Lisbon School of Economics & Management
Universidade de Lisboa

MESTRADO EM
CIÊNCIAS EMPRESARIAIS

DISSERTAÇÃO

OS POTENCIAIS PERIGOS DA INTELIGÊNCIA ARTIFICIAL: ESTUDO PARA O USO RESPONSÁVEL NAS EMPRESAS

ELISA MARIA MOREIRA MORGADO

JANEIRO - 2024

Resumo

A rápida evolução da inteligência artificial (IA) tem tido um impacto considerável na sociedade. No entanto, essa expansão rápida, não permitiu uma análise detalhada dos potenciais riscos envolvidos. Como medida para reduzir esses riscos, particularmente no contexto empresarial, este estudo destaca a crescente influência da IA na sociedade e reforça a importância de uma abordagem responsável.

O problema central abordado nesta pesquisa diz respeito ao estado de preparação das empresas para a adoção da IA e aos fatores fundamentais na adoção da tecnologia IA das empresas. O objetivo deste estudo é investigar e compreender os perigos associados à tecnologia, não com o propósito de desencorajar o uso, mas sim de promover práticas responsáveis.

Para atingir esse objetivo, a metodologia a utilizar envolve entrevistas com programadores e representantes de empresas. Essas entrevistas permitem a recolha de diversas perspetivas sobre os desafios da IA e a identificação de possíveis soluções para esses desafios.

Deste modo, concluiu-se que as empresas menos tecnológicas não estão devidamente preparadas para a adoção da IA. Muitas dessas organizações tendem a focar-se mais no lucro a curto prazo, ignorando os potenciais riscos que a tecnologia pode causar. Essa postura cria vulnerabilidades, onde a falta de uma abordagem ética e preventiva na adoção da IA pode aumentar os riscos, comprometendo a segurança e o impacto social a longo prazo.

Palavras-Chave: Inteligência Artificial, IA, Impacto, Perigos, Legislação, *ChatBots*, LLMs.

Abstract

The rapid evolution of artificial intelligence (AI) has had a considerable impact on society. However, this rapid expansion has not allowed for a detailed analysis of the potential risks involved. As a measure to reduce these risks, particularly in the business context, this study highlights the growing influence of AI on society and reinforces the importance of a responsible approach.

The central problem addressed in this research concerns the state of readiness of companies for the adoption of AI and the key factors in companies' adoption of AI technology. The aim of this study is to investigate and understand the dangers associated with the technology, not for the purpose of discouraging use, but rather to promote responsible practices.

To achieve this goal, the methodology to be used involves interviews with programmers and company representatives. These interviews allow for the collection of diverse perspectives on the challenges of AI and the identification of possible solutions to these challenges.

As a result, it was concluded that less technological companies are not adequately prepared for the adoption of AI. Many of these organizations tend to focus more on short-term profit, ignoring the potential risks that technology can cause. This attitude creates vulnerabilities, where the lack of an ethical and preventative approach to AI adoption can increase the risks, compromising safety and social impact in the long term.

Keywords: Artificial Intelligence, AI, Impact, Legislation, ChatBots, LLMs.

Índice

Índice de Tabelas

Lista de Abreviaturas

IA – Inteligência Artificial

UE – União Europeia

EUA – Estados Unidos da América

LLMs – *Large Language Models*

CAPÍTULO 1 - INTRODUÇÃO

1.1 Contexto do estudo

Nos últimos anos, a inteligência artificial (IA) surgiu como um marco significativo na história da tecnologia. Desde a conferência seminal realizada no Dartmouth *College* em 1956, que cunhou o termo "IA" (McCarthy et al., 1956), até aos avanços contemporâneos representados pelo ChatGPT-4, a IA percorreu um caminho notável (Bubeck et al., 2023).

A IA está rapidamente a tornar-se uma tecnologia disruptiva, com o potencial de transformar significativamente todos os setores da sociedade, incluindo as empresas oferecendo benefícios nos processos e tomadas de decisão (Olan et al., 2022).

No entanto, a rápida ascensão da IA também levanta preocupações sobre os seus potenciais perigos, os quais podem não estar sendo adequadamente ponderados (Turchin et al., 2020). Neste contexto, a responsabilidade refere-se à obrigação de indivíduos e organizações de responderem pelos seus atos, especialmente no desenvolvimento e implementação de tecnologias como a IA, garantindo que os seus impactos sejam éticos.

1.2 Formulação do problema

O desenvolvimento acelerado da IA trouxe consigo enormes benefícios para empresas em todo o mundo. No entanto, essa mesma velocidade criou preocupações consideráveis sobre os seus potenciais perigos, como: *Black Box*, em que existe uma falta de transparência em como os sistemas de IA chegam às suas conclusões (Setzu et al, 2021), desinformação com a criação de imagens, vídeos, documentos e até mesmo voz humana falsa de forma rápida e eficiente (Zhou et al, 2023), alucinações que criam

informações falsas ou enganosas pela IA (Monteith, 2023), cibercrime que aumentou a dificuldade de detetar e combater crimes cibernéticos (Treleaven et al, 2023), maior risco de roubo de dados confidenciais (Heister et al, 2021), manipulação da IA para fornecer informações falsas a indivíduos mal intencionados (Gupta et al, 2023) e a inteligência artificial geral que tem o potencial de superar a inteligência humana (Roli et al, 2021).

1.3 Questões centrais de investigação e objetivos

Este estudo tem como objetivos explorar e compreender a preparação das empresas para a tecnologia de IA, através de uma análise aprofundada das estratégias, desafios, riscos, competências necessárias, diretrizes éticas e impactos nos empregos relacionados à IA, de modo a contribuir para um futuro com IA responsável e benéfica para todos. Para alcançar estes objetivos, a pesquisa procura responder às seguintes questões:

1. Qual é o estado de preparação das empresas para a adoção da IA?
2. Quais são os fatores fundamentais na adoção da tecnologia IA das empresas?

1.4 Justificação do estudo

Este estudo justifica-se pela necessidade de compreender os perigos inerentes à IA, um tema amplamente discutido na atualidade. Ao fornecer orientação para o uso responsável dessa tecnologia nas empresas, a pesquisa abordará os aspetos técnicos, mas também se aprofundará nos impactos éticos, sociais e organizacionais, beneficiando toda a sociedade e promovendo práticas responsáveis e conscientes no uso da IA. Num contexto em que a IA se torna protagonista nas discussões sobre o futuro da tecnologia,

esta pesquisa assume um papel importante ao esclarecer alguns dos pontos mais cruciais relacionados com a sua implementação.

1.5 Motivação do estudo

O interesse em estudar a inteligência artificial (IA) nasceu de um conjunto de experiências e convicções, através da programação e de trabalhos práticos, houve a oportunidade de ter contacto direto com a IA, o que me proporcionou um vislumbre do seu potencial transformador e despertou interesse pelas suas capacidades. Ao mesmo tempo, foi desenvolvida a convicção de que a IA, pela sua força, precisa ser utilizada de forma responsável, pois, além dos benefícios, a tecnologia também apresenta perigos que devem ser cuidadosamente considerados.

1.6 Metodologia de investigação

A pesquisa será realizada através de uma abordagem qualitativa, com foco nas entrevistas com empresas ativas no campo da Inteligência Artificial e especialistas experientes na área. Este método permitirá uma análise detalhada e contextualizada dos potenciais perigos da IA, proporcionando uma compreensão mais ampla. Além das entrevistas, a revisão de literatura e análise de estudos complementarão a recolha de dados, assegurando uma investigação abrangente e fundamentada sobre o tema.

1.7 Estrutura

Com o propósito de oferecer uma visão mais abrangente sobre o tema em estudo, a pesquisa será cuidadosamente estruturada em seis capítulos.

O primeiro capítulo inicia com uma introdução que contextualiza e destaca a relevância do estudo, estabelecendo os objetivos e delineando a estrutura da pesquisa.

No segundo capítulo, a revisão da literatura aborda a definição e a evolução da Inteligência Artificial, analisando perspetivas sobre os potenciais perigos, os impactos nas empresas e a legislação relacionada.

O terceiro capítulo do estudo abrange a metodologia, oferecendo uma descrição detalhada da abordagem qualitativa. Destaca-se a escolha de entrevistas e outras estratégias de recolha de dados. No quarto capítulo, os resultados provenientes das entrevistas e da revisão da literatura são apresentados, enquanto o quinto capítulo se dedica à respetiva análise e interpretação.

O sexto e último capítulo, dedicado às conclusões, resume as descobertas da pesquisa, discute as suas limitações e propõe sugestões para investigações futuras. Esta estrutura procura proporcionar uma leitura lógica e detalhada, permitindo uma compreensão abrangente dos potenciais perigos da Inteligência Artificial no ambiente empresarial e incentivando reflexões sobre práticas responsáveis e futuras investigações.

CAPÍTULO 2 – REVISÃO DA LITERATURA

2.1 Definição e Evolução da IA

O surgimento da inteligência artificial (IA) foi marcado pela realização da Conferência de Dartmouth em 1956, que reuniu investigadores de diversas disciplinas, como informática, matemática e física, para explorar o potencial da IA. A conferência foi importante por vários motivos. Em primeiro lugar, ela ajudou a definir a área da IA e a estabelecer os seus objetivos, de forma a compreender como as máquinas poderiam utilizar a linguagem, formar pensamentos e resolver problemas que eram anteriormente exclusivos dos seres humanos, aprimorando as suas próprias capacidades. Em segundo lugar, ela estimulou o desenvolvimento de pesquisas em IA em todo o mundo e chamou a atenção do público para as capacidades da IA (McCarthy et al., 1956; Mijwil et al., 2023; Pale et al., 2023).

A Conferência de Dartmouth foi influenciada pelo Teste de Turing, que propôs um método para avaliar a inteligência das máquinas (Sejnowski, 2023). O teste consiste num jogo em que um humano e uma máquina conversam com um juiz humano, se o juiz não conseguir distinguir a máquina do humano, então a máquina é considerada inteligente. A ausência de uma definição formal única pode ser explicada por vários fatores. Um fator importante é o foco predominantemente voltado para aplicações práticas em detrimento de questões mais teóricas (Turing, 1950).

Após a conferência, um considerável número de investigadores começou a estudar IA, e desenvolveram investigações importantes. Entre esses investigadores, o psicólogo Rosenblatt, cujo trabalho consistiu na criação de uma rede neural artificial, mostrou que as redes neurais artificiais têm a capacidade intrínseca de aprender a partir de dados, abrindo, assim, caminho para avanços significativos na área da IA (Rosenblatt, 1958).

As primeiras linguagens de programação dedicadas à IA surgiram na década de 1950, entre as quais se destaca o LISP, uma linguagem de programação criada especificamente para IA, que se concentra na manipulação de funções e listas. Foi fundamental para o desenvolvimento de investigações em IA nas últimas duas décadas, pois permitiu aos pesquisadores implementar algoritmos de IA complexos (Glasgow et al., 1985). Outra linguagem de programação importante para o desenvolvimento da IA é o Prolog, uma linguagem popular que se concentra na declaração de factos e regras, essenciais para a representação de conhecimento e resolução de problemas (Dwivedi et al., 2023).

Na década de 60, foram desenvolvidas novas linguagens de programação que foram fundamentais para o avanço da IA. Um dos sistemas desenvolvidos nessa época foi o ELIZA, baseado num modelo linguístico que permite ao programa reconhecer padrões em entradas de texto e substituir esses padrões por respostas predefinidas (Rajaraman, 2023), sendo um marco significativo na trajetória da IA, pois foi o primeiro sistema a simular uma interação conversacional com um ser humano, à semelhança do ChatGPT (Weizenbaum, 1966).

O desenvolvimento de sistemas especialistas (*Expert systems*) foi um marco importante na história da IA. Estes sistemas tentaram igualar o raciocínio humano por meio de um conjunto de regras pré-definidas. Inicialmente, os sistemas especialistas foram bem-sucedidos ao abordar uma variedade de desafios, incluindo diagnósticos médicos e estratégias de xadrez, no entanto, as suas limitações tornaram-se evidentes ao longo do tempo, contribuindo para um declínio na área da IA (Cowan, 2001). Além disso, o relatório Lighthill, questionou a capacidade da IA de alcançar a inteligência humana, que argumentou que a área da IA era uma pesquisa muito difícil e que seria

necessário um esforço significativo para atingir esse nível de inteligência. Após este relatório, houve uma diminuição na área da IA devido a uma combinação de fatores (Cambridge University, 2020), incluindo as abordagens mais simples de redes neurais artificiais, que foram criticadas por serem limitadas e inadequadas para a resolução de problemas complexos (Minsky et al., 1988).

Um conceito que surgiu em 1997 refere-se à recolha e análise de grandes volumes de dados: o *big data,* trouxe uma nova perspetiva sobre os dados, que podem ser usados para criar conhecimento e exercer controlo sobre as pessoas (Beer, 2016). O big data permitiu o renascimento da IA, já que esta ferramenta depende de dados para treinar os seus modelos, fornecendo os dados necessários para o treino de modelos mais complexos, que podem analisar e interpretar grandes conjuntos de dados. No entanto, os desenvolvimentos de novas tecnologias de IA também levaram preocupações sobre o tratamento de dados e a substituição da mão de obra humana (Duan et al., 2019).

Ainda no século XX, surgiram duas novas ferramentas de IA que tiveram um impacto importante na área: o *machine learning* e o *deep learning.* O *machine learning* concentra-se na construção de sistemas que podem aprender com os dados, permitindo que os sistemas de IA se adaptem e melhorem o seu desempenho ao longo do tempo, sem a necessidade de serem explicitamente programados para cada tarefa específica (Forootan et al., 2022). Por sua vez, o *deep learning* é uma subárea do *machine learning* que se baseia em redes neurais artificiais, permitido avanços significativos na área, com uma potencial transformação em diversos domínios, nomeadamente pela capacidade de superar os seres humanos em algumas tarefas (Pallai, 2020).

O avanço do *deep learning* permitiu o desenvolvimento de modelos de IA generativa, que são capazes de criar novos conteúdos com base em dados existentes.

Esses modelos são treinados em grandes conjuntos de dados de conteúdos existentes, o que lhes permite desenvolver conteúdos que se assemelham à realidade, tendo o potencial de criar material de alta qualidade, praticamente indistinguível do trabalho criado por seres humanos (Banhe et al., 2023). Já são usados em produtos comerciais, como o ChatGPT, um *chatbot* desenvolvido pela OpenAI (Nah et al., 2023).

Contudo, os modelos de IA generativa, que são capazes de criar dados, trazem consigo o desafio da falta de transparência, sendo muitas vezes complexos e incompreensíveis para os humanos. Esse fenómeno é chamado como *Black Box.* A falta de transparência pode criar riscos, dificultando a compreensão das razões pelas quais os modelos tomam determinadas decisões, o que pode levar a erros, discriminação e até mesmo danos (Setzu et al., 2021). À medida que a dependência em sistemas de IA cresce, a necessidade de compreensão desses modelos torna-se crucial para garantir a fiabilidade e a aceitação social (Weitz et al., 2022).

2.2 Potenciais Perigos da IA

Nos últimos anos, especialistas da área, da política e da ética têm expressado preocupações crescentes sobre os potenciais perigos da utilização da inteligência artificial. Argumentam que os avanços rápidos da IA estão a ultrapassar a capacidade de compreensão e controlo dos potenciais riscos, o pode conduzir a ameaças existenciais, como a possibilidade de superar a inteligência humana, substituição de empregos, propagação de desinformação e preocupações sobre a segurança dos dados pessoais (Turchin et al., 2020). Apresentam-se nos anexos I e II, um sumário, respetivamente dos perigos e dos impactos da IA.

Em resposta a essas preocupações, um grupo de especialistas introduziu uma carta aberta em 2023, pedindo uma pausa no desenvolvimento de sistemas de IA. Defendem

que é necessário mais tempo para entender os riscos da IA e desenvolver mecanismos para os mitigar (Samuel, 2023). No entanto, nem todos os especialistas concordam com a necessidade de uma pausa. Alguns argumentam que a IA pode ser utilizada para resolver problemas urgentes da sociedade, como as alterações climáticas e a pobreza. Afirmam que, sem a tecnologia, não seremos capazes de avaliar e corrigir os riscos já presentes na sociedade (Clarke, 2023).

Entre os defensores da pausa no desenvolvimento da IA, há muitos especialistas que desempenharam um papel fundamental no seu progresso. Um exemplo é Geoffrey Hinton, conhecido como o "Padrinho da IA" pelas suas contribuições para o avanço de redes neurais. Hinton assinou a carta aberta (Future of Life, 2023) e deu entrevistas a várias revistas após deixar o cargo na Google. A sua saída da empresa demonstra o seu compromisso em expressar publicamente as suas preocupações sobre os riscos da IA. Hinton reconhece que os avanços recentes estão a transformar as máquinas em seres mais inteligentes do que o esperado (Heaven, 2023). Muitas vezes essa inteligência não é facilmente explicável, fenómeno conhecido como *Black Box* (Hassija et al., 2023), o que as torna altamente perigosas, visto que poderão aprender sozinhas e reescrever o seu próprio código de modo a modificar os objetivos para as quais foram criadas. Para evitar esses riscos, é necessário realizar mais testes para encontrar explicações sobre o funcionamento das máquinas e aumentar a regulamentação da IA, de forma a garantir que seja usada de forma responsável (Pelley, 2023).

Porém, os fundadores e *chatbots* como o Bard e o ChatGPT também expressaram a sua opinião em relação a estes perigos. O CEO da Google, Sundar Pichai, expressou preocupações sobre a IA, chegando a afirmar que a "sociedade não está preparada" para os riscos potenciais. Essa declaração reflete a expectativa de um aumento significativo

de notícias, imagens ou vídeos falsos criados por IA, como o Bard (Pelley, 2023). Além disso, Sam Altman, criador do ChatGPT, também manifestou as suas preocupações na presença do Comité do Senado dos EUA, pedindo uma análise de questões éticas relacionadas com a tecnologia, que podem prejudicar ou manipular pessoas (Ivanova, 2023).

Uma das principais preocupações com a IA é o seu uso inadequado para espalhar notícias falsas (*fake news*). O rápido avanço da IA tornou mais fácil manipular conteúdo e criar notícias falsas em larga escala, seja por profissionais, amadores ou *bots* que utilizam formatos como vídeos, áudios ou imagens reais (Zhou et al., 2023). Nos últimos anos, o número de artigos falsos, produzidos em vários idiomas, criados por *bots* com IA sem supervisão humana, aumentou exponencialmente (Sadeghi, 2023). Alguns desses *bots*, são os modelos de linguagem mais avançados, como o ChatGPT da OpenAI e o Bard da Google (Brewster et al., 2023), que podem ser usados para criar texto ou imagens de forma rápida e barata (Chesterman, 2023).

Muitas vezes, empresas criam artigos falsos com o objetivo de atingir um grande número de pessoas com sua publicidade. Utilizando métodos de rastreio e tecnologias de IA para influenciar diálogos nas redes sociais, como a desinformação, direcionada a pessoas especificas que, à partida, acreditam em determinados artigos, gastando tempo lê-los nas plataformas ou páginas *online* e proporcionando ganhos financeiros através da publicidade (Hajli et al., 2021).

Para facilitar a divulgação de desinformação nas plataformas, as pessoas ou empresas criam *bots* sociais, que são contas de utilizadores automatizadas que usam plataformas de media social (Bontridder et al., 2021). Os *bots* sociais podem ser usados para espalhar desinformação de forma eficaz, promovendo conflitos, dividindo a

opinião pública e impulsionando ainda mais desinformação ou até mesmo a manipulação do sentimento público (Ferrara, 2023).

No entanto, após o lançamento dos *chatbots*, surgiram várias questões sobre a exatidão dos textos produzidos por essas ferramentas, que tornam mais fácil a produção de notícias falsas (Węcel et al., 2023). Esse fenómeno acontece porque, por vezes, a IA não entende completamente o significado das palavras que lhe são fornecidas. Como resultado, pode fornecer respostas que não são factualmente corretas ou que se desviam do conteúdo real. Esse fenómeno é conhecido como "alucinações" (Monteith, 2023).

As alucinações representam uma limitação significativa à funcionalidade da IA e podem levantar preocupações de segurança. Isto acontece porque os dados usados para treinar a IA podem ser inadequados ou insuficientes. Como resultado, a IA pode ser enganada para criar conteúdo falso (Hill, 2023).

Além da desinformação, o cibercrime e a fraude também são preocupações crescentes. A IA pode ser usada para criar imagens ou vídeos manipulados, contratos fictícios, ataques a empresas e até mesmo imitar voz humana (Shahzad et al., 2022). Esses ataques são facilitados por *chatbots*, que podem ser usados para criar *malware* e automatizar tarefas. Como resultado, o cibercrime está a tornar-se mais sofisticado e industrializado.

Para combater o cibercrime no futuro, as autoridades policiais estão a usar a IA (Treleaven, et al., 2023). No entanto, o problema também é humano. As fraudes autorizadas, em que os utilizadores legitimamente autorizam transações fraudulentas, podem ser difíceis de detetar. Isto acontece porque os utilizadores podem ser enganados com golpes de persuasão que usam IA para estabelecer uma relação de confiança (Ma et al., 2023).

De maneira a efetuar essas fraudes, frequentemente as bases de dados são invadidas para roubar informações pessoais, como nome, endereço, número de cartão de crédito e senhas de acesso (Heister et al., 2021). Esses dados também podem ser fornecidos voluntariamente por pessoas que aceitam os termos e condições de sites (*cookies*), redes sociais e plataformas de vídeo *online*. Assim, esses dados podem ser partilhados ou vendidos a terceiros, que podem usá-los para cometer fraudes, manipular opiniões ou realizar outros crimes. A IA pode ser usada para automatizar esses processos e torná-los mais difíceis de detetar (Bartneck, 2021).

Mesmo que alguns dados não revelem diretamente a identidade de uma pessoa, é relativamente fácil identificar indivíduos ao combinar conjuntos de dados com outras fontes *online*. Isso pode levar à desconfiança dos clientes em relação às empresas, que se podem recusar a partilhar os seus dados. Isso pode prejudicar as relações comerciais (Carmody et al., 2021), pois as empresas precisam de dados de qualidade para maximizar os seus lucros (Majeed, 2023).

A maioria dos ataques cibernéticos é desenvolvida com a ajuda de ferramentas de IA, como o ChatGPT. Essas ferramentas fornecem pedaços de código que podem ser usados para criar *malware*. Embora a OpenAI, proprietária do ChatGPT, tenha implementado políticas de segurança para impedir a disseminação desse tipo de informação, é possível contornar essas restrições (Monje et al., 2023). A natureza flexível da linguagem de programação permite que indivíduos mal-intencionados convençam a IA a fornecer informações, usando técnicas psicológicas, de modo que a IA forneça as informações desejadas, que consequentemente, podem ser usadas para desenvolver *malware* de forma eficaz (Gupta et al., 2023).

Os riscos discutidos até agora referem-se à IA estreita, que é programada para realizar tarefas específicas de forma direcionada. No entanto, um perigo ainda maior é a inteligência artificial geral, que é programada para igualar a inteligência humana em termos de amplitude e adaptabilidade (Schlegel et al., 2021). Essa inteligência seria um operador autónomo capaz de aprender sem supervisão, o que levanta preocupações sobre como controlar um programa substancialmente mais inteligente do que os humanos. É possível que essa inteligência tenha a capacidade de se autoaperfeiçoar de forma repetitiva, criando versões mais inteligentes de si mesma e alterando os seus objetivos pré-programados (Roli et al., 2021). As suas criações poderiam criar inúmeros desafios e ameaças sociais, desde a substituição da força de trabalho até a manipulação de sistemas políticos e militares, e até mesmo a possibilidade de extinção humana (McLean et al., 2021).

Existem opiniões divergentes sobre a criação desta inteligência comparável à inteligência humana. Alguns especialistas acreditam que ela está muito longe de acontecer ou que será produtiva para as organizações (Kshetri, 2023), enquanto outros acreditam que há 50% de hipóteses de ser inventada até 2060, tendo já sido discutido os problemas de controlo e política associados a essa inteligência. Além disso, a corrida para desenvolver uma IA superinteligente pode levar à criação de ferramentas defeituosas, pois existe uma maior preocupação com o avanço tecnológico do que com questões éticas e de segurança. Isso pode resultar no uso da tecnologia para fins ilegais ou prejudiciais, como ganhar vantagem no mercado (Naudé, 2019)

Embora existam preocupações sobre a viabilidade de alcançar essa inteligência, ela não está tão distante quanto se imagina. O ChatGPT-4 já mostra evidências de inteligência artificial geral, embora com algumas limitações. No entanto, o sistema

demonstrou avanços significativos na capacidade de realizar e compreender a leitura de texto e progrediu consideravelmente em termos de senso comum (Bubeck et al., 2023; Brezinski et al., 2023; Haase et al., 2023).

*2.3 Impacto n*as empresas

O interesse pela implementação de IA nas organizações aumentou significativamente, com empresas a procurarem obter benefícios substanciais em termos de valor agregado (Olan et al., 2022). Para implementar a IA com sucesso, é necessário dispor de uma infraestrutura adequada, conjuntos de dados complexos e treinados, além investimentos financeiros (Lee et al., 2022). Além disso, é importante que os funcionários estejam abertos a aprender informações novas ao trabalhar com soluções de IA. Para facilitar isso, as empresas devem garantir que os funcionários tenham as qualificações necessárias e estejam atualizados com os novos requisitos associados à IA. Esses novos profissionais são responsáveis por criar, implementar e realizar soluções baseadas nessa tecnologia (Enholm et al., 2022).

A adoção da IA por empresas ainda é um processo cauteloso, com muitas preocupações sobre a implementação e os riscos associados (Anexo I) (Curtis et al., 2022). Cada vez mais, as empresas procuram estratégias para reduzir os riscos associados à IA, tanto para o negócio quanto para os clientes que exigem cada vez mais essa segurança. No entanto, ainda há muitas dúvidas sobre como gerir a IA, devido à falta de regulamentação, governação adequada e até saber quem são os responsáveis pela gestão desses riscos (Schuett, 2023). Apesar dessas preocupações, as organizações continuam a usar a IA, principalmente as empresas de grandes dimensões, mas ainda estão a experimentar e não se comprometem com grandes investimentos. Elas propõem

algumas regras e práticas éticas para a adoção, desenvolvimento e, especialmente, no fornecimento de aplicações ou plataformas de IA (Harvard Business Review, 2021).

A Microsoft é uma das empresas impactadas pela IA, sendo impulsionada a adotar melhores práticas (Schuett, 2023). A empresa implementou em toda a sua estrutura processos e princípios como justiça, confiabilidade e segurança. Isso pretende garantir que as ferramentas da Microsoft tenham impactos positivos para a sociedade, acompanhando-as desde o início e ao longo de sua vida útil. No entanto, a Microsoft reconhece que as medidas voluntárias não são suficientes para reduzir todos os riscos existentes. A empresa acredita que a criação de leis é necessária para proteger a indústria e a sociedade (Microsoft, 2023). Ainda assim, mesmo que os regulamentos legais exijam que as empresas garantam a confiabilidade das tecnologias de IA, elas podem não ser totalmente transparentes e compreensíveis, como no caso dos modelos de *Black Box* (Weitz et al., 2022).

Sundar Pichai, CEO da Google, também se manifestou sobre a IA, afirmando que ela terá impacto em todos os produtos de todas as empresas, assim como nos empregos. Profissionais como escritores, arquitetos, engenheiros de software e outros trabalhadores do conhecimento terão de lidar com a presença significativa de sistemas de IA nas suas tarefas cotidianas (Pelley, 2023). Contudo, embora reconheça que haverá alguns desafios, a Google também se compromete com o desenvolvimento responsável de tecnologia, realizando múltiplos testes para garantir que as suas tecnologias sejam benéficas para a sociedade, reduzam possíveis preconceitos, aumentam a privacidade e não sejam usadas para prejudicar (Google, 2023).

O uso de sistemas de IA para dinamizar os mercados está a tornar-se cada vez mais comum. As empresas utilizam algoritmos para ajustar automaticamente os preços e

descontos, o que pode criar mais receita e lucros do que os concorrentes (Krakowski et al., 2022). No entanto, essa prática também pode levar a uma potencial ameaça de cooperação automática e autónoma entre os algoritmos, que podem unir-se para atingir um objetivo comum, muitas vezes prejudicando terceiros ou violando regras éticas e legais. Nesse contexto, o uso de *deepfakes* direcionados às empresas para obter benefícios financeiros de forma ilegal torna-se ainda mais preocupante. A falsificação de identidade, por exemplo, através da criação de vídeos ou áudios manipulados com a ajuda de IA, pode levar a graves consequências, como fraudes, golpes e danos à reputação (Mustak et al., 2023). Esta prática pode escapar à supervisão das autoridades reguladoras, pois a complexidade dos algoritmos de IA dificulta a compreensão (Sanchez-Cartas et al., 2022). Além disso, à medida que as empresas utilizam mais algoritmos de IA em diferentes áreas e competem entre si para oferecer melhores serviços, a qualidade dos dados utilizados para treinar esses algoritmos pode ser afetada. Isso prejudica imediatamente o desempenho dos algoritmos, mas também afeta a forma como eles aprendem ao longo do tempo, comprometendo a sua eficácia e orientação futura, o que é essencial para o seu funcionamento adequado (Ginart et al., 2021).

A IA está a ser utilizada para automatizar tarefas rotineiras de recrutamento e seleção, com o objetivo de aumentar a eficiência e atrair os melhores candidatos para as organizações. No entanto, a implementação da IA nesses processos enfrenta preocupações significativas, incluindo questões éticas, legais, de privacidade e morais (Ore et al., 2021).

Estes problemas surgem porque a IA pode criar tendências discriminatórias, como género, raça, etnia e cor, podem levar ao risco de multas por práticas de recrutamento discriminatórias (Ore et al., 2021). Por exemplo, se os dados utilizados

para treinar um algoritmo de recrutamento forem tendenciosos, o algoritmo pode produzir resultados discriminatórios. Isso pode acontecer involuntariamente, pois os algoritmos são treinados em dados históricos que refletem as desigualdades existentes na sociedade (Zou et al., 2018). Contudo, esses resultados discriminatórios são muitas vezes ignorados devido ao pensamento equívoco de que os processos de IA são sempre imparciais e neutros. Caso esses problemas não sejam corrigidos adequadamente, os algoritmos podem agravar as desigualdades existentes e continuar a discriminar grupos minoritários (Chen, 2023).

2.4 Legislação na Inteligência Artificial

O rápido avanço da IA fez da sua regulação uma questão central na política da UE. Como dados são essenciais para o desenvolvimento da IA, a UE criou um quadro legislativo para proteger dados e privacidade, além de regular a exportação de dados para países terceiros (Hadzovic, 2023).

Em 2021, a Comissão Europeia apresentou a primeira proposta para regulamentar a IA na UE. A proposta classifica os sistemas de IA com base no risco que representam para os utilizadores. Sistemas considerados ameaças serão proibidos, enquanto os sistemas de alto risco enfrentarão uma série de obrigações para poderem ser usados no mercado da UE, e os sistemas de risco limitado devem cumprir requisitos mínimos de transparência, de modo a permitir que os utilizadores tomem decisões informadas. O não cumprimento dessas regras pode resultar na proibição ou retirada do sistema de IA do mercado, sujeitando-o a sanções significativas pelo desrespeito das normas (Madiega, 2023).

No entanto, o início dessas regulamentações ainda não está completamente adaptado à realidade da IA, considerando a possibilidade da IA desenvolver convicções

morais e legais no futuro (Edwards, 2022). Mesmo que os fabricantes e fornecedores de produtos de IA se dediquem a criar sistemas seguros, não seriam considerados culpados em casos de defeitos nos produtos. A complexidade em condenar um sistema de IA pode exigir alterações na legislação para atribuir personalidade jurídica à IA (Khan, 2021). Adicionalmente, é crucial ressaltar que, embora o sistema em si possa ser considerado seguro do ponto de vista do *software,* a origem dos dados utilizados, provenientes de bases de dados e plataformas de pesquisa como o Google, pode introduzir distorções significativas. Essas distorções nos dados podem resultar em impactos indiretos, prejudicando determinados grupos, como mulheres ou homens, no contexto de processos de recrutamento. É importante observar que, embora essa influência não represente uma violação direta e evidente da legislação da União Europeia sobre igualdade de género, ela destaca uma falha na legislação existente em abordar questões relacionadas com a IA e o seu potencial impacto nas práticas de recrutamento (Lütz, 2022).

No entanto, a UE não está sozinha na elaboração de leis para esses sistemas. Em 2022, os Estados Unidos da América (EUA) propuseram a Lei de Responsabilidade Algorítmica, abordando o uso geral de sistemas de decisão automatizados. Esta legislação indica que as organizações que implementam esses sistemas devem adotar várias medidas concretas para identificar e reduzir os riscos sociais, éticos e legais. Esta lei abrange todos os tipos de sistemas de decisão automatizados, mas também impõe obrigações às empresas que os utilizam. Devem garantir que, ao tomar decisões com efeitos legais significativos na vida de um consumidor, seja assegurado um tratamento justo. No entanto, alguns desafios persistem, como o facto da aplicação ser exclusiva a grandes empresas e a questão da possibilidade de discriminação nas empresas ainda não

estar clara, pois a legislação pode não ser suficientemente específica nesse aspeto. As empresas podem optar por deixar de recolher informações sobre classes protegidas para evitar a necessidade de análises, o que não é um problema do sistema e não viola a legislação existente (Mökander et al.,2022; Gursoy et al., 2022).

Em 1 de agosto de 2024, foi aprovado o primeiro Regulamento de Inteligência Artificial, que estabelece diretrizes ainda mais claras para o uso de IA na União Europeia. O regulamento reforça a classificação de sistemas de IA em quatro categorias de risco: mínimo (recomendações baseadas em IA e filtros de spam), específico de transparência (revelar aos utilizadores que estão a interagir com IA), elevado (requisitos rigorosos no âmbito dos dados) e inaceitável (manipulam o comportamento humano**)**. Além disso, as empresas que não cumprirem as regras podem enfrentar multas severas, que podem atingir até 7% do volume de negócios anual global. Os Estados-Membros da União Europeia têm até 2 de agosto de 2025 para nomear autoridades nacionais responsáveis por supervisionar a aplicação das regras relacionadas com os sistemas de IA e realizar a fiscalização do mercado. A nível da UE, o Serviço de IA da Comissão Europeia será o principal órgão encarregado de implementar o Regulamento de Inteligência Artificial, especialmente no que diz respeito aos modelos de IA de finalidade geral. A maioria das regras começará a vigorar em 2 de agosto de 2026, contudo, as proibições de sistemas de IA com risco inaceitável entrarão em vigor seis meses antes, enquanto as regras para os modelos de IA de finalidade geral serão aplicáveis após 12 meses.

CAPÍTULO 3 – METODOLOGIA

Neste capítulo, são delineadas as estratégias e abordagens selecionadas para a recolha e análise dos dados, proporcionando uma compreensão clara do processo de orientação da investigação, visando responder de maneira eficaz às questões de investigação propostas.

Para o presente estudo, foi escolhido uma análise qualitativa, que se concentra num método de investigação que procura produzir uma descrição rica e detalhada do objetivo de estudo, realizando uma análise de dados com suporte em entrevistas, documentos, imagens e observações de como as pessoas interpretam o tema (Aspers et al., 2019). Um dos principais objetivos é a compreensão do mundo social através de um exame da interpretação desse mundo pelos seus participantes, o que implica tentar identificar padrões inerentes, em vez de impor ideias preconcebidas aos dados (Bell et al., 2019).

Sendo assim, para uma melhor recolha de dados e dada a complexidade do tema e as diferentes perspetivas entre aqueles que criam a IA e aqueles que utilizam, é importante criar grupos distintos de entrevistados. Por um lado, serão entrevistados programadores e especialistas que têm conhecimento e contacto direto com a tecnologia e que compreendem os detalhes do seu funcionamento. Por outro lado, serão entrevistados gestores que utilizam a IA como ferramenta nas suas atividades, partilhando as suas perspetivas. Esta divisão metodológica permite uma melhor recolha de dados mais eficaz, de modo a comparar as visões e experiências dos dois grupos de *stakeholders* no contexto da IA.

Dessa forma, adotou-se uma abordagem de entrevista semiestruturada. Neste método, as perguntas não são rigidamente predefinidas, proporcionando flexibilidade

para explorar as respostas dos participantes de maneira mais personalizada. Embora algumas perguntas principais sejam estabelecidas previamente, é possível formular perguntas adicionais com base nas respostas dos participantes, enriquecendo a recolha de dados ao permitir uma exploração mais detalhada dos tópicos em discussão (Bell et al., 2019). Todas as entrevistas serão realizadas através da metodologia *cross-sectional*, escolhida pela sua agilidade e capacidade de investigar associações entre múltiplas exposições e resultados num ponto específico do tempo (Wang et al., 2020). É crucial destacar que a recolha de dados adere aos princípios éticos, garantindo a prevenção de danos, a obtenção de consentimento dos entrevistados, a proteção da privacidade através da confidencialidade e evitando qualquer forma de engano. Além disso, todas as questões seguirão um guião de entrevista elaborado com base na revisão da literatura, que se apresenta no Anexo III, mantendo uma abordagem uniforme de entrevistado para entrevistado (Bell et al., 2019).

Na etapa inicial do tratamento e análise qualitativa, é imperativo realizar uma transcrição cuidadosa dos dados em formato de áudio das entrevistas gravadas para o software Word. Este software servirá como a plataforma apropriada para a análise, utilizando o processo de codificação (Wilson, 2014).

A codificação, nesse contexto, envolve a revisão das transcrições das entrevistas, atribuindo etiquetas ou nomes às partes que revelam potencial significado em relação ao tema em estudo (Bell et al., 2019). Para efetuar essa codificação, é utilizado a ferramenta MAXQDA, uma aplicação que auxilia na identificação e destaque das perguntas e reflexões relevantes (Saldana, 2021). Esta abordagem sistemática permitirá uma interpretação detalhada dos dados, identificando padrões e tendências fundamentais para a compreensão do tema em questão (Wilson, 2014).

CAPÍTULO 4 – APRESENTAÇÃO DOS RESULTADOS

A recolha de dados para o estudo foi realizada em abril de 2024, envolvendo nove participantes cuidadosamente selecionados para garantir uma ampla variedade de perspetivas. O processo de seleção abrangeu programadores e investigadores experientes em IA, além de representantes de organizações que já implementaram essa tecnologia nas suas operações.

Para identificar os participantes, foram utilizados dois métodos principais. Um deles foi a utilização da plataforma *LinkedIn* para pesquisar e contactar indivíduos que se encaixassem nos critérios de seleção. O segundo método consistiu em recorrer à rede de contactos da licenciatura e outros contactos pessoais relevantes com experiência em IA. Esta abordagem resultou num grupo diversificado, abrangendo diversas áreas da comunidade de IA, desde o desenvolvimento técnico até a aplicação prática em organizações reais, com o objetivo de obter resultados mais abrangentes e fiáveis.

O contacto com os potenciais entrevistados foi realizado de forma transparente e com preparação adequada para as entrevistas. As mensagens enviadas continham uma breve descrição pessoal, área de estudo e grau académico, bem como uma explicação clara e concisa sobre o tema da dissertação, os objetivos do estudo e os detalhes da metodologia, incluindo a duração e formato das entrevistas.

Para garantir a preparação dos participantes e a obtenção de respostas completas e informativas, foi disponibilizado um guião de entrevista para alguns deles. As entrevistas foram conduzidas online, principalmente via Microsoft Teams, e num caso específico, foi utilizado a plataforma online da empresa para se adaptar às suas necessidades. Todas as entrevistas foram gravadas em áudio com dispositivos móveis, com consentimento dos entrevistados, e tiveram duração média entre 20 e 50 minutos.

A Tabela 1 oferece uma caracterização completa do painel de entrevistados, englobando informações como formação académica, tipo de empresa e função profissional, em conjunto com a Tabela 2, que apresenta um resumo da duração das entrevistas realizadas na plataforma online *Microsoft Teams.*

Tabela 1 - Descrição dos participantes

ENT.	CARGO	FORMAÇÃO ACADÉMICA	EMPRESA	CARGO
ENT 1	Gestor	Administração de empresas	Empresa Multinacional de Telecomunicações	Diretor de Cuidados Digitais
ENT 2	Gestor	Sistemas de informação	Empresa Multinacional de Tecnologia	Diretor Nacional de Tecnologia
ENT 3	Gestor	Engenheiro de Eletrónica e Computadores	Empresa Multinacional de Tecnologia da Informação e Comunicação	Diretor, Engenharia de Soluções
ENT 4	Gestor	Engenharia física	Empresa de Atividades dos serviços de tecnologia da informação	Cofundador e sócio da empresa
ENT 5	Gestor	Gestão e Ciências da Computação	Empresa multinacional portuguesa de retalho alimentar	Diretor de Tecnologia
ENT 6	Programador	Engenharia Informática	Universidade A	Diretor e Professor de programação
ENT 7	Programador	Engenharia Eletrotécnica e de Robótica	Universidade B	Professor de IA
ENT 8	Programador	Línguas e Literatura Moderna	Universidade C	Investigador de IA
ENT 9	Programador	Engenharia Informática	Universidade D	Professor de programação e IA

Fonte: Elaboração própria

Tabela 2 - Duração das entrevistas

Tempo máximo de entrevista	48 minutos
Tempo mínimo de entrevista	22 minutos
Tempo médio de entrevista	33 minutos

Fonte: Elaboração própria

Após a gravação e transcrição das entrevistas, foi utilizado o *software* de análise qualitativa MAXQDA para investigar as perspetivas dos entrevistados. Através dessa ferramenta, padrões, temas recorrentes e relações entre as diversas ideias expressas pelos participantes foram identificados e explorados com rigor.

Com o objetivo de enriquecer ainda mais a compreensão dos resultados e proporcionar uma visão abrangente da análise qualitativa realizada, é disponibilizado no Anexo IV um resumo detalhado dos códigos utilizados durante a análise. Cada código é definido e descrito com precisão, fornecendo uma compreensão clara de como os entrevistados percebem o tema em questão. Em conjunto com a matriz visual de códigos, apresentada no Anexo V, facilita-se a compreensão da presença dos códigos em cada entrevista.

CAPÍTULO 5 – ANÁLISE E DISCUSSÃO DOS RESULTADOS

A análise dos resultados visa compreender as perspetivas dos entrevistados, contribuindo para a identificação de estratégias de redução dos riscos da inteligência artificial.

5.1 Infraestruturas e Funcionários Adequados

A implementação bem-sucedida da IA depende de três pilares fundamentais: infraestrutura forte, conjuntos de dados complexos e treinados e investimento financeiro estratégico. Esta trilogia crucial, confirmada por Lee et al. (2022) e pelas entrevistas realizadas, revela-se como a chave para desbloquear o potencial transformador da IA.

Entre os entrevistados, o tema mais recorrente em relação à infraestrutura foi o processamento computacional, reconhecido por todos os entrevistados, exceto o ENT 4. Gestores e programadores de IA concordam que esta etapa é vital para o bom funcionamento e treino dos modelos de IA, como constata:

> […] a capacidade que nós temos de ter computacional para treinar esses modelos é de facto, bastante grande e, portanto, a despesa para manter essa infraestrutura é elevada. (ENT. 9)

No entanto, a aquisição dessa infraestrutura exige um investimento significativo (ENT. 4, ENT. 5, ENT. 6, ENT. 9). Curiosamente, dois desses entrevistados representam empresas com menor nível de maturidade tecnológica dentro da amostra. Esta realidade, ausente na literatura, pode representar um obstáculo para a implementação da IA em empresas pequenas ou com recursos limitados, que podem estar menos familiarizadas com a ferramenta IA e, consequentemente, menos dispostas a investir em infraestruturas robustas.

O ENT. 6 sugere uma alternativa para minimizar os custos associados à infraestrutura de IA, que passa pela utilização de modelos pré-existentes oferecidos por grandes empresas do setor:

> […] No caso das empresas, se elas quiserem, elas próprias criam os modelos e então vou ter esse custo, se elas quiserem usar os modelos existentes por ofertas comerciais, […] eu acho que até são bastante acessíveis para o benefício, cerca de 20€ por utilizador. (ENT. 6)

O uso de dados externos pelas empresas apresenta perigos reais, como alucinações (Hill, 2023) e algoritmos discriminatórios (Zou et al., 2018). Esta preocupação intensifica-se quando é observado nas entrevistas, por quase todos os participantes que referem o uso de dados externos, com exceção do ENT. 2, ENT. 3 e ENT. 5. Esta prevalência pode ser parcialmente explicada pela procura por dados a baixo custo, conforme mencionado. No entanto, essa procura por economia pode ter um preço alto em termos de ética e precisão: *[…] esses dados estão de certa maneira "sujos", no sentido em que é preciso limpar muito desses dados.* (ENT. 7), tornando-os altamente propensos a vieses e ao aumento do risco de discriminação algorítmica.

Embora os dados internos sejam mencionados com menos frequência (ENT. 1, ENT. 3, ENT. 6 e ENT. 9), há um reconhecimento claro dos seus benefícios:

> […] Sempre que possível usamos modelos desenvolvidos na empresa, ou seja, modelos em que somos nós que controlamos exclusivamente os dados que são colocados de maneira que sempre que é necessário usamos os dados de tráfego reais, naturalmente, mas que permitem ter um controlo total do lado da empresa em que não envolve qualquer entidade externa. (ENT. 1)

Ao discutir as infraestruturas, dois dos entrevistados (ENT. 2 e ENT. 5) mencionaram a importância de que os profissionais que lidam com IA possuam *[…] as competências digitais para poder tirar o uso destas ferramentas. Portanto, saibam fazer*

prompts, saibam que tipo de ferramentas é que existem, saibam que os algoritmos são probabilísticos e que podem errar […] (ENT. 2).

Destaca-se, assim, a importância da capacitação dos funcionários, que precisam ter as qualificações necessárias para responder aos requisitos da IA (Enholm et al., 2022). A maioria dos entrevistados está disposta a oferecer treino aos colaboradores, visando aprimorar as suas habilidades no uso dessa tecnologia, em colaboração com profissionais do setor. No entanto, alguns entrevistados (ENT.4, ENT.5, ENT.6, ENT.9) acreditam que a autoaprendizagem seja suficiente para que os funcionários adotem e utilizem essa tecnologia.

5.2 Perigos da IA

No que diz respeito aos perigos da IA, foram discutidos temas como tendências discriminatórias, IA geral, *Black Box*, manipulação e qualidade dos dados, incluindo *deepfakes*.

A IA, apesar dos seus avanços promissores, também gera preocupação entre especialistas, que alertam para os perigos que esta tecnologia pode representar (Turchin et al., 2020). Esta preocupação reflete-se nas entrevistas, onde a maioria dos participantes demonstra apreensão com o tema, com exceção do ENT. 4, que se destaca por minimizar os problemas e priorizar os benefícios da IA: *[…] nós como estamos numa empresa em que somos treinados para ter primeiro a física da coisa e depois dizer se os dados são válidos ou não. […]* (ENT. 4), reforçando essa visão divergente.

Em relação ao problema da discriminação na IA, todos os especialistas presentes (ENT. 6, ENT. 7, ENT. 8 e ENT. 9) reconheceram o desafio de solucioná-lo, alertando que: *[…] eu tenho algum receio que não seja possível controlar estes comportamentos estranhos, que às vezes nós vemos nos LLMs e no IA, que alguns deles têm a ver*

precisamente com esses vieses, que existem nos dados que deram origem àquele modelo. […] (ENT. 6). Como aponta Zou et al. (2018), esses vieses são característicos aos algoritmos, pois são treinados em conjuntos de dados históricos que refletem as desigualdades da sociedade, o que torna a resolução do problema um desafio significativo. Para combater o problema da discriminação na IA, grande parte dos participantes, exceto ENT. 4, ENT. 6 e ENT. 8, sugeriram a validação da informação como uma possível solução, de modo a detetar tendências presentes nos conjuntos de dados. Além disso, alguns participantes (ENT. 2 e ENT. 3) destacaram a importância do pensamento crítico para controlar mais atentamente os resultados produzidos.

Para combater problemas de *deepfakes*, as opiniões dividem-se entre diferentes abordagens. Uma parte dos participantes (ENT. 1, ENT. 5 e ENT. 6) defende a regulamentação das plataformas digitais, com mecanismos de controlo para identificar e remover conteúdos falsos. No entanto, segundo Hajli et al. (2021), por vezes, as próprias empresas criam conteúdos falsos com o objetivo de atingir um grande número de pessoas para fins publicitários. Outros participantes (ENT. 2 e ENT. 8) defendem a sinalização dos produtos produzidos por IA, como *[…] na possibilidade de colocar uma marca de água nos vídeos ou nas imagens que são produzidas pela inteligência artificial.* […] (ENT. 8), o que poderia auxiliar na identificação e controlo da informação falsa.

Finalmente, alguns dos entrevistados (ENT. 4, ENT. 8, ENT. 9) propõem o desenvolvimento de uma cultura de desconfiança, com o objetivo de estimular o senso crítico dos utilizadores na análise de conteúdos online.

A procura por dados para se destacar no mercado e oferecer serviços aprimorados, como apontam Ginart et al. (2021), pode prejudicar a qualidade dos dados,

impactando negativamente o desempenho das aplicações. Os participantes da pesquisa (ENT. 2 e ENT. 7) alertam para esse limite: *[…] está-se a chegar a um limite em que os modelos não têm dados novos, eles não conseguem desenvolver novas capacidades. A única maneira é criar dados realistas […]* (ENT. 7), afirmando que uma das soluções é a criação de dados próprios, mas também a validação com IA (ENT. 2 e ENT. 3) sugerindo a utilização de filtros de IA para validar os dados ou a utilização de outros programas para essa validação (ENT. 1, ENT. 5 e ENT. 8).

A manipulação de ferramentas de IA representa um dos desafios mais complexos da era digital, pois essas ferramentas fornecem blocos de código que podem ser facilmente utilizados para a criação de *malware* (Monje et al., 2023). Apesar dos esforços para reduzir esse risco, como testes de manipulação (ENT. 1, ENT. 2 e ENT. 6), as empresas simulam cenários de manipulação para detetar pontos fracos e falhas nos sistemas, permitindo a correção preventiva de problemas antes que causem prejuízos. Outra estratégia para combater a manipulação de ferramentas de IA é a validação das informações inseridas (ENT. 3 e ENT. 7), garantindo a fiabilidade dos dados utilizados nos sistemas de IA.

Contudo, essas técnicas podem ser insuficientes, e os investigadores alertam para a utilização de técnicas psicológicas para convencer a IA a fornecer informações, abrindo portas para fins maliciosos (Gupta et al., 2023). Essa dificuldade é demonstrada pela maioria dos entrevistados, com exceção do ENT. 4, que reconhecem a natureza complexa do problema: *[…] Nós nunca conseguimos garantir 100% isso e quem está a dizer que consegue naturalmente deve estar a mentir […]* (ENT. 7) e até alguns entrevistados (ENT. 4, ENT. 6, ENT. 8 e ENT. 9) comparam a manipulação da IA aos perigos da internet, reconhecendo a similaridade dos riscos.

A falta de transparência na IA, que dificulta a compreensão do processo decisório dos modelos, também conhecido como *Black Box* (Setzu et al., 2021), não parece ser um grande impedimento para o uso dessas ferramentas pelos participantes do estudo. No entanto, algumas preocupações foram levantadas, como a necessidade do controlo da informação (ENT. 2, ENT. 3, ENT. 7, ENT. 9) para avaliar a fiabilidade dos resultados e até mesmo a importância de partilhar as fontes utilizadas pela IA (ENT. 2). É importante destacar que alguns entrevistados mencionaram que a decisão de usar ou não ferramentas de IA depende do custo de risco associado aos resultados (ENT. 1, ENT. 5), avaliando se os benefícios superam os potenciais riscos antes de decidirem se continuam a utilizar as ferramentas de IA na empresa.

A ideia da IA Geral, planeada para alcançar a inteligência humana em termos de amplitude (Schlegel et al., 2021), ainda é vista como algo distante e desconhecido pela maioria dos entrevistados, com exceção do ENT. 3. Esta visão é partilhada por Kshetri (2023), que aponta para um futuro incerto em relação à IA Geral. Apesar da dúvida geral, um participante (ENT. 3) já utiliza a ferramenta ChatGPT-4, que demonstra algumas características da IA Geral. Esta experiência prática vai ao encontro das pesquisas de Bubeck et al. (2023), Brezinski et al. (2023) e Haase et al. (2023), que apresentam evidências do potencial da IA Geral, como a capacidade de realizar e entender a leitura de textos e conceitos de senso comum.

5.3 Uso responsável da IA

O rápido avanço da IA cria debates sobre a sua regulamentação. Especialistas defendem diretrizes claras para um uso ético da tecnologia (Hadzovic, 2023). As entrevistas revelam que todos reconhecem a necessidade de regulamentar a IA, no entanto, há uma diversidade de visões sobre a melhor forma de abordar a questão.

Alguns participantes (ENT.1 e ENT.3) alertam que os riscos da IA são subestimados na legislação atual, reforçando as preocupações levantadas por Edwards (2022) sobre a inadequação das regulamentações face à realidade da IA, enquanto outros (ENT.2, ENT.3 e ENT.8) defendem uma regulamentação que oriente o desenvolvimento da IA sem sufocar a inovação, procurando um equilíbrio entre segurança e progresso. Por outro lado, alguns participantes (ENT.4 e ENT.5) posicionam-se contra a regulamentação, temendo que esta traga mais problemas do que benefícios, argumentando que *[…] tornar as coisas mais difíceis e mais complicadas para toda a gente. […]* (ENT. 4) e que *[…] a questão é mais social que empresarial […]* (ENT. 5).

Embora as visões sobre a regulamentação da IA divergissem, todos os participantes nas entrevistas concordaram num ponto crucial: a importância do uso responsável da IA nas empresas. A grande maioria dos entrevistados (ENT.1, ENT.2, ENT.3, ENT.6, ENT.8 e ENT.9) reconheceu que a responsabilidade pelas ações da IA recai sobre as empresas que a implementam, e não sobre a tecnologia em si. Isto liga-se à questão da proteção de dados, também mencionada (ENT. 2, ENT. 3, ENT. 5, ENT. 9), já que muitas aplicações de IA utilizam dados pessoais. Os entrevistados demonstraram a intenção de garantir e proteger essas informações. Outro ponto importante mencionado por muitos participantes, com exceção do ENT. 4, ENT. 5 e ENT. 6, foi a importância da inclusividade, destacando que ninguém deve ser deixado para trás na era da IA e que todos devem ter acesso à tecnologia.

A transparência também foi um tema fundamental identificado nas entrevistas (ENT. 2, ENT. 3 e ENT. 9). É crucial entender que os resultados da IA são derivados

dos dados utilizados, por isso é importante disponibilizar as referências, de forma a evitar violações de direitos autorais que podem ocorrer em algumas respostas criadas.

Embora outros temas tenham recebido menos atenção para melhorar o uso da IA, a sua importância não deve ser subestimada. Entre eles, destaca-se a integridade da aplicação (ENT. 2 e ENT. 6), que é essencial para garantir a fiabilidade e a filtragem de possíveis vieses nos sistemas de IA. A sustentabilidade (ENT. 3) e a participação dos grupos éticos (ENT. 3) são cruciais para garantir que o desenvolvimento e a aplicação da tecnologia estejam alinhados com princípios éticos e valores sociais. Um dos participantes (ENT. 3) expressou preocupação com a negligência no acompanhamento ético por parte das empresas que desenvolvem tecnologias de IA: *[…] Preocupa-me quando entidades que gerem este tipo de tecnologia dispensam as equipas éticas para acompanharem todo esse processo e validarem se é correto. Isso é uma preocupação.* (ENT. 3)

Por outro lado, um dos entrevistados (ENT.4) mostrou-se indiferente tanto à legislação sobre IA quanto a regras internas: "*[…] Nós não temos nenhuma regra especial aplicada à inteligência artificial que não seja aplicável a qualquer outra coisa que se passe na empresa. […]"* (ENT.4).

As entrevistas revelaram diferenças significativas nas perceções sobre os perigos da IA entre programadores e gestores. Os programadores demonstraram um nível mais elevado de consciência dos riscos potenciais da IA, reconhecendo, inclusive, a existência de problemas que podem ser insolúveis com as tecnologias atuais. Em contraste, os gestores apresentaram uma visão mais focada em soluções e técnicas para reduzir os perigos da IA, mesmo que algumas dessas soluções fossem consideradas inviáveis pelos programadores.

CAPÍTULO 6 – CONCLUSÕES, LIMITAÇÕES E INVESTIGAÇÃO FUTURA

Na primeira pergunta da pesquisa, "Qual é o estado de preparação das empresas para a adoção de IA", a análise das respostas às entrevistas sobre infraestrutura revela um consenso geral quanto à sua relevância. A maioria dos entrevistados reconhece a importância do investimento em infraestrutura computacional para o treino de dados, assim como os seus custos inerentes. No entanto, há uma preferência por dados externos de outras entidades, em detrimento dos dados internos, que, embora mais controlados, exigem maior investimento e podem representar um risco. Esta preferência por dados externos é preocupante, pois os dados são a principal fonte dos riscos mencionados posteriormente.

As perceções sobre os perigos da IA variaram entre os participantes, especialmente entre aqueles que representavam empresas não tecnológicas. Estes participantes, que geralmente ocupam cargos de tomada de decisão em IA e possuem formação em gestão e física, podem ter sido influenciados pela sua formação. Demonstraram certa dificuldade em identificar alguns problemas, questionando a sua existência, e mostraram-se mais preocupados com o impacto financeiro nos clientes do que com o uso adequado da tecnologia. Esta ênfase nos custos e nas opiniões dos clientes reflete a experiência e as prioridades deste grupo, menos familiarizado com o campo tecnológico.

Em relação à regulamentação da IA, a maioria dos participantes concorda com a sua necessidade. No entanto, os entrevistados com menor familiaridade tecnológica discordam, expressando preocupação de que a regulamentação possa prejudicar o bom funcionamento da empresa. Esta divergência de opiniões pode estar relacionada ao

perfil profissional dos participantes. Aqueles com menor familiaridade com a tecnologia tendem a ser mais resistentes, enquanto os entrevistados mais tecnológicos, que defendem a regulamentação e as boas práticas, possuem formação e experiência na área de IA, permitindo-lhes uma visão mais abrangente dos desafios e benefícios desta tecnologia. A perspetiva de investigadores na área confirma a necessidade de regulamentação, evidenciando a importância do equilibrar o desenvolvimento da IA com a redução dos seus riscos.

A adoção da IA de forma ética e responsável por todas as empresas continua a ser um desafio a superar. Esta dificuldade torna-se ainda mais evidente em empresas onde a área de decisão ou o grupo responsável pela implementação da IA não possui experiência em tecnologia. Priorizar a ética da tecnologia desde o início, em vez de focar apenas nas vantagens e no lucro a curto prazo, é crucial para evitar problemas futuros.

A segunda pergunta da pesquisa, "Quais são os fatores fundamentais na adoção da tecnologia IA pelas empresas?", evidenciou, através das entrevistas, a importância de diversas boas práticas para o uso eficaz desta tecnologia. Entre essas práticas, a responsabilidade destacou-se como tema central, com os entrevistados a realçarem a necessidade de atribuir claramente a responsabilidade por eventuais falhas ou erros à empresa que produz ou implementa a tecnologia.

A qualidade dos dados utilizados em ferramentas de IA é fundamental para garantir a privacidade, a transparência, a robustez e o bom funcionamento da tecnologia. Esta questão está intimamente relacionada com a infraestrutura, dado que os dados são o principal fator para o funcionamento da IA. No entanto, há um longo caminho a

percorrer na limpeza e harmonização destes dados, especialmente quando se trata de dados externos, que muitas vezes são a principal fonte de problemas.

Embora alguns entrevistados as tenham mencionado, práticas menos óbvias, como a disponibilização das fontes dos sistemas de IA e a criação de grupos éticos para a elaboração desses programas, revelam-se ainda mais relevantes para promover a transparência, justiça e responsabilidade na IA.

Apesar da relevância dos resultados, este estudo apresenta algumas limitações. Uma das mais desafiantes foi a dificuldade em encontrar representantes de empresas que já tivessem experiência com IA. Por ser uma tecnologia relativamente nova e ainda em fase de adoção por muitas empresas, o leque de potenciais entrevistados foi consideravelmente reduzido.

Outra limitação, que também abre caminho para futuras pesquisas, reside na questão da legislação e da responsabilidade no uso de dados para o treino de sistemas de IA. A falta de clareza neste aspeto da sociedade cria dúvidas sobre a autorização e o consentimento para a utilização desses dados. Esta questão jurídica complexa exige atenção e investigação aprofundada.

Referências bibliográficas

Agar, J. (2020). What is science for? The Lighthill report on artificial intelligence reinterpreted. *The British Journal for the History of Science*, 53(3), 289 – 310. doi: https://doi.org/10.1017/s0007087420000230.

Aspers, P. e Corte, U. (2019). What Is Qualitative in Qualitative Research. *Qualitative Sociology*, 42(2), 139–160. doi: https://doi.org/10.1007/s11133-019-9413-7.

Banh, L. e Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, 33(1), 1–17. doi: https://doi.org/10.1007/s12525-023-00680-1.

Bartneck, C., Lütge, C., Wagner, A. e Welsh, S. (2020). Privacy Issues of AI. *An Introduction to Ethics in Robotics and AI*, 61–70. doi: https://doi.org/10.1007/978-3-030-51110-4_8.

Beer, D. (2016). How should we do the history of Big Data?. *Big Data & Society*, 3(1), 1–10. doi: https://doi.org/10.1177/2053951716646135.

Bell, E., Bryman, A. e Harley, B. (2022). Business Research Methods. 6ª Ed. S.L.: Oxford University Press.

Bontridder, N. e Poullet, Y. (2021). The role of artificial intelligence in disinformation. *Data & Policy*, 3, 1–21. doi: https://doi.org/10.1017/dap.2021.20.

Brewster, J. e Sadeghi, M. (2023). *Red-Teaming Finds OpenAI's ChatGPT and Google's Bard Still Spread Misinformation*. [Em Linha]. Disponível em: https://www.newsguardtech.com/special-reports/red-teaming-finds-openai-chatgpt-google-bard-still-spread-misinformation/ [Acesso em: 2023/12/26].

Brezinski, H. e Jurek, W. (2023). Editorial Introduction. *The Poznań University of Economics Review*, 9(2), 1–12. doi: https://doi.org/10.18559/ebr.2023.2.735.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y.T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M.T. e Zhang, Y. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. *arXiv*, 5, 1–154. doi: https://doi.org/10.48550/arxiv.2303.12712.

Campesato, O. (2020). *Artificial intelligence, machine learning and deep learning*, 1 ª Ed. Virginia: Mercury Learning And Information.

Carmody, J., Shringarpure, S. e Van de Venter, G. (2021). AI and privacy concerns: a smart meter case study. *Journal of Information, Communication and Ethics in Society*, 19(4), 492–505. doi: https://doi.org/10.1108/jices-04-2021-0042.

Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities and Social Sciences Communications*, 10(567), 1–12. doi: https://doi.org/10.1057/s41599-023-02079-x.

Chesterman, S. (2023). AI-Generated Content is Taking over the World. But Who Owns it?. *SSRN Electronic Journal*, 1–10. doi: http://dx.doi.org/10.2139/ssrn.4321596.

Clarke, L. (2023). Call for AI pause highlights potential dangers. Science, 380(6641), 120–121. doi: https://doi.org/10.1126/science.adi2240.

Cowan, R. (2001). Expert systems: aspects of and limitations to the codifiability of knowledge. *Research Policy*, 30(9), 1355–1372. doi: https://doi.org/10.1016/S0048-7333(01)00156-1.

Curtis, C., Gillespie, N. e Lockey, S. (2022). AI-deploying organizations are key to addressing 'perfect storm' of AI risks. *AI and Ethics*, 3, 145–153. doi: https://doi.org/10.1007/s43681-022-00163-7.

Duan, Y., Edwards, J. S. e Dwivedi, Y. K. (2019). Artificial Intelligence for Decision Making in the Era of Big Data – evolution, Challenges and Research Agenda. *International Journal of Information Management*, 48, 63–71. doi: https://doi.org/10.1016/j.ijinfomgt.2019.01.021.

Edwards, L. (2022). Regulating AI Europe: four problems and four solutions. Ada Lovelace Institute. ISBN: 978-1-7397950-0-9.

Enholm, I. M., Papagiannidis, E., Mikalef, P. e Krogstie, J. (2021). Artificial Intelligence and Business Value: a Literature Review. *Information Systems Frontiers*, 24, 1709–1734. doi: https://doi.org/10.1007/s10796-021-10186-w.

Ferrara, E. (2023). Social bot detection in the age of ChatGPT: Challenges and opportunities. *First Monday*, 28(6), 1–30. doi: https://doi.org/10.5210/fm.v28i6.13185.

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K. e Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of information technology case and application research*, 25(3), 277–304. doi: https://doi.org/10.1080/15228053.2023.2233814.

Forootan, M. M., Larki, I., Zahedi, R. e Ahmadi, A. (2022). Machine Learning and Deep Learning in Energy Systems: A Review. *Sustainability*, 14(8), 1–49. doi: https://doi.org/10.3390/su14084832.

Frank, Darius-Aurel., Jacobsen, L. F., Søndergaard, H. A. e Otterbring, T. (2023). In companies we trust: consumer adoption of artificial intelligence services and the role of trust in companies and AI autonomy. *Information Technology & People*, 36(8), 155–173. doi: https://doi.org/10.1108/itp-09-2022-0721.

Ginart, T., Zhang, E., Kwon, Y. e Zou, J. (2021). Competing AI: How does competition feedback affect machine learning?. *Proceedings of Machine Learning Research*, 130, 1693–1701.

Glasgow, J. e Browse, R. (1985). Programming languages for artificial intelligence. *Computers & Mathematics with Applications*, 11(5), 431–448. doi: https://doi.org/10.1016/0898-1221(85)90049-5.

Google (2023). *Google AI Principles*. [Em Linha]. Disponível em: https://ai.google/responsibility/principles/ [Acesso em: 2024/01/17].

Gupta, M., Akiri, C., Aryal, K., Parker, E. e Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy. *IEEE Access*, 11, 80218–80245. doi: https://doi.org/10.1109/ACCESS.2023.3300381.

Gursoy, F., Kennedy, R. e Kakadiaris, I. (2022). A Critical Assessment of the Algorithmic Accountability Act of 2022. *SSRN Electronic Journal*, 1–8. doi: https://doi.org/10.2139/ssrn.4193199.

Haase, J. e Hanel, P. H. P. (2023). Artificial muses: Generative Artificial Intelligence Chatbots Have Risen to Human-Level Creativity. *Journal of Creativity*, 33(3), 1–7. doi: https://doi.org/10.1016/j.yjoc.2023.100066.

Hadzovic, S., Mrdovic, S. e Radonjic, M. (2023). A Path Towards an Internet of Things and Artificial Intelligence Regulatory Framework. *IEEE Communications Magazine*, 61(7), 90–96. doi: https://doi.org/10.1109/mcom.002.2200373.

Hajli, N., Saeed, U., Tajvidi, M. e Shirazi, F. (2021). Social Bots and the Spread of Disinformation in Social Media: The Challenges of Artificial Intelligence. *British Journal of Management*, 33(3), 1238–1253. doi: https://doi.org/10.1111/1467-8551.12554.

Harvard Business Review (2021). *How Organizations Can Mitigate the Risks of AI* [Em linha]. Disponível em: https://hbr.org/sponsored/2021/12/how-organizations-can-mitigate-the-risks-of-ai [Acesso em: 2024/01/01].

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M. e Hussain, A. (2023). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation*, 16, 45–74. doi: https://doi.org/10.1007/s12559-023-10179-8.

Heaven, W. D. (2023). *Geoffrey Hinton tells us why he's now scared of the tech he helped build* [Em Linha]. Disponível em: https://www.technologyreview.com/2023/05/02/1072528/geoffrey-hinton-google-why-scared-ai/ [Acesso em: 2023/12/26].

Heister, S. e Yuthas, K. (2021). How Blockchain and AI Enable Personal Data Privacy and Support Cybersecurity. *Blockchain Potential in AI*, 1–14. doi: https://doi.org/10.5772/intechopen.96999.

Hill, M. (2023). Hallucinating Machines. *Te Herenga Waka*, 1–106.

Ivanova, I. (2023). *Father of ChatGPT: AI could "go quite wrong"* [Em Linha]. Disponível em: https://www.cbsnews.com/news/sam-altman-senate-chatgpt-ai-could-go-quite-wrong/ [Acesso em: 2023/12/26].

Ma, K. W. F., Dhot, T. e Raza, K. (2023). Considerations for Using Artificial Intelligence to Manage Authorized Push Payment (APP) Scams. *IEEE Engineering Management Review*, 51(3), 166–179. doi: https://doi.org/10.1109/emr.2023.3288432.

Khan, K., Ali, A., Khan, Z. e Siddiqua, H. (2021). Artificial Intelligence and Criminal Culpability. *2021 International Conference on Innovative Computing (ICIC)*, 1–7. doi: https://doi.org/10.1109/icic53490.2021.9692954.

Krakowski, S., Luger, J. e Raisch, S. (2022). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal*, 44, 1425–1452 . doi: https://doi.org/10.1002/smj.3387.

Węcel, K., Sawiński, M., Stróżyna, M., Lewoniewski, W., Księżniak, E., Stolarski, P., Abramowicz, W. (2023). Artificial intelligence—friend or foe in fake news campaigns. *Economics and Business Review*, 9(2), 41–70. doi: https://doi.org/10.18559/ebr.2023.2.736.

Lee, Y. S., Kim, T., Choi, S. e Kim, W. (2022). When does AI pay off? AI-adoption intensity, complementary investments, and R&D strategy. *Technovation*, 118, 1–47. doi: https://doi.org/10.1016/j.technovation.2022.102590.

Lütz, F. (2022). Gender equality and artificial intelligence in Europe. Addressing direct and indirect impacts of algorithms on gender-based discrimination. ERA Forum, 23, 33–52. doi: https://doi.org/10.1007/s12027-022-00709-6.

Madiega, T. (2023). Artificial intelligence act [Em Linha]. Disponível em: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf [Acesso em: 2024/01/10].

Majeed, A. e Hwang, S. O. (2023). When AI Meets Information Privacy: The Adversarial Role of AI in Data Sharing Scenario. *IEEE Access*, 11, 76177–76195. doi: https://doi.org/10.1109/ACCESS.2023.3297646.

Sanchez-Cartas, J. M. e Katsamakas, E. (2022). Artificial Intelligence, Algorithmic Competition and Market Structures. *IEEE Access*, 10, 10575–10584. doi: https://doi.org/10.1109/access.2022.3144390.

McCarthy, J., Minsky, M. L., Rochester, N. e Shannon, C. E. (1955). *A proposal for the Dartmouth summer research project on artificial intelligence*, 1 ª Ed. Hanover, NH: Dartmouth College.

McLean, S., Read, G. J. M., Thompson, J., Baber, C., Stanton, N. A. e Salmon, P. M. (2021). The risks associated with Artificial General Intelligence: A systematic review. *Journal of Experimental & Theoretical Artificial Intelligence*, 35(5), 1–17. doi: https://doi.org/10.1080/0952813x.2021.1964003.

Microsoft (2023). What is Microsoft's Approach to AI? | Microsoft Source. [Em Linha] Microsoft. Disponível em:

https://news.microsoft.com/source/features/ai/microsoft-approach-to-ai/ [Acesso em: 2024/01/10].

Mijwil, M. M., Ali, G. e Sadıkoğlu, E. (2023). The Evolving Role of Artificial Intelligence in the Future of Distance Learning: Exploring the Next Frontier. *Mesopotamian Journal of Computer Science*, 2023, 92–99. doi: https://doi.org/10.58496/MJCSC/2023/012.

Minsky, M. L. e Seymour, P. (1969). *Perceptrons*, 1ª Ed. Massachustts: MIT Press.

Mökander, J., Juneja, P., Watson, D. S. e Floridi, L. (2022). The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: what can they learn from each other?. *Minds and Machines*, 32, 751–758. doi: https://doi.org/10.1007/s11023-022-09612-y.

Monje, A., Monje, A., Hallman, R. e Cybenko, G. (2023). Being a Bad Influence on the Kids: Malware Generation in Less Than Five Minutes Using ChatGPT. *Reliability and Security (ARES '23)*, 19, 1–6. doi: https://doi.org/10.13140/RG.2.2.29391.97448.

Monteith, S., Glenn, T., Geddes, J. R., Whybrow, P. C., Achtyes, E. e Bauer, M. (2023). Artificial intelligence and increasing misinformation. *The British Journal of Psychiatry*, 224(2), 33–35. doi: https://doi.org/10.1192/bjp.2023.136.

Mustak, M., Salminen, J., Mäntymäki, M., Rahman, A. e Dwivedi, Y. K. (2023). Deepfakes: Deceptions, Mitigations, and Opportunities. *Journal of Business Research*, 154, 1–15. doi: https://doi.org/10.1016/j.jbusres.2022.113368.

Naudé, W. (2019). The Race Against the Robots and the Fallacy of the Giant Cheesecake: Immediate and Imagined Impacts of Artificial Intelligence. *IZA Discussion*, 1–32. doi: https://doi.org/10.2139/ssrn.3390207.

Nguyen, Q. P. e Vo, D. H. (2022). Artificial intelligence and unemployment:An international evidence. *Structural Change and Economic Dynamics*, 63, 40–55. doi: https://doi.org/10.1016/j.strueco.2022.09.003.

Kshetri, N. (2023). Generative Artificial Intelligence and the Economics of Effective Prompting. *IEEE Computer*, 56(12), 112–118. doi: https://doi.org/10.1109/mc.2023.3314322.

Olan, F., Arakpogun, E. O., Suklan, J., Nakpodia, F., Damij, N. e Jayawickrama, U. (2022). Artificial intelligence and knowledge sharing: Contributing factors to organizational performance. *Journal of Business Research*, 145, 605–615. doi: https://doi.org/10.1016/j.jbusres.2022.03.008.

Ore, O. e Sposato, M. (2022). Opportunities and risks of artificial intelligence in recruitment and selection. *International Journal of Organizational Analysis*, 30(6), 1771–1782. doi: https://doi.org/10.1108/ijoa-07-2020-2291.

Pal, S., Kumari B M, K., Razauddin, Kadam, S. e Saha, A. (2023). *The AI Revolution: Future Unveiled* [Em Linha]. Disponível em: https://www.iarapublication.com/books/pdf/the-ai-revolution-future-unveiled.pdf [Acesso em: 2024/01/21].

Pelley, S. (2023). *Geoffrey Hinton on the promise, risks of artificial intelligence | 60 Minutes - CBS News* [Em Linha]. Disponível em:

https://www.cbsnews.com/news/geoffrey-hinton-ai-dangers-60-minutes-transcript/ [Acesso em: 2024/01/07].

Pelley, S. (2023). *Google's AI experts on the future of artificial intelligence | 60* Minutes - CBS News [Em Linha]. Disponível em: https://www.cbsnews.com/news/google-artificial-intelligence-future-60-minutes-transcript-2023-06-11/ [Acesso em: 2023/12/26].

Raikes, J. (2023). *AI Can Be Racist: Let's Make Sure It Works For Everyone* [Em Linha]. Disponível em: https://www.forbes.com/sites/jeffraikes/2023/04/21/ai-can-be-racist-lets-make-sure-it-works-for-everyone/?sh=1005c5c12e40 [Acesso em: 2023/12/26].

Dwivedi, R. K., Nand, P. e Pal, O. (2023). Evolution of Machine Translation for Indian Regional Languages using Artificial Intelligence. *2023 International Conference on Disruptive Technologies (ICDT)*, 64–768. doi: https://doi.org/10.1109/icdt57929.2023.10150776.

Roli, A., Jaeger, J. e Kauffman, S. A. (2022). How Organisms Come to Know the World: Fundamental Limits on Artificial General Intelligence. *Frontiers in Ecology and Evolution*, 9, 1–14. doi: https://doi.org/10.3389/fevo.2021.806283.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386-408. doi: https://doi.org/10.1037/h0042519.

Sadeghi, M., Arvanitis, L., Padovese, V., Pozzi, G., Badilini, S., Vercellone, C., Roache, M., Wang, M., Brewster, J., Huet, N., Schimmel, B., Slomka,

A., Pfaller, L. e Vallee, L. (2023). *Tracking AI-enabled Misinformation: 623 'Unreliable AI-Generated News' Websites (and Counting), Plus the Top False Narratives Generated by Artificial Intelligence Tools*. [Em Linha]. Disponível em: https://www.newsguardtech.com/special-reports/ai-tracking-center/ [Acesso em: 2023/12/26].

Saldana, J. (2021). *Coding Manual For Qualitative Researchers*, 2ª Ed. S.L.: Sage Publications.

Samuel, J. (2023). Response to the March 2023 'Pause Giant AI Experiments: An Open Letter' by Yoshua Bengio, signed by Stuart Russell, Elon Musk, Steve Wozniak, Yuval Noah Harari and others…. *SSRN Electronic Journal*, 1-3. doi: https://doi.org/10.2139/ssrn.4412516.

Schlegel, D. e Uenal, Y. (2021). A Perceived Risk Perspective on Narrow Artificial Intelligence. *Pacific Asia Conference on Information Systems*, 44, 1-15.

Schuett, J. (2023). Three lines of defense against risks from AI. *AI & Society*, 1-15. doi: https://doi.org/10.1007/s00146-023-01811-0.

Sejnowski, T. J. (2023). Large Language Models and the Reverse Turing Test. *Neural Computation*, 35(3), 309–342. doi: https://doi.org/10.1162/neco_a_01563.

Setzu, M., Guidotti, R., Monreale, A., Turini, F., Pedreschi, D. e Giannotti, F. (2021). GLocalX - From Local to Global Explanations of Black Box AI Models. *Artificial Intelligence*, 294, 1–15. doi: https://doi.org/10.1016/j.artint.2021.103457.

Shahzad, H. F., Rustam, F., Flores, E. S., Mazón, J. L. V., Diez, I. T. e Ashraf, I. (2022). A Review of Image Processing Techniques for Deepfakes. *Sensors*, 22(12), 1-28. doi: https://doi.org/10.3390/s22124556.

IBM Data and AI Team (2023). Shedding light on AI bias with real world examples [Em Linha]. Disponível em: https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples/ [Acesso em: 2023/12/26].

Treleaven, P., Barnett, J., Brown, D., Bud, A., Fenoglio, E., Kerrigan, C., Koshiyama, A., Sfeir-Tait, S. e Schoernig, M. (2023). The Future of Cybercrime: AI and Emerging Technologies Are Creating a Cybercrime Tsunami. *Social Science Research Network*, 1–34. doi: https://doi.org/10.2139/ssrn.4507244.

Turchin, A. e Denkenberger, D. (2018). Classification of global catastrophic risks connected with artificial intelligence. *AI & Society*, 35, 147–163. doi: https://doi.org/10.1007/s00146-018-0845-5.

Turing, A. M. (1950). Computing Machinery and intelligence. *Mind*, 59(236), 433–460. doi: https://doi.org/10.1093/mind/LIX.236.433.

V. Rajaraman (2023). From ELIZA to ChatGPT. *Resonance*, 28, 889–905. doi: https://doi.org/10.1007/s12045-023-1620-6.

Wang, X. e Cheng, Z. (2020). Cross-Sectional Studies: Strengths, Weaknesses, and Recommendations. *Chest Journal*, 158(1), 65–71 doi: https://doi.org/10.1016/j.chest.2020.03.012.

Weitz, K., Dang, C. T. e André, E. (2023). Do We Need Explainable AI in Companies? Investigation of Challenges, Expectations, and Chances from

Employees' Perspective. *Human-Computer Interaction*, 2, 1–11. doi: https://doi.org/10.48550/arXiv.2210.03527.

Weizenbacm, J. (1966). ELIZA – A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36 – 45. doi: https://doi.org/10.1145/365153.365168.

Wilson, J. (2014). *Essentials of Business Research: A Guide to Doing Your Research Project*, 2ª Ed. Los Angeles: Sage.

Zhou, J., Zhang, Y., Luo, Q., Parker, A. G. e Munmun De Choudhury (2023). Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 426, 1-20. doi: https://doi.org/10.1145/3544548.3581318.

Zou, J. e Schiebinger, L. (2018). AI can be sexist and racist — it's time to make it fair. *Nature*, 559(7714), 324–326. doi: https://doi.org/10.1038/d41586-018-05707-8.

Zubrow, K. (2023). *How Google's 'Don't be evil' motto has evolved for the AI age | 60 Minutes - CBS News*. [Em Linha]. Disponível em: https://www.cbsnews.com/news/how-googles-dont-be-evil-motto-has-evolved-for-ai-age-60-minutes-2023-07-09/ [Acesso em: 2023/12/27].

Anexos

Anexo I – Perigos da IA

Perigo	Breve definição	Referência
Black Box	Sistema cujo funcionamento interno é difícil de compreender, ou seja, não é possível entender como o sistema chega às suas conclusões.	(Setzu et al., 2021)
Noticias falsas	A IA pode ser usada para criar conteúdo falso, como imagens, vídeo, documentos ou voz humana, de forma mais rápida e eficiente do que os humanos, e pode ser usada para atingir um público mais amplo.	(Zhou et al., 2023).
***Bots* sociais**	Programas de computador projetados para imitar o comportamento humano em plataformas de *social media*.	(Ferrara, 2023)
Alucinações	Resultados criados por sistemas de IA que não correspondem à realidade	(Monteith, 2023)
Cibercrime	Uso de inteligência artificial (IA) para cometer crimes, fraudes usando computadores, tornando o cibercrime mais difícil de detetar e combater.	(Treleaven, et al., 2023)
Roubo de informação	Bases de dados são invadidas para roubar informações pessoais, como nome, endereço, número de cartão de crédito e senhas para produzir fraudes.	(Heister et al., 2021)
Manipulação de *bots*	Permite que indivíduos mal-intencionados convençam a IA a fornecer informações de pedaços de código, que podem ser usadas para desenvolver malware de forma eficaz	(Gupta et al., 2023)
Inteligência artificial geral	Capaz de aprender sem supervisão, autoaperfeiçoar-se, alterar os seus objetivos pré-programados e ser substancialmente mais inteligente do que os humanos.	(Roli et al., 2021)

Fonte: Elaboração própria

Anexo II – Impactos da IA

Impacto	Breve definição	Referência
Infraestruturas adequadas	Infraestrutura adequada para suportar a IA, conjuntos de dados complexos, treinados e investimentos financeiros.	(Lee et al, 2022)
Funcionários adaptados	Funcionários abertos a aprender informações novas e com qualificações adequadas para trabalhar com soluções de IA.	(Enholm et al., 2022)
Mudança no funcionamento das empresas	Introdução de algumas regras e práticas éticas para a adoção, desenvolvimento e especialmente no fornecimento de aplicações ou plataformas de IA.	(Harvard Business Review, 2021)
Falta de regulamentação	Mesmo com alguns cuidados por partes das empresas, a criação de leis também ajudaria a reduzir os riscos para a indústria, a academia e a sociedade.	(Microsoft, 2023)
Competição	Á medida que as empresas utilizam mais algoritmos de IA em diferentes áreas e competem entre si para oferecer melhores serviços, a qualidade dos dados usados para ensinar esses algoritmos pode ser afetada.	(Ginart et al., 2021).
Black Box	Tecnologias de IA que não são totalmente transparentes e compreensíveis para os funcionários.	(Weitz et al., 2022).
Questões éticas no recrutamento	A IA pode criar tendências discriminatórios, como género, raça, etnia e cor, que levam ao risco de multas por práticas de recrutamento discriminatórias.	(Ore et al., 2021)

Fonte: Elaboração própria

Anexo III – Guião de entrevista

Tema	Suporte da Revisão da Literatura	Referência	Perguntas para os gestores	Perguntas para os programadores
Infraestruturas adequadas	Infraestrutura adequada para suportar a IA, conjuntos de dados complexos, treinados e investimentos financeiros.	(Lee et al., 2022)	Quais são os principais desafios que a empresa enfrenta na construção de uma infraestrutura/aplicações/ ferramentas para IA?	Que infraestrutura/ferramentas/aplicações são necessárias para ter um bom suporte para fazer ou usar de IA nas empresas?
Funcionários adaptados	Funcionários abertos a aprender informações novas e com qualificações adequadas para trabalhar com soluções de IA.	(Enholm et al., 2022)	Como a empresa está a preparar os funcionários para trabalhar com soluções de IA? Principalmente aqueles que não usam tecnologia	Acha que qualquer pessoa pode trabalhar com inteligência artificial?
Questões éticas no recrutamento	A IA pode criar tendências discriminatórios, como género, raça, etnia e cor, que levam ao risco de multas por práticas de recrutamento discriminatórias.	(Ore et al., 2021)	Como a empresa está a controlar e avaliar os seus sistemas de IA para identificar e corrigir quaisquer tendências discriminatórias?	É possível garantir que os sistemas de IA sejam justos e imparciais?
Fakenews/ DeepFakes	A IA pode ser usada para criar conteúdo falso, como imagens, vídeo, documentos ou voz humana, de forma mais rápida e eficiente do que os humanos, e pode ser usada para atingir um público mais amplo.	(Domenico et al., 2021)	A empresa está preparada para proteção das *deepfakes*?	É possível garantir a deteção e proteção de *fakenews* ou *deepfakes*?
Competição	Á medida que as empresas utilizam mais algoritmos de IA em diferentes áreas e competem entre si para oferecer melhores serviços, a qualidade dos dados usados para ensinar esses algoritmos pode ser afetada.	(Ginart et al., 2021).	Quais são os principais desafios que a empresa enfrenta na garantia da qualidade dos dados usados para treinar os seus sistemas de IA?	É possível garantir que os dados usados para competir sejam de boa qualidade?
Alucinações	Resultados criados por sistemas de IA que não correspondem à realidade	(Monteith, 2023)	É possível confiar em ferramentas IA nas empresas mesmo tendo alucinações?	A ferramentas IA com alucinações é confiável?
Manipulação de *bots* / Cibercrime / Roubo de informação	Permite que indivíduos mal-intencionados convençam a IA a fornecer informações de pedaços de código, que podem ser usadas para desenvolver *malware* de forma eficaz. Uso de inteligência artificial (IA) para , fraudes usando computadores, tornando o cibercrime mais difícil de detetar e combater. Bases de dados são invadidas para roubar informações pessoais, como nome, endereço, número de cartão de crédito e senhas para produzir fraudes.	(Gupta et al., 2023)	É possível garantir que as aplicações feitas para os utilizadores não são usadas pelas más razões?	É possível garantir que as aplicações feitas para os utilizadores não são usadas pelas más razões?
Black Box	Tecnologias de IA que não são totalmente transparentes e compreensíveis para os funcionários.	(Weitz et al., 2022).	É seguro ter aplicações no mercado para os utilizadores, mesmo não tendo uma certeza clara de como ela funciona?	É seguro ter aplicações no mercado para os utilizadores, mesmo não tendo uma certeza clara de como ela funciona?
Inteligência artificial geral	Capaz de aprender sem supervisão, autoaperfeiçoar-se, alterar os seus objetivos pré-programados e ser substancialmente mais inteligente do que os humanos.	(Roli et al., 2021)	Existe algum receio por parte da empresa em usar inteligência artificial geral?	Existe algum receio da inteligência artificial geral e o que ela pode fazer?
Falta de regulamentação	Mesmo com alguns cuidados por partes das empresas, a criação de leis também ajudaria a reduzir os riscos para a indústria, a academia e a sociedade.	(Microsoft, 2023)	Sente que é necessário existir regulamentação, para as empresas sentirem mais seguras em usar IA?	Sente que a regulamentação sobre a IA é suficiente para ser usada de forma mais segura?
Mudança no funcionamento das empresas	Introdução de algumas regras e práticas éticas para a adoção, desenvolvimento e especialmente no fornecimento de aplicações ou plataformas de IA.	(Harvard Business Review, 2021)	Quais são os principais princípios éticos que a empresa segue na adoção, desenvolvimento e fornecimento de aplicações ou plataformas de IA?	Que conselhos éticos dava às empresas de modo a usar a IA o melhor possível?

Fonte: Elaboração própria

Anexo IV – Códigos e subcódigos presentes nas entrevistas

Código	Explicação	Freq.	Excerto dos Entrevistados
Q1. Infraestruturas	Sustenta o funcionamento de sistemas e serviços.		
Preço	Custo da infraestrutura	4	(…) custa 5.000.000 de dólares, só a treinar. (ENT. 4) Depois não é barato também, esse tipo de infraestruturas (…) (ENT. 5) No caso das empresas, se elas quiserem, elas próprias criaram os modelos, então vou ter esse custo, se elas quiserem usar os modelos existentes por ofertas comerciais, (…) eu acho que até são bastante acessíveis para o benefício, cerca de 20€ por utilizador. (ENT. 6) (…) se pensarmos montar uma infraestrutura de raiz os custos são dispendiosos, em particular no que toca a treinar modelos. (ENT. 9)
Competências	Competência dos funcionários	2	(…) é garantir que as pessoas têm as competências digitais para poder tirar o uso destas ferramentas. Portanto, saibam fazer promts, saibam que tipo de ferramentas é que existem, saibam que os algoritmos são probabilísticos e que podem errar. (ENT. 2) (…) é difícil encontrar pessoas que consigam pegar nos dados, digamos assim, que estão em sistemas operacionais (…) (ENT. 5)
Dados Externos	Dados que são comprados a outras entidades	6	(…) sempre que usamos modelos externos o que conseguimos garantir aqui, aos nossos clientes é que nenhum conteúdo da conversa é usado para a aprendizagem de modelos fora do universo da empresa (…) (ENT. 1) Na realidade, hoje em dia, nós podemos usar os dados que nós temos em conjunto com dados de outras entidades (…) (ENT. 4) (…) se elas quiserem usar os modelos existentes por ofertas comerciais, como a Open AI ou da Google, (…) (ENT. 6) Já existem *datasets* públicos, (…) e agora existem mais dados, mais alguns websites que dão *open source* de dados, mas mesmo assim, esses dados estão de certa maneira "sujos", no sentido em que é preciso limpar muito desses dados. (ENT. 7) (…) as pessoas têm acesso a estes modelos numa versão gratuita, que têm uma qualidade bastante degradada, digamos assim, por comparação com os modelos pagos, porque isto tem a ver com o modelo (…) (ENT. 8) Parece que a larga maioria opta por utilizar esses serviços fornecidos pelos *big players*. (ENT. 9)
Processamento	Hardware ou Software necessários para suportar IA	8	(…) os dados, a forma como nós lidamos com os dados e interagimos com ele, precisam de ter uma capacidade de processamento (…) (ENT. 1) (…) outro é uma capacitação técnica (…) (ENT. 2) (…) é preciso uma infraestrutura computacional com muito desempenho para poder criar e fazer todo o processo, quer de aprendizagem, quer de análise, quer da própria criação de conteúdo. (ENT. 3) Data Wherehouse e a capacidade processamento (…) (ENT. 5) (…) isso envolve normalmente um grande custo, quer a nível de hardware e estamos a falar se calhar de milhares de GPUS (…) (ENT. 6) (…) obviamente nas indústrias, essa capacidade torna-se muito mais fácil se nomeadamente se tivermos numa Google ou numa open IA, eles já têm servidores e que aguentam essa quantidade de RAM de dados para conseguirem treinar. (ENT. 7) Isto exige uma capacidade de computação gigantesca, uma coisa que as pessoas não têm noção, exige CPUs potentíssimos, data centres gigantescos e que consomem uma energia (…) (ENT. 8) (…) a capacidade que nós temos de ter computacional para treinar esses modelos é de facto, bastante grande e, portanto, a despesa para manter essa infraestrutura é elevada (ENT. 9)
Dados Internos	Dados feitos pela organização	4	Sempre que possível usamos modelos desenvolvidos na empresa, ou seja, modelos em que somos nós que controlamos exclusivamente os dados que são colocados de maneira que sempre que é necessário usamos os dados de tráfego reais em que permite ter um controlo total (ENT. 1) (…) por isso estão a usar ferramentas internas e não ferramentas públicas com dados que tipicamente não são controlados ou que são públicos de Internet, nem sempre é positivo (…) (ENT. 3) (…) infraestrutura é criada pela própria empresa, nesse caso, há um investimento bastante avultado. (ENT. 6) (…) por algum requisito muito específico, existir a necessidade de fazer as coisas de raiz numa infraestrutura própria, por exemplo, sei de alguns casos em que essa restrição existe por questões de privacidade dos dados e não ser possível de treinar modelos na *cloud* com dados por questões de relacionadas com a sensibilidade dos dados. (ENT. 9)
Q2. Colaboradores	Adaptação de novas tecnologias		
Aprendizagem	Investir em formações e cursos para a aprendizagem	6	(…) conhecer melhor a tecnologia e apostar na aprendizagem do conhecimento, mas também no talento e ir buscar talento que nos permita usar melhor a tecnologia, isto naturalmente vai implicar ir buscar mais talento e olhar para as pessoas que temos atualmente na empresa e consigamos fazer com sucesso uma transição ou uma conversão das competências. (ENT. 1) (…) acho que as empresas têm de ter planos de capacitação e de adoção (…). Acho que é mesmo uma decisão de avançar e perceber como isto é estratégico para o futuro das organizações. (ENT. 2) Criar um conjunto de formações, treinos a explicar essa utilização da IA e que foram criadas um conjunto de *guidelines* para utilização da IA. (ENT. 3) Portanto, nós na universidade estamos agora preocupados em ensinar essas tecnologias. (ENT. 7) (…) é possível e desejável dar formação. Há empresas que estão atentas a estas questões e portanto, parece-me que vai ter até impacto a nível do rendimento das empresas. (ENT. 8) Acho que qualquer preparação ou necessidade de utilizar uma ferramenta nova num contexto que seja independentemente do tipo de empresa, requer sempre algum nível de formação. (ENT. 9)
Aprendizagem Autodidata	Adapta-se sozinho à nova tecnologia	4	Portanto com a inteligência artificial vai ser uma coisa semelhante, vai demorar algum tempo a ser adotada, as pessoas vão adotar e vão trabalhar (ENT. 4) (…) vão aprendendo on the job, começa a fazer coisas mais simples e depois fazer coisas mais complicadas com o tempo. (ENT. 5) (…) eventualmente toda a gente chegou lá e acho que aqui vai acontecer o mesmo. (ENT. 6) (…) eu acho que a própria utilização destas ferramentas tem caminhada a passos largos para ser amplamente adotada, ou seja, para que a sua facilidade de utilização também seja cada vez maior. (ENT. 9)
Q3. Discriminação	As aplicações de IA podem ter tendências discriminatórias e prejudicar alguns grupos		
Identificação do problema	Ter uma visão clara do problema da IA	8	(…) portanto estas questões relacionadas com viés ou até mesmo dos erros, temas relacionados com a privacidade e a segurança daquilo que a empresa tem a seu cargo, são absolutamente relevantes. (ENT. 1) (…) é colocar os produtos com salvaguardas que permitam olhar para temas de enviesamento e para temas de justiça dos algoritmos. (ENT. 2) A empresa desde o primeiro momento criou uma nota quer interna, quer pública sobre a ética e sobre a sua aplicabilidade e o seu requisito na aplicabilidade, à inteligência artificial. (ENT. 3)

Código	Explicação	Freq.	Excerto dos Entrevistados
			Nas aplicações que nós temos de recomendações, eu não estou a ver que isto se concretize (ENT. 5) (…) ainda hoje os próprios cientistas que trabalham nisto há anos e anos, não conseguem perceber porque é que acontece essas alucinações. (ENT. 6) (…) mas pode sim acontecer, obviamente, porque nós não temos o controle 100%, porque não temos recursos humanos para conseguir limpar essa informação. (ENT. 7) Eu penso que é uma questão extremamente difícil porque a discriminação não é só a nível, por exemplo, das imagens que produzem. (ENT. 8) Dentro da inteligência artificial, há um chapéu muito grande e dentro desse chapéu, há muitas subáreas que atacam diferentes problemas dentro da inteligência artificial. Uma das subáreas, que inclusive é uma das áreas muito quentes, com muita atividade de investigação e com muito interesse por parte da comunidade é a área do *fairness* que vem exatamente desta questão da IA ser justa. (ENT. 9)
Validação de informação	Garantir que os sistemas operem com base em dados precisos e confiáveis, proporcionando resultados confiáveis.	6	(…) temos um conjunto de especialistas que se juntaram para avaliar estes riscos especificamente, (…) mas esta parcialidade ou viés nos dados de treinos possam ser assegurados. (ENT. 1) (…) definir salvaguardas dos produtos e dos serviços para tentar limitar esses erros que existem nos dados e é por isso que hoje em dia quando falamos com o Chat GPT, existem várias coisas que ele estes modelos já não respondem porque existem filtros de moderação, de alguma forma que tentam perceber se aquilo é conteúdo ofensivo. (ENT. 2) (…) temos também uma ferramenta que nos ajuda a validar se a informação que realmente é depois trocada, se também está de acordo com essa política. (ENT. 3) (…) há provedores de clientes que tipicamente até se o tema acontecesse de forma sustentável, seria uma intervenção deles (…) (ENT. 5) Dependendo da resposta que ele dá, antes de enviar isso para um utilizador, conseguirmos limpar à mão de uma forma de engenharia essa resposta (…) (ENT. 7) (…) eu próprio tenho que ser cuidadoso o suficiente e perceber que se calhar tenho que adotar algumas destas estratégias que vão ajudar a que o modelo seja o mais *fair* possível e tenho que tentar monitorizar o modelo para perceber se isso está a ser conseguido ou não (ENT. 9)
Questionamento	Ponderar o valor, a confiabilidade e a relevância de informações, ideias e argumentos.	2	(…) o pensamento crítico e pensamento analítico e portanto, a esta nova onda de inteligência artificial legislativa não tirou o humano do *loop*, pelo contrário, coloca com a responsabilidade final de decidir e perceber e para isso é preciso cultivar o sentido crítico do pensamento analítico. (ENT. 2) Não é só uma questão de segurança e privilégio das pessoas à informação, é também a tal ideia de analisarmos a resposta da IA e percebermos se é uma resposta que é condizente com o seu propósito, (…) (ENT. 3)
Desafios	Dificuldade significativa em encontrar uma resposta eficaz para o problema.	4	(…) eu tenho algum receio que não seja possível controlar estes comportamentos estranhos, que às vezes nós vemos nos LLMs e no IA, que alguns deles têm a ver com precisamente com esses vieses, que existem nos dados que deram origem àquele modelo. (ENT. 6) Nós não conseguimos limpar todas elas a tempo útil, portanto, nós assumimos já à priori, que estamos a alimentar uma rede com dados "sujos". Vai sempre haver a possibilidade (…), porque nós não temos o controle 100%, porque não temos recursos humanos para conseguir limpar essa informação. (ENT. 7) Eu penso que é uma questão que, para já, não se vai resolver e isso implica na formação que se dá às pessoas, elas terem a perceção de que esta tecnologia tem estas falhas e que são falhas graves, mas refletem a humanidade. (ENT. 8) (…) resolver na essência da palavra que é fazer com que o problema deixe de existir é muito difícil porque o 100% não existe, mas há claramente estratégias para mitigar esses problemas e mais uma vez é uma área que tem tido muito trabalho nos últimos anos. (ENT. 9)
Q4. DeepFakes	Manipulados de dados usando inteligência artificial para substituir informação real.		
Identificação do problema	Ter uma visão clara do problema da IA.	9	Nós do ponto de vista de deepfakes que não estão relacionados com estas questões e que têm mais a ver com temas de fraude, nós também estamos preocupados com este acompanhamento muito personalizado em todos os projetos. (ENT. 1) Acho que hoje em dia esse é realmente um risco que existe, (…) (ENT. 2) Bem as soluções de IA que utilizamos estão disponíveis internamente, não permitem esse tipo de utilização (ENT. 3) (…) isso não é novidade, sempre foi com ou sem inteligência artificial (…) (ENT. 4) Sim, sim, pois, (…) (ENT. 5) Eu acho que é um problema bastante complicado porque se há uns anos as *fake news* e esse tipo de informação falsa, desde que apareceram praticamente as redes sociais, (…) (ENT. 6) (…) é um risco em que vai chegar a uma altura em que os modelos estão tão bem-criados que vai ser muito difícil nós conseguirmos quase distinguir o que é real ou não. (ENT. 7) É difícil identificar uma notícia verdadeira de uma falsa e nós já estamos no ponto em que (…) neste momento nós não conseguirmos resolver isso do lado de útil, (…) (ENT. 8) (…) não é um problema que está resolvido nem pouco, nem mais ou menos, mas há trabalho a ser feito nesse domínio (…) (ENT. 9)
Desconfiança	Falta de confiabilidade, honestidade ou integridade em algo.	3	A minha empresa, não tem exposição a esse risco, mas aquilo que acontece na sociedade é que as pessoas já estão desconfiadas, portanto vão criar notícias falsas (ENT. 4) (…) temos que desconfiar de tudo neste momento (…) (ENT. 8) Aquilo que se pode fazer é desenvolver estratégias que permitam distinguir aquilo que é fake daquilo que é real, (…) (ENT. 9)
Averiguar informação	Verificar a veracidade das informações	3	As nossas equipas que fazem basicamente a auscultação do que se passa na comunicação social e nas redes sociais que é onde este tipo de fenómenos acontecem com mais frequência. Nós temos uma equipa que faz esse trabalho de forma mais automatizadas (…) (ENT. 1) Acho que fazemos alguma monitorização relativamente próxima das redes sociais (…) (ENT. 5) (…) mesmo que se arranjem humanos para verificar essa informação, nunca vamos arranjar humanos suficientes para monitorizar toda a informação que uma IA consegue produzir, (…) (ENT. 6)
Sinalização do texto	Adicionar informações explicativas e interpretativas aos modelos de inteligência artificial, permitindo uma melhor compreensão dos dados que foram utilizados nos modelos.	2	(…) por exemplo, se usares o Copilot até hoje em dia é sempre dito que aquela informação é gerada por inteligência artificial e, portanto, ter sempre essa indicação (…) (ENT. 2) (…) na possibilidade de colocar uma marca de água nos vídeos ou nas imagens que são produzidas pela inteligência artificial (…) (ENT.8)
Q4. Qualidade dos dados	Integridade, consistência e relevância das informações armazenadas em sistemas e base de dados essenciais para o funcionamento eficiente de empresas e organizações		
Identificação do problema	Ter uma visão clara do problema da IA	7	(…) é muitas vezes difícil de identificar o erro porque eles estão a reproduzir padrões lógicos, (…) torna-se muito evidente que isto é um obstáculo. (ENT. 1) (…) eles basicamente usaram toda a informação pública que é visto pela Internet e essa informação tem os problemas que a nossa sociedade tem de enviesamento, injustiça, de conteúdo ofensivo e portanto há esse problema na qualidade dos dados. (ENT. 2) (…) nós focamos para que, a partir desses dados ou dessa informação, cada vez a informação seja melhor e não pior (…) (ENT. 3) (…) eu desconfio que vai ser muito difícil garantir que os dados vêm sempre com boa qualidade (…) (ENT. 6)

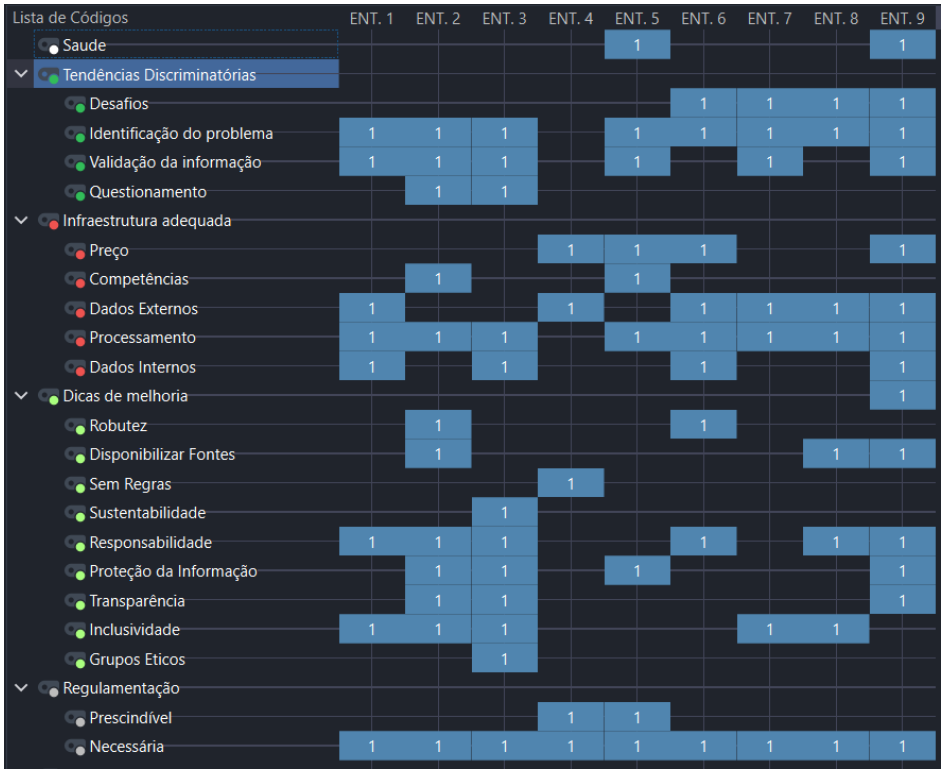
Código	Explicação	Freq.	Excerto dos Entrevistados
			(…) vão estar só a treinar apenas para um determinado assunto e já não vão ser capazes de olhar mais para isso, porque só estão, viciados com aqueles dados e vai ser um problema enorme quando as empresas apenas usarem os dados sintéticas e realmente não tirarem dados realistas. (ENT. 7) Os próprios dados começaram já a ser utilizados pelo Google em 2016, quando começou, as primeiras bases de dados, baseiam-se em dados muito fracos, nomeadamente, fóruns da Internet, que são bastante preconceituosos (…) (ENT. 8) Se os dados de entrada que são utilizados para utilizar o modelo forem maus, o modelo vai ser mau. Nós temos obviamente de tentar maximizar a qualidade dos dados que utilizamos para treinar o modelo. (ENT. 9)
Não identificação do problema	Não têm uma visão clara do problema da IA	2	Nós não temos esse problema, naquilo que é muito grave, naquilo que é positivo, nas coisas que fazemos todos os dias, isso não nos preocupa, eu acho que até é positivo. (ENT. 4) Eu acho que a questão não é bem assim, não acho que os dados necessariamente fiquem piores com o tempo. (ENT. 5)
Dados Próprios	Dados produzidos pela empresa	2	(…) que é treinar ou refinar os modelos com dados próprios, ou seja, onde a organização tem controlo sobre os dados em que o modelo linguístico está a responder (…) (ENT.2) Portanto, está-se a chegar a um limite em que os modelos não têm dados novos, eles não conseguem desenvolver novas capacidades. A única maneira é de criar dados realistas. (ENT. 7)
Utilização de outros programas	Verificar a informação da IA em conjunto com outros programas sem IA	3	(…) ter um sistema central que faz um controlo mais científico do tema e usar os LLMs com essa perspetiva é naturalmente uma das formas de podermos controlar este risco que falávamos aqui (…) (ENT. 1) (…) é tomar as decisões apropriadas para fazer esta limpeza (…) (ENT. 5) (…) há, de facto, pessoas que são pagas para fazerem a validação de um de marcação de informação nas bases de dados, mas terá de haver sempre pessoas que façam esta validação, que validem a informação das bases de dados (ENT. 8)
Aceitação do erro	Assumir erros nas aplicações	2	Tentarmos ter algum conforto e depois de vermos que o nível de erro é residual, podemos aceitar determinado erro e depois tomarmos as nossas decisões (ENT. 1) (…) nós como estamos numa empresa que somos treinados para ter primeiro a física da coisa e depois dizer se os dados são válidos ou não. (ENT. 4)
Validação dos dados	Validar a veracidade das informações	4	(…) e por isso é são feitos muitos destes filtros de moderação que são modelos de *machine learning* em cima dos modelos dos LLMs (…) (ENT.2) (…) mas também existe um processo de validação contínua não só feito pela IA, mas também feito pela larga utilização que depois é feita pelas pessoas (…) (ENT. 3) (…) os dados deviam passar por um processo de curadoria, apenas os dados com qualidade é que deviam entrar para os modelos (…) (ENT. 6) Isso passa muito por uma questão de monitorização, ou seja, o modelo quando está a ser utilizado num contexto real, ele deve ser monitorizado com uma determinada periodicidade e nessa monitorização, nós avaliamos uma série de questões, em particular, questões relacionadas com alterações significativas quer aos dados, quer aos conceitos de dados. (ENT. 9)
Q5. Manipulação	Uso indevido de inteligência artificial para criar e controlar *bots* automatizados que executam ações online com o objetivo de prejudicar pessoas ou sistemas.		
Identificação do problema	Ter uma visão clara do problema da IA	7	Então diria que sim. hoje em dia, as nossas soluções são robustas nessa perspetiva, porque desde cedo temos essa preocupação, desde o desenho do processo do projeto, a conceção até o desenvolvimento e avaliação (…) mas não fazemos compromissos em temas que ponham em causa esta confiança que os clientes têm em nós, naturalmente. (ENT. 1) (…) eu acho que temos que estar preparados enquanto sociedade para perceber que estes algoritmos podem alucinar e que são algoritmos probabilísticos, não são algoritmos determinísticos (…) (ENT. 2) (…) e esse é um desafio, mas é um desafio alcançável, mas continuo, ou seja, sempre de trabalho constante de melhoria das soluções. Claramente é um problema (…) (ENT. 3) Nunca se vai conseguir segurança a 100% que não há ataques (…) (ENT. 6) Nós nunca conseguimos garantir 100% isso e quem está a dizer que consegue naturalmente deve estar a mentir (…) (ENT. 7) Eu não estou a ver como é que isso seja possível, aliás, é comum aparecerem notícias de estratégias de como conseguir manipular o algoritmo, há sempre maneira de contornar isso (…) (ENT. 8) (…) essa questão de nós os conseguirmos ludibriar é de facto possível e por muito que se tente proteger o modelo, haverá sempre uma forma de conseguir dar a volta a isso, porque, de facto, aqueles modelos não olham para aquilo que nós estamos a dizer de uma perspetiva crítica com o raciocínio crítico. (ENT. 9)
Não identificação do problema	Não têm uma visão clara do problema da IA	1	Eu sei que se fizermos umas coisas de forma que o Chat GPT faça uma coisa racista qualquer, está bem, mas isso foi para a pessoa que está a forçar que ele dissesse isso. (ENT. 4)
Testar apps	Avaliar aplicações de inteligência artificial para garantir que eles satisfazem os requisitos	3	implica que todos esses temas relacionados com testes, com vulnerabilidades, encontrar os diferentes cenários de fraude, de extração de informação possível sejam endereçados com medidas muito concretas (…) e sempre que não é possível, nós decidimos não lançar como uma solução final. (ENT. 1) uma das técnicas que hoje em dia se faz muitos chama-se *rating*, que não é mais do que ter equipas a tentar fazer o *break* destes algoritmos, a tentar encontrar pontos de vulnerabilidade. (ENT. 2) (…) arranjam-se mecanismos de perceber que o sistema foi atacado e tenta-se que da próxima vez, esse ataque já não tenha resultado e é assim que vai evoluindo. (ENT. 6)
Internet	Tanto a IA quanto a Internet podem ser ferramentas para a propagação de desinformação e manipulação da opinião pública.	4	Desde que as pessoas sejam desconfiadas à partida das fontes, não interessa o que aquilo que é dito. Só o facto de haver uma aplicação que produz uma sequência de texto que pode estar certo ou pode estar errado, acho que não é diferente da internet, não vejo aí nenhuma questão problemática. (ENT. 4) (…) da mesma forma que desde os sistemas informáticos, nunca se conseguiu garantir que há segurança a 100% em qualquer sistema informático. (ENT. 6) Isso é como a Internet em geral, que tem um uso para o bem, mas também tem uso para tudo o que é de atividades ilegais da parte do ser humano. (ENT. 8) (…) ela pode sempre ser utilizada para o bem e para o mal. Quer dizer, isso acontece com a Internet, isso acontece com qualquer ferramenta tecnológica que tenha sido desenvolvida desde sempre. (ENT. 9)
Validação de informação	Validar a veracidade das informações	2	Foi uma preocupação desde o ponto inicial, por isso é que nós também não libertámos uma plataforma de IA ou sequer a aprovámos a utilização de plataformas externas da IA para efeito profissional até termos este contexto definido, portanto, de validação da informação, se a informação era fidedigna, se está a ser utilizada e se estão a usar as fontes certas para a informação que queremos obter, (…) (ENT. 3) Todas as respostas estão a ser enviadas pelo Chat GPT antes de mostrar ao utilizador, ela passa por uma caixa preta e essa caixa valida se aquela informação não é discriminatória, (…) (ENT. 7)
Q6. Black Box	Modelo de IA complexo onde o processo interno de tomada de decisão é difícil ou impossível de entender para os humanos.		
Identificação do problema	Ter uma visão clara do problema da IA	8	(…) acho que criou-se uma ideia muito errada do que é uma caixa negra, com respostas certas para todos os problemas, longe de ser isso, (…) (ENT. 1) Eu diria que sim, com alguns pressupostos. No caso do Chat GPT ou do Gpt 4, nós sabemos quais são os dados que foram usados genericamente, mas na verdade, aquele algoritmo é uma caixa preta, portanto uma Black Box. (ENT. 2) A inovação, traz sempre desafios que inicialmente não existiam e é importante nós seguirmos sempre com os compromissos éticos muito vincados, mas não impedir o desenvolvimento tecnológico, seria sempre um erro ou até impossível. (ENT. 3) Não ser possível explicar, não é *clean ability*, (…) (ENT. 5) Atualmente, não é possível perceber porque é que os LLMs dão certas respostas, (…) (ENT. 6) Elas prometem coisas que obviamente não o fazem, que não conseguem fazer pelas suas limitações. (ENT. 7)

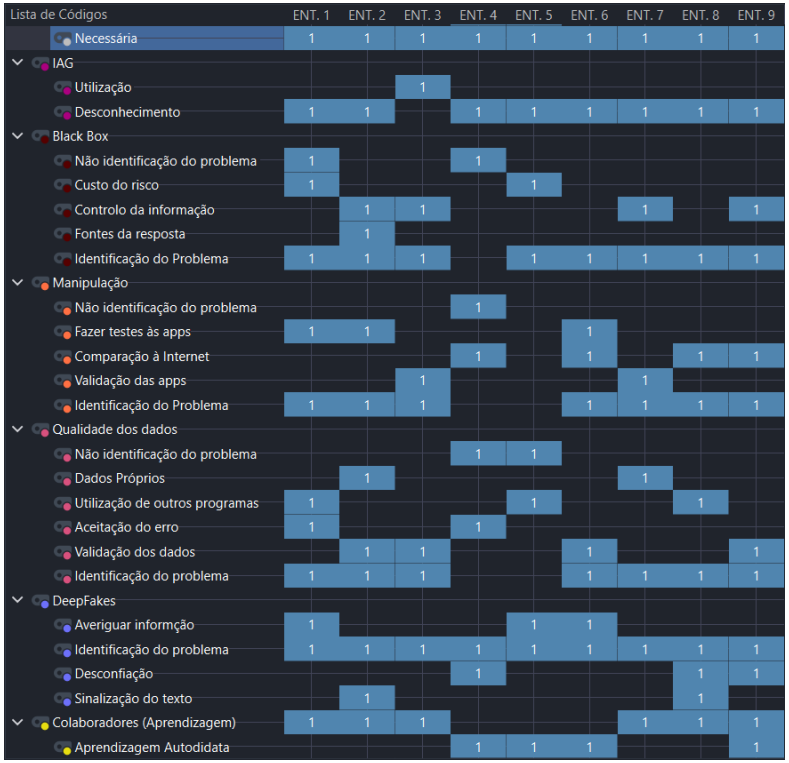
Código	Explicação	Freq.	Excerto dos Entrevistados
			Eu acho que aquilo que me parece é que toda a tecnologia tem aspetos positivos e negativos e agora tudo depende da perspetiva que os seres humanos dão a tecnologia. (ENT. 8) (…) de facto os modelos que nós temos, em particular os modelos *deep learning*, que estão na base da tecnologia, são modelos com resultados muito bons, mas não há uma capacidade de perceber o racional, (…) (ENT. 9)
Não identificação do problema	Não têm uma visão clara do problema da IA	2	(…) portanto, não concordaria muito com a afirmação de que achamos que não há uma compreensão clara do funcionamento (…) (ENT. 1) Não é bem assim, ou seja, tem mais a ver com limitações do cérebro humano do que propriamente com a forma como aquilo trabalha, (…) (ENT. 4)
Custo do risco	A quantificação das perdas financeiras potenciais que podem ser incorridas devido a eventos de inteligência artificial adversos.	2	(…) nós gerimos os riscos de uma forma muito controlada e muito consciente. Primeiro tentamos assegurar, e não colocamos disponíveis soluções cujo funcionamento fique aplicado determinado tipo de riscos e minimizando esses riscos de uma forma muito clara. Portanto, diria que a gestão de risco existe em qualquer projeto e naturalmente que nestes projetos, apesar destas componentes diferentes, ele é feito de uma forma estruturalmente diferente (…) (ENT. 1) A primeira coisa que eu acho é que a principal questão da Black Box é qual é o custo das previsões falhadas e quanto custa se as previsões falhadas começam a ser muitas (…) (ENT. 5)
Controlo da informação	Medidas implementadas para gerir e assegurar a qualidade e disponibilidade da informação.	4	(…) ele vai enviar as fontes de informação para poderemos ver e isso é muito importante, precisamente para isso, para conseguir dar alguma transparência de onde veio aquela resposta efetivamente. (ENT. 2) (…) nós conseguimos ter noção de quanto é que cada pessoa usa a inteligência oficial e para quê, com que tipo de dados é que são consultados e para que são utilizados, que permite-nos criar um mapeamento muito forte sobre isso, o que nos deixa de alguma forma perceber do prejuízo versus benefício (ENT. 3) (…) ele é um modelo que apenas gere informação e essa informação pode ser fidedigna ou não, porque ele está a aprender com a Internet. As pessoas quando escrevem nos fóruns podem responder de uma forma acertada ou não e ele aprende com essa informação basicamente. (ENT. 7) Eventualmente conseguimos validar se aquele resultado está correto ou não, (…) (ENT. 9)
Fontes de resposta		3	Eu não consigo perceber nem fundamentar, tenho que ir validar por mim de fato se alguma daquela resposta está errada, mas isso apanha o tema da transparência, (…) é providenciar as fontes de informação sempre que houver uma resposta. (ENT. 2)
Q7. Inteligência Artificial Geral	Modelo de IA complexo onde o processo interno de tomada de decisão é difícil ou impossível de entender para os humanos.		
Utilização		1	Nós acreditamos que a utilizam responsável do IA é importante, portanto não vemos isso como uma barreira para o seu desenvolvimento, para a continuação de evolução pelas pessoas e até agora tem os resultados da sua utilização que nós medimos internamente têm sido positivos. (ENT. 3)
Desconhecimento		8	Abordagem é uma coisa da qual nós não abdicamos na empresa. Portanto, a resposta sobre este tema de inteligência artificial mais geral, digamos assim que tem este tipo de implicações, vai sempre passar por uma avaliação do risco (…) (ENT. 1) Eu acho que nós ainda não sabemos o que é essa inteligência artificial. (…) Dito isto, ainda estamos muito longe de tentar ter algo que funcione como os humanos funcionam e não sei se vai demorar 3 anos, 5 anos ou 20 anos, mas eu acho que temos que nos preparar para essa possibilidade ou pelo menos para ficarmos mais próximos. (ENT. 2) (…) a inteligência artificial geral, pois é uma abstração, ninguém sabe exatamente como é que funciona o cérebro humano, quanto mais andar a perceber coisas desse género (…) (ENT. 4) (…) ainda não existe hoje, portanto, acho que é uma pergunta demasiado hipotética para estar-lhe a dizer aqui o que é que seja. (ENT. 5) Acho que a inteligência artificial geral ainda vai demorar muitos anos. Se conseguir, há muitos desafios pela frente e eu acho que nós não devemos ter medo de enfrentar esses desafios. (ENT. 6) (…) eu não sei se o mundo atualmente está preparado para isso, porque ainda se desconhece muita coisa. (ENT. 7) Eu penso que sim, pode haver e a resposta é a mesma de há pouco, só que já estamos noutro nível do jogo, digamos assim, (ENT. 8) (…) porque nós ainda não estamos, parece-me, e acho que é um consenso mais ou menos aceite pela comunidade, ainda não estamos assim tão perto de ter essa inteligência artificial geral. (ENT. 9)
Q8. Regulamentação	Processo de estabelecer regras e padrões que visam controlar o comportamento de indivíduos, empresas e outras entidades na área da inteligência artificial.		
Prescindível	Lei ou regulamentação que pode ser dispensado ou que não é essencial.	2	Daquilo que se prevê que possa ser eu sou absolutamente contra, até porque já passei pela regulamentação da banca e eu sei que a regulamento traz asneira e a tornar as coisas mais difíceis e mais complicadas para toda a gente. (ENT. 4) (…) dentro da empresa, do ponto de vista da empresa acho que não (…), acho que a questão é mais social que empresarial. (ENT. 5)
Necessário	Lei ou Regulamentação que é essencial e fundamental.	9	Acho que a regulamentação é sempre uma coisa bem-vinda e acho que é principalmente em áreas em que existe algum desconhecimento (…) e sabemos que essas preocupações serão sempre muito maiores do que aquelas que foram os requisitos da legislação sobre este tema e da regulamentação. (ENT. 1) (…) é importante regulamentar sem travar a inovação. Mas com uma perspetiva de inovação responsável e orientada aos riscos da tecnologia e não ir ao detalhe da tecnologia, mas sim o que é que pode vir pela utilização da tecnologia (…) Portanto, acho que são iniciativas muito relevantes, em que já existem no mercado, a IA tem aqui vários meses de trabalho (…) (ENT. 2) Acho que até um certo ponto, a regulamentação é necessária e deverá existir. Pessoalmente o que eu vejo é que o ritmo da regulamentação é diferente do ritmo do desenvolvimento tecnológico e portanto, é difícil de chegar a um ponto onde, ao nível do que já vai hoje em dia a tecnologia, a regulamentação estar a par e não é o caso (…) as entidades que criam esses modelos de inteligência artificial, têm alguma forma transpor em esses princípios, mais abrangente possível todos esses princípios éticos para dentro do IA. (ENT. 3) O Chat GPT deu uma frase que o autor, tinha no livro dele e ninguém vai pagar sobre aquilo que ele escreveu, é uma questão importante e isso tem que ser regulamentado. (ENT. 4) (…) se eu perguntasse do ponto de vista do consumidor, eu diria que eventualmente sim (…) (ENT. 5) Sim, acredito. É isso que tem acontecido em todas as áreas de tecnologia e nesta também tem que acontecer. (ENT. 6) Sem dúvida, (…) Acho que obviamente deveriam ser criadas essas leis para proibir esse tipo de ações, porque já estamos a criar algo em que nós conseguimos meter supostamente a fazer com que essa pessoa faça coisas que ela nunca fez (…) (ENT. 7) Portanto, é preciso legislação, sem dúvida, mas não no sentido de limitadora, mas no sentido de orientadora, no sentido de quase linhas orientadoras de como utilizar isto para que todos estejamos minimamente confortáveis. (ENT. 8) (…) parece que sim, que vai ter que surgir agora, onde exatamente, acho que isso não é claro e há muito trabalho para ser feito e vai ter que misturar gente técnica com gente não técnica, (…) (ENT. 9)
Q9. Dicas de melhoria	Sugestões, ideias ou recomendações com o objetivo de aperfeiçoar o uso da inteligência artificial.		
Robustez	Capacidade de um sistema de inteligência artificial resistir a	2	(…) a robustez destes algoritmos (…) porque essa robustez é o que nos permite ter mais confiança na sua utilização e, portanto, quando falamos de *red team* e quando falamos de testes, quando falamos de filtros de moderação, tudo isto são camadas que devem ser adicionadas em cima destes modelos de inteligência artificial generativa, para que possam ser mais robustos e possam lidar

Código	Explicação	Freq.	Excerto dos Entrevistados
	perturbações externas sem sofrer danos significativos ou perder sua funcionalidade.		com todos os riscos (…) (ENT. 2) que é as próprias pessoas votarem ou darem feedback sobre as respostas que obtêm. A resposta relativa se for qualidade, então vão dar muitas pessoas vão dar um feedback positivo. Se a resposta não tiver qualidade, muitas pessoas vão dar um feedback negativo e o modelo vai aprendendo com isso (…) (ENT. 6)
Fontes Reveladas	Revelar as fontes que deram origem a um texto, imagem ou vídeo.	3	(…) se usares o copilot, é sempre dito que aquela informação gerada por inteligência artificial e, portanto, ter sempre essa indicação (…) (ENT. 2) Mas sempre que a tecnologia for utilizada, tem de haver lá um emblema, um aviso e isto devia ser obrigatório (ENT. 8) (…) ao nível da questão dos direitos de autor, mas ao nível da IA (…) (ENT. 9)
Sem regras		1	Nós não temos nenhuma regra especial aplicada à inteligência artificial que não seja aplicável a qualquer outra coisa que se passe na empresa. (ENT. 4)
Sustentabilidade	Capacidade de persistir e se renovar ao longo do tempo, sem comprometer nada.	1	(…) outro também é sustentabilidade. (ENT. 3)
Responsabilidade	Obrigação de responder pelos atos da inteligência artificial.	6	(…) a contribuição para essencialmente utilização da tecnologia no sentido positivo (…) (ENT. 1) (…) ter a responsabilidade sobre quem produz essa informação e de quem vê essa informação (…) (ENT. 2) (…) também temos que ter responsabilidade não só profissional, mas também social, porque muitas das nossas soluções têm impacto e promovem o suportam de serviços críticos das sociedades mundiais. (ENT. 3) (…) nós somos os responsáveis por aquilo que fazemos com a informação que obtemos, nunca vamos poder dizer, "Ah, mas eu fiz isto, porque a IA disse para fazer isto". Não, a IA sugeriu, e eu agora aceito ou não aceito o que a IA me diz (…) (ENT. 6) (…) depois a questão da identificação quando ela é utilizada e não utilizar esta tecnologia para coisas ilegais, como é evidente. (ENT. 8) (…) desta responsabilidade sobre o que é que o modelo faz (…) (ENT. 9)
Privacidade	Controlar o acesso às informações pessoais e decidir como elas são usadas.	4	O Tema da privacidade e segurança de dados, e este é fundamental, especialmente quando nós vimos, é muitos utilizadores quando começaram a usar o Chat GPT que partilhavam dados pessoais e, portanto, o tema da privacidade e da segurança é um tema fundamental (…) (ENT. 2) A ideia da informação e da proteção da informação de quem é o dono da informação que nós estamos a utilizar, (…) (ENT. 3) A coisa que a gente leva muito a sério, tem a ver tudo com privacidade, minimizamos os dados, mais cedo é possível. (ENT. 5) (…) a questão da privacidade das pessoas, (…) (ENT. 9)
Transparência	Abertura em relação à informação sobre os perigos da inteligência artificial.	3	O tema da transparência que já falámos, perceber como é que estes algoritmos que são caixas pretas, ter noção de qual é a informação de base e para que possamos fazer *fact check* dessa informação. (ENT. 2) (…) tidos como conta no princípio da transparência, (…) (…) questões relacionadas com a parte da justiça, garantir ao máximo possível que o resultado e que aquilo que o modelo produz é o resultado em função dos dados que são apresentados (…) (ENT. 9)
Inclusividade	Incluir e valorizar todas as pessoas.	5	(…) temas de inclusão, (…) (ENT. 1) A parte da inclusão, essa é também uma parte muito relevante, não deixar ninguém para trás, garantir que aquilo que nós estamos a fazer não é apenas para uma parte muito reduzida da população, todos vão conseguir usar esta tecnologia de forma responsável, (…) (ENT. 2) Acima de tudo, de inclusividade, (…) (ENT. 3) Portanto, qualquer pessoa consegue agora e as empresas conseguem implementar isso nas suas infraestruturas, conseguem ter os grandes resultados que estão a ser obtidos atualmente na comunidade científica. (ENT. 7) Depois há outras questões se a empresa quiser adotar esta tecnologia tem de formar os seus funcionários, (…) (ENT. 8)
Grupos éticos	Equipas formados por especialistas em ética que se reúnem para analisar questões éticas complexas sobre inteligência artificial.	1	Preocupa-me quando entidades que gerem este tipo de tecnologia dispensam as equipas éticas para acompanharem todo esse processo e validarem se é correto. Isso é uma preocupação. (ENT. 3)
Saúde	Usar inteligência artificial para questões de saúde ainda não é confortável.		(…) se tiver uma sequência de recomendações do género, que você tem um cancro, já tem outro impacto, (…) (ENT. 5) tem um modelo que vai assistir um médico na tomada de decisão, aí a questão já é mais complexa, (…) a falta da lógica e do pontual de perceber porque é que se chega àquela decisão, pode, de facto ser um problema, (…) (ENT. 9)

Fonte: Elaboração própria

Anexo V – Matriz de Códigos





Fonte: MAXQDA

Anexo VI – Resultado das entrevistas

Tema	Subtema	Resultado	Entrevistados que mencionaram
Infraestrutura e Funcionários Adequados	Processamento computacional	Essencial para treino e funcionamento de modelos de IA. Exige investimento significativo.	ENT1, ENT2, ENT3, ENT5, ENT6, ENT7, ENT8, ENT9
	Modelos Pré-existentes	Alternativa para minimizar custos, mas com riscos de dados externos e vieses.	ENT1, ENT4, ENT6, ENT7, ENT8, ENT9
	Dados internos	Permite um controlo total da empresa sobre os dados.	ENT1, ENT3, ENT6, ENT9
	Aptidão dos Funcionários	Essencial para *prompt* de modelos, compreensão de algoritmos probabilísticos e conhecimento de riscos da IA.	ENT1, ENT2, ENT.3, ENT.7, ENT.8, ENT.9 (formações) ENT.4, ENT.5, ENT.6, ENT.9 (autoaprendizagem)
Perigos da IA	Identificação dos problemas	A adoção da IA apresenta benefícios promissores, mas também exige atenção aos perigos potenciais, como vieses algorítmicos, *deepfakes*, manipulação de dados e falta de transparência, para garantir um uso responsável dessa tecnologia.	ENT. 1, ENT.2, ENT.3, ENT.5, ENT.6, ENT.7, ENT.8, ENT.9
		A empolgação com a IA é compreensível, mas ignorar seus perigos potenciais, pode levar a consequências indesejadas.	ENT. 4 ENT.1, ENT.5 (mostram algumas dúvidas na identificação dos problemas)
	Discriminação	Origem em vieses nos dados históricos. Sugestões: validação dos dados e o pensamento critico.	ENT.1, ENT.2, ENT.3, ENT.5, ENT.7, ENT.9
		Dificuldades em combater a discriminação na IA.	ENT.6, ENT.7, ENT.8, ENT.9
	Deepfakes	Preocupação com desinformação. Sugestões: regulamentação de plataformas e sinalização de conteúdo criado pela IA.	ENT.1, ENT.2, ENT.4, ENT5, ENT6, ENT.8, ENT.9
	Qualidade dos Dados	Impactada pela procura por dados a baixo custo. Sugestões: Criação de dados próprios e validação com e sem IA.	ENT2, ENT.3, ENT.5, ENT6, ENT.7, ENT.8, ENT.9 ENT.1, ENT.4, ENT.5 (algumas dúvidas na identificação do problema e custo risco)
	Manipulação de Ferramentas de IA	Risco presente. Sugestões: testes de manipulação e validação de dados.	ENT.1, ENT.2, ENT.3, ENT.6, ENT.7 ENT.4, ENT.6, ENT.8, ENT.9 (comparação com a internet)
	Black Box	Não impede uso da IA, mas levanta necessidade de controlo da informação e compartilhamento de fontes.	ENT.2, ENT.3, ENT.7, ENT.9 ENT.1, ENT.5 (custo do risco)
	IA Geral	Vista como distante e incerta, exceto por um participante que utiliza Chat GPT 4.	Todos exceto ENT.3
Uso Responsável da IA	Regulamentação da IA	Necessária, mas com diferentes visões sobre a melhor abordagem.	Todos ENT.4, ENT.5 (não estão totalmente de acordo com a necessidade de regulamentação)
Boas praticas da IA	Disponibilização das fontes	Tornar a fonte e os dados utilizados no desenvolvimento de sistemas de IA acessíveis ao público.	ENT.2, ENT.8, ENT.9
	Robustez	Capacidade dos sistemas de IA de lidar com situações inesperadas e dados maliciosos sem comprometer o seu desempenho ou criar resultados incorretos.	ENT.2, ENT.6
	Responsabilidade	Ter a responsabilidade de quem produz e implementar a IA, assim como os resultados que dá.	ENT.1, ENT.2, ENT.3, ENT.6, ENT.8, ENT.9
	Proteção da informação	Privacidade da informação que é utilizada em ferramentas de IA.	ENT.2, ENT.3, ENT.5, ENT.9
	Transparência	Clareza sobre como os sistemas de IA funcionam, como tomam decisões e como utilizam os dados.	ENT.2, ENT.3, ENT.9
	Inclusividade	Tornar a tecnologia IA acessível a todas as pessoas.	ENT.1, ENT.2, ENT.3, ENT.7, ENT.8
	Grupos Éticos	Elaboração de diretrizes que promovam a transparência, a justiça e a responsabilidade na IA.	ENT.3

Fonte: Elaboração própria