições públicas portuguesas.....	24
3.3 Quadro normativo e legal em Portugal.....	26

3.3.1	Comparação com outros modelos europeus.....	28
3.3.2	Desafios reguladores e lacunas existentes	31
Capítulo IV	– Perspetivas futuras da IA nas instituições	34
4.1	Tendências emergentes em IA e ética	34
4.2	O papel das tecnologias emergentes na governação.....	36
4.3	Formação e capacitação para a ética na IA	39
4.3.1	Importância da formação em ética na IA.....	39
4.3.2	Boas práticas para a formação em ética.....	40
4.3.3	Abordagens recomendadas.....	40
4.3.4	Exemplos de iniciativas em Portugal.....	41
4.4	Propostas para um modelo regulatório inovador.....	42
Capítulo V	– Estudo de Caso: Assistente Virtual da AT	45
5.1	Contextualização do estudo de caso	45
5.2	Metodologia	45
5.3	Prontidão multidimensional	47
5.3.1	Avaliação da prontidão tecnológica.....	48
5.3.2	Avaliação social.....	48
5.3.3	Considerações legais.....	48
5.4	Prontidão organizacional	49
5.5	Necessidades, expectativas e “dores” dos cidadãos	49
5.6	Interação entre humanos e máquinas.....	51
5.7	Métricas do assistente virtual.....	51
5.8	Resultados e aprendizagens	52
Capítulo VI	– Conclusões.....	54
Referências Bibliográficas	56

Índice de Figuras

Figura 1: Proposta metodológica para uma visão sistémica	53
Figura 2: Modelo da ferramenta de prontidão multidimensional	54
Figura 3: Metodologia utilizada nas sessões de grupos focais	57
Figura 4: Mapa de jornada tipo para um dado perfil de utilizador	57
Figura 5: Modelo de análise sequencial de conversação	58
Figura 6: Tabela de métricas do AV	59

Lista de abreviaturas

IA – Inteligência Artificial

PLN – Processamento de Linguagem Natural

RGPD – Regulamento Geral de Proteção de Dados

OCDE - Organização para a Cooperação e Desenvolvimento Económico

AMA - Agência para a Modernização Administrativa

SNS - Serviço Nacional de Saúde

DGAEP - Direção-Geral da Administração e do Emprego Público

AT - Autoridade Tributária e Aduaneira

IGFSS - Instituto de Gestão Financeira da Segurança Social

IoT - Internet das Coisas

ONG – Organizações não governamentais

AV – Assistente Virtual

1.1 A Inteligência Artificial nas Instituições Públicas

A IA tem vindo a afirmar-se como uma solução importante nas instituições do setor público, introduzindo novas metodologias e recursos na gestão e na prestação de serviços. Esta inovação está a transformar a forma como estas entidades operam e se relacionam com os cidadãos. Esta dissertação analisa a influência da Inteligência Artificial nas instituições públicas, sublinhando as suas vantagens, os desafios que enfrenta e o contexto atual em que se insere.

Existem vários tipos de IA, tais como o *machine learning*, o Processamento de Linguagem Natural (PLN) e as redes neuronais. Estas tecnologias estão a ser utilizadas em diversos setores, incluindo a saúde, segurança pública e gestão de recursos. O conceito de IA surgiu pela primeira vez nos anos 50, impulsionada pelo avanço dos algoritmos e pelo aumento da capacidade de processamento computacional. Atualmente, a sua aplicação tem vindo a tornar-se uma realidade nas instituições do setor público (Russell & Norvig, 2016).

1.2 Justificação da escolha do tema

A escolha deste tema é justificada por diversos fatores que evidenciam a importância crescente da tecnologia na administração pública e no quotidiano dos cidadãos. Investigar esta área é crucial não apenas para compreender os benefícios e desafios relacionados com a implementação da IA, mas também para garantir que o avanço tecnológico se realize de maneira responsável e em consonância com os valores éticos da sociedade.

A IA está a transformar a forma como os serviços públicos são prestados, proporcionando maior eficiência, personalização e agilidade nas respostas às necessidades dos cidadãos. Com o aumento da automação e da análise de dados, as instituições públicas conseguem tomar decisões fundamentadas em evidências, o que pode levar a políticas públicas mais eficazes (Meyer et al., 2021). Contudo, esta mudança também apresenta desafios que precisam ser enfrentados. Compreender as implicações

éticas e sociais da IA é, portanto, crucial para assegurar que as tecnologias sejam utilizadas em prol do bem comum.

1.3 Objetivos

Para a realização desta dissertação foram delineados um objetivo principal e alguns objetivos secundários.

O objetivo principal passa por analisar a importância da ética e da regulação da IA nas instituições públicas, identificando os desafios e as oportunidades que surgem com a implementação desta tecnologia neste setor.

Quanto aos objetivos secundários podemos considerar os seguintes:

- Análise do estado atual da IA nas instituições públicas, destacando as principais áreas de aplicação;
- Perceção das repercussões éticas da aplicação da IA nas instituições públicas, tais como a análise do viés algorítmico, da privacidade, da transparência e da responsabilidade nas decisões automatizadas;
- Avaliação da eficácia do atual quadro regulatório em Portugal e a sua aptidão para enfrentar os desafios impostos pela IA, incluindo a conformidade com a legislação nacional e europeia;
- Elaboração de sugestões que incentivem uma abordagem ética e responsável na adoção da IA nas instituições públicas, assegurando a salvaguarda dos direitos dos cidadãos;
- Descobrir formas de incluir a sociedade civil na regulação da IA, promovendo maior transparência e aceitação das tecnologias em uso.

1.4 Estrutura da dissertação

A dissertação está organizada em sete capítulos, da seguinte forma:

- **Capítulo I – Introdução:** Apresenta o contexto, a justificação do tema e os objetivos da dissertação;
- **Capítulo II – Conceitos fundamentais e o papel da ética:** Fornece uma base teórica para a compreensão da IA no contexto das instituições públicas e explora a importância da ética no desenvolvimento e na implementação da IA;
- **Capítulo III - O guia da AMA e regulação da IA na Administração Pública Portuguesa:** Analisa o "Guia para a Inteligência Artificial, Ética, Transparente e Responsável na Administração Pública" e o quadro regulatório atual da IA em Portugal, enfatizando as normas existentes, as lacunas identificadas e as propostas para uma regulação mais eficaz;
- **Capítulo IV - Perspetivas futuras da IA nas Instituições Públicas:** Discute as direções futuras para a IA nas instituições públicas, analisando tendências emergentes, oportunidades de inovação e os desafios que ainda precisam ser enfrentados;
- **Capítulo V – Metodologia Qualitativa – Estudo de caso:** O estudo avalia um Assistente Virtual da Autoridade Tributária, destacando melhorias necessárias e garantindo conformidade ética e regulatória.
- **Capítulo VI – Conclusão:** Resume os principais resultados da pesquisa, discute as implicações práticas e apresenta as lacunas e recomendações para futuras investigações e desenvolvimentos no campo da IA e ética;

2.1 Definição de Inteligência Artificial

A IA diz respeito ao conjunto de técnicas e métodos que permitem que máquinas e sistemas computacionais realizem tarefas que normalmente exigem a intervenção humana e habilidades cognitivas. Esta área da ciência da computação visa desenvolver algoritmos e modelos que consigam reproduzir a inteligência humana em diversos contextos, desde a resolução de problemas até à tomada de decisões.

A IA não se limita apenas à execução de tarefas específicas, mas também inclui a capacidade de interagir de forma adaptativa e eficaz com o ambiente. A IA é dividida em duas categorias principais: IA fraca e IA forte. A IA fraca refere-se a sistemas desenvolvidos para realizar tarefas concretas, como o reconhecimento de voz ou a tradução automática, sem qualquer consciência ou compreensão verdadeira. Por outro lado, a IA forte é apenas uma concepção teórica de inteligência que possui uma compreensão total e uma capacidade de raciocínio semelhante à humana (Searle, 1980).

2.1.1 Tipos de Inteligência Artificial

A IA pode ser categorizada em várias tipologias, cada uma com características e aplicações distintas que têm relevância particular no âmbito da Administração Pública. As principais categorias de IA incluem a IA reativa, a IA com memória limitada, a IA da teoria da mente e a IA auto-consciente. Essas categorias diferem em termos de complexidade e capacidade de adaptação, refletindo variados níveis de inteligência computacional.

- **IA Reativa**

Esta forma de IA é desenvolvida para reagir a estímulos específicos de forma previsível, sem a capacidade de aprender ou evoluir ao longo do tempo. Um exemplo marcante é o jogo de xadrez, onde sistemas como o *Deep Blue* são programados para avaliar posições e calcular movimentos sem qualquer conhecimento sobre partidas

anteriores. Na administração pública, este tipo de IA pode ser aplicado em sistemas de monitorização, onde a resposta a eventos em tempo real é fundamental, como em sistemas de alerta para desastres naturais.

- **IA Limitada à Memória**

Este modelo tem a capacidade de aprender com experiências anteriores e aplicar esse conhecimento a novas circunstâncias. Na administração pública, essa abordagem pode ser utilizada na análise de dados para prever tendências sociais, avaliar a eficácia de políticas públicas ou otimizar a distribuição de recursos. Por exemplo, modelos preditivos podem auxiliar na identificação de áreas que necessitam de maior intervenção social, aumentando a eficácia dos programas governamentais (Binns, 2018).

- **IA da Teoria da Mente**

Apesar de ainda estar em desenvolvimento, a IA da teoria da mente possui o potencial de compreender emoções, crenças e intenções humanas. Essa capacidade poderia transformar a maneira como os serviços públicos se relacionam com os cidadãos, possibilitando uma comunicação mais empática e personalizada. Um exemplo prático seria a utilização de assistentes virtuais que não só respondem a perguntas, mas também captam o contexto emocional do utilizador, ajustando as suas respostas (Davenport & Ronanki, 2018).

- **IA Auto-Consciente**

Trata-se de ainda de uma construção teórica que se refere a sistemas capazes de ter consciência de si mesmos e das suas emoções. Embora as aplicações práticas na administração pública sejam ainda escassas, as discussões sobre as implicações éticas e sociais deste tipo de IA são fundamentais, pois podem levantar questões importantes sobre a responsabilidade e a autonomia dos sistemas inteligentes (Searle, 1980).

2.2 Aplicações da IA nas instituições públicas

As aplicações da IA nas instituições públicas são amplas e variadas, apresentando um impacto significativo na eficiência e na qualidade dos serviços oferecidos.

Os sistemas de IA podem ser empregues para automatizar tarefas repetitivas, como o tratamento de documentos e a organização de arquivos, aliviando a carga de trabalho administrativo.

Relativamente aos serviços prestados ao cidadão, estão a ser integrados *chatbots* e assistentes virtuais para fornecer informações e responder a perguntas dos cidadãos de forma rápida e eficaz.

A IA possibilita a análise de grandes quantidades de dados para detetar padrões e obter insights que podem orientar a formulação de políticas públicas. Por exemplo, a análise de dados demográficos e sociais pode ser fundamental na criação de políticas de saúde públicas mais direcionadas e eficazes (Marr, 2021).

Algumas ferramentas de IA podem melhorar a distribuição de recursos em serviços públicos. Por exemplo, algoritmos preditivos podem ser aplicados para assegurar que as ambulâncias e outras unidades de emergência estejam posicionadas adequadamente para responder rapidamente a chamadas de socorro (Brynjolfsson & McAfee, 2014).

Sistemas inteligentes podem ser utilizados para vigiar e analisar transações em tempo real, auxiliando na deteção de comportamentos suspeitos e na prevenção de fraudes. A utilização da IA neste âmbito fortalece a integridade e a transparência das operações públicas, o que contribui para aumentar a confiança dos cidadãos nas instituições.

2.2.1 Benefícios da implementação da IA

A aplicação da IA nas instituições públicas traz consigo benefícios tais como:

- **Eficiência Operacional:** A IA possibilita a execução de tarefas repetitivas com maior rapidez e precisão, levando a uma diminuição de custos e a uma utilização mais eficiente dos recursos públicos (Meyer et al., 2021). Por exemplo, no contexto médico, sistemas de IA podem auxiliar no diagnóstico de doenças com maior celeridade e precisão, examinando testes e históricos clínicos (Esteve et al., 2019).
- **Tomada de Decisão Informada:** A análise sofisticada de dados apoia a tomada de decisões fundamentadas em evidências, promovendo políticas públicas mais eficazes e em sintonia com as necessidades da população.
- **Melhoria na Prestação de Serviços:** Através da adoção de tecnologias de IA, as instituições têm a capacidade de proporcionar serviços mais personalizados e orientados para o cidadão, o que resulta numa maior satisfação e participação da comunidade. Isso não só enriquece a experiência do utilizador, mas também pode resultar numa maior satisfação do cliente e na sua lealdade (Lambrecht & Tucker, 2019).
- **Otimização de Recursos:** Em contextos governamentais, os algoritmos podem ser aplicados para melhorar a distribuição de recursos públicos, auxiliando na identificação de áreas que requerem mais intervenção, como na saúde e na educação. Isso possibilita uma utilização mais eficiente dos orçamentos e uma resposta mais eficaz em situações de crise (Meyer et al., 2021).

2.2.2 Riscos da implementação da IA

Apesar dos benefícios, a adoção da IA também pode apresentar alguns riscos:

- **Ética e Privacidade:** A implementação de dados pessoais na construção de sistemas de IA suscita preocupações significativas em relação à privacidade dos cidadãos e à proteção dos seus dados. A existência de regulamentações, como o Regulamento Geral da Proteção de Dados (RGPD), é essencial para assegurar que

a aplicação da IA respeite os direitos dos indivíduos (Voigt & Von dem Bussche, 2017).

- **Discriminação e Viés:** Os algoritmos podem reforçar preconceitos já existentes se forem treinados com dados tendenciosos. Isso pode levar a discriminações em áreas sensíveis, como emprego, concessão de crédito e justiça penal. Pesquisas indicam que sistemas de reconhecimento facial, por exemplo, frequentemente apresentam taxas de erro mais altas para indivíduos de etnias minoritárias, levantando preocupações significativas sobre a justiça e a equidade (Buolamwini & Gebru, 2018).
- **Capacitação de Recursos Humanos:** A introdução da IA exige que os profissionais possuam um conjunto adequado de competências e formação. A ausência de formação específica pode comprometer a eficácia das tecnologias e criar resistência à sua adoção.
- **Falta de Transparência:** Os processos de decisão não são facilmente acessíveis para os utilizadores. Esta falta de transparência pode provocar desconfiança e indiferença em relação à tecnologia, particularmente quando as decisões afetam diretamente a vida das pessoas. A dificuldade em questionar ou contestar decisões automatizadas é um desafio crescente que requer soluções eficazes (O'Neil, 2016).
- **Impacto na Autonomia:** A utilização crescente de algoritmos em decisões que influenciam diretamente os indivíduos pode restringir a autonomia e a liberdade de escolha. Quando as decisões são tomadas de forma automatizada, os cidadãos podem ter a sensação de que têm menos controlo sobre os seus próprios destinos, especialmente em situações em que a IA decide o acesso a serviços fundamentais (Zuboff, 2019).

2.3 O papel da ética

A regulação da IA e das tecnologias digitais, de forma geral, está a ganhar uma importância crescente no âmbito das políticas públicas, especialmente à medida que essas tecnologias se integram na vida quotidiana. A interligação entre regulação, políticas públicas e ética é crucial para assegurar que o desenvolvimento e a implementação da IA respeitem princípios como justiça, transparência e responsabilidade. Esta secção examina a importância da ética na elaboração de políticas e regulamentos, sublinhando o impacto que estes podem ter nas instituições públicas.

Com o crescimento da utilização da IA em várias áreas, desde a saúde até à segurança pública, surgem apreensões sobre os potenciais impactos negativos destas tecnologias. A ausência de diretrizes éticas bem definidas pode resultar em práticas que coloquem em risco a privacidade dos cidadãos, perpetuem discriminações ou resultem em decisões automatizadas sem a supervisão apropriada (Comissão Europeia, 2020). A regulação ética é, assim, uma resposta essencial para minimizar esses riscos e incentivar um desenvolvimento tecnológico responsável.

Os sistemas regulatórios em vigor muitas vezes não conseguem acompanhar a complexidade e a velocidade das inovações tecnológicas. As leis tendem a ser lentas na adaptação a novas realidades, o que resulta em lacunas que podem ser aproveitadas. Por exemplo, a ausência de uma estrutura legal abrangente para a IA pode permitir que empresas atuem sem supervisão, aumentando o risco de abusos, como a recolha indevida de dados ou a utilização de algoritmos discriminatórios (Binns, 2018).

Um exemplo positivo é o RGPD, que define princípios estritos sobre a recolha e o tratamento de dados pessoais. Este regulamento salvaguarda a privacidade dos cidadãos e exige que as organizações sejam transparentes nas suas práticas, promovendo um ambiente que incentiva a responsabilidade ética na utilização de tecnologias (Voigt & Von dem Bussche, 2017).

A incorporação de princípios éticos nas políticas públicas é essencial para garantir que a tecnologia traga benefícios à sociedade em geral. A ética deve atuar como um guia na formulação de políticas que regulam a IA, assegurando que as decisões tomadas por algoritmos reflitam valores sociais importantes, como a equidade e a inclusão. Este

enfoque ético pode ser ilustrado pela criação de conselhos consultivos que envolvam especialistas em ética, representantes da sociedade civil e cidadãos, promovendo um diálogo aberto sobre as implicações das tecnologias (Jobin et al., 2019).

Para além disso, as políticas públicas precisam de ser suficientemente flexíveis para se adaptarem às rápidas transformações tecnológicas. A realização de revisões regulares e de consultas públicas pode facilitar a atualização de regulamentos e assegurar que estes reflitam os valores de uma sociedade em constante evolução. Este método participativo reforça a legitimidade das políticas e fomenta uma maior aceitação por parte da população.

Diversas iniciativas a nível global evidenciam a incorporação de princípios éticos nas políticas públicas ligadas à IA. A Organização para a Cooperação e Desenvolvimento Económico (OCDE) elaborou diretrizes que sublinham a importância de os sistemas serem concebidos de modo a respeitar os direitos humanos e a promover a inclusão (OECD, 2019). Estes princípios servem de referência para países que desejam criar ou rever as suas legislações sobre IA.

Outro exemplo relevante é a atuação da Comissão Europeia, que, para além do RGPD, está a criar um quadro regulatório para a IA que integra valores éticos essenciais. As diretrizes sugeridas pela Comissão visam assegurar que a IA seja utilizada de maneira a respeitar a dignidade humana, promovendo a confiança e a segurança (Comissão Europeia, 2021).

2.3.1 Princípios éticos para o desenvolvimento da IA

Para assegurar que a IA seja utilizada de forma responsável e vantajosa, vários princípios éticos têm surgido como essenciais. Estes princípios podem contribuir para a construção de um futuro em que a tecnologia respeite os valores humanos e fomente o bem-estar social.

- **Transparência**

A transparência é um princípio fundamental no desenvolvimento da IA. Refere-se à capacidade de compreender e rastrear como os sistemas de IA fazem as suas escolhas. A ausência de transparência pode gerar desconfiança entre os utilizadores, especialmente em áreas críticas como a saúde ou a justiça. É crucial que as organizações clarifiquem os processos algorítmicos e os dados utilizados, facilitando assim a compreensão das decisões automatizadas (Comissão Europeia, 2020). Promover a transparência não só aumenta a confiança, mas também permite responsabilizar as entidades que desenvolvem e utilizam IA.

- **Justiça e Não Discriminação**

A justiça é um princípio que garante que os sistemas de IA não perpetuem ou amplifiquem desigualdades já existentes. O viés algorítmico, que pode surgir de dados tendenciosos ou de escolhas de *design* inadequadas, pode resultar em discriminações em áreas como recrutamento, concessão de crédito ou aplicação da lei. É fundamental aplicar métodos rigorosos para identificar e reduzir o viés nos dados e algoritmos, assegurando que a IA funcione de forma justa e equitativa (Barocas et al., 2019). Portanto, promover a justiça deve ser um objetivo central em todas as fases do desenvolvimento da IA.

- **Responsabilidade e Prestação de Contas**

A responsabilidade diz respeito à importância de atribuir consequências pelas ações e decisões. É fundamental que existam mecanismos bem definidos para responsabilizar tanto os desenvolvedores quanto as organizações que aplicam estas tecnologias. Isso inclui a criação de estruturas que garantam que as decisões automatizadas possam ser auditadas e, se necessário, contestadas (Binns, 2018). A prestação de contas é crucial para assegurar que os direitos dos cidadãos sejam respeitados e que os interesses da sociedade estejam protegidos.

- **Privacidade e Proteção de Dados**

A salvaguarda da privacidade é um princípio fundamental, especialmente num contexto onde a IA requer grandes quantidades de dados pessoais. É crucial que os sistemas de IA sejam desenvolvidos com o objetivo de proteger a privacidade dos indivíduos, assegurando que os dados sejam recolhidos, armazenados e utilizados de

maneira responsável. Isso implica a implementação de medidas de segurança eficazes e a adesão a regulamentos como o RGPD. A proteção da privacidade cumpre uma função legal e reforça a confiança dos cidadãos nas tecnologias.

- **Beneficência e Não Maleficência**

Os princípios de beneficência e não maleficência requerem que o desenvolvimento e a utilização da IA tenham como objetivo promover o bem-estar humano e evitar causar prejuízos. Isso significa que as tecnologias devem ser projetadas para proporcionar benefícios sociais, elevando a qualidade de vida e fomentando a inclusão (Floridi et al., 2018). Ao mesmo tempo, é crucial evitar aplicações que possam causar danos, como nos casos de vigilância invasiva ou em sistemas de IA aplicados a armamentos.

- **Sustentabilidade**

A sustentabilidade é um princípio que se torna progressivamente mais importante à medida que as preocupações ambientais se tornam mais evidentes. O desenvolvimento da IA deve ter em conta não só os impactos sociais, mas também os efeitos ambientais. Isso implica a redução da pegada de carbono associada ao treino de modelos de IA e a análise de como as tecnologias podem ser utilizadas para incentivar práticas sustentáveis em diferentes setores (Marr, 2021). A adoção de abordagens sustentáveis no desenvolvimento da IA pode contribuir para um futuro mais equilibrado e responsável.

2.3.2 Transparência e Responsabilidade

A transparência e a responsabilidade são dois princípios essenciais na ética da IA, tendo um papel determinante na construção da confiança dos cidadãos e na redução dos riscos associados a esta tecnologia. Com a crescente complexidade dos sistemas de IA e a sua integração nas operações sociais e administrativas, torna-se indispensável assegurar que estas tecnologias sejam compreendidas e passíveis de auditoria.

- **Transparência dos Dados**

Para garantir que a IA seja desenvolvida de maneira responsável, é fundamental que as fontes de dados utilizadas sejam claras e compreensíveis. A utilização de dados

tendenciosos ou de baixa qualidade pode resultar em resultados injustos e discriminatórios. Portanto, as organizações devem ser transparentes sobre a proveniência dos dados, os métodos de recolha e as práticas de processamento, assegurando que estão em conformidade com as normas de proteção de dados, como o RGPD da União Europeia.

- **Transparência Algorítmica**

Os algoritmos de IA frequentemente funcionam como “caixas-pretas”, tornando difícil entender como as decisões são tomadas. Para melhorar a transparência, é crucial que os desenvolvedores integrem mecanismos que expliquem claramente os fundamentos das decisões algorítmicas. Ferramentas de explicação, que revelem os critérios que levam a resultados específicos, são fundamentais para aumentar a confiança nas decisões automatizadas (Doshi-Velez & Kim, 2017).

- **Transparência Organizacional**

As instituições devem implementar uma abordagem de transparência em toda a sua estrutura, abrangendo também a gestão e supervisão dos projetos de IA. A divulgação de relatórios de impacto, auditorias e avaliações independentes pode contribuir para assegurar que os sistemas de IA estejam em conformidade com os valores éticos e sociais, promovendo um ambiente de responsabilização (Binns, 2018).

- **Responsabilidade dos criadores**

Os profissionais responsáveis pelo design e pela implementação de sistemas de IA devem ter plena consciência das implicações éticas das suas decisões. Isso requer formação em ética e uma compreensão das repercussões sociais e legais das tecnologias que desenvolvem. A responsabilidade deve ser integrada em todo o ciclo de vida do desenvolvimento da IA, desde a fase de conceção até à implementação e à manutenção (Comissão Europeia, 2020).

- **Mecanismos de Responsabilização**

É essencial que haja mecanismos legais e institucionais que possibilitem a responsabilização por decisões tomadas por sistemas de IA. Estes mecanismos devem contemplar a possibilidade de contestar decisões automatizadas e de solicitar reparações

quando estas não respeitam direitos ou causam danos. O avanço de legislações que abordem claramente a responsabilidade na IA é um passo crucial para garantir a proteção dos cidadãos (Jobin et al., 2019).

- **Governança e Supervisão**

O estabelecimento de estruturas de governança sólidas que incluam diversos intervenientes é fundamental para promover a responsabilidade na IA. Isso pode incluir a formação de comités de ética, a realização de auditorias periódicas e a criação de plataformas de feedback onde os utilizadores possam manifestar preocupações ou propor melhorias. A monitorização contínua dos sistemas de IA é essencial para assegurar que estes se mantenham em conformidade com os padrões éticos e legais definidos (OECD, 2019).

2.3.3 Questões de justiça e não discriminação

As questões de justiça e não discriminação são fundamentais nas conversas sobre a IA, particularmente no que diz respeito à sua utilização em situações que impactam diretamente a vida das pessoas. Com a crescente integração dos sistemas de IA nos processos de decisão, as preocupações com a justiça e a imparcialidade tornam-se cada vez mais pertinentes.

A presença de viés algorítmico pode levar a discriminação sistemática e reforçar desigualdades sociais já existentes, tornando crucial a análise crítica desses fatores. Por exemplo, uma investigação realizada por Buolamwini e Gebru (2018) revelou que sistemas de reconhecimento facial apresentam taxas de erro muito mais altas para mulheres de pele escura em comparação com homens de pele clara. Este tipo de viés não só afeta a eficácia dos sistemas, como também provoca sérias consequências sociais e legais, especialmente em contextos onde as decisões algorítmicas influenciam oportunidades de emprego ou o acesso a serviços.

As consequências das decisões algorítmicas discriminatórias vão além dos impactos individuais, afetando comunidades inteiras e podendo reforçar estigmas sociais. Quando grupos marginalizados são sistematicamente prejudicados por decisões automatizadas,

isso pode agravar as desigualdades existentes e perpetuar ciclos de pobreza e exclusão social. Um exemplo significativo é a aplicação de algoritmos em sistemas de justiça criminal, onde decisões relativas a fianças e penas podem ser influenciadas por dados que refletem preconceitos raciais e socioeconómicos (Angwin et al., 2016).

Para enfrentar as questões de justiça e não discriminação, é essencial implementar uma série de garantias e práticas durante o desenvolvimento e a utilização da IA:

- **Auditoria de Algoritmos:** A realização de auditorias regulares nos sistemas de IA é uma medida essencial para identificar e reduzir o viés. Essas auditorias devem analisar não só a precisão dos algoritmos, mas também as suas consequências sociais e éticas. A transparência nos resultados das auditorias é vital para fortalecer a confiança do público nas tecnologias em uso (Raji & Buolamwini, 2019).
- **Diversidade nas Equipas de Desenvolvimento:** A participação de profissionais de diferentes origens e experiências nas equipas responsáveis pelo desenvolvimento de sistemas de IA pode ajudar a detetar preconceitos que poderiam passar despercebidos. A diversidade não só enriquece a abordagem dos problemas, mas também favorece a criação de soluções mais justas e equitativas (Smith et al., 2018).
- **Definição de Diretrizes Éticas:** É fundamental estabelecer diretrizes claras que enfatizem a justiça e a não discriminação. As organizações e entidades governamentais devem elaborar políticas que incentivem o desenvolvimento responsável da IA, garantindo que todos os sistemas sejam projetados para reduzir viés e discriminação. Isso abrange a promoção de melhores práticas na recolha de dados e na conceção de algoritmos (Comissão Europeia, 2020).

A participação da comunidade é uma estratégia eficaz para assegurar que as preocupações relacionadas com justiça e discriminação sejam reconhecidas e levadas em conta. Consultas públicas, fóruns e *workshops* que integrem a participação de grupos frequentemente marginalizados podem proporcionar informações valiosas sobre como os sistemas de IA afetam a vida das pessoas. Este envolvimento não apenas promove a inclusão, mas também contribui para o desenvolvimento de soluções mais justas e equilibradas (Wang et al., 2021).

2.4 Ética na tecnologia: um panorama global

A ética na tecnologia tem-se afirmado como uma área fundamental, especialmente com o crescimento e a difusão da IA e outras inovações digitais. O cenário global da ética tecnológica revela uma diversidade de abordagens e práticas que variam conforme o contexto cultural, político e económico de cada país.

O acelerado avanço da IA e da automação suscita apreensões acerca dos impactos que estas tecnologias podem causar na sociedade. A capacidade da IA de analisar vastas quantidades de dados e tomar decisões automatizadas pode levar a situações de discriminação, violações de privacidade e resultados imprevistos. Assim, é essencial criar diretrizes éticas que orientem tanto o desenvolvimento quanto a implementação dessas tecnologias, garantindo que contribuam para o bem-estar social e respeitem os direitos humanos (Jobin et al., 2019).

Várias instituições e países têm assumido a responsabilidade de criar princípios éticos que orientem a implementação da IA. Um exemplo disso é o conjunto de diretrizes publicado pela Comissão Europeia em 2019, que sublinha a importância de um desenvolvimento "ético, seguro e centrado no ser humano" (Comissão Europeia, 2019). Este documento realça valores como a transparência, a responsabilidade e a não discriminação, sugerindo que as tecnologias sejam concebidas tendo em conta a inclusão e a equidade.

Nos Estados Unidos, a ética tecnológica é frequentemente discutida através de grupos de trabalho e iniciativas de investigação, como a *Partnership on AI*, que junta

empresas do setor tecnológico, organizações da sociedade civil e acadêmicos para debater e promover práticas responsáveis na utilização da IA (Partnership on AI, 2020). Estas colaborações têm como objetivo estabelecer um espaço para diálogo e partilha de melhores práticas que ajudem a reduzir os riscos relacionados com o uso da IA.

As perspetivas éticas sobre a tecnologia podem variar consideravelmente entre diferentes regiões. Na Ásia, por exemplo, países como a China adotaram uma abordagem mais reguladora, focando na segurança e na estabilidade social nas suas orientações éticas. O governo chinês implementou normas que não só incentivam o progresso da IA, mas também garantem que essas tecnologias estejam em consonância com os valores e objetivos sociais do país (Zhang, 2021).

Em contrapartida, na Europa, a ética tecnológica está frequentemente ligada a um forte foco nos direitos individuais e na salvaguarda de dados. O RGPD, que entrou em vigor em 2018, exemplifica como a Europa tem dado prioridade à privacidade e à autonomia dos cidadãos em face do avanço tecnológico (Voigt & Von dem Bussche, 2017). Este regulamento estabelece normas rigorosas sobre a forma como os dados pessoais podem ser recolhidos e utilizados, refletindo um compromisso com a ética na era digital.

Apesar das iniciativas em andamento, ainda existem muitos desafios a serem enfrentados. A ausência de normas globais coesas e a dificuldade em aplicar diretrizes éticas em contextos variados são questões urgentes. A interconexão das tecnologias a nível mundial implica que as consequências de ações realizadas num país podem impactar outras regiões, tornando imprescindível um esforço conjunto para estabelecer padrões éticos universais.

A educação e a sensibilização sobre a ética tecnológica são questões fundamentais. Profissionais da tecnologia, formuladores de políticas e cidadãos devem ser dotados do conhecimento necessário para compreender as implicações éticas das tecnologias que utilizam ou desenvolvem. A integração de temas éticos na formação académica e profissional pode contribuir para a formação de uma nova geração de tecnólogos mais conscientes e responsáveis (Goggin, 2019).

2.4.1 Panorama atual da IA nas instituições públicas em Portugal

Em Portugal, a adoção da IA nas entidades públicas está a crescer. O "Guia para a IA, Ética, Transparente e Responsável na Administração Pública", elaborado pela Agência para a Modernização Administrativa (AMA), define diretrizes e princípios que visam assegurar a utilização ética e responsável da IA nas instituições. A crescente ênfase nas questões éticas e sociais demonstra uma consciencialização crescente sobre a necessidade de harmonizar a inovação tecnológica com os direitos e as necessidades dos cidadãos.

Há uma interação complexa entre inovação, eficiência e responsabilidade ética. Com o avanço tecnológico, é crucial que as instituições se dediquem a enfrentar os desafios que surgem, assegurando que a IA traga benefícios à sociedade de maneira justa e equitativa. A formação, a regulamentação e a transparência são elementos fundamentais para uma implementação eficaz da IA nas administrações públicas.

CAPÍTULO III – O GUIA DA AMA E REGULAÇÃO DA IA NA ADMINISTRAÇÃO PÚBLICA PORTUGUESA

3.1 Análise do Guia da AMA

O "Guia para a Inteligência Artificial, Ética, Transparente e Responsável na Administração Pública", criado pela Agência para a Modernização Administrativa (AMA) de Portugal, oferece um conjunto de orientações e princípios destinados a incentivar o uso responsável da IA nas entidades públicas. Este documento surge como uma resposta à crescente adoção da IA na administração pública e à necessidade de assegurar que esta tecnologia seja empregue de maneira ética, transparente e respeitadora dos direitos dos cidadãos.

O guia divide-se em várias seções que abordam desde os princípios fundamentais da ética na IA até recomendações práticas para a sua implementação nas instituições públicas. Os principais objetivos são:

- **A promoção da confiança** dos cidadãos nas tecnologias de IA utilizadas pelos serviços públicos;
- **Assegurar a analogia** com os direitos e liberdades fundamentais, em conformidade com a legislação nacional e europeia;
- **Orientar as entidades públicas** na implementação de soluções de IA que respeitem os princípios éticos e que visem a melhoria da qualidade dos serviços prestados.

3.1.1 Princípios Fundamentais

Os princípios delineados no guia são essenciais para a construção de uma abordagem ética à IA:

- **Legalidade e Legitimidade:** As soluções de IA precisam ser criadas e implementadas em conformidade com a legislação, garantindo o respeito pelos

direitos dos cidadãos. Isso significa uma adesão rigorosa ao RGPD e a outras normas pertinentes;

- **Transparência e Explicabilidade:** O guia sublinha a importância de os cidadãos entenderem o processo de tomada de decisões automatizadas. A transparência é essencial para assegurar que essas decisões não sejam vistas como arbitrárias, e a explicabilidade é fundamental para que os cidadãos possam questionar e compreender os resultados das ações algorítmicas;
- **Não Discriminação e Inclusão:** É salientado que os sistemas de IA não devem reforçar desigualdades ou preconceitos. O guia sugere a adoção de práticas que garantam a equidade no acesso aos serviços públicos e que levem em conta as necessidades dos grupos vulneráveis;
- **Responsabilidade e Prestação de Contas:** As organizações que adotam soluções de IA devem ser responsáveis pelas suas ações e pelos resultados da utilização dessas tecnologias. Isso envolve o estabelecimento de mecanismos que possibilitem auditorias e a responsabilização em situações de decisões duvidosas ou injustas.

3.1.2 Recomendações práticas

Além dos princípios éticos, o guia também fornece recomendações práticas que as instituições públicas devem seguir para implementar a IA de forma responsável:

- **Desenvolvimento de Competências:** O guia recomenda que as entidades públicas apostem na formação de colaboradores sobre a IA, ética e transparência, preparando-os para fazer escolhas informadas em relação à adoção de tecnologias.

- **Criação de Mecanismos de *Feedback*:** Recomenda-se a criação de canais de comunicação com os cidadãos, permitindo-lhes manifestar preocupações ou sugestões sobre a utilização da IA nos serviços públicos. O retorno da comunidade pode ser um recurso importante para ajustar e aprimorar as práticas existentes.
- **Avaliação de Impacto:** É fundamental realizar avaliações de impacto ético e social antes de introduzir novas tecnologias. Essas análises devem ter em conta não só a eficácia técnica, mas também as implicações sociais e éticas das soluções sugeridas.

3.1.3 Recomendações e orientações

O Guia apresenta um conjunto de recomendações e orientações destinadas a assegurar a utilização responsável da IA nas entidades públicas. Estas diretrizes são fundamentais para a implementação de tecnologias que respeitem os direitos dos cidadãos e promovam um serviço público eficiente e equitativo. As recomendações e orientações incluem:

- Transparência;
- Explicabilidade;
- Auditoria de Viés;
- Design Inclusivo;
- Programas de Formação;
- Criação de Especialistas Internos;
- Criação de um canal de comunicação;
- Consultas Públicas;
- Avaliações de Impacto Ético;
- Monitorização Pós-Implementação;

Estas orientações delineadas são fundamentais para garantir que a IA seja utilizada de forma responsável nas instituições públicas.

3.1.4 Desafios e oportunidades da aplicação do guia

- **Falta de Formação e Capacitação**

Um dos maiores desafios para a implementação eficaz do guia é a falta de formação especializada em IA e ética nas instituições públicas. Muitos colaboradores podem não possuir o conhecimento necessário sobre a tecnologia, as suas implicações e os princípios éticos a serem respeitados. Esta carência de formação pode resultar em resistência à adoção de novas tecnologias e em falhas na sua implementação (DGAEP, 2020).

- **Complexidade Técnica e Algorítmica**

A complexidade técnica dos sistemas de IA pode dificultar a sua compreensão e utilização ética. Esta falta de clareza pode provocar desconfiança entre os cidadãos e levantar preocupações acerca da justiça das decisões automatizadas.

- **Viés e Discriminação**

A deteção e a redução do viés algorítmico permanecem um desafio importante. Assegurar que os dados utilizados sejam representativos e não discriminatórios é fundamental para promover a igualdade (Buolamwini & Gebru, 2018).

- **Mudança Cultural nas Instituições**

A implementação de um guia ético exige não só a adoção de novas tecnologias, mas também uma transformação cultural nas instituições públicas. Isso envolve fomentar uma mentalidade que valorize a ética e a transparência em todas as ações, o que pode ser um processo moroso e desafiante.

- **Recursos Limitados**

A aplicação das diretrizes sugeridas pode exigir investimentos consideráveis em tecnologia, formação e processos. As restrições orçamentais nas instituições públicas podem dificultar a disponibilização dos recursos necessários para uma implementação completa e eficaz (Melo et al., 2021).

- **Aperfeiçoamento da Eficiência Administrativa**

A implementação de soluções de IA que estejam em conformidade com as orientações do guia pode levar a melhorias substanciais na eficiência administrativa. A automação de tarefas rotineiras pode libertar recursos humanos para se concentrarem em atividades mais complexas que requerem um toque pessoal, elevando assim a eficácia dos serviços públicos.

- **Fortalecimento da Confiança Pública**

Ao adotar os princípios de transparência e responsabilidade, as instituições podem reforçar a confiança dos cidadãos na administração pública. A aplicação de práticas éticas e a partilha de informações sobre o processo de tomada de decisões podem aprimorar a relação entre o governo e a população, fomentando uma cultura de participação e envolvimento.

- **Inovação e Modernização**

A implementação do guia pode estimular a inovação nas entidades públicas, estabelecendo um ambiente favorável ao desenvolvimento e à integração de novas tecnologias. A modernização dos serviços públicos, mediante o uso da IA, pode resultar na criação de soluções mais eficazes e orientadas para o cidadão, promovendo o avanço social e económico.

- **Desenvolvimento de Parcerias**

A aplicação do guia pode facilitar a formação de colaborações entre o setor público e privado, assim como com instituições de ensino. Estas parcerias podem gerar projetos inovadores que tragam vantagens para ambas as partes e incentivem o desenvolvimento de soluções tecnológicas que estejam em sintonia com os princípios éticos.

3.2 Exemplos de implementação da IA em instituições públicas portuguesas

A adoção da IA nas instituições públicas em Portugal tem aumentado, evidenciando um interesse crescente em otimizar a eficiência, a transparência e a qualidade dos serviços oferecidos aos cidadãos. Diversos exemplos demonstram como a IA está a ser aplicada de maneira inovadora nas administrações públicas, com um foco especial na ética e na responsabilidade.

- **Serviço Nacional de Saúde (SNS)**

Um dos exemplos mais destacados da utilização da IA no setor público em Portugal é a sua aplicação no Serviço Nacional de Saúde (SNS). O SNS tem recorrido a sistemas de IA para otimizar o atendimento médico e aprimorar os diagnósticos. Em particular, ferramentas de IA estão a ser utilizadas na análise de exames médicos, como radiografias e ressonâncias magnéticas, facilitando a deteção precoce de doenças, como o cancro (Esteva et al., 2019). Estas tecnologias aumentam a precisão dos diagnósticos e também aceleram o tempo de resposta, beneficiando os pacientes.

Para além disso, o SNS tem utilizado *chatbots* para simplificar o agendamento de consultas e esclarecer dúvidas comuns, oferecendo um atendimento mais eficiente e acessível aos cidadãos. Estes sistemas automatizados contribuem para reduzir a carga de trabalho dos profissionais de saúde, permitindo que se concentrem em cuidados mais críticos e personalizados (Melo et al., 2021).

- **Direção-Geral da Administração e do Emprego Público (DGAEP)**

A DGAEP tem impulsionado iniciativas que incorporam a IA com o objetivo de aprimorar a gestão de recursos humanos na administração pública. Um exemplo disso é a criação de sistemas de análise preditiva para gerir competências e identificar as necessidades de formação dos funcionários públicos. Estes sistemas utilizam dados relativos ao desempenho e à formação para antecipar quais as áreas que requerem

intervenção, facilitando a programação de formações e o desenvolvimento de carreira dos colaboradores (DGAEP, 2020).

A DGAEP tem investigado o uso de algoritmos para otimizar a eficiência nos processos de recrutamento. A adoção de ferramentas de triagem automatizadas para analisar as candidaturas possibilita a identificação dos perfis mais adequados para determinadas funções, aumentando a eficácia do processo e diminuindo o tempo necessário para a seleção de candidatos (DGAEP, 2021).

- **Autoridade Tributária e Aduaneira (AT)**

A Autoridade Tributária e Aduaneira tem aplicado IA para prevenir a evasão fiscal e melhorar a arrecadação de impostos. A utilização de algoritmos de machine learning possibilita a identificação de padrões de comportamento que podem sinalizar fraudes fiscais. A análise de vastas quantidades de dados financeiros permite à AT focar auditorias e investigações nos contribuintes com maior risco, aumentando a eficácia das suas ações de fiscalização (AT, 2020).

A AT tem apostado em sistemas que automatizam a análise das declarações fiscais, tornando o processo de revisão mais ágil e permitindo uma resposta mais rápida aos contribuintes. Estas inovações não apenas aumentam a eficiência operacional, mas também criam um ambiente mais justo e equitativo na cobrança de impostos (AT, 2021).

- **Instituto de Gestão Financeira da Segurança Social (IGFSS)**

O IGFSS tem adotado soluções de IA para aprimorar a gestão e a análise de dados ligados à segurança social. Um exemplo é a utilização de algoritmos para antecipar a necessidade de apoio social em áreas específicas, o que possibilita uma alocação mais eficiente dos recursos disponíveis. Ao analisar dados demográficos e socioeconómicos, o IGFSS consegue identificar comunidades que precisam de mais intervenção, aumentando a eficácia dos programas sociais (IGFSS, 2020).

O instituto tem recorrido à IA para melhorar o atendimento ao público, implementando assistentes virtuais que auxiliam os cidadãos a obter informações sobre benefícios e serviços disponíveis, facilitando a navegação por processos administrativos complexos (IGFSS, 2021).

Desde a saúde até à administração pública, as iniciativas em curso demonstram um compromisso com a inovação e a melhoria contínua, alinhadas com os princípios éticos delineados no "Guia para a Inteligência Artificial, Ética, Transparente e Responsável na Administração Pública".

3.3 Quadro normativo e legal em Portugal

O panorama normativo e legal em Portugal sobre a IA está em constante desenvolvimento, refletindo a crescente incorporação desta tecnologia em várias áreas da sociedade. Este contexto é moldado não apenas pela legislação nacional, mas também por regulamentos europeus e diretrizes internacionais que procuram garantir uma utilização ética e responsável da IA. Neste segmento, analisaremos as principais leis e regulamentos que definem o ambiente jurídico para a IA em Portugal.

- **Regulação da Proteção de Dados**

Uma das legislações mais relevantes que impactam a utilização da IA em Portugal é o RGPD, que está em vigor desde maio de 2018. Este regulamento da União Europeia define normas estritas para o tratamento de dados pessoais, exigindo que as organizações que empregam a IA assegurem a privacidade e a proteção das informações dos cidadãos. O RGPD estabelece responsabilidades, como a necessidade de obter consentimento explícito para o processamento de dados e a implementação de medidas de segurança apropriadas.

O RGPD, em particular, impõe que as organizações sejam claras sobre a utilização de algoritmos nos processos de tomada de decisão, especialmente quando esses processos

possam impactar negativamente os direitos e liberdades dos indivíduos (Comissão Europeia, 2020).

- **Lei da Proteção de Dados Pessoais**

Em Portugal, a Lei n.º 58/2019, de 8 de agosto, que garante a implementação do RGPD, regula a proteção de dados pessoais e define normas específicas para o tratamento de dados a nível nacional. Esta legislação oferece uma estrutura adicional que complementa o RGPD, abordando questões como os direitos dos titulares de dados e as obrigações das entidades que gerem informações pessoais.

A legislação portuguesa realça a relevância da proteção de dados desde a fase de conceção dos sistemas (*privacy by design*) e a necessidade de realizar avaliações de impacto sobre a proteção de dados, especialmente em contextos que empregam tecnologias de IA (CNPD, 2020).

- **Código Penal e Responsabilidade Penal**

A questão da responsabilidade penal associada a decisões automatizadas é outra área que requer atenção no âmbito da IA. O Código Penal português não aborda de forma direta as consequências das decisões algorítmicas, mas é crucial que as entidades que utilizam IA considerem as repercussões legais das suas ações. Em casos onde um algoritmo possa causar danos a indivíduos, a responsabilidade legal pode ser atribuída à organização que desenvolveu ou implementou a tecnologia, especialmente se houver negligência em assegurar que a IA funcione de maneira ética e segura (Martins, 2021).

- **Regulação da Inteligência Artificial na União Europeia**

Portugal também está a alinhar-se com as iniciativas da União Europeia para regular a IA. Em abril de 2021, a Comissão Europeia apresentou uma proposta de Regulamento sobre IA, com o objetivo de estabelecer um quadro jurídico abrangente para a utilização da IA na Europa. Esta proposta classifica as aplicações de IA de acordo com o seu nível

de risco, definindo requisitos específicos para sistemas considerados de alto risco, que abrangem áreas como a saúde, os transportes e a segurança pública (Comissão Europeia, 2021).

Portugal, na sua qualidade de membro da União Europeia, terá de integrar esta regulamentação no seu sistema jurídico, o que implicará uma revisão das normas atuais e a criação de novas políticas para assegurar a conformidade com os requisitos estabelecidos.

- **Diretrizes Éticas e Normas Técnicas**

Para além da legislação, as instituições públicas em Portugal estão a adotar diretrizes éticas que incentivam a utilização responsável da IA. O Guia elaborado pela AMA oferece um conjunto de normas e princípios que as entidades devem cumprir para assegurar que a implementação das tecnologias de IA respeite os direitos dos cidadãos e fortaleça a confiança do público (AMA, 2020).

À medida que a IA continua a evoluir, será essencial que o quadro legal se adapte para abordar novos desafios e garantir que as tecnologias sejam utilizadas de forma ética e responsável.

3.3.1 Comparação com outros modelos europeus

A regulação da IA na Europa está a avançar de forma acelerada, com diversos países a criarem as suas próprias estratégias e modelos para enfrentar os desafios e oportunidades que esta tecnologia traz. A seguir, será realizada uma comparação entre o modelo português e outros modelos de referência na Europa, sublinhando as semelhanças e diferenças nas abordagens adotadas.

- **Modelo da União Europeia**

Um dos esforços mais significativos para regular a IA a nível europeu é a proposta de Regulamento sobre IA, apresentada pela Comissão Europeia em abril de 2021. Este regulamento pretende estabelecer um quadro harmonizado que define regras claras para o desenvolvimento e a utilização de sistemas de IA em todos os Estados-Membros. A proposta classifica as aplicações de IA de acordo com o seu nível de risco, estabelecendo requisitos específicos para sistemas considerados de alto risco, como os utilizados na saúde, na segurança pública e na gestão de infraestruturas críticas (Comissão Europeia, 2021).

A abordagem da União Europeia enfatiza a necessidade de garantir a segurança e a proteção dos direitos fundamentais dos cidadãos, promovendo a transparência, a responsabilidade e a ética no uso da IA. Este regulamento está alinhado com a legislação existente, como o RGPD, e complementa as iniciativas dos Estados-Membros para regulamentar a IA de forma responsável.

- **Modelo francês**

A França tem adotado uma postura proativa em relação à regulação da IA, culminando na apresentação do "Plano Nacional para a IA" em 2018. Este plano não apenas incentiva o avanço das tecnologias de IA, mas também define diretrizes éticas que devem ser observadas durante a sua implementação. O país realça a importância de uma IA que respeite a dignidade humana, promova a diversidade e reduza preconceitos (Gouvernement Français, 2018).

O governo francês estabeleceu a "Comissão Nacional de Ética Digital", cuja função é aconselhar o governo em questões éticas ligadas à IA e à transformação digital. Esta comissão tem exercido um papel relevante na promoção do debate público sobre as implicações éticas da IA, uma prática que poderia ser considerada no contexto português, onde a participação da sociedade civil é igualmente crucial.

- **Modelo alemão**

A Alemanha adota uma abordagem focada na ética, e em 2020, o governo apresentou a "Estratégia Nacional de IA", que incorpora princípios éticos para o desenvolvimento e a utilização de sistemas de IA. Esta estratégia enfatiza a relevância da transparência, da responsabilidade e da proteção de dados, além de defender uma "IA voltada para o bem comum" (Bundesministerium für Wirtschaft und Energie, 2020).

Um aspeto característico do modelo alemão é a sua intensa valorização da consulta pública e do envolvimento das partes interessadas na formulação de políticas. A Alemanha organizou uma série de workshops e consultas com representantes da indústria, da academia e da sociedade civil para assegurar que as opiniões de diferentes grupos sejam consideradas no processo de regulação.

- **Modelo dinamarquês**

A Dinamarca também se tem destacado na regulação da IA, adotando uma abordagem abrangente que tem em conta não só os aspetos técnicos, mas também os sociais e éticos. O "Plano de Ação para a IA" do país, apresentado em 2019, sublinha a importância de uma IA que seja "fiável, transparente e responsável", colocando o ser humano no centro das decisões algorítmicas.

A Dinamarca fomenta uma colaboração próxima entre o governo, as empresas e as instituições de investigação para estimular a inovação em IA, enquanto garante a proteção dos direitos dos cidadãos. Esta abordagem colaborativa pode servir de modelo para Portugal, especialmente no que se refere à criação de parcerias entre os setores público e privado.

- **Comparação com o Modelo Português**

O modelo português destaca-se pelo seu foco na ética, na transparência e na inclusão. No entanto, ao comparar com os modelos franceses, alemães e dinamarqueses, observa-se o seguinte:

Envolvimento da Sociedade Civil: Embora Portugal tenha definido diretrizes éticas, poderia tirar proveito de uma maior participação da sociedade civil na elaboração de políticas relacionadas com a IA, assim como acontece na França e na Alemanha.

Estratégia Nacional de IA: Enquanto na França e na Alemanha existem estratégias nacionais bem definidas, Portugal ainda está a trabalhar na criação de um quadro mais estruturado para a implementação da IA, o que poderia reforçar a sua posição no contexto europeu.

Consulta Pública: A Dinamarca sobressai pela sua abordagem colaborativa, envolvendo diferentes partes interessadas nas discussões sobre IA. Uma prática semelhante poderia ser mais aprofundada em Portugal para enriquecer o debate público e a aceitação das tecnologias de IA.

A análise do modelo português em relação a outros modelos europeus revela tanto semelhanças como oportunidades de aprimoramento. Ao tirar lições das experiências de outros países e ao promover a participação da sociedade civil, Portugal pode progredir na criação de um quadro normativo sólido.

3.3.2 Desafios reguladores e lacunas existentes

Um dos maiores desafios regulatórios é a rápida evolução das tecnologias de IA. As soluções que utilizam IA estão a ser criadas e implementadas a um ritmo que excede a capacidade das legislações atuais de se adaptarem. Este descompasso pode levar a situações em que a legislação não consegue cobrir todas as particularidades da nova tecnologia, resultando em incertezas jurídicas e em dificuldades na aplicação de normas adequadas (Binns, 2018).

A intrincada natureza técnica dos algoritmos de IA dificulta uma regulação eficaz. A escassez de conhecimento técnico entre legisladores e reguladores pode prejudicar a eficácia das normas estabelecidas, tornando mais complicado o processo de supervisão e a aplicação de sanções em casos de violação das diretrizes éticas e legais (Zuboff, 2019).

Embora o RGPD defina diretrizes claras para a proteção de dados pessoais, existem falhas na aplicação prática dessas normas em contextos específicos de IA. Por exemplo, os algoritmos que tomam decisões automatizadas podem não ser totalmente auditáveis, o que torna difícil a detecção de preconceitos ou erros (Comissão Europeia, 2021).

A dificuldade em assegurar que os indivíduos tenham acesso a informações claras sobre a utilização dos seus dados por sistemas de IA pode levar a uma falta de transparência. Essa falta de clareza é prejudicial, pois os cidadãos precisam compreender os processos que influenciam as suas vidas (Voigt & Von dem Bussche, 2017).

A participação da sociedade civil na criação de políticas e regulamentos é essencial para garantir que as preocupações e necessidades dos cidadãos sejam levadas em conta. Contudo, em muitos casos, essa participação tem sido restrita. A ausência de consultas públicas eficazes e de mecanismos que possibilitem a participação ativa da comunidade nas discussões sobre IA pode levar a uma desconexão entre os reguladores e a população (Meyer et al., 2021).

A falta de uma representação adequada na elaboração de políticas pode resultar em soluções que não capturam a diversidade de perspectivas e necessidades da sociedade. Isso é especialmente importante em matérias de justiça e não discriminação, onde a inclusão de diversos grupos sociais pode ajudar a reduzir preconceitos algorítmicos.

Apesar de o Guia da AMA oferecer orientações relevantes, é ainda necessária uma estrutura normativa mais sólida que trate especificamente das práticas éticas da IA. A falta de normas éticas obrigatórias pode levar a práticas inconsistentes entre diversas instituições e setores, comprometendo a integridade dos serviços oferecidos (AMA, 2020).

A elaboração de códigos de ética dedicados ao uso da IA poderia contribuir para uniformizar as práticas e assegurar que todos os operadores adiram aos mesmos princípios, fomentando a confiança e a aceitação das tecnologias por parte dos cidadãos.

A execução de auditorias em sistemas de IA para analisar o seu desempenho e impacto social é uma prática que ainda enfrenta desafios consideráveis. A escassez de recursos, tanto a nível humano como técnico, pode restringir a capacidade das instituições de realizar auditorias eficazes. Além disso, a complexidade dos algoritmos dificulta a condução de avaliações que sejam claras e acessíveis (Binns, 2018).

4.1 Tendências emergentes em IA e ética

Conforme as tecnologias se tornam mais avançadas e se integram no dia a dia, várias tendências emergentes surgem, trazendo consigo implicações importantes para a ética na IA. Esta secção analisa algumas dessas tendências, enfatizando a forma como elas influenciam o futuro da IA e a necessidade de um enquadramento ético sólido.

- **Aumento da Automação e Impacto no Trabalho**

Uma das tendências mais notáveis na IA é o aumento da automação de tarefas anteriormente executadas por seres humanos. Embora essa automação possa melhorar a eficiência e diminuir os custos operacionais, também levanta questões éticas sobre o futuro do trabalho e o impacto nas profissões. A substituição de postos de trabalho por sistemas automatizados pode resultar em desigualdades sociais, afetando particularmente os trabalhadores de setores mais vulneráveis (Brynjolfsson & McAfee, 2014).

Implicações Éticas: É fundamental que as políticas públicas considerem a transição para um mercado de trabalho mais automatizado. Isso envolve a necessidade de formação e requalificação dos trabalhadores, garantindo que as oportunidades de emprego sejam justas e acessíveis a todos. A ética deve orientar a forma como empresas e governos enfrentam essas transformações, priorizando o bem-estar dos indivíduos afetados.

- **IA e Tomada de Decisão**

Sistemas baseados em IA estão a ser utilizados para antecipar diagnósticos médicos, avaliar riscos em processos judiciais e melhorar operações de segurança. Contudo, a crescente dependência de algoritmos para decisões que impactam a vida das pessoas levanta preocupações relacionadas com a responsabilidade e a transparência.

Implicações Éticas: A ausência de transparência nos algoritmos empregados pode levar a decisões que os cidadãos não compreendem ou não conseguem contestar. A implementação de sistemas que assegurem a explicabilidade das decisões algorítmicas é

crucial para garantir que os indivíduos possam entender e, se necessário, questionar os resultados que os afetam diretamente.

- **Vigilância e Privacidade**

A crescente adoção da IA em sistemas de vigilância, como câmaras de reconhecimento facial e monitorização digital, levanta importantes questões éticas relacionadas com a privacidade e as liberdades individuais. Embora estas tecnologias possam contribuir para a segurança pública, o seu uso indiscriminado pode resultar em abusos de poder e numa situação de vigilância que ameaça a liberdade dos cidadãos (Zuboff, 2019).

Implicações Éticas: É crucial que exista uma regulamentação clara que defina os limites da vigilância. As instituições devem assegurar que a utilização de tecnologias de vigilância respeite os direitos à privacidade, promovendo um equilíbrio entre segurança e liberdade. A ética deve orientar as decisões sobre a implementação e uso desses sistemas, garantindo que sejam aplicados apenas quando necessário e de forma proporcional.

- **Ética Algorítmica e Viés**

O viés algorítmico permanece como um dos principais desafios éticos na IA. A crescente consciência sobre esta questão tem levado muitas organizações a implementar práticas de ética algorítmica, que incluem a análise contínua dos dados utilizados e a validação dos resultados obtidos.

Implicações Éticas: É fundamental desenvolver diretrizes e estruturas éticas para a criação de IA, a fim de minimizar o viés. As organizações devem adotar uma abordagem proativa para identificar e corrigir preconceitos, promovendo a equidade e a justiça nas decisões algorítmicas (Barocas et al., 2019).

- **Regulação e Governança**

Com o aumento das preocupações éticas, há uma pressão crescente para que governos e organismos internacionais definam normas que regulamentem o desenvolvimento e a aplicação de tecnologias de IA. A proposta de Regulamento da IA pela Comissão Europeia exemplifica como a governança da IA pode ser organizada para assegurar a proteção dos direitos dos cidadãos, ao mesmo tempo que se incentiva a inovação (Comissão Europeia, 2021).

Implicações Éticas: A regulação deve ser guiada por princípios éticos que reflitam as preocupações da sociedade. É fundamental que as regulamentações sejam flexíveis e capazes de responder rapidamente às inovações tecnológicas, garantindo que os direitos humanos sejam sempre uma prioridade na utilização da IA.

4.2 O papel das tecnologias emergentes na governança

As tecnologias emergentes, como a IA, *blockchain*, Internet das Coisas (IoT) e *big data*, estão a alterar radicalmente a forma como os governos funcionam e interagem com os cidadãos. Estas inovações oferecem novas oportunidades para melhorar a eficiência, aumentar a transparência e reforçar a responsabilidade na governança. Esta secção explora o papel dessas tecnologias emergentes na governança, destacando suas aplicações e implicações.

- **Inteligência Artificial na Tomada de Decisão**

A IA tem o potencial de transformar a tomada de decisão nas instituições públicas. Sistemas de IA podem analisar grandes conjuntos de dados para fornecer informações que orientem políticas e práticas. Por exemplo, algoritmos de machine learning podem ser utilizados para antecipar padrões sociais e económicos, permitindo que os governos planeiem intervenções mais eficazes em áreas como a saúde, a educação e a segurança (DGS, 2021).

Exemplo de Aplicação: A utilização de IA para a análise de dados em tempo real permite que os governos ajustem políticas em resposta a crises, como pandemias ou desastres

naturais. Isso não só aumenta a eficácia das respostas, como também possibilita uma melhor alocação de recursos.

- **Blockchain e Transparência**

A tecnologia *blockchain* é reconhecida pela sua capacidade de proporcionar um registo imutável e transparente de transações. Na governação, pode ser usada para aumentar a transparência em processos administrativos, como a gestão de contratos públicos e a distribuição de fundos. A implementação do *blockchain* pode ajudar a prevenir fraudes e a aumentar a confiança do público nas instituições governamentais (Catalini & Gans, 2016).

Exemplo de Aplicação: Em várias jurisdições, o uso de *blockchain* para rastrear a origem de bens e serviços, bem como para monitorizar a execução de contratos, está a ser explorado como uma forma de garantir que os recursos públicos sejam utilizados de maneira responsável.

- **IoT na Gestão Urbana**

A IoT refere-se à interligação de dispositivos através da Internet, permitindo a recolha e a troca de dados. No âmbito da gestão urbana, o conceito de cidades inteligentes está a emergir, utilizando a IoT para otimizar serviços públicos, como a gestão de tráfego, a monitorização da qualidade do ar e a eficiência energética (Komninos, 2013).

Exemplo de Aplicação: Cidades como Barcelona e Copenhaga implementaram sensores para monitorizar o tráfego em tempo real, ajustando semáforos e melhorando a fluidez do trânsito. Esta abordagem não só melhora a mobilidade urbana, como também contribui para a redução das emissões de carbono.

- **Big Data e Análise Preditiva**

A análise de *big data* permite que os governos processem grandes quantidades de informação de várias fontes para identificar tendências e prever comportamentos futuros. Esta capacidade pode ser utilizada para aumentar a eficácia das políticas públicas e dos serviços sociais, proporcionando uma compreensão mais profunda das necessidades da população (Meyer et al., 2021).

Exemplo de Aplicação: A análise de dados demográficos, de saúde e económicos pode auxiliar as instituições na identificação de áreas que necessitam de intervenção, permitindo que os recursos sejam direcionados para onde são mais necessários. Projetos de IA na saúde pública, como a previsão de surtos de doenças, exemplificam como o *big data* pode ser aplicado de forma eficaz.

- **Desafios Éticos e Considerações de Governança**

Embora as tecnologias emergentes apresentem oportunidades significativas, também acarretam desafios éticos que devem ser cuidadosamente ponderados. A implementação de IA e outras tecnologias pode levantar questões sobre privacidade, segurança e viés algorítmico, que precisam de ser geridos para proteger os direitos dos cidadãos (Binns, 2018).

Recomendações:

- **Desenvolvimento de Políticas Éticas:** Os governos devem estabelecer diretrizes que orientem a utilização ética das tecnologias emergentes, garantindo que a privacidade e os direitos humanos sejam respeitados.
- **Aliciamento da Sociedade Civil:** Incluir a sociedade civil nas discussões sobre a implementação de tecnologias emergentes pode alinhar as políticas públicas com as expectativas e necessidades dos cidadãos.

As tecnologias emergentes têm o potencial de transformar a governança, oferecendo novas ferramentas para melhorar a eficiência, a transparência e a responsabilidade das instituições públicas. No entanto, é essencial que a implementação dessas tecnologias seja acompanhada de um compromisso ético e de uma análise cuidadosa das implicações sociais. Ao adotar uma abordagem proativa e responsável, os governos podem maximizar os benefícios das tecnologias emergentes e construir uma administração pública mais eficaz e justa.

4.3 Formação e capacitação para a ética na IA

A formação em ética relacionada com a IA é crucial para assegurar que os profissionais que desenvolvem e implementam estas tecnologias o façam de forma responsável, respeitando os direitos dos cidadãos e promovendo práticas éticas. Com a crescente integração da IA na administração pública e na vida diária, a necessidade de uma base sólida em ética torna-se ainda mais urgente. Este segmento examina a relevância da formação em ética, as melhores práticas e as abordagens recomendadas.

4.3.1 Importância da formação em ética na IA

A utilização crescente da IA levanta questões complexas de natureza ética, incluindo viés algorítmico, privacidade, transparência e responsabilidade. Sem uma formação adequada, os profissionais podem não estar plenamente cientes das implicações éticas das suas decisões, resultando em práticas que perpetuam injustiças ou causam danos. Assim, a formação em ética é essencial para:

- **Promover a Consciencialização:** A formação ética ajuda a aprofundar a compreensão das consequências sociais e éticas da IA, permitindo que os profissionais identifiquem e abordem questões problemáticas desde as etapas iniciais do desenvolvimento de um projeto.
- **Desenvolver Competências Críticas:** Profissionais bem formados em ética são mais capazes de tomar decisões informadas que considerem não apenas os resultados técnicos, mas também o impacto social e ético das tecnologias implementadas (Floridi et al., 2018).

- **Fomentar uma Cultura Organizacional Ética:** A formação em ética contribui para a criação de uma cultura organizacional que valoriza a responsabilidade e a integridade, estabelecendo um ambiente de trabalho onde as preocupações éticas são discutidas abertamente.

4.3.2 Boas práticas para a formação em ética

A implementação de programas de formação em ética na IA deve seguir algumas melhores práticas:

- **Currículo Interdisciplinar:** A formação deve incorporar uma abordagem que una aspectos técnicos da IA com filosofia, direito e ciências sociais. Isso permitirá que os participantes compreendam a complexidade das questões éticas que surgem no desenvolvimento de sistemas de IA (Binns, 2018).
- **Formação Prática:** É importante que os cursos incluam estudos de caso e cenários reais que ilustrem os desafios éticos encontrados nas aplicações de IA. A formação prática ajuda os profissionais a aplicar conceitos éticos em situações concretas e a desenvolver soluções criativas para dilemas éticos.
- **Feedback e Discussão:** A realização de sessões de discussão e feedback após a formação pode enriquecer a experiência de aprendizagem. O diálogo aberto sobre questões éticas permite que os participantes compartilhem as suas experiências e aprendam uns com os outros, fortalecendo a compreensão coletiva das implicações éticas da IA.

4.3.3 Abordagens recomendadas

Para garantir uma formação eficaz em ética na IA, várias abordagens podem ser implementadas:

- **Programas de Formação Contínua:** A formação deve ser um processo contínuo e não um evento isolado. As organizações devem investir em programas que

mantenham os profissionais atualizados sobre as últimas tendências e desafios éticos relacionados com a IA.

- **Colaboração com Universidades e Centros de Pesquisa:** Parcerias com instituições académicas podem enriquecer os programas de formação. As universidades podem oferecer conhecimento especializado em ética, direito e tecnologia, desenvolvendo programas adaptados às necessidades do setor público e privado.
- **Inclusão de Diversas Perspetivas:** A formação deve incluir uma variedade de vozes e perspetivas, especialmente de grupos que podem ser afetados por decisões algorítmicas. Isso assegura que a formação em ética na IA seja inclusiva e represente as preocupações de diversas partes interessadas (Goggin, 2019).

4.3.4 Exemplos de iniciativas em Portugal

Em Portugal, várias iniciativas estão a ser desenvolvidas para promover a formação em ética na IA:

- **Cursos e *Workshops*:** Instituições públicas e ONGs têm promovido cursos e *workshops* que abordam a ética na IA. Estas iniciativas visam capacitar profissionais em áreas como a administração pública, saúde e educação, permitindo que lidem com a tecnologia de forma informada e responsável.
- **Iniciativas Académicas:** As universidades têm começado a integrar disciplinas sobre ética em IA nos seus currículos, preparando a nova geração de profissionais para enfrentar os desafios éticos associados à utilização da tecnologia.

A formação em ética relacionada com a IA é essencial para garantir que os profissionais atuem de maneira responsável e ética. Ao adotar melhores práticas e abordagens eficazes, as instituições podem criar um ambiente em que as preocupações

éticas sejam prioritárias, contribuindo para o desenvolvimento de tecnologias que respeitem os direitos dos cidadãos e promovam o bem-estar social. Com um compromisso contínuo com a formação ética, é possível moldar um futuro em que a IA beneficie a sociedade de forma justa e equitativa.

4.4 Propostas para um modelo regulatório inovador

À medida que a IA se torna mais presente na vida pública e privada, é essencial que os governos estabeleçam um modelo regulatório inovador que aborde os desafios éticos e sociais que surgem, ao mesmo tempo que promove a inovação e a confiança do público. A seguir, apresentam-se propostas para um modelo regulatório que seja eficaz, flexível e centrado no cidadão.

- **Abordagem Baseada em Risco**

Uma proposta chave para um modelo regulatório inovador é a adoção de uma abordagem que considere os riscos. Isso envolve classificar as aplicações de IA de acordo com o seu potencial de impacto negativo sobre os cidadãos e a sociedade. Sistemas de IA considerados de alto risco, como aqueles utilizados em decisões de saúde, justiça e segurança pública, devem estar sujeitos a exigências regulatórias mais rigorosas, incluindo avaliações de impacto ético e auditorias periódicas (European Commission, 2021).

Implicações: Esta abordagem permitirá uma utilização mais eficiente dos recursos regulatórios, assegurando uma supervisão mais atenta das tecnologias com maior potencial de causar danos.

- **Criação de um Quadro de Supervisão Adaptativo**

Dada a natureza em constante mudança da IA, é necessário um sistema de supervisão que possa evoluir rapidamente para acompanhar novas tecnologias e desafios. Isso pode ser alcançado através da criação de uma entidade reguladora dedicada, capaz

de atualizar diretrizes e regulamentos com base em consultas públicas e feedback contínuo das partes interessadas.

Proposta: A formação de um conselho consultivo que inclua especialistas em tecnologia, ética, direito e representantes da sociedade civil pode ajudar a orientar a supervisão regulatória e garantir que as políticas permaneçam relevantes e eficazes (Meyer et al., 2021).

- **Fomento à Transparência e à Responsabilidade**

Para cultivar a confiança nas tecnologias de IA, é crucial enfatizar a transparência e a responsabilidade. Desenvolvedores e utilizadores de sistemas de IA devem ser obrigados a divulgar informações sobre a utilização de dados, os algoritmos aplicados e os processos de tomada de decisão.

Proposta: Implementar requisitos de transparência que obriguem as organizações a publicar relatórios regulares sobre o desempenho dos seus sistemas de IA, incluindo qualquer viés identificado e as ações tomadas para o mitigar. Além disso, deve haver mecanismos claros de contestação e reparação para responsabilizar as entidades por decisões que resultem em danos aos cidadãos (Binns, 2018).

- **Incentivo à Inovação e Desenvolvimento Sustentável**

O modelo regulatório deve não só proteger os direitos dos cidadãos, mas também incentivar a inovação e o desenvolvimento sustentável. Isso pode ser alcançado através de incentivos para o desenvolvimento de tecnologias de IA que priorizem a ética e o bem-estar social.

Proposta: O governo poderia introduzir subsídios e incentivos fiscais para empresas que demonstrem compromisso com a IA responsável, estimulando o desenvolvimento de soluções que respeitem os direitos dos cidadãos e tenham um impacto positivo na sociedade (Floridi et al., 2018).

- **Formação e Capacitação Contínua**

Um modelo regulatório inovador deve incluir uma forte componente de formação, não apenas para os profissionais que lidam com IA, mas também para reguladores e legisladores. O objetivo é garantir que todos os envolvidos compreendam as complexidades das tecnologias emergentes e as suas implicações éticas.

Proposta: Estabelecer programas de formação contínua sobre IA, ética e regulamentação, acessíveis a todos os níveis de governo e à sociedade civil, assegurando que as políticas sejam fundamentadas numa sólida compreensão das tecnologias em questão (Goggin, 2019).

- **Envolvimento da Sociedade Civil**

O envolvimento ativo da sociedade civil no processo regulatório é essencial para assegurar que as preocupações dos cidadãos sejam ouvidas e levadas em consideração. A criação de mecanismos de consulta e participação pública pode ajudar a promover uma maior aceitação das políticas e aumentar a transparência.

Proposta: Fomentar fóruns de discussão e plataformas de feedback onde os cidadãos possam expressar as suas preocupações e sugestões em relação à implementação de tecnologias de IA. Isso não apenas fortalece a confiança pública, mas também garante que as políticas reflitam as necessidades e os valores da sociedade (Meyer et al., 2021).

Um modelo regulatório inovador para a IA deve ser flexível, baseado em risco e centrado na promoção da transparência e da responsabilidade. Ao adotar essas propostas, os governos podem assegurar que a IA seja utilizada de forma ética e que beneficie a sociedade como um todo. O compromisso com a formação contínua, o envolvimento da sociedade civil e a promoção da inovação são fundamentais para criar um ambiente onde a tecnologia possa prosperar, respeitando os direitos e a dignidade dos cidadãos.

CAPÍTULO V – ESTUDO DE CASO: ASSISTENTE VIRTUAL DA AT

Este capítulo irá explorar um estudo de caso concreto envolvendo a avaliação do Assistente Virtual (AV) da AT, desenvolvido pelo LabX. A análise baseia-se num modelo sistémico que abrange múltiplos domínios — tecnológico, social, legal e organizacional e enfatiza a centralidade das necessidades e expectativas dos cidadãos.

5.1 Contextualização do estudo de caso

O LabX, laboratório de experimentação do setor público, foi encarregue de avaliar o desempenho do AV da AT, que funcionava em fase piloto. O objetivo era desenvolver um conjunto de recomendações para otimizar o desempenho do *chatbot* e adaptar as suas funcionalidades às exigências éticas e regulatórias vigentes em Portugal. Para isso, o LabX colaborou com dois centros de investigação: o Digital Transformation CoLab (DTx), especializado em tecnologias emergentes, e a Behavioral Research Unit, da Universidade Católica, com foco na análise comportamental dos cidadãos.

O estudo visou não só melhorar o AV existente, mas também validar uma metodologia de avaliação que poderia ser escalada e aplicada a outras soluções de IA no setor público. A abordagem desenvolvida destaca-se por capturar dimensões críticas, como a prontidão tecnológica, a interação humano-máquina e as perceções dos utilizadores, incluindo aspetos éticos e de privacidade.

5.2 Metodologia

A metodologia aplicada foi estruturada em várias vias de exploração que permitiram uma avaliação integral do AV. As principais foram:

- **Prontidão Multidimensional:** Uma ferramenta que avalia o grau de maturidade da solução tecnológica, considerando aspetos legais, organizacionais, sociais e tecnológicos.

- **Necessidades, Expectativas e "Dores" dos Utilizadores:** Uma componente fundamental que procura mapear as interações dos cidadãos com o AV, identificando pontos de fricção e áreas de melhoria;
- **Interação Humano-Máquina:** Uma análise detalhada da experiência dos cidadãos durante a interação com o AV, medindo fatores como confiança, empatia, e clareza das respostas;
- **Métricas do AV:** A plataforma do *chatbot* gera dados essenciais, como o número de encaminhamentos e a intensidade de utilização, que são fundamentais para avaliar e melhorar o desempenho AV.
- **Parcerias com o Sistema Científico:** A colaboração com entidades académicas garantiu a robustez da metodologia e o alinhamento com as melhores práticas internacionais.

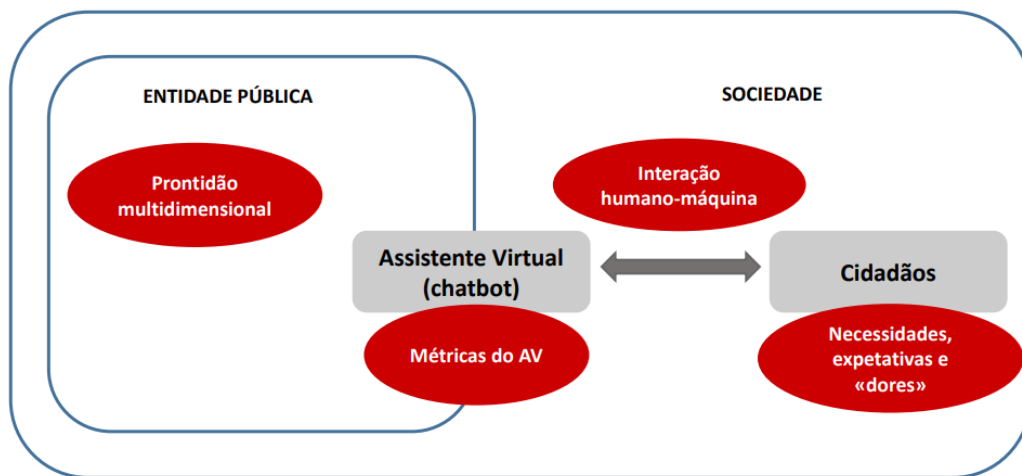


Figura 1 - Proposta metodológica para uma visão sistémica

Fonte: <https://labx.gov.pt/projetos-posts/inteligencia-artificial-centrada-nos-cidadaos/>
(30/10/2024)

Este diagrama descreve a interação entre uma entidade pública, um assistente virtual (*chatbot*) e a sociedade. O diagrama é composto por:

- **Entidade Pública:** Inclui elementos como "Prontidão multidimensional" e "Métricas do AV", que são fatores internos usados para otimizar o funcionamento e a avaliação do assistente virtual;
- **Assistente Virtual (*chatbot*):** Serve de intermediário entre a entidade pública e a sociedade;
- **Sociedade:** Foca-se na "Interação humano-máquina" e nas "Necessidades, expectativas e 'dores'" dos cidadãos, refletindo a importância da experiência e das percepções dos utilizadores na avaliação da eficácia do *chatbot*.

O diagrama destaca a interdependência entre a gestão do *chatbot* pela entidade pública e a experiência dos cidadãos na sociedade.

5.3 Prontidão multidimensional

Foi desenvolvida uma versão adaptada ao contexto português de um modelo de avaliação da Comissão Europeia para medir o nível de prontidão de soluções de IA nos serviços públicos. Em colaboração com o parceiro DTx, o modelo considera quatro domínios: tecnológico, social, legal e organizacional, oferecendo uma visão multidimensional da tecnologia e permitindo a definição de estratégias que aproveitem oportunidades, enquanto se reconhecem e gerem potenciais obstáculos.



Figura 2 - Modelo da ferramenta de prontidão multidimensional

Fonte: <https://labx.gov.pt/projetos-posts/inteligencia-artificial-centrada-nos-cidadaos/> (30/10/2024)

5.3.1 Avaliação da prontidão tecnológica

A prontidão tecnológica é um aspeto crucial na regulação ética da IA. O AV foi avaliado com base em três indicadores principais:

- **Ferramentas de Conversação:** As tecnologias utilizadas devem garantir uma interação fluida e adaptável às necessidades dos utilizadores;
- **Processamento de Linguagem:** O AV precisa de compreender as intenções dos cidadãos com precisão, o que exige métodos avançados de processamento de linguagem natural;
- **Metodologia de Conversação:** A eficácia na deteção de intenções e na geração de respostas é crítica para assegurar um diálogo coerente.

5.3.2 Avaliação social

O AV deve atender às necessidades do público-alvo, sendo necessário um autodiagnóstico que avalie o seu impacto, utilidade, e as barreiras que pode impor. Foram utilizados três indicadores:

- **Requisitos básicos**, que medem a facilidade de uso e acessibilidade;
- **Desempenho social**, que avalia o esforço cognitivo e a usabilidade;
- **Barreiras sociais**, que consideram equidade, transparência e confiança.

5.3.3 Considerações legais

A conformidade legal é fundamental no contexto das instituições públicas. Foram usados três indicadores principais:

- **Privacidade dos Dados:** A proteção dos dados dos cidadãos é uma preocupação central. O AV deve garantir que as informações sejam tratadas com o máximo rigor e respeito pelas normas de proteção de dados;
- **Ética e Responsabilidade:** O desenvolvimento do AV inclui assegurar que os algoritmos não introduzam vieses ou prejudiquem grupos vulneráveis.

- **Conformidade Regulamentar:** A solução deve estar alinhada com a legislação portuguesa e europeia, facilitando auditorias e inspeções externas.

5.4 Prontidão organizacional

Este domínio avalia o impacto do AV na organização, considerando capacidades internas e recursos tecnológicos. Os indicadores incluem:

- **Infraestrutura**, que mede a segurança e disponibilidade de dados;
- **Capacidades**, que avaliam a maturidade da organização para otimizar processos;
- **Habilidades**, que se referem às competências técnicas e digitais dos colaboradores necessárias para implementar e gerir o AV.

5.5 Necessidades, expectativas e “dores” dos cidadãos

Para que as tecnologias de IA sirvam de forma eficaz o público, é essencial que atendam às necessidades e expectativas dos cidadãos e que se compreendam as dificuldades enfrentadas durante a sua utilização. O objetivo desta abordagem foi analisar os diferentes perfis de cidadãos e as suas interações com o AV. Para isso, realizaram-se grupos focais que mapearam as jornadas de utilização, identificando pontos problemáticos e as rotas percorridas pelos utilizadores entre os vários canais de atendimento.

A identificação dos perfis dos utilizadores deve considerar fatores como padrões de uso, dados sociodemográficos e a condição de utilização (particular ou profissional). Os grupos focais ajudam a revelar as necessidades, expectativas e dificuldades de cada perfil, permitindo que os participantes interajam com o AV em tarefas simuladas do seu dia-a-dia.

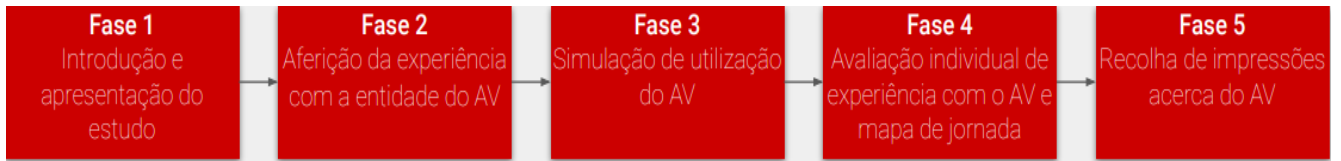


Figura 3 - Metodologia utilizada nas sessões de grupos focais

Fonte: <https://labx.gov.pt/projetos-posts/inteligencia-artificial-centrada-nos-cidadaos/> (30/10/2024)

Os mapas de jornada dos grupos focais visualizam a experiência dos cidadãos ao utilizar o AV, destacando etapas antes, durante e após o serviço. Estes mapas identificam pontos de “dor” e áreas que precisam de melhorias para otimizar a experiência do utilizador.

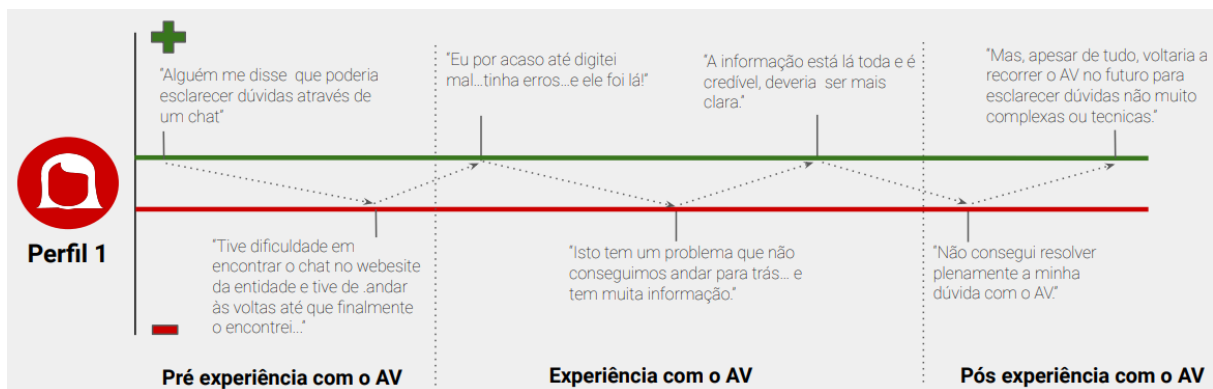


Figura 4 - Mapa de jornada tipo para um dado perfil de utilizador

Fonte: <https://labx.gov.pt/projetos-posts/inteligencia-artificial-centrada-nos-cidadaos/> (30/10/2024)

O mapa de jornada ilustra a experiência de um perfil de utilizador, destacando pontos de “dor” (em vermelho) e comentários positivos (em verde) durante a utilização do AV. Esses pontos críticos devem ser abordados pela entidade responsável para melhorar a interação e as respostas do AV.

Além de mapear a experiência com o AV, é crucial entender o seu papel e ligações com outros serviços oferecidos pela entidade. A metodologia utilizada analisa como os cidadãos interagem com esses serviços e o fluxo até chegarem ao AV, ajudando a

identificar se o AV é a primeira opção de contato e o percurso realizado pelos utilizadores. Isso esclarece o enquadramento do AV nos diferentes canais disponíveis.

5.6 Interação entre humanos e máquinas

De forma a avaliar a experiência dos cidadãos com o AV, foi fornecido um questionário digital que recolhe o feedback sobre três momentos principais: o contacto com o website, a interação com o AV (confiança e utilidade), e a interação com o assistente humano quando há encaminhamento. Este questionário permite comparar o desempenho do AV com o suporte humano e obter opiniões imediatas dos utilizadores.

O questionário de satisfação foi desenvolvido com base num modelo que capta quatro momentos essenciais da interação com o AV: receção, reação, resolução do problema e retorno. Este modelo assegura uma avaliação completa da experiência do utilizador.

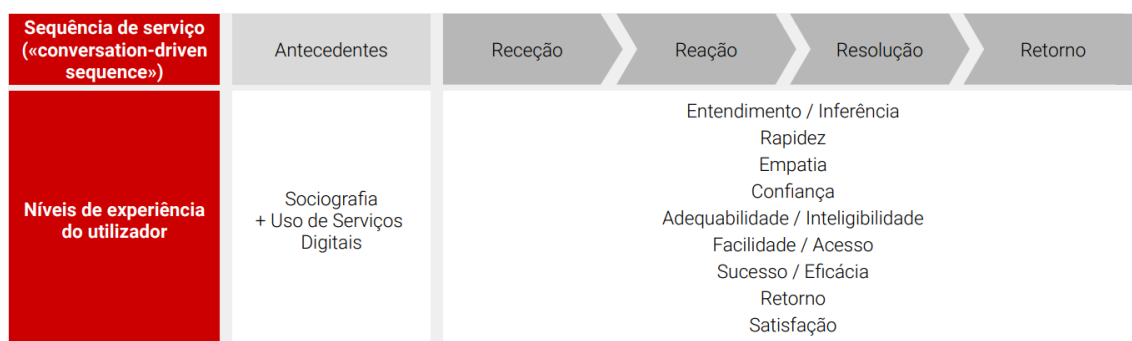


Figura 5 - Modelo de análise sequencial de conversação

Fonte: <https://labx.gov.pt/projetos-posts/inteligencia-artificial-centrada-nos-cidadaos/> (30/10/2024)

5.7 Métricas do assistente virtual

A plataforma do AV gera métricas essenciais para monitorizar continuamente o comportamento dos cidadãos e o desempenho do AV, abrangendo aspetos como eficiência (encaminhamentos e utilizadores ativos), temas mais procurados e a intensidade de utilização.

No quadro seguinte são dados alguns exemplos de métricas fundamentais que a plataforma do AV deve disponibilizar:

Métrica	Objetivo
N.º de utilizadores (interações) por hora e dia (total users)	Permite conhecer a distribuição da procura e os seus identificar e outros comportamentos da procura.
N.º de utilizadores ativos que recorrem ao AV	Possibilita conhecer se os utilizadores são recorrentes na utilização do AV (se responde ao que pretendem os utilizadores) e o grau de adoção do AV (novos utilizadores).
N.º de novos utilizadores	
Duração média das interações;	Estas métricas permitem identificar o nível de envolvimento (engagement) entre os cidadãos e o AV.
N.º total de mensagens;	
N.º de mensagens médio trocados com os utilizadores;	
N.º de mensagens de fraco entendimento,	Permite identificar a quantidade de interações que são encaminhadas pelo AV para o assistente humano.
Taxa de sucesso da interação (goal completion rate)	Captura a taxa de sucesso do AV nas respostas aos seus utilizadores (interações que não foram abandonadas pelos utilizadores nem encaminhadas para o assistente humano).
Temas/assuntos mais procurados	Captura quais os temas de interesse dos utilizadores do AV.

Figura 6 - Tabela de métricas do AV

Fonte: <https://labx.gov.pt/projetos-posts/inteligencia-artificial-centrada-nos-cidadaos/> (30/10/2024)

5.8 Resultados e aprendizagens

As principais aprendizagens desta abordagem sistémica para avaliar o desempenho de soluções de IA e a sua interação com os cidadãos são:

Os AVs em fase inicial funcionam melhor quando focados em áreas delimitadas, permitindo respostas mais estáveis e eficazes. No início, é preferível evitar temas complexos ou que exijam respostas garantidas, dado que os cidadãos não esperam essa sofisticação numa fase inicial.

O AV deve funcionar como um ponto inicial de atendimento, complementando outros canais e ajudando a reduzir a pressão sobre serviços presenciais e telefónicos. É crucial continuar o desenvolvimento do AV com ciclos experimentais para melhorar a compreensão das solicitações e a qualidade das respostas.

O design do AV como um "parceiro de conversa" para esclarecer dúvidas está alinhado com os progressos registados nas aplicações desta tecnologia a nível internacional. Identificar os níveis de maturidade do AV permite situar o seu

desenvolvimento e criar um plano de melhorias contínuas com base nos resultados das avaliações.

A abordagem promove a partilha e a articulação entre iniciativas dentro da Administração Pública, assegurando a integração de sistemas e o desenvolvimento de uma infraestrutura digital comum. É essencial manter o AV focado nas necessidades dos cidadãos, realizando testes, monitorizações e avaliações constantes, com investimento contínuo para promover a evolução incremental da solução.

A metodologia experimental mostrou-se eficaz e promissora para o uso na Administração Pública, comprovando, através de um caso real, que pode ser ampliada e utilizada por outras entidades para avaliar e otimizar os seus AVs.

A aplicação no AV do portal das Finanças permitiu realizar um diagnóstico atual e propor melhorias para melhor atender às necessidades dos cidadãos. Os utilizadores aceitam estas soluções, mesmo sem interação humanizada, reconhecendo o seu potencial de melhoria. No entanto, enfatizam a importância da transparência e credibilidade. Para além disso, os AVs precisam de supervisão humana contínua, pois não funcionam de forma totalmente autónoma, exigindo treino constante para aperfeiçoar as respostas.

O sucesso do projeto deve-se em parte às parcerias estratégicas com instituições científicas. Estas colaborações permitiram a adaptação do modelo de avaliação da IA ao contexto português e garantiram uma análise comportamental rigorosa dos cidadãos. A cooperação entre o LabX e os centros de investigação assegurou que a metodologia fosse robusta e que as soluções propostas fossem viáveis e sustentáveis a longo prazo.

Este estudo de caso ilustra a importância de uma abordagem ética e regulatória abrangente na implementação de IA nas instituições públicas em Portugal. A metodologia desenvolvida pelo LabX provou ser eficaz não só na avaliação da prontidão tecnológica, mas também na identificação das expectativas e necessidades dos cidadãos. As recomendações resultantes incluem a melhoria contínua da usabilidade, a transparência no tratamento de dados, e a implementação de um modelo de governação que assegure o cumprimento das normas éticas e legais.

O estudo sobre a ética e a regulação da IA no âmbito das instituições públicas em Portugal expõe uma realidade complexa e multifacetada, caracterizada por avanços importantes, mas também por desafios éticos e práticos que necessitam de ser abordados. Ao longo desta pesquisa, foram destacados vários temas relevantes, como a necessidade de transparência, a responsabilidade no uso da IA, e o papel crucial da sociedade civil na supervisão e no fornecimento de feedback sobre as políticas implementadas.

A incorporação da IA nas instituições públicas tem o potencial de aumentar a eficiência e a qualidade dos serviços, mas exige uma abordagem ética rigorosa. Exemplos bem-sucedidos, como a utilização de *chatbots* e sistemas de análise preditiva na área da saúde, mostraram como a tecnologia pode revolucionar a administração pública. No entanto, também surgiram casos de insucesso, como os desafios encontrados na automação de processos de recrutamento, evidenciando a necessidade de uma supervisão crítica e de formação adequada em ética.

A sugestão de um modelo regulatório inovador, fundamentado em risco, transparência e envolvimento da sociedade civil, é vital para assegurar que a implementação da IA ocorra de maneira responsável e justa.

As consequências práticas deste estudo são relevantes para os responsáveis políticos, gestores e profissionais da administração pública. É aconselhável que as instituições públicas implementem uma abordagem organizada e sistemática para a formação em ética na IA, centrando-se na sensibilização para temas como viés algorítmico, privacidade e transparência. Deve-se promover a formação contínua, capacitando os profissionais para enfrentar as complexidades relacionadas com a adoção de tecnologias emergentes.

A criação de mecanismos para consulta e participação pública deve ser uma prioridade, permitindo que os cidadãos partilhem as suas preocupações e colaborem na elaboração de políticas que impactam as suas vidas. A transparência nas decisões e nas operações relacionadas com a IA deve ser assegurada, promovendo a confiança e a aceitação das tecnologias.

Resumindo, a utilização da IA nas instituições públicas em Portugal oferece um caminho promissor para a modernização da administração pública, mas deve ser acompanhado por uma atenção cuidadosa às questões éticas. Este estudo destaca a necessidade de um compromisso coletivo com a ética, a transparência e a inclusão.

Esta dissertação enfrenta várias limitações que podem impactar a generalização dos resultados e a profundidade da análise. Primeiramente, a escolha de uma metodologia qualitativa, embora adequada para explorar as complexidades da ética e regulação da IA nas instituições públicas, pode introduzir subjetividades. Além disso, a seleção de casos práticos pode não representar a totalidade das experiências de implementação da IA em diferentes contextos ou instituições, limitando a aplicabilidade das conclusões.

O acesso a informações sensíveis sobre a implementação da IA pode ter sido restrito, limitando a profundidade da análise. A evolução rápida da tecnologia da IA também representa um desafio, pois as informações recolhidas podem rapidamente tornar-se obsoletas, o que pode impactar a relevância dos dados e análises apresentados.

Por fim, existem limitações teóricas que restringem a abrangência do estudo. A dissertação concentra-se no contexto português, o que pode reduzir a generalização das conclusões a outros países com diferentes realidades sociais, políticas e culturais. Embora se tenha procurado fazer uma análise profunda das questões éticas e regulatórias, a complexidade das problemáticas abordadas implica que muitas outras dimensões da ética em IA poderiam ser exploradas mais a fundo. Portanto, a investigação sobre a ética e regulação da IA é um campo em constante evolução, e futuras pesquisas podem ampliar a compreensão das implicações sociais e éticas associadas à tecnologia.

Para investigações futuras, sugere-se a realização de estudos empíricos que analisem a eficácia das políticas e práticas atualmente em vigor, além de investigar o impacto da IA nas comunidades vulneráveis. A ética na IA é um domínio em constante desenvolvimento, e a reflexão contínua sobre as suas implicações é fundamental para assegurar que a tecnologia beneficie o bem comum e respeite a dignidade humana.

REFERÊNCIAS BIBLIOGRÁFICAS

- AMA (Agência para a Modernização Administrativa). (2020). Guia para a Inteligência Artificial, Ética, Transparente e Responsável na Administração Pública.
- Angwin, J., Larson, J., Mattu, S., Kirchner, L. (2016). Machine Bias. ProPublica.
- AT (Autoridade Tributária e Aduaneira). (2020). Relatório Anual de Atividades.
- AT (Autoridade Tributária e Aduaneira). (2021). Iniciativas de Combate à Evasão Fiscal.
- Autoridade da Concorrência. (2021). Relatório de Atividades.
- Barocas, S., Hardt, M., Narayanan, A. (2019). Fairness and Machine Learning.
- Binns, R. (2018). Fairness in Machine Learning: Lessons from Political Philosophy.
- Brynjolfsson, E., McAfee, A. (2014). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies.
- Bundesministerium für Wirtschaft und Energie. (2020). AI Strategy of the Federal Government.
- Buolamwini, J., Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency.
- Catalini, C., Gans, J. S. (2016). Some Simple Economics of Blockchain.
- CNPD (Comissão Nacional de Proteção de Dados). (2020). Relatório Anual de Atividades.
- Danish Government. (2019). Denmark's Strategy for Artificial Intelligence.
- Davenport, T. H., & Ronanki, R. (2018). How Intelligent Automation is Transforming the Workplace.
- DGAEP (Direção-Geral da Administração e do Emprego Público). (2020). Relatório de Atividades.
- DGAEP (Direção-Geral da Administração e do Emprego Público). (2021). Inovações na Gestão de Recursos Humanos.

DGS (Direção-Geral da Saúde). (2021). Relatório sobre o Uso de Dados em Saúde Pública.

Doshi-Velez, F., Kim, P. (2017). Towards a rigorous science of interpretable machine learning. Proceedings of the 34th International Conference on Machine Learning.

European Commission. (2019). Ethics Guidelines for Trustworthy AI.

European Commission. (2021). Proposal for a Regulation laying down harmonised

Esteva, A., Kuprel, B., et al. (2019). Dermatologist-level classification of skin cancer with deep neural networks.

Floridi, L., et al. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, and Responsibilities. The Italian Review of AI Ethics.

Goggin, G. (2019). Disability, Human Rights and Information Technology.

Gouvernement Français. (2018). Plan National pour l'Intelligence Artificielle.

IGFSS (Instituto de Gestão Financeira da Segurança Social). (2020). Relatório de Atividades.

IGFSS (Instituto de Gestão Financeira da Segurança Social). (2021). Soluções Inovadoras na Segurança Social.

Jobin, A., Ienca, M., Andorno, R. (2019). Artificial Intelligence: The Global Landscape of Ethics Guidelines. Nature Machine Intelligence.

Komninos, N. (2013). Intelligent Cities: Between Digital and Urban Social Networks.

LABX. Inteligência Artificial centrada nos cidadãos: Uma abordagem sistémica para avaliação de assistentes virtuais nos serviços públicos. Fevereiro 2021.

Lambrecht, A., Tucker, C. (2019). Algorithmic Bias? An Empirical Study of Apparent Bias in Amazon Alexa. Proceedings of the 2019 ACM Conference on Fairness, Accountability, and Transparency.

Marr, B. (2021). Forecasting, Predicting and Analyzing the Future of Business and Technology.

Martins, A. (2021). A Responsabilidade Penal em Decisões Automatizadas: Desafios e Perspectivas. Revista de Direito Penal e Criminologia.

Melo, R., et al. (2021). The role of AI in public health: Applications and challenges.

Meyer, M., et al. (2021). How algorithms are shaping the future of public services: Towards a framework for analysis. *Journal of Public Policy*.

OECD. (2019). *OECD Principles on Artificial Intelligence*.

OECD. (2020). *Artificial Intelligence in Society*.

O’Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*.

Partnership on AI. (2020). *A Guide to AI Ethics*.

Raji, I. D., Buolamwini, J. (2019). Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. *Proceedings of the 2019 ACM Conference on Fairness, Accountability, and Transparency*.

Russell, S., Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*. Pearson.

Searle, J. R. (1980). *Minds, brains, and programs*. *The Behavioral and Brain Sciences*,

Smith, A., et al. (2018). *Discrimination in AI: A Human Rights Perspective*. *Artificial Intelligence and Human Rights*.

Voigt, P., Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR)*.

Wang, H., et al. (2021). *AI for Everyone: The Role of Public Engagement in Building Trust in AI*.

Zhang, K. (2021). The Ethics of AI in China: Opportunities and Challenges. *Journal of AI Ethics*.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. *PublicAffairs*.