Inteligência Artificial, Segurança e Direitos

Filipe Gonçalves Reina Amaral Fernandes

Tese para obtenção de Grau Científico de Mestrado em

Mestrado de Segurança da Informação e Direito do Ciberespaço

Orientadores: Prof. Dr. Miguel Nuno Dias Alves Pupo Correia
Coronel da GNR, Pedro Manuel Sequeira Estrela Moleirinho

Júri

Presidente: Prof. Dr. Carlos Caleiro

Arguente: Prof. Dr. Sérgio Nunes

Vogal: Prof. Dr. Miguel Nuno Dias Alves Pupo Correia

Dezembro de 2020

“The threat of malevolent machines with monstrous artificial intelligence dominating humanity is imaginary, the threat of humanity misusing machines is real.”

“A ameaça de máquinas monstruosas a dominarem a humanidade é imaginária, o risco de a humanidade usar mal as máquinas é real.”

Luciano Floridi

Agradecimentos

A elaboração do presente trabalho de investigação levou a que fizesse um percurso por diversas instituições e auscultasse especialistas que contribuíram de forma decisiva para a sua realização.

Presto a minha homenagem a todos que de alguma forma me ajudaram ao longo deste projeto, destacando o orientador, Coronel da GNR, Pedro Manuel Sequeira Estrela Moleirinho, onde a sua experiência, conhecimentos e competência permitiram orientar, centralizar e dar um enfoque pragmático relativo a um tema complexo e multifacetado.

O meu agradecimento ao Professor Doutor Miguel Correia, do Instituto Superior Técnico de Lisboa, pela sua dedicação e apoio na identificação e contacto de diversos entrevistados, pela sua disponibilidade constante e orientação para a metodologia adequada para o estudo do problema.

A todos os entrevistados, que partilharam do seu tempo e conhecimentos para contribuir para a consciencialização das capacidades e riscos de uma tecnologia verdadeiramente inovadora como a Inteligência Artificial. O seu contributo foi fundamental para sustentar, do ponto de vista teórico e prático, todo o trabalho de investigação documental, e ainda materializar estes avanços pela demonstração de aplicações desta tecnologia.

Por último à minha família e amigos que me apoiaram e motivaram para a aceitação deste desafio, sem os quais não seria possível terminar.

Resumo

A Inteligência Artificial (IA) é uma tecnologia com potencial transformador da nossa vida em sociedade, afigurando-se como a resposta a diversos desafios que vivemos e como sendo um verdadeiro motor de desenvolvimento económico. No entanto, existem preocupações legais, éticas e socioeconómicas que devem ser devidamente estudadas e consideradas.

O presente estudo pretende identificar qual o impacto da aplicação dos sistemas de IA na Segurança, nos dados e na liberdade dos cidadãos enquanto parte integrante da sociedade.

É importante rever os conceitos necessários para compreender o contexto e o modo de funcionamento dos sistemas de IA, assim como o estudo de exemplos concretos de IA para fins de segurança, concretamente na Cibersegurança, no Policiamento Preditivo e na Videovigilância com recurso ao Reconhecimento Facial.

Há indicadores claros que a IA, aplicada à segurança, irá tornar diversos processos mais eficazes e eficientes. No entanto, não podemos deixar de considerar possíveis limitações tecnológicas e a possibilidade de ocorrência de ingerências na esfera dos direitos dos cidadãos. Torna-se também importante refletir que medidas devem ser tomadas para salvaguardar que o desenvolvimento e a utilização de sistemas de IA seja feito de forma responsável, ética e segura.

O desafio, perante nós exposto, obriga a trilhar um caminho de consciencialização das capacidades e impactos do desenvolvimento da IA, identificando linhas orientadoras para assegurar o respeito pelos Direitos, Liberdades e Garantias dos cidadãos, enquanto se beneficia dos frutos desta tecnologia.

Palavras-chave: Inteligência Artificial, Segurança, Direitos, Cibersegurança, Policiamento Preditivo, Reconhecimento Facial;

Abstract

Artificial Intelligence is a technology with the potential to transform our life and our society, it appears to be the answer to several societal challenges that we face and being a real engine of economic development. However, we must properly study and address the legal, ethical, and socio-economic concerns that arise from its development and use.

The present study aims to identify the effects of AI systems on citizens' security, data, and freedoms as an integral part of society.

We give an overview of the concepts needed to understand the context and the inner workings of AI, then, studied concrete examples of AI applications in the security domain, specifically in the domains of Cybersecurity, Predictive Policing and Video Surveillance using Facial Recognition.

There are clear indicators that AI will bring about added efficacy and efficiency to the processes that the systems address, but we also concluded that there may be interference on citizens' freedoms, rights and guarantees in the use of AI systems. We reflected on the measures that assure the development and use of AI systems in a responsible, ethical, and safe manner.

This challenge, requires us to follow a path of awareness of the capabilities and impacts of AI development, identifying guidelines to ensure respect for the citizens' freedoms, rights and guarantees, while benefiting from the fruits of this technology.

Keywords: Artificial Intelligence, Security, Rights, Cybersecurity, Predictive Policing, Facial Recognition

Conteúdos

Lista de Figuras

Lista de Siglas

API – *Application programming interface* – Interface de Programação de Aplicações
AV – Antivírus
C2 – Comando e Controlo
CC – Código Civil
CCTV – *Closed-circuit television* – Circuito fechado de televisão
CEDH – Convenção Europeia dos Direitos do Homem
CEDN – Conceito Estratégico de Defesa Nacional
CNCS – Centro Nacional de Cibersegurança
CNPD – Comissão Nacional de Proteção de Dados
CPP – Código do Processo Penal
CRP – Constituição da República Portuguesa
DL – *Deep Learning* – Aprendizagem profunda
DLG – Direitos, Liberdades e Garantias
ENISA – *European Union Agency for Cybersecurity* – Agência da União Europeia para a Cibersegurança
UE – União Europeia
EUA – Estados Unidos da América
FS – Forças de Segurança
GNR – Guarda Nacional Republicana
HVS – Hitachi Visualization Suite
IA – Inteligência Artificial
IoT – *Internet of Things* – Internet das Coisas
ISO – International Organization for Standardization – Organização Internacional para a Standartização
IST – Instituto Superior Técnico
ITU – *International Telecommunication Union* – União Internacional de Telecomunicações
LDN – Lei da Defesa Nacional
LFM – Hitachi Live Face Matching
LSI – Lei de Segurança Interna
ML – *machine learning* – Aprendizagem automática
NATO – OTAN – Organização Tratado do Atlântico Norte
NLP – *Natural Language Processing* – Processamento de linguagem natural
OECD – *Organisation for Economic Co-operation and Development* – Organização para a Cooperação e Desenvolvimento Económico
ONU – Organização das Nações Unidas

OSCE – *The Organization for Security and Co-operation in Europe* – Organização para a Segurança e Cooperação na Europa

PIDCP – Pacto Internacional sobre os Direitos Civis e Políticos

PIIC – Plataforma de Intercâmbio de Informação Criminal

PJ – Polícia Judiciária

PM – Polícia Marítima

PP – Policiamento Preditivo

PSP – Polícia de Segurança Pública

RGPD – Regime Geral da Proteção de Dados

SEF – Serviço de Estrangeiros e Fronteiras

SIEM – *Security information and event management* – Sistema de Informações de segurança e gestão de eventos

SIIC – Sistema Integrado de Informação Criminal

SIS – Serviço de Informações de Segurança

SOC – *Security Operations Center* – Centro de Operações de Segurança

SSVI – Smart Spaces and Video Intelligence

UNICRI – *The United Nations Interregional Crime and Justice Research Institute* – Instituto Inter-regional de Investigação de Crime e Justiça das Nações Unidas

USD – Dólares Americanos

Capítulo 1 – Introdução

1.1 Enquadramento

A presente dissertação de mestrado, subordinada ao tema Inteligência Artificial, Segurança e Direitos, é o culminar do ciclo de estudos do Mestrado em Segurança da Informação e Direito do Ciberespaço. Este mestrado constitui-se um exemplo de uma abordagem multidisciplinar a um tema transversal à sociedade, a Segurança da Informação.

Este mestrado conjunto do Instituto Superior Técnico, da Faculdade de Direito da Universidade de Lisboa e da Escola Naval, proporcionou a transmissão de competências e de conhecimentos que nos preparam para enfrentar os desafios da era da informação.

Estes desafios espelham uma realidade onde é difícil definir as fronteiras entre o mundo físico e o mundo digital, pela crescente digitalização da sociedade. As tarefas que preenchiam o nosso quotidiano físico e *offline* passam a ser possíveis de realizar à distância de um clique (NEGREIRO, et al., 2019 p. 2).

A transformação vivida desde o início do milénio foi rápida e profunda, no entanto, o futuro que se avizinha promete uma aceleração e desenvolvimento verdadeiramente disruptivo. Com o contributo dominante da Inteligência Artificial (IA).

A par destas inovações tecnológicas, mantêm-se constante uma necessidade primordial, a Segurança. Os desafios de segurança que marcaram o início deste século, como por exemplo o terrorismo, evoluíram com o avanço da tecnologia. Esta evolução obriga o Estado e as Forças de Segurança (FS) a estudar estes fenómenos para que seja possível adaptarem-se e darem uma resposta adequada e eficaz.

É precisamente esse o objetivo deste trabalho de investigação, estudar o surgimento da IA como vetor de segurança e as implicações que advêm da sua utilização numa sociedade que se pretende democrática, constituída por cidadãos de plenos direitos.

A IA evoluiu de uma pequena área de investigação para dezenas de aplicações que usamos diariamente. Assim que as grandes empresas tecnológicas vislumbraram o seu potencial, rapidamente entraram na presente “corrida às armas” da IA.

A comunidade científica alertou, e continua a alertar, para os perigos que esta nova tecnologia acarreta. Acredita-se que irá continuar a desenvolver-se e atingir um nível de inteligência equivalente à do ser humano, ou até mesmo superior. Algo que se afigura distante, mas real (LEE, 2018 p. 5).

Para desenvolver estes sistemas, temos de perceber o nosso próprio funcionamento, a nossa forma de pensar, de articular o pensamento ou até mesmo como é que nos tornámos conscientes, desafios que, mesmo não sendo impeditivos, constringem os avanços da IA.

Paulatinamente, o monopólio da capacidade preditiva está a passar das nossas mãos para os bits que constituem os algoritmos[1] e sistemas de IA. O futuro que se perspetiva, com a generalização das aplicações desta tecnologia, varia conforme os autores, escolas de pensamento e interesses dos investigadores. Para uns, este frenesim de avanços constantes e últimos desenvolvimentos das capacidades da IA são augúrios da nova realidade que está para chegar. Outros acreditam que apenas estamos a colher os frutos da era da informação e que uma transformação profunda da sociedade ainda está a uma geração de acontecer (LEE, 2018 p. 8) .

Este despertar de inteligência não é ameaçador por ainda estarmos longe da criação de um agente autónomo dotado de inteligência equiparável à do ser humano. Por outro lado, o presente ritmo de investimento e desenvolvimento asseguram que a IA de nível humano ou até superior seja um futuro previsível. Ainda assim, é necessário discernir as potencialidades desta tecnologia e conhecer as suas limitações e potenciais consequências negativas para a nossa sociedade – como a questão ética da perpetuação da discriminação através dos algoritmos.

Cabe-nos guiar esta integração da IA na sociedade, fazer previsões sobre o futuro, identificar ameaças, mitigar os riscos, aproveitar as oportunidades e trilhar o caminho do desenvolvimento humano.

Problema de investigação e formulação de hipóteses

As implicações da IA na sociedade estão a ser estudadas por diversos organismos e instituições internacionais com o objetivo de prever e se prepararem para o choque tecnológico que a IA acarreta. No entanto, as suas aplicações à segurança pelas FS é uma realidade ainda pouco estudada (ITU p. 38).

Consideramos que a escolha de um tema tão pouco maduro compreende em si um risco do presente trabalho ser considerado ultrapassado a curto prazo. Os exemplos identificados de limitações da tecnologia podem ser suprimidos pela inovação. Por outro lado, pode ser avançada legislação e regulamentos que solucionem alguns dos problemas apresentados. Perspetiva-se que as entidades certificadoras criem quadros de referência, *frameworks*, normas e *standards* que implementem a segurança, a privacidade e a ética por desenho nas diversas tecnologias que recorrem a sistemas de IA.

Quando nos referimos à IA estamos a referir-nos a sistemas de IA restrita, desenhados para resolver problemas específicos, em contraste com a IA Geral que se propõem a desenvolver um sistema de IA com capacidades cognitivas equiparadas ou superiores às do ser humano. A IA Geral será sem dúvida um grande desafio para a humanidade, devendo ser um assunto amplamente discutido.

[1] Sequência de instruções que diz a um computador o que fazer. (DOMINGOS, 2017 p. 25)

O valor acrescentado do presente trabalho é dissipar alguns mitos que pairam sobre esta tecnologia, demonstrar as suas capacidades atuais, alertar para os impactos possíveis e incentivar a discussão pública.

O presente trabalho tem como questão central:

Qual é o impacto da aplicação dos sistemas de IA pelas FS na Segurança e nos nossos direitos?

Centralizado nas questões da aplicação de IA à segurança, visando o impacto que esta tecnologia terá no cidadão enquanto parte integrante do Estado e este último como garante da segurança do cidadão.

Como questões derivadas, surgem então:

QD1 - Que sistemas de IA existem para apoiar as FS?

QD2 - Que ingerências poderão ocorrer na esfera dos direitos, liberdades e garantias com o recurso a sistemas de IA pelas FS?

QD3 - Que medidas devem ser tomadas para salvaguardar que o desenvolvimento e a utilização de sistemas de IA pelas FS é feito de forma responsável, ética e segura?

Responder às questões apresentadas obriga a trilhar um caminho de consciencialização das capacidades e impactos do desenvolvimento da IA, identificando linhas orientadoras para assegurar o respeito pelos Direitos, Liberdades e Garantias (DLG) dos cidadãos.

Definidos os objetivos da investigação, iniciou-se este projeto pela revisão da literatura, procedendo à pesquisa e análise documental de obras literárias, artigos de revistas científicas, monografias, dissertações de mestrado, teses de doutoramento, notícias de órgãos de comunicação social de referência, bem como de outros documentos relevantes.

Foram realizadas entrevistas a vários especialistas[2] da área da Segurança, da Tecnologia e Investigação em Portugal, que personificam o conhecimento desta área de estudo, aplicando-se o método inquisitivo, por intermédio de entrevistas[3] semi-diretivas. Este trabalho de campo permitiu sustentar a componente teórica pela demonstração de exemplos de aplicações de IA em Portugal e no mundo. Contribuiu ainda para respondermos às questões formuladas que motivam esta investigação.

O trabalho é constituído por cinco capítulos. O primeiro capítulo introduz o tema e proporciona-nos uma visão e alcance do trabalho. No segundo capítulo é feita uma revisão da literatura, a definição dos vários conceitos que constituem a esfera de influência da IA e a procura do estabelecimento de um estádio de conhecimento relativo à IA que permita a compreensão das tecnologias apresentadas no capítulo subsequente. No terceiro capítulo expomos o resultado do trabalho de campo, pela apresentação de três áreas de aplicação da IA, Cibersegurança, Policiamento Preditivo (PP) e Videovigilância com recurso ao

[2] Ver anexo I – Lista de Entidades Entrevistadas
[3] Ver anexo II – Guião de Entrevista

Reconhecimento Facial, respondendo à QD1. No quarto capítulo apresenta-se a resposta às QD2 e QD3 integrando a síntese dos contributos recolhidos pelas entrevistas e a investigação realizada. O quinto capítulo apresenta as reflexões finais e recomendações do presente trabalho.

Capítulo 2 – Segurança, Privacidade, Dados e Inteligência Artificial

A segurança, a privacidade e os dados são elementos que constituem a nossa realidade e que impreterivelmente produzem efeitos no nosso quotidiano. Se a segurança e a privacidade são uma preocupação tão antiga como a coexistência humana, os Dados e a IA são conceitos mais recentes.

2.1 Segurança

A segurança é um conceito de caráter polissémico e em permanente evolução, não sendo fácil definir de forma consensual e inequívoca. No entanto, consideramos que uma definição abrangente engloba os conceitos de segurança pública, proteção civil, prevenção e investigação criminal – que por sua vez são absorvidos por um conceito mais alargado, o de segurança nacional (PEREIRA, 2018 p. 3).

A conceção de segurança nacional assenta em quatro pilares: "*a Defesa Nacional e as Forças Armadas; a Segurança Interna e as Forças Policiais; a Segurança do Estado e os Serviços de Informações e a Segurança Comunitária e a Proteção Civil"*. Este paradigma de Segurança Nacional parece-nos mais ajustado à realidade em que vivemos, pelo esbatimento das fronteiras entre os riscos e a necessidade de os combater, e pela fluidez da distinção entre as ameaças externas e as ameaças internas (GOUVEIA, 2018 p. 366).

No entanto, podemos observar que a Defesa Nacional tem os seus objetivos definidos no n.º 2 do artigo 273.º da Constituição da República Portuguesa (CRP), que os explicita nos seguintes termos: *"garantir a independência nacional, a integridade do território e a liberdade e a segurança das populações contra qualquer agressão ou ameaças externas*" e que tem como documento enquadrador, a Lei da Defesa Nacional[4] (LDN) (ALMEIDA, 2013 p. 29).

Por outro lado, a Segurança Interna é definida nos termos da Lei de Segurança Interna (LSI), Lei nº 53/2008, de 29 de Agosto, como *"a atividade desenvolvida pelo Estado para garantir a ordem, a segurança e a tranquilidade públicas, proteger pessoas e bens, prevenir e reprimir a criminalidade e contribuir para assegurar o normal funcionamento das instituições democráticas, o regular exercício dos direitos, liberdades e garantias fundamentais dos cidadãos e o respeito pela legalidade democrática*." As medidas previstas nesta lei *"destinam-se, em especial, a proteger a vida e a integridade das pessoas, a paz pública e a ordem democrática, designadamente contra o terrorismo, a criminalidade violenta ou altamente*

[4] Declaração de Retificação n.º 52/2009 - Diário da República n.º 138/2009, Série I de 2009-07-20

organizada, a sabotagem e a espionagem, a prevenir e reagir a acidentes graves ou catástrofes, a defender o ambiente e a preservar a saúde pública." (RÊGO, 2013 p. 91).

A segurança pode também ser definida como "*um esforço de governação concertado, envolvendo todos os agentes e as capacidades públicas e privadas que contribuem para um clima de paz social e de tranquilidade pública num país*." (ALMEIDA, 2013 p. 7).

De facto, ao Estado incumbe-se a tarefa de proteger os cidadãos: *"A segurança pública, dever do Estado, direito e responsabilidade de todos, é exercida para a preservação da ordem pública e da incolumidade das pessoas e do património*", mas entendemos que "*a segurança é uma questão de Estado, mas, mais do que isso, é um bem público. Sem segurança não há desenvolvimento económico. Sem segurança não há democracia.*" (TEIXEIRA, 2002 p. 10).

Do ponto de vista doutrinário, a segurança pode ser entendida como um desígnio que visa reduzir o risco. É difícil atingir o nível de risco zero, visto ser ilusória a possibilidade de proteger todas as vulnerabilidades de forma absoluta. Não é exequível alcançar um estado de completa segurança de forma generalizada para todos os cidadãos, em todas as situações. Aproximar o grau de risco ao valor zero implicaria a utilização de recursos e meios policiais numa escala omnipresente, obrigando a estabelecer um Estado Policial – possibilidade que deve ser afastada por razões lógicas de respeito pelos princípios democráticos.

Perante esta lógica da impossibilidade da segurança total, podemos assumir a impossibilidade de acabar completamente com o crime. Assim, considera-se que existe uma continuidade do crime devido à incapacidade de antever todos os objetivos e meios de possíveis ataques (ZEDNER, 2003 p. 156). Consequentemente, o crime pode ser entendido como um risco à segurança, exigindo assim a adoção de medidas preventivas, porque "*onde os riscos podem ser calculados, é mais económico prevenir a perda do que punir retrospetivamente*" (ZEDNER, 2009 p. 71).

Importante também é a noção de *"direito à segurança"*, que materializa a garantia do exercício dos direitos, liberto de agressões ou ameaças. A segurança significa neste contexto, cumulativamente, duas coisas: "*o direito de defesa perante agressões dos poderes públicos e o direito de proteção conferido pelos poderes públicos contra agressões ou ameaças de outrem."* (SILVA, 2010 p. 5).

Para materializar o conceito de segurança ao nível do Estado, surgem as políticas de segurança que visam alcançar três objetivos primordiais: proteger as pessoas, as instituições democráticas, as infraestruturas críticas; a prevenção os riscos simétricos e assimétricos de se manifestarem; e, por último, conter os impactos e/ou os efeitos de um acontecimento catastrófico, gerindo as suas consequências, recuperando a ordem e a lei, e facilitando o regresso às condições de normalidade anteriores à ocorrência de uma crise (ALMEIDA, 2013 p. 7).

2.2 Forças de Segurança e Modelos de Policiamento

Para compreender o conceito de FS devemos começar por compreender o conceito que subjaz na sua génese, o conceito de polícia.

"[…] *pode definir-se a Polícia como modo de atuar da autoridade administrativa que consiste em intervir no exercício das atividades individuais suscetíveis de fazer perigar interesses gerais, tendo por objeto evitar que se produzam, ampliem ou generalizem os danos sociais que a lei procura prevenir*" (CAETANO, 1977 p. 339).

Para Bacelar Gouveia, a segurança corresponde a uma necessidade básica. Desse prisma, a atividade de polícia e a atividade de segurança podem ser consideradas serviço público. A referência anterior manifesta-se relevante, uma vez que as definições de segurança interna e forças policiais estão profundamente enraizadas no conceito de Polícia. Facto em grande medida cunhado pelo Direito Administrativo do século XIX. É comum divisar-se três aceções de polícia: a polícia em sentido funcional, ou seja, polícia como atividade; polícia em sentido orgânico designadamente polícia como organização; e ainda polícia em sentido formal, como manifestação de poder (GOUVEIA, 2018 p. 560).

As missões e funções de polícia têm a sua consagração na Lei Fundamental e no seu Artigo 272.º (CRP) sob a epígrafe "*Polícia*", definindo como funções "*defender a legalidade democrática e garantir a segurança interna e os direitos dos cidadãos*" e ainda a *"prevenção dos crimes, incluindo a dos crimes contra a segurança do Estado, só pode fazer-se com a observância das regras gerais sobre polícia e com o respeito pelos direitos, liberdades e garantias dos cidadãos*".

As entidades incumbidas de executar tais funções são designadas na LSI como FS, constituindo organismos públicos, estando exclusivamente ao serviço do povo português e com carácter apartidário, que concorrem para a garantia da segurança interna, nomeadamente[5]: a Guarda Nacional Republicana (GNR), a Polícia de Segurança Pública (PSP), a Polícia Judiciária (PJ), o Serviço de Estrangeiros e Fronteiras (SEF), a Polícia Marítima (PM) e o Serviço de Informações de Segurança (SIS). Para além das entidades referidas anteriormente, acrescentam-se outras entidades Administrativas que exercem ainda funções de segurança, mas com competências circunscritas às respetivas áreas de atuação e legislação habilitante, como por exemplo os órgãos da Autoridade de Segurança Alimentar. Todas estas entidades são designadas Órgãos de Polícia Criminal. A organização, as atribuições e as competências das forças de segurança constam das respetivas leis orgânicas e demais legislação complementar.

Um dos objetos de estudo do presente trabalho é o Policiamento Preditivo – um modelo de policiamento. Estes modelos pretendem representar as relações e a postura que

[5] Cf. o n. º2 e n. º3 do artigo 25.º. da Lei n.º 53/2008 (LSI).

determinados Estados adotam relativamente à sua organização policial. O estabelecimento destes modelos, que agrupam abordagens ao policiamento – umas mais reativas, outras mais proativas – pretendem obter legitimidade na atuação e conquistar a confiança dos cidadãos. Importante também referir que estes modelos derivam e convergem em diversos pontos e representam diversas abordagens que cada Estado adotou e desenvolveu, considerando o contexto da sua evolução histórica, sociopolítica e económica.

À imagem das restantes estruturas públicas, as organizações policiais também foram evoluindo e adotando novas formas de organização, de táticas e técnicas, e também de metodologias e de processos de gestão (MOLEIRINHO, 2018 p. 102).

Podemos encontrar diversas correntes e metodologias de policiamento, sendo geralmente divididos em cinco grandes modelos: o policiamento tradicional[6], o policiamento comunitário, o policiamento orientado para os problemas, o policiamento orientado pelas informações, e mais recentemente, o PP[7]. Cada um destes modelos possui diferentes objetivos estratégicos, vantagens e vulnerabilidades, podendo complementar-se quando usados simultaneamente.

O policiamento tradicional é provavelmente o mais conhecido e utilizado dos modelos. Refere-se a um estilo reativo e orientado à resposta de incidentes, sendo transmitida uma imagem autocrática e repressiva das FS. O *modus operandi* está assente no patrulhamento aleatório e a realização de outras tarefas – resposta a ocorrências ou emergências baseadas em denúncias, ou pelo patrulhamento de visibilidade. Enquanto este modelo tradicional de policiamento vê a reação a questões de segurança e segurança pública como tarefa da polícia, os modelos mais recentes procuram uma aproximação entre a comunidade e as FS, numa abordagem proativa às questões de segurança. O modelo tradicional de policiamento tem demonstrado sérias dificuldades em lidar com os riscos e ameaças atuais, como a globalização, a permeabilidade das fronteiras, o surgimento de novos fluxos migratórios, o aumento da mobilidade, a livre circulação de bens e serviços e a crescente desigualdade social. Acresce ainda o surgimento de movimentos de radicalização como porta de entrada para as atividades terroristas, cuja repercussão se tem sentido na UE (OSCE, 2017 p. 10 e 14).

Estas definições permitem enquadrar as aplicações da IA em matéria de segurança em análise no capítulo 3 do presente trabalho. Numa perspetiva global, as FS procuram constantemente aumentar a sua a eficácia e eficiência. Uma forma de cumprir este desiderato é acompanhar e promover as inovações tecnológicas.

[6] Também designado policiamento profissional-burocrático, profissional ou militar-burocrático.
[7] Também designado de policiamento com base em estatística.

A nível internacional, as FS procuram explorar o potencial da IA no desempenho da sua missão. Esta tendência poderá ser replicada em Portugal ao abrigo da modernização da administração pública e transição digital.

A quantidade crescente de dados obtidos e armazenados pelas polícias também exigiu a adoção de métodos e ferramentas mais sofisticados para a gestão, partilha e análise de dados, culminando na criação de sistemas de informação internos em cada entidade. Para alem destas, foram criadas plataformas de partilha de informação – designadamente, o Sistema Integrado de Informação Criminal (SIIC) e a Plataforma de Intercâmbio de Informação Criminal[8] (PIIC). As potencialidades da IA são variadas e o seu espetro de aplicação poderá aumentar nos próximos anos com os avanços nas áreas da robótica, veículos não tripulados (*drones*), veículos autónomos, entre outros.

Atualmente, existem aplicações de IA para a identificação de padrões, previsão de riscos, análise de imagem e vídeo, sistemas de reconhecimento facial ou ferramentas de análise preditiva criminal que apresentam resultados relevantes. Portugal está a dar os primeiros passos no recurso a IA. O estudo destas aplicações à realidade portuguesa deve ser aprofundado e divulgado. É fundamental informar e esclarecer a sociedade das vantagens, potencialidades, desvantagens e riscos desta tecnologia. Pretende-se que este trabalho contribua para este desiderato.

[8] Cf. o artigo 1.º. da Lei n.º 73/2009.

2.3 Desafios de Segurança 2.0

Consideramos que a segurança absoluta é uma realidade inatingível. No entanto, para mitigar o risco é necessário conhecê-lo. Dependendo do país ou organização internacional, a avaliação das principais ameaças e riscos à segurança divergem. Indubitavelmente, a transformação proporcionada pelo desenvolvimento tecnológico que se verificou nos últimos anos fez surgir novas ameaças, novos vetores de ataques e até proporcionou novos domínios de confrontação, passando do tradicional, terra, mar e ar, para as contendas também no espaço e no ciberespaço[9] (WELCH, 2011).

O nosso Conceito Estratégico de Defesa Nacional[10] (CEDN) aponta como principais riscos e ameaças à segurança nacional:

1º - terrorismo;

2º - proliferação de armas de destruição massiva;

3º - criminalidade transnacional organizada;

4º - cibercriminalidade;

5º - pirataria.

É importante ter em consideração que o CEDN foi aprovado em 2013 e já considerava a ameaça *ciber* relacionada ao crime.

Em 2013, a Comunidade de Informações dos Estados Unidos da América (USA)[11] apresentava ameaça *ciber* como a primeira prioridade, seguida do terrorismo e do crime organizado de natureza transnacional (CLAPPER, 2013).

Este organismo mantém a ameaça *ciber* como sendo a primeira prioridade motivada pela capacitação dos seus principais adversários e competidores estratégicos – em matérias de ciberataques, ciberespionagem e influência para obter vantagens a nível político, económico e militar (COATS, 2019). Constata-se que a China, Rússia, Irão e Coreia do Norte estão a operar no ciberespaço com o intuito de roubar informação e afetar as infraestruturas críticas. A segunda prioridade prende-se com a capacidade de ingerência nas eleições.

---

[9] "O termo Ciberespaço é usado para referir o conjunto dos sistemas informáticos como *hardware, software,* redes de comunicação, equipamentos e meios de comunicação e a informação neles processados e armazenados. Sinónimos comumente usados para Ciberespaço são expressões tais como: espaço virtual, mundo virtual, reino eletrónico, esfera da informação, etc. (FERNANDES, 2013 p. 13) O termo Ciberespaço (na sua versão anglo-saxónica *cyberspace*) foi cunhado em 1984, pelo escritor de ficção científica William Gibson, no seu romance, como *"a consensual hallucination experienced daily by billions of legitimate operators, in every nation, by children being taught mathematical concepts [...] A graphical representation of data abstracted from the banks of every computer in the human system. Unthinkable complexity. Lines of light ranged in the non-space of the mind, clusters and constellations of data.*" (Gibson,1984).

[10] Aprovada pela resolução do Conselho de Ministros n.º 19/2013.

[11] *United States Intelligence Community*

Verificaram-se interferências externas em eleições de 2016 nos EUA com os ataques ao partido Democrata (CBS 60 Minutes, 2020) no referendo relativo ao *Brexit* no Reino Unido (The Guardian, 2018) e relativamente às eleições europeias de maio de 2019 (New York Times, 2019).

Na perspetiva da Estratégia Global para a Política Externa e de Segurança da União Europeia (EEAS, 2016), o terrorismo continua a ser a principal ameaça, pelo passado recente dos vários atentados terroristas em solo europeu que visaram a França, Bélgica Alemanha, Reino Unido e Espanha. Para além do terrorismo, a Cibersegurança é uma área de preocupação da UE, que definiu como objetivo reforçar as capacidades tecnológicas destinadas a atenuar ameaças e o aumento da resiliência das infraestruturas, redes e serviços críticos, bem como a diminuição da cibercriminalidade. Para tal, é fundamental a cooperação e a partilha de informações entre Estados-Membros, instituições, o setor privado e a sociedade civil de forma a promover uma cultura de Cibersegurança comum e aumentar a preparação contra eventuais ciberataques.

Podemos constatar que as principais ameaças globais evoluíram e IA passará a ter um papel de destaque. No campo da interação social, a capacidade NLP[12], ou seja, leitura, interpretação e escrita da linguagem natural, poderá ser utilizada para gerar e difundir desinformação em larga escala, conhecidas por *Fake News*[13]. Para corroborar esta afirmação, a OpenAI[14] – conhecida por partilhar livremente o código fonte das aplicações que desenvolve – afirmava que o seu novo modelo de IA, intitulado GPT2 especializado em NLP, era tão bom que o risco de uso para fins maliciosos era demasiado elevado para ser divulgado na sua total capacidade (The Guardian, 2019). Optaram por uma divulgação sequencial de versões cada vez mais eficazes, sendo que a última versão designada GPT3 já foi publicada e, à data, é possível interagir com o modelo através de uma API[15] (OpenAI, 2019).

[12] *Natural Language Processing* - É um ramo da IA que tem como objeto de estudo a interação entre os sistemas informáticos e os seres humanos através da utilização de linguagem natural. Tem como objetivo dotar um sistema informático com a capacidade de ler, interpretar, entender e reproduzir a linguagem humana de forma a transformar a linguagem numa ferramenta ou meio de comunicação entre o Homem e a máquina.

[13] *Fake News* – São consideradas histórias falsas que se assemelham a notícias verídicas, difundidas pela internet, redes sociais ou outros meios, geralmente com o intuito de desinformar, influenciar opiniões políticas ou apenas como divertimento. (CAMBRIDGE, 2020)

[14] OpenAI – Empresa de investigação e desenvolvimento com sede em San Francisco, Califórnia, USA, que tem como principal missão assegurar que a criação de IAG beneficie toda a humanidade. Para tal, investiga e desenvolve projetos diretamente relacionados com a criação de IAG segura e benéfica. Conta com dirigentes e investigadores de renome como Greg Brockman, Ilya Sutskever, Sam Altman, Adam D'Angelo, Holden Karnofsky, Reid Hoffman, Sue Yoon, Tasha McCauley assim como investidores de peso como a Microsoft, a fundação Reid Hoffman e a Khosla Ventures. (OpenAI, 2020)

[15] *Application programming interface* ou Interface de Programação de Aplicações permite partilha de funcionalidades ou serviços de determinada aplicação de forma pública ou privada. Um exemplo, caso pretenda apresentar a localização de um restaurante no seu site poderia usar a API da GoogleMaps para gerar uma janela interativa com a localização do restaurante.

Para além do texto, aplicações de IA como a Face2Face (THIES, et al., 2016) designada de *deepfake* também causam apreensão pela sua verosimilhança[16], permitindo que futuramente seja quase impercetível a diferença ao olho humano.

O uso de veículos aéreos, terrestres ou aquáticos não tripulados, comumente conhecidos como *drones*, cujas capacidades de deteção de alvos, planeamento, autonomia e navegação autónoma serão ampliadas com recurso a IA aumentando assim as suas potenciais aplicações por grupos terroristas. Exemplo disso foi o ataque às instalações da *Saudi Aramco*[17] em Abqauiq e Khurais, Arábia Saudita. Este ataque reivindicado pelos rebeldes *Houthis* confirma uma já antecipada ameaça da utilização de *drones* em atos de terrorismo (The Guardian, 2019).

Se em 2017 a EUROPOL[18] considerava previsível o uso de *drones* por organizações criminosas nas suas atividades ilícitas, em especial no tráfico de estupefacientes (EUROPOL, 2017 p. 34), atualmente já existem relatos desta utilização – como por exemplo, para a introdução de estupefacientes no valor de 550.000€ em vários estabelecimentos prisionais no Reino Unido (The Guardian, 2018) e ainda para operações de contra vigilância e reconhecimento, como aconteceu em Espanha (The Guardian, 2019).

Espera-se que os computadores quânticos ofereçam um aumento sem precedentes no poder de processamento computacional e que ajudem a alcançar avanços significativos em várias áreas científicas. Garantidamente, o advento da computação quântica será altamente disruptivo para o cenário atual da segurança informática, pelo facto de os problemas criptográficos que servem de base para esta segurança, mesmo sendo demasiados complexos para os computadores clássicos, serão facilmente ultrapassados pelos computadores quânticos. Inevitavelmente, terá de se recorrer a padrões de segurança e à criptografia quântica por forma a manter um nível de segurança aceitável (EUROPOL, 2019 p. 35).

Os dispositivos da Internet das Coisas (IoT)[19] estão a ser desenvolvidos para fins comerciais e industriais, e avizinha-se a sua integração nas próprias infraestruturas com o aparecimento de cidades inteligentes[20]. Os dispositivos de IoT vão beneficiar da convergência

[16] https://tinyurl.com/deepfakevideo ou https://youtu.be/cQ54GDm1eL0
[17] *Saudi Arabian Oil Company* – Empresa petrolífera detida pelo Reino da Arábia Saudita.
[18] Europol é a agência da UE responsável por garantir o cumprimento da lei. Tem como missão principal ajudar a construir uma Europa mais segura em benefício de todos os cidadãos da UE. Sediada na Haia, nos Países Baixos, presta apoio aos 28 Estados-Membros da União no âmbito à luta contra as formas graves de criminalidade internacional e de terrorismo. Além disso, colabora com países terceiros e organizações internacionais. (EUROPOL, 2020)
[19] Compreende todos os aparelhos e objetos que se encontram habilitados a estarem permanentemente ligados à Internet, sendo capazes de se identificar na rede e de comunicar entre si. Podem ter o seu estado alterado através daquele meio, com ou sem o envolvimento ativo do ser humano e têm capacidade para recolher uma vasta quantidade de informação sobre o que os rodeia (CNCS, 2019).
[20] Designadas como *Smart City* aplicam as tecnologias da comunicação para pessoas, informações e elementos da cidade para criar uma cidade mais sustentável, aumentar a qualidade de vida dos seus habitantes pelo aumento da eficiência e eficácia de vários serviços da cidade como por exemplo, o tráfego, saneamento, gestão de resíduos, iluminação (SCHAFFERS, et al., 2012 p. 5).

e integração de tecnologias, como a IA, a rede móvel 5G e a computação – na nuvem e localmente[21] designada de *edge.* A implementação incorreta de protocolos de segurança nestes dispositivos poderá originar novos vetores de ataques ou converter estes dispositivos em servos digitais de atores mal-intencionados – como sucedido em 2016 com o *malware Mirai* que desencadeou um ataque massivo de negação de serviços de forma distribuída (DDoS) que atingiu mais de 600.000 dispositivos (ANTONAKAKIS, et al., 2017).

[21] Computação *in loco,* próximo do local de recolha e utilização dos dados.

2.4. A Privacidade

A palavra em latim *Privates* significa separado do resto. A privacidade materializa-se na capacidade de uma pessoa se separar do resto e assim revelar-se de forma livre e espontânea. Independentemente das divergências culturais no gozo deste direito, há um entendimento básico comum: "*Uma pessoa cuja privacidade seja significativamente afetada não pode viver uma vida sem medo e sem privação. Pressupõe-se a garantia da proteção básica dos direitos de privacidade para que se possa viver uma vida com segurança humana*" (MOREIRA, et al., 2012 p. 387) .

O direito à privacidade, também conhecido como direito à reserva da vida privada, encontra-se plasmado em diversos diplomas internacionais e nacionais. A nível internacional, consagra-se na Declaração Universal dos Direitos do Homem (DUDH) concretamente, no seu artigo 12.º que *"Ninguém sofrerá intromissões arbitrárias na sua vida privada, na sua família, no seu domicílio ou na sua correspondência, nem ataques à sua honra e reputação. Contra tais intromissões ou ataques toda a pessoa tem direito a proteção da lei."*

O Pacto Internacional sobre os Direitos Civis e Políticos (PIDCP) no seu art.ºº 17º estabelece que: *"Ninguém será objeto de intervenções arbitrárias ou ilegais na sua vida privada, na sua família, no seu domicílio ou na sua correspondência, nem de atentados ilegais à sua honra e à sua reputação."* E que *"Toda e qualquer pessoa tem direito à proteção da lei contra tais intervenções ou tais atentados."*

E ainda na Convenção Europeia dos Direitos do Homem (CEDH) onde refere no artigo 8.º que "*Qualquer pessoa tem direito ao respeito da sua vida privada e familiar, do seu domicílio e da sua correspondência."* Continuando no seu número 2: "*Não pode haver ingerência da autoridade pública no exercício deste direito senão quando esta ingerência estiver prevista na lei e constituir uma providência que, numa sociedade democrática, seja necessária para a segurança nacional, para a segurança pública, para o bem-estar económico do país, a defesa da ordem e a prevenção das infrações penais, a proteção da saúde ou da moral, ou a proteção dos direitos e das liberdades de terceiros."*

No ordenamento jurídico português, o direito à privacidade está consagrado na CRP artigo 26.º, ditando que *"A todos são reconhecidos os direitos à identidade pessoal, ao desenvolvimento da personalidade, à capacidade civil, à cidadania, ao bom nome e reputação, à imagem, à palavra, à reserva da intimidade da vida privada e familiar e à proteção legal contra quaisquer formas de discriminação.*" E ainda que "*A lei estabelecerá garantias efetivas contra a obtenção e utilização abusivas, ou contrárias à dignidade humana, de informações relativas às pessoas e famílias."*

O Código Civil (CC) no seu artigo 80.º sob a epígrafe "Direito à reserva sobre a intimidade da vida privada" acrescenta que "*Todos devem guardar reserva quanto à*

intimidade da vida privada de outrem." E ainda que *"A extensão da reserva é definida conforme a natureza do caso e a condição das pessoas."*

Verifica-se uma íntima relação entre o exercício da liberdade individual e o direito à privacidade, uma vez que a privacidade é o direito à tranquilidade, ao isolamento contra os olhares curiosos, à reserva e proteção de algo que é apenas do indivíduo, um escudo ao escrutínio da sua vida privada. Assim, considera-se que a privacidade é um meio para garantir o desenvolvimento pleno da personalidade individual, sem intromissões de terceiros e gozando algo que nos diferencia e nos caracteriza – o livre arbítrio (JABUR, 2000 p. 260).

Ora, este livre arbítrio também garante que "*do ponto de vista jurídico-constitucional, uma pessoa que decide tornar públicos comportamentos geralmente protegidos pela reserva de intimidade da vida privada não está, por esse motivo, a renunciar a esse direito, mas sim a exercê-lo autonomamente de acordo com as suas próprias preferências. O direito à intimidade é compatível com diferentes modos de utilização"* (CANOTILHO, et al., 2003 p. 56).

"Numa sociedade pluralista, em que os indivíduos têm incomensuráveis e antagônicas visões de mundo, bem como interesses e objetivos completamente distintos, o direito à privacidade deverá ser analisado de acordo com a conceção do próprio titular, evitando-se, assim, que o direito à privacidade se transforme num unidimensional dever de privacidade" (PIRES, et al., 2014).

Este raciocínio conduz-nos à ideia de que o direito à privacidade consiste numa faculdade inerente a qualquer pessoa, um direito que está sujeito à vontade da própria pessoa de o gozar em toda a sua plenitude, protegendo-se contra intromissões de terceiros e preferindo a não divulgação de informações de carácter pessoal, ou de abdicar do mesmo, expondo a sua vida privada.

Este balanceamento entre o que queremos partilhar e o queremos manter privado é nos dias de hoje um exercício extremamente complicado, pois na era da informação aquilo que consideramos pessoal, privado e íntimo poderá não o ser. O grau de certeza com que afirmamos que uma informação é privada está inversamente relacionada com a digitalização da mesma.

Quando os nossos avós tinham um diário na cabeceira da cama, fechado a cadeado, tudo o que era lá escrito era visto por quem detinha a chave e tinha acesso ao quarto. O repositório *offline* dos dados de eventos marcantes que iam construindo era mantida na sua esfera privada. E o seu acesso era restrito e controlado. A divulgação da informação que continha, ou até mesmo a sua própria existência, dependia da vontade do seu proprietário.

Se hoje for escrito um diário digital, no programa *Microsoft Office 365*, o mesmo não se poderá dizer. Um relatório[22] comprovou que este recolhia informação referente, à data que o

[22] Produzido pelo Ministério da Justiça e Segurança, do governo dos Países Baixos para a avaliação do impacto da proteção de dados da aplicação *Microsoft Office 365*. (NAS, et al., 2019)

documento foi criado, data e horas a que era acedido, a conta, o nome e email do utilizador, endereço de IP[23], localização geográfica (aproximada) , identificação do dispositivo que estava a ser utilizado, identificação do navegador de internet (*web browser*), sistema operativo, entre vários outros (NAS, et al., 2019 p. 34). Para além deste facto, também detetaram que a aplicação envia, secretamente, dados de diagnóstico para uma empresa de telemarketing[24] sediada nos EUA, sem qualquer informação da sua existência e finalidade do processamento de dados (NAS, et al., 2019 p. 8).

Este é apenas um dos muitos exemplos que espelham uma nova realidade da era da informação. A nossa vida quotidiana deixa um vasto rasto digital. Os dispositivos que nos abrem as portas para todos os benefícios desta nova era são também responsáveis pelo registo e monitorização constante dos mais variados tipos de dados.

Assim, é necessário equilibrar a segurança com a privacidade, numa ótica compreensível de que quanto mais segurança se procura, invariavelmente, maior será a compressão da nossa liberdade e privacidade.

Porém, esta tese da resposta equilibrada apresenta um aspeto problemático logo no seu nome, visto que "equilibrar" pressupõe o conhecimento do resultado que se deseja alcançar. Tendo a noção exata do ponto de equilíbrio, pode insistir-se mais em segurança ou mais em liberdade (ASHWORTH, 2007).

[23]IP – Internet Protocol Address, é um protocolo de comunicação em rede que atribui um número identificativo a cada dispositivo conectado com o propósito de o identificar, localizar e permitir que comunique com os restantes dispositivos via esse mesmo protocolo.
[24] Especializada na criação de perfis preditivos de pessoas para fins comerciais.

2.5 Dados e Dados Pessoais

Referimos anteriormente a existência de um rasto digital da nossa atividade, gerado por uma miríade de dispositivos que preenchem o nosso quotidiano. Se os computadores pessoais foram dos primeiros interfaces entre o ser humano e o mundo digital, hoje, os sistemas ciberfísicos[25] transformam a relação entre o mundo físico e o ciberespaço. Para concretizar esta afirmação basta observar um catálogo de qualquer loja de eletrodomésticos e produtos eletrónicos de consumo para constatarmos a existência de todo um mundo *Smart*[26]*:* telemóveis, relógios, assistentes pessoais, câmaras de vigilância, videoporteiros, fechaduras, entre muitos outros que existem atualmente e ainda mais que surgirão. Todos estes dispositivos têm sensores responsáveis por obter informações sobre o ambiente que os rodeia, e podem ser geridos pelo utilizador ou por um sistema geralmente apelidado de "casa inteligente".

O volume de dados gerados mundialmente é difícil de colocar em perspetiva, mas espera-se que em 2025 sejam gerados 463 exabytes[27] de dados diariamente, o equivalente a aproximadamente 213 milhões de DVDs. Fatores como a taxa de penetração da internet que chegou a 57% em 2019, a tecnologia 5G e as cidades inteligentes vão contribuir para a geração desta grande quantidade de dados (VISUAL CAPITALIST, 2019).

Esta vastidão e variedade de dados informáticos obriga a que sejam categorizados de formas diferentes em virtude da informação que representam. Podemos encontrar várias definições de dados informáticos no normativo português, começando pela Lei n.º 109/2009 de 15 de setembro que aprova a Lei do Cibercrime, transpondo para a ordem jurídica interna a Decisão Quadro n.º 2005/222/JAI, do Conselho, de 24 de fevereiro, relativa a ataques contra sistemas de informação e que adapta o direito interno à Convenção sobre Cibercrime do Conselho da Europa. Esta lei estabelece as disposições penais materiais e processuais, bem como as disposições relativas à cooperação internacional em matéria penal, relativas ao domínio do cibercrime e da recolha de prova em suporte eletrónico.

Na alínea b) do artigo 2.º, estabelece que *Dado Informático "é qualquer representação de factos, informações ou conceitos sob uma forma suscetível de processamento num sistema informático, incluindo os programas aptos a fazerem um sistema informático executar uma função"*. Assim, os dados parecem ser definidos pela sua finalidade e não pelas suas características técnicas. Desta forma, futuras formas de representação de informação continuarão abrangidas. Um exemplo, o ADN está em vias de se tornar um meio de armazenamento de dados (LANGSTON, 2019). Neste contexto, o ADN utilizado para este fim

[25] Sistemas eletrónicos conectados em rede constituídos por um conjunto de unidades de processamento lógico, sensores e atuadores que permitem a interação com o mundo físico através de comandos transmitidos por telemetria. (YING, et al., 2008)
[26] Alusivo a sistemas com capacidade de comunicação via *Internet.*
[27] Para visualizar o volume de dados gerados, observe a Figura 1 Dados gerados num dia

deverá ser considerado como *Dado Informático* para, à partida, gozar das mesmas proteções que os outros dados gozam e que se encontram previstas nesta lei.

Na alínea seguinte, define dados de tráfego "*como os dados informáticos relacionados com uma comunicação efetuada por meio de um sistema informático, gerados por este sistema como elemento de uma cadeia de comunicação, indicando a origem da comunicação, o destino, o trajeto, a hora, a data, o tamanho, a duração ou o tipo do serviço subjacente*". Delimitando e distinguindo dados de tráfego e dados de conteúdo. Recorrendo à analogia, os dados de tráfego constituem o envelope que leva a carta (dados de conteúdo). Visto estes dados de tráfego conterem informação respeitante às circunstâncias das comunicações e não do próprio conteúdo[28] podem ser intitulados de dados sobre dados – metadados.

Mais recentemente, uma categoria de dados tomou grande destaque no espaço mediático. Referimo-nos aos Dados Pessoais. Não pretendendo fazer uma análise histórica do direito à proteção dos dados pessoais, importa referir que este não surgiu com a aprovação do Regulamento Geral da Proteção de Dados[29](RGPD) mas já está consagrado na CRP desde 1976, mesmo que não apresentasse o mesmo alcance da sua mais recente formulação. Por conseguinte, a integração deste direito no conjunto de direitos fundamentais demonstra a vontade do legislador em edificar o direito à proteção de dados e à não discriminação através do tratamento informático de dados.

Segundo o artigo 4º do RGPD, considera-se dado pessoal toda informação relativa a uma pessoa singular identificada, a que corresponde a uma pessoa física, ou dados que permitam, direta ou indiretamente, identificar essa pessoa. Uma pessoa pode ser considerada identificável pela presença de um ou pela correlação de vários identificadores, como por exemplo um "*nome, um número de identificação, dados de localização, identificadores por via eletrónica[30] ou a um ou mais elementos específicos da identidade física, fisiológica, genética, mental, económica, cultural ou social dessa pessoa singular"*. Para determinar se há uma probabilidade razoável de os meios serem utilizados para identificar a pessoa singular, importa considerar todos os fatores objetivos, como os custos e o tempo necessário para a identificação. Assim como a tecnologia disponível à data do tratamento dos dados e a evolução tecnológica.[31]

---

[28] O acesso a estes dados obedece aos critérios mais protecionistas e que visam salvaguardar os DLG dos cidadãos.

[29] Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016 relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados). Estabelece normas relativas à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais pelas instituições, órgãos e organismos da União, bem como pelos Estados-Membros no exercício de atividades relativas à aplicação do direito da União e à livre circulação desses dados.

[30] Determinados tipos de *cookies* podem ser considerados como dados pessoais caso sejam usados para identificar o utilizador quando entra num *website*. Este pode conter informações relativas ao nome do utilizador, email, que dispositivo, localização geográfica relativa, IP e *web browser* utilizado.

[31] Considerando (26) do RGPD.

De entre os dados pessoais, os dados biométricos requerem especial atenção – *dados pessoais resultantes de um tratamento técnico específico relativo às características físicas, fisiológicas ou comportamentais de uma pessoa singular que permitam ou confirmem a identificação única dessa pessoa singular, nomeadamente imagens faciais ou dados dactiloscópicos.* Estes dados biométricos podem ser utilizados para identificar uma pessoa através da observação do seu corpo ou partes dele. Os métodos de identificação mais comuns são os que recorrem a dados dactiloscópicos, mais comumente designadas como impressões digitais – no entanto também podem ser impressões palmares ou outro segmento específico. Outro método de identificação biométrico a ser utilizado de forma generalizada e que tem observado grandes avanços tecnológicos, quer na precisão, quer na velocidade da comparação e confirmação, é designado como reconhecimento facial.

Outra definição importante é a de tratamento, que consiste numa "*operação ou um conjunto de operações efetuadas sobre dados pessoais ou sobre conjuntos de dados pessoais, por meios automatizados ou não automatizados, tais como a recolha, o registo, a organização, a estruturação, a conservação, a adaptação ou alteração, a recuperação, a consulta, a utilização, a divulgação por transmissão, difusão ou qualquer outra forma de disponibilização, a comparação ou interconexão, a limitação, o apagamento ou a destruição"*.[32] Um sistema de IA tem como principal função o tratamento de dados, uma vez que opera de forma automatizada, com diversas formas de leitura, interpretação e transformação de dados.

O RGPD adensa a necessidade de autorização ou o consentimento do titular dos dados, entende-se como *uma manifestação de vontade, livre, específica, informada e explícita, pela qual o titular dos dados aceita, mediante declaração ou ato positivo inequívoco, que os dados pessoais que lhe dizem respeito sejam objeto de tratamento*[33]*.* Este ato positivo, idealmente, poderá ser materializado numa declaração escrita, oral ou mesmo pela validação de uma opção ao visitar um sítio na *web*. Independentemente da forma, é obrigatório a manifestação de vontade livre e informada. Estes dois pontos poderão gerar bastante debate, nomeadamente no que se considera uma decisão informada. Os formulários com opções pré-validadas ou avisos que a utilização do serviço comprova a aceitação das políticas de privacidade não deverão constituir um consentimento lato. O consentimento deverá abranger todas as atividades de tratamento realizadas com a mesma finalidade.

Outro conceito importante, e por ser considerado um meio efetivo de garantir a privacidade do titular dos dados, é a pseudonimização, que se define como *o tratamento de dados pessoais de forma que deixem de poder ser atribuídos a um titular de dados específico sem recorrer a informações suplementares, desde que essas informações suplementares*

[32] 2) do Artigo 4.º do RGPD.
[33] 11) do Artigo 4.º do RGPD.

sejam mantidas separadamente e sujeitas a medidas técnicas e organizativas para assegurar que os dados pessoais não possam ser atribuídos a uma pessoa singular identificada ou identificável[34].

A principal novidade deste regulamento são as coimas avultadas para as entidades que não respeitam diversos direitos. Sumariamente: o direito a ser informado – saber se os seus dados pessoais, são ou não objeto de tratamento; o direito de aceder aos seus dados; o direito do conhecimento das finalidades do tratamento dos dados; o direito de retirada do consentimento e o direito de ordenar a sua eliminação. O titular dos dados tem o direito de não ficar sujeito a nenhuma decisão tomada exclusivamente com base no tratamento automatizado, incluindo a definição de perfis, que produza efeitos na sua esfera jurídica ou que o afete significativamente de forma similar podendo requerer a intervenção de uma pessoa no processo.

Este regulamento entrou em vigor em maio de 2018 e foi posteriormente transposto para o ordenamento jurídico português pela aprovação da Lei n.º 58/2019 com alguma controvérsia pela desaplicação algumas normas constantes desta Lei. A Comissão Nacional de Proteção de Dados (CNPD), designada de Autoridade de Controlo, à qual é atribuída a função de fiscalizar a aplicação do RGPD, argumentam que algumas destas normas contradizem manifestamente o estatuído no regulamento, violando o princípio do primado da União Europeia preceituado na Constituição da República Portuguesa e pondo em causa a sua aplicabilidade direta, a sua eficácia e consistência da aplicação.

É importante referir que o RGPD não se aplica ao tratamento de dados recolhidos pelas FS. Neste contexto, é aplicável a Diretiva EU 2016/680, posteriormente adaptada pela Lei n.º 59/2019. Esta visou a extensão das garatias de protecção dos dados pessoais quando são tratados pelas autoridades competentes para efeitos de prevenção, investigação, deteção ou repressão de infrações penais ou execução de sanções penais, e à livre circulação destes dados.

Este regulamento e a respetiva lei visam estender as mesmas garantias de proteção previstas no RGPD quanto à licitude e segurança do tratamento de dados, ressalvando a principal particularidade que distingue os dois normativos: a limitação do direito de acesso, retificação ou apagamento dos dados. Estes direitos podem ser recusados ou restringidos na medida necessária e proporcional para evitar prejuízo para investigações, inquéritos ou processos judiciais, a prevenção, deteção, investigação ou repressão de infrações penais ou para a execução de sanções penais, para proteger a segurança pública, a segurança nacional ou proteger os direitos, liberdades e garantias de terceiros. Esta restrição não limita a possibilidade de o titular dos dados recorrer destas limitações junto da CNPD.

[34] 5) do Artigo 4.º do RGPD.

O RGPD constitui-se como um exemplo de regulamentação essencial para equilibrar a relação de poder entre o cidadão, os estados e as grandes empresas tecnológicas. No entanto, não é uma solução definitiva para todos os problemas relacionados com a temática da recolha e protecção de dados. Para a consciencialização do problema, contribuíram diversas personalidades que não podemos deixar de referir:

Edward Snowden expos a recolha indiscriminada e massiva de dados pelas Agências de Informações dos EUA. Que criaram as fundações para uma *"Eternidade Digital".*

Brittany Kaiser delatou a utilização abusiva de dados pessoais de utilizadores do *Facebook* pela empresa *Cambridge Analytica* expondo as suas capacidades de interferência na sociedade e na manipulação de massas.

Shoshana Zuboff que sintetiza no seu livro *"The Age of Surveillance Capitalism"* [35] o surgimento, evolução e domínio atual de um sistema económico de exploração dos nossos dados, e, por conseguinte, da nossa personalidade.

Jaron Lanier com os seus" Ten *Arguments for Deleting Your Social Media Accounts Right Now*" demonstra a forma como deixamos de ser unicamente clientes das nossas redes sociais e nos transformam em produtos.

Todos estes contributos são reveladores de uma capacidade de recolha e tratamento de dados inimaginável. Ficamos a compreender o verdadeiro poder das grandes tecnológicas e as suas implicações na sociedade, assim como a importância de regulamentação como RGPD.

[35] A Era do Capitalismo de Vigilância

2.6 A Inteligência Artificial

A IA está a reconfigurar a nossa sociedade, começando pela forma como nos relacionamos, a forma como trabalhamos e até a nossa economia, prometendo gerar ganhos de produtividade, reduzir custos e tornar todos os processos mais eficientes.

Idealmente, a IA contribui para uma vida melhor e perspetiva-se que, com a sua ajuda, as pessoas poderão fazer melhores previsões e tomar decisões mais informadas. Estas tecnologias ainda estão num estágio de maturação, sendo possível observar novas e melhores aplicações de IA ao nosso quotidiano a cada dia que passa.

Há grande expectativa de que a IA seja a chave para enfrentar vários desafios globais e promover a inovação e o crescimento. À medida que as aplicações da IA permeiam as nossas vidas, o seu poder transformativo deve ser colocado ao serviço das pessoas e do planeta e não somente das grandes empresas que as desenvolvem.

Como qualquer ferramenta, a IA pode ser aplicada a fins contrários aos do bem comum, logo, o seu desenvolvimento acarreta preocupações éticas que não podem ser relativizadas por benefícios económicos. Um dos principais pilares do problema é a confiabilidade dos sistemas de IA. Onde se inclui o perigo de codificar e reforçar preconceitos existentes relacionados com o género e com a raça, ou com a violação de direitos liberdades e garantias – como a privacidade ou liberdade de expressão ou credo.

Crescem as preocupações relativas aos sistemas de IA que exacerbam as desigualdades sociais, as alterações climáticas, o desemprego generalizado, a consolidação de mercados e o fosso digital entre países.

Nenhum país ou ator tem todas as respostas para estes desafios. Portanto, precisamos de cooperação internacional e respostas multissetoriais para orientar o desenvolvimento e o uso da IA para o bem comum.

2.6.1 Breve História da IA

Em 1950, o matemático britânico Alan Turing publicou um artigo sobre computadores e inteligência (TURING, 1950), colocando a questão se as máquinas poderiam pensar. Propôs que os computadores poderiam aprender com a experiência e adaptar o seu comportamento de forma inteligente, tal como uma criança (OLIVEIRA, 2019).

Para testar a sua hipótese, desenvolveu uma heurística simples: poderá um computador manter uma conversa e responder a perguntas de uma maneira que induza um ser humano a pensar que conversa com um humano? O teste de Turing, apesar das suas limitações em avaliar a inteligência de forma geral, foi modificado e é ainda hoje utilizado para avaliar *chatbots* ou assistentes pessoais digitais. Paralelamente, Claude Shannon propôs a criação de uma máquina que pudesse ser ensinada a jogar xadrez. A máquina podia ser treinada

usando o método de *brute force* – força bruta, mapeando todas as jogadas possíveis – ou avaliando um pequeno conjunto de movimentos estratégicos do oponente. Outro exemplo interessante é o *Logic Theory,* um programa de computador escrito em 1956 por Allen Newell, Herbert A. Simon e Cliff Shaw. Este foi o primeiro programa deliberadamente codificado para aplicar um raciocínio automatizado, sendo considerado o primeiro programa de IA (ANYOHA, 2017).

Muitos consideram que a IA nasceu na *Dartmouth Summer Research Project on Artificial Intelligence*, uma conferência que decorreu no verão de 1956 em *Dartmouth College*. Nesta conferência, o conceito de IA foi escalpelizado por John McCarthy, Alan Newell, Arthur Samuel, Herbert Simon e Marvin Minsky.

Os avanços na área de IA foram inconstantes nos últimos 60 anos. As promessas iniciais destes pais da IA provaram-se excessivamente otimistas. Este facto culminou no agora designado "*inverno da IA*", onde se verificou um corte de financiamento e redução do interesse na investigação de IA nos anos 70.

O "*inverno da IA"* terminou na década de 90, com o aumento do poder computacional e da capacidade de armazenamento de dados, que permitiu a resolução de tarefas complexas e demonstrar o verdadeiro potencial da IA (UW, 2006). Um destes exemplos surgiu em 1995, quando o investigador Richard Wallace desenvolveu a *Artificial Linguistic Internet Computer Entity* (A.L.I.C.E), que podia manter conversas básicas – precursor dos *Chatbots*. Também na década de 90, a IBM desenvolveu o computador *Deep Blue* que usou uma abordagem de *brute force* para jogar, por duas vezes, contra o Grão-Mestre de xadrez e campeão mundial da altura, Gary Kasparov. A sua grande capacidade computacional (11.38 Gigafolps[36]) conseguia prever seis ou mais passos futuros e podia calcular 330 milhões de posições por segundo. No segundo encontro derrotou o Grão-Mestre com o resultado de 3½ – 2½ (SOMERS, 2013).

Nos últimos anos, verificaram-se diversos desenvolvimentos: a grande disponibilidade de dados – *big data*; a computação em nuvem – *cloud computing;* a capacidade computacional; e a capacidade de armazenamento. Estes desenvolvimentos geraram grandes sinergias com as inovações técnicas de IA como a *machine learning* (ML) – traduzido como aprendizagem automática. Estes fatores aumentaram drasticamente o poder, a disponibilidade, o crescimento e o impacto da IA.

O progresso tecnológico contínuo também resulta no aperfeiçoamento de sensores que capturam dados mais confiáveis, imprescindíveis para o desenvolvimento e utilização pelos sistemas de IA. A quantidade de dados disponíveis para os sistemas de IA continua a crescer à medida que esses sensores se tornam menores e mais baratos de implementar, onde se

[36] Medida padrão de capacidade computacional que indica o número operações ajustadas de vírgula flutuante por segundo medido (Conselho da União Europeia, 2007).

podem enquadrar os dispositivos da IoT. O resultado é um avanço significativo em muitas áreas principais de pesquisa em IA, como o NLP, veículos autónomos, robótica, visão computacional, avaliações de risco, entre outras. Alguns dos desenvolvimentos mais interessantes da IA são aplicações em áreas fora da esfera da ciência da computação de forma direta, como na área da saúde, medicina, biologia e finanças.

2.6.2. A procura por uma definição

É indiscutível que a definição de IA tem sofrido evoluções desde a sua conceção. No entanto, os conceitos basilares permanecem constantes. Considera-se que a IA é um ramo da Ciência da Computação que tem como finalidade o estudo de *agentes inteligentes.* Estes são qualquer dispositivo que reconhece o seu ambiente e executa ações que maximizam a sua probabilidade de atingir os seus objetivos (POOLE, et al., 1998).

Enquanto disciplina científica fundada no pressuposto de que a inteligência humana "*pode ser descrita com tanta precisão que pode ser feita uma máquina para simulá-la*" (TURING, 1950), a IA inclui diversas abordagens e técnicas.

Os sistemas de IA, genericamente, são sistemas de *software* e de hardware concebidos por seres humanos que, tendo recebido um objetivo complexo, atuam na dimensão física ou digital. Tem sensores para percecionar o seu ambiente mediante a aquisição de dados e interpretam os dados recolhidos processando-os em informações e conhecimento. Este conhecimento permite decidir qual a melhor ação a adotar para atingir o objetivo estabelecido (EU High-Level Expert Group on Artificial Intelligence, 2019).

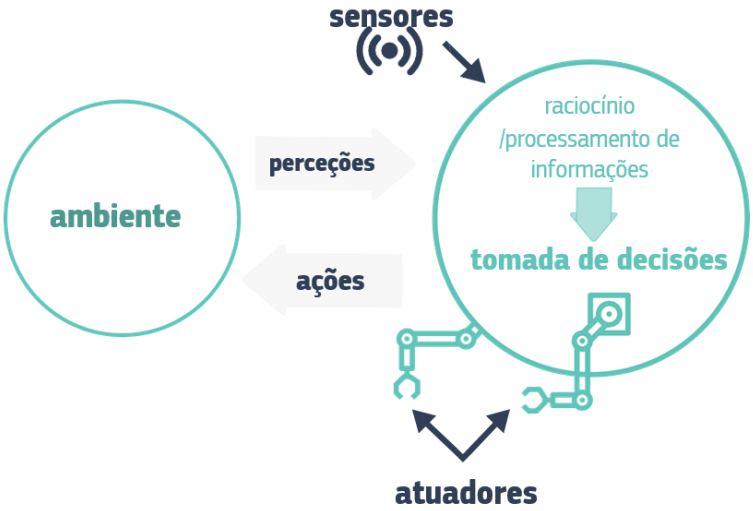

Figura 1 Representação de um sistema de IA.

Um sistema de IA (Figura 1)[37] é, de forma genérica, constituído por três elementos fundamentais: sensores, lógica operacional e atuadores. Os sensores recolhem dados em

[37] Fonte: EU High-Level Expert Group on Artificial Intelligence, 2019

bruto do ambiente em que se inserem, enquanto os atuadores agem para alterar o ambiente. A principal capacidade de um sistema de IA reside na sua lógica operacional. Consoante o conjunto de objetivos determinados e com base nos dados recolhidos pelos sensores, a lógica operacional toma decisões que se materializam pela ação dos atuadores. Estas decisões assumem a forma de recomendações, previsões ou decisões que podem influenciar o estado do ambiente (RUSSEL, et al., 2009).

O objetivo da IA é ensinar os computadores a fazer as coisas que atualmente os seres humanos fazem melhor, e aprender é incontestavelmente a mais importante dessas coisas: sem aprendizagem, nenhum computador consegue acompanhar um ser humano durante muito tempo (LEE, 2018). Nos últimos anos, os avanços significativos vieram da subdisciplina da IA, a ML e ainda de uma subespecialização designada *Deep Learning* (DL) – aprendizagem profunda. Esta concentra-se no ensino de máquinas pela aplicação de algoritmos aos dados. Muitas vezes, os termos AI, ML e DL são usados de forma indiscriminada e errónea.

Como podemos observar na Figura 2[38], a aprendizagem automática é um subcampo da IA, mas aos olhos do público a divisão pode não ser evidente. Isto acontece devido à grande corrida e mediatização destes termos para fins de mercado.

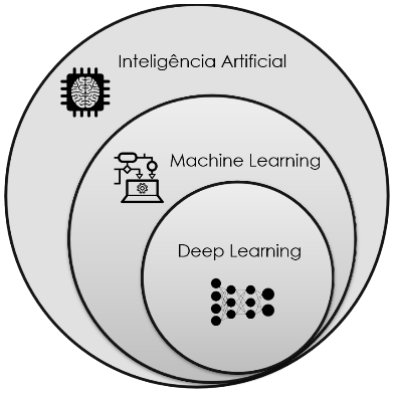

Figura 2 Ramos da IA

A aprendizagem automática assume diversas formas e é conhecida por muitos nomes diferentes: reconhecimento de padrões, modelação estatística, exploração de dados, descoberta de conhecimento, análise preditiva, ciência dos dados, sistemas adaptativos, sistemas auto-organizados, e muito mais. Cada um destes sistemas é usado por diferentes comunidades e têm diferentes associações (DOMINGOS, 2017).

O processo de ML materializa-se na aplicação de algoritmos a dados por forma a estabelecer regras que são representadas por um modelo.

Domingos explica que "*todos os algoritmos têm um input e output: os dados entram no computador; o algoritmo faz o que quer com eles, e o resultado sai. A ML inverte esta situação: entram dados e o resultado desejado, e sai o algoritmo que transforma os primeiros no segundo. Os algoritmos de aprendizagem são aqueles que criam outros algoritmos. Com a ML, os computadores escrevem os seus próprios programas, para que não tenhamos de ser nós a fazê-lo.*" (DOMINGOS, 2017).

Das diferentes abordagens de ML é possível identificar um subgrupo ímpar no que se refere aos resultados e a aplicações desta forma de aprendizagem.

Este subgrupo começou por ser designado como "redes neuronais", por ter adotado uma abordagem inspirada no cérebro humano. Os cientistas procuraram emular a estrutura

[38] Fonte: Elaboração própria

que consideramos fulcral para a inteligência humana, o neurónio. Esta abordagem mimetiza a arquitetura subjacente ao cérebro, construindo camadas de neurónios artificiais capazes de receber e transmitir informação numa estrutura aproximada das nossas redes de neurónios biológicas. Estes algoritmos são alimentados com exemplos de um dado fenómeno – imagens, jogos de xadrez, sons – e deixam que as próprias redes neuronais identifiquem padrões nos dados em causa. Por outras palavras, quanto menos interferência humana, melhor (LEE, 2018).

Para Lee (2018), as diferenças entre as duas abordagens podem ser observadas no modo como tratam um problema simples: verificar se aparece um gato numa dada imagem. A abordagem das regras tentará estabelecer modelos do tipo "se *x*, então *y*" para ajudar o programa a tomar uma decisão: "Se houver duas formas triangulares em cima de uma forma circular, então provavelmente há um gato na imagem." A abordagem das redes neuronais, pelo contrário, alimentará o programa com milhares de fotografias catalogadas como "gato" ou "não gato", deixando que o programa descubra por si quais as feições distintivas do rótulo "gato".

Com a evolução desta segunda abordagem e a entrada na era da informação as redes neuronais culminaram no DL. Esta abordagem requer uma vasta quantidade de dados de um domínio específico de forma a conseguir tomar uma decisão em condições otimizadas para a obtenção do objetivo desejado. O DL permite fazê-lo ao treinar-se para reconhecer padrões e correlações profundas que ligam muitos pontos informativos a um objetivo. Este processo de deteção de padrões é mais fácil quando os dados são rotulados com esse efeito desejado – "gato" ou "não gato". Pode então apoiar-se no seu modelo de correlações, muitas das quais invisíveis ou irrelevantes para o observador humano, para tomar decisões melhores do que as de um ser humano (LEE, 2018).

É importante distinguir outros dois conceitos de IA restrita ou fraca e IA geral ou forte. Considera-se um sistema de IA geral quando é concebido como sendo capaz de executar a maioria das atividades que os seres humanos conseguem realizar. Tem a capacidade de resolver a generalidade dos problemas que lhe sejam apresentados. Ainda subsistem muitos desafios de natureza ética, científica, tecnológica e de segurança em aberto no que respeita à criação das capacidades necessárias para a IA geral se tornar realidade. Designadamente o raciocínio de senso comum, a consciência, o melhoramento recursivo e a capacidade de a máquina definir os seus próprios objetivos. Devido à sua natureza complexa e incerta, o desenvolvimento de IA Geral não se prevê imediato, ficando por esse motivo excluído do escopo deste trabalho.

Os sistemas de IA restritos, pelo contrário, só conseguem executar uma ou poucas tarefas específicas. Atingem resultados surpreendentes por explorar a capacidade superior de um computador para processar grandes quantidades de dados, detetar padrões e

correlacionar eventos que seriam difíceis ou impossíveis de serem detetados por um ser humano. Assim, estes sistemas centrados em dados, são capazes de superar os humanos apenas em tarefas específicas (The Cylance Data Science Team, 2017).

2.6.3. Algoritmos e Modelos

Os computadores são constituídos por milhares de milhões de pequenos interruptores, designados de transístores, que os algoritmos podem ligar e desligar milhões de vezes nas designadas operações.

Ao combinar diversas operações, é possível estabelecer cadeias de raciocínio lógico elaborado. Um algoritmo é uma sequência de instruções que diz a um computador o que fazer (DOMINGOS, 2017). No caso da IA, os algoritmos de aprendizagem são desenvolvidos por programadores através de linguagens de programação – como Python, R e Java –, e diversas plataformas como *Tensorflow*, *Keras* e *Pythorch*. A aplicação destes algoritmos a dados de treino culmina na criação de um modelo que permite fazer predições ou tomar decisões sem que tenham sido explicitamente programados para isso.

O modelo é uma representação total ou parcial do ambiente externo do sistema que descreve as suas relações e dinâmicas. O processo de treino culmina na definição de um modelo numérico (equivalente a uma fórmula matemática) utilizado para calcular uma decisão a partir dos dados, depois de determinados os objetivos.

As principais formas de aprendizagem estão categorizadas de acordo com a relação que estabelecem com o seu ambiente e se os dados utilizados estão ou não rotulados. São elas: a aprendizagem supervisionada, a aprendizagem não-supervisionada e a aprendizagem por reforço.

Figura 3 Modelo de Treino Supervisionado

Os algoritmos de ML treinados (Figura 3)[39] com recurso a aprendizagem supervisionada enquadram-se com o exemplo anterior do detetor de gatos. Procura-se que o algoritmo aprenda por generalização a partir da análise de um conjunto de dados de treino rotulados. O algoritmo de aprendizagem infere uma função que mapeia uma entrada para uma saída com base em exemplos de pares entrada-saída, para fazer previsões sobre os valores de saída. Durante a análise dos exemplos (a fase de treino), as ponderações das ligações são ajustadas para corresponderem o mais possível ao que é mostrado nos exemplos disponíveis, ou seja, a minimizar o erro entre o resultado esperado e o que é obtido. Se os exemplos dados forem em número suficiente e suficientemente variados, incluindo a maioria das situações possíveis, obter-se-á um modelo treinado. No final da fase de treino, sujeita-se o modelo a uma fase de testes perante exemplos que nunca tenha visto anteriormente e verifica-se se a tarefa foi bem aprendida.

Por oposição, os algoritmos de ML treinados de forma não supervisionada utilizam dados de treino não rotulados com o objetivo de detetar padrões ou semelhanças entre os pontos de dados, com o mínimo de intervenção humana. Um dos principais métodos usados na aprendizagem não supervisionada é a análise de *clusters,* que permite agrupar ou segmentar conjuntos de dados com atributos compartilhados, a fim de extrapolar relacionamentos algorítmicos.

Os algoritmos de ML de reforço recorrem a um método de aprendizagem por experimentação. Permite-se que o agente interaja com o seu ambiente realizando ações aleatórias. O agente é recompensado caso as ações contribuam para cumprir o objetivo determinado pelo programador. O processo de treino leva a que o algoritmo determine automaticamente qual o comportamento ideal para maximizar a obtenção da recompensa (EU High-Level Expert Group on Artificial Intelligence, 2019).

[39] Fonte: www.pngfind.com/mpng/hooRJmb_7wdata-machine-learning-training-algorithm-hd-png-download/

Quando uma empresa anuncia que o seu novo produto recorre à tecnologia de IA, o produto emprega variantes destes algoritmos e modelos, depois de devidamente treinados para os fins comerciais para que são desenvolvidos. Por exemplo: análise e classificação de imagens médicas como apoio ao diagnóstico, gestão de fundos de investimento em bolsas e na avaliação de risco para obtenção de um seguro.

Um dos principais problemas atuais da IA é a sua reduzida transparência e interpretabilidade. Deste facto advém a necessidade de investimento no desenvolvimento de uma IA explicável. É frequente observar a analogia da IA se tratar de uma "Caixa Negra", comumente designada por *"Black Box Paradox"*. Podemos observar os *inputs* e os *outputs* dos sistemas de IA, no entanto, podemos não conseguir explicar ou justificar como chegaram ao resultado (Comissão Europeia, 2020).

Estes algoritmos e modelos variam consoante os dados e formas como são treinados. Procuram encontrar correlações probabilísticas complexas através do processamento de dados, com diversas dimensões e variáveis, que são codificados de forma autónoma. Desta forma, as diversas aplicações podem evoluir de forma imprevisível e as suas decisões podem ser difíceis de replicar. Estes atributos são pouco favoráveis em sistemas que interfiram com os DLG dos cidadãos (OECD, 2019).

Capítulo 3 – Aplicações de IA à Segurança

As aplicações de tecnologias de IA proliferam no mercado e nas nossas vidas. E a cada momento é revelado um novo produto revolucionário graças às capacidades da IA. Seja um telefone com capacidade de visão noturna, carros autónomos, robôs humanoides ou assistentes pessoais digitais que permitem controlar vários elementos da nossa vida.

Esta era de inovação que dá os seus primeiros passos pode ser equiparada ao surgimento da internet e todo o universo de avanços que se verificaram na década subsequente. Especialistas argumentam que também a IA será uma nova plataforma de desenvolvimento e de expansão.

Surge então a questão:

QD1 - Que sistemas de IA existem para apoiar as FS?

Existem sistemas de IA de apoio a áreas transversais de qualquer entidade – como a gestão de recursos humanos, gestão de processos internos, gestão do ambiente de trabalho – e sistemas exclusivos à segurança. Neste contexto estão em desenvolvimento e utilização, sistemas de: análise e tradução de voz para texto de telecomunicações, análise de texto para produção de informações, agentes virtuais para recolha de depoimentos ou receção de queixas, veículos autónomos terrestres e aéreos para patrulhamento de fronteiras, identificação de fraude fiscal, identificação de publicações de conteúdo proibido nas redes sociais, entre outros (INTERPOL & UNICRI, 2019). As FS desempenham várias missões, correspondendo cada uma delas a uma potencial aplicação de IA. Por este motivo, consideramos necessária uma redução epistemológica para três áreas de especial interesse e com desenvolvimentos promissores: a Cibersegurança, o Policiamento Preditivo e a Videovigilância com recurso a Reconhecimento Facial.

A área da Cibersegurança, por coabitar a esfera digital com a IA, observou desenvolvimentos notáveis e diversificados, em particular nas aplicações defensivas. Como exemplo, foram desenvolvidos: classificadores de *malware* com funções semelhantes a antivírus; sistemas de análise comportamental para deteção de uma ameaça interna; *Honeypots* com capacidade adaptativa e que simulam uma "organização fachada" digital para despistar ataques.

Outra aplicação de IA é o PP. Idealizado pela ficção científica, implementado em diversos países de formas diferentes, o PP procura antecipar ou prever o acontecimento de crimes, pessoas em risco de cometer crimes, identificar e caracterizar criminosos e ainda identificar pessoas em risco de serem vítimas de crime.

Por fim, as aplicações da IA à Videovigilância com recurso ao Reconhecimento Facial, que se enquadram na disciplina de reconhecimento biométrico. Esta é uma das aplicações mais promissora da IA para fins de segurança. Atingiu uma capacidade vastamente superior

à dos humanos. No entanto, é a aplicação que, em potência, poderá ser mais disruptiva para uma sociedade tipicamente ocidental.

3.1. Cibersegurança

Alguns especialistas preveem que até ao final do ano de 2025 haverá 75 mil milhões de dispositivos conectados[40]. Cada vez mais surgem dispositivos inteligentes que efetivamente mudam a maneira como vivemos, como nos comportamos e interagimos com o mundo que nos rodeia. À medida que a tecnologia se enraíza e se integra nas nossas vidas, tornamos-mos cada vez mais dependentes dela. Mas essa dependência torna-nos vulneráveis na eventualidade da tecnologia falhar, em especial, quando provocada por ações criminosas.

No mundo de hoje, é importante que a tecnologia esteja disponível, protegida e segura. Caso contrário, será uma questão de tempo até sermos vítimas de um ciberataque. Tal facto verifica-se agora em Portugal, onde são também comuns os relatos de ataques a empresas na área das telecomunicações[41], energia[42], grupos de saúde[43] ou mesmo a administração pública[44].

Qualquer tecnologia apresenta fragilidades, mas pelo facto destes novos dispositivos estarem ligados em rede, expõe estas vulnerabilidades à vontade dos criminosos em qualquer parte do mundo. Em Cibersegurança é verdade absoluta que nenhum dispositivo ou organização é categoricamente seguro, é um processo dinâmico, em constante evolução, quer preditiva quer reativa. A Cibersegurança é fundamental para garantir o estabelecimento de medidas de segurança, geralmente dispostas em profundidade[45], e para a adoção de procedimentos que elevem a resiliência do dispositivo ou da organização a um nível de segurança adequado ao objeto a proteger. A norma ISO / IEC 27032 da Organização Internacional para a Standartização define Cibersegurança como a preservação da confidencialidade, integridade e disponibilidade da informação no ciberespaço. Por sua vez, o ciberespaço é definido como o ambiente complexo resultante da interação de pessoas, *software* e serviços na Internet por meio de dispositivos e redes. A ENISA[46], Agência da União Europeia para a Cibersegurança que visa reforçar as capacidades de Cibersegurança da UE e dos seus Estados Membros – desenvolvendo políticas, processos, serviços e capacidades

[40] https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/
[33] https://jornaleconomico.sapo.pt/noticias/altice-portugal-foi-alvo-de-ataque-informatico-mas-consequencias-foram-praticamente-nulas-576828
[42] https://www.dn.pt/pais/edp-alvo-de-ataque-informatico-12066017.html
[43] https://www.dn.pt/sociedade/hospitais-e-centros-de-saude-sem-acesso-a-e-mail-para-evitar-ataque-informatico-8477221.html
[44] https://www.cmjornal.pt/portugal/detalhe/junta-foi-alvo-de-ataque-informatico
[45] *Defense in depth*: Implementação de diversas tecnologias e procedimentos que aumentam a resiliência de um sistema, por ex: *Firewalls*, IDS - *Intrusion Detection System* – sistemas de deteção de intrusão, IPS – Intrusion Prevention System – sistemas de prevenção de intrusão, *antivírus*, e SIEMs - *Security Information and Event Management* – sistema de gestão de eventos de segurança, etc
[46] https://www.enisa.europa.eu/

para responder aos desafios relacionados com a Cibersegurança –, refere que a Cibersegurança abrange cinco domínios:

A proteção das comunicações, que se traduz na proteção contra uma ameaça à infraestrutura técnica de um sistema que pode levar a uma alteração das suas características, a fim de desempenhar tarefas fora do âmbito para que foram adquiridos pelos seus proprietários ou utilizadores;

A proteção das operações, que consiste em evitar alterações ao normal funcionamento, procedimentos ou fluxos de trabalho, que alterem os resultados pretendidos pelos seus proprietários, designers ou utilizadores;

A segurança da informação, que se resume na proteção contra ameaça de roubo, eliminação ou alteração de dados armazenados ou transmitidos num sistema digital;

A segurança física, proteção contra ameaças físicas que podem influenciar ou afetar o bem-estar de um sistema informático, tendo como exemplos: o acesso físico a servidores, introdução de hardware malicioso numa rede ou a coerção de colaboradores e familiares;

E ainda a segurança pública, que se considera a proteção contra uma ameaça cuja origem é do ciberespaço, mas que pode ameaçar ativos físicos ou ciberfísicos, de forma a proporcionar um ganho político, militar ou estratégico para o atacante, como por exemplo um ataque a infraestruturas críticas (ENISA, 2015).

Considerando os recursos limitados que as empresas dispõem para a Cibersegurança, não é possível que uma organização consiga eliminar todos os vetores de ataque possíveis. A crença de que se pode prevenir os ataques concentrando-se exclusivamente nos objetos ou bens que procuramos proteger tem-se manifestado insuficientes.

A IA poderá ser parte do problema e também da solução. Existem aplicações para aumentar a segurança dos dispositivos, sistemas e aplicações (CYLANCE, 2018). Em contrapartida, também poderá passar a integrar o arsenal de ferramentas ao dispor de atores maliciosos que podem recorrer às capacidades da IA para personalizar e automatizar a distribuição de conteúdo malicioso, ataques de engenharia social ou e-mails de *phishing*[47]. Desta forma, os ataques que recorrem a IA serão altamente personalizados, independentemente da escala da operação. Este m*alware*[48] será capaz de aprender maneirismos, nuances comportamentais e o idioma do alvo, analisando as comunicações por e-mail e redes sociais. Poderá usar este conhecimento para replicar o estilo de escrita do

[47] *Phishing* - É uma tática de engenharia social usada para convencer um indivíduo ou utilizador, a fornecer informações confidenciais e/ou tomar uma ação através de comunicações aparentemente seguras e confiáveis. Os atores mal-intencionados empregam técnicas de *phishing* por vários motivos, incluindo o roubo de identidade, acesso a informações confidenciais, transmissão de *malware* (incluindo o *ransomware*), acesso remoto não autorizado, e realização de transações financeiras não autorizadas. (Homeland Security, 2018 p. 2)

[48] *Malware. – Software* com propósitos malignos, seja a comprometer a privacidade dos utilizadores ou destruir informação. É um termo muito abrangente que engloba um conjunto de características presentes em software malicioso tais como: recolha de informação privada, acesso a recursos restritos ou apenas para causar danos. (FERREIRA, et al., 2013 p. 17)

utilizador, criando mensagens que parecem altamente credíveis. As mensagens escritas por *malware* com recurso a AI serão, portanto, quase indistinguíveis das comunicações genuínas (DIXON, et al., 2019).

Agentes mal-intencionados podem recorrer a algoritmos de IA personalizados para automatizar o processo de descoberta de novas vulnerabilidades ou vetores de ataque. Estes algoritmos podem também ser utilizados para apoiar a seleção e a priorização de alvos, e ainda reagir e adaptar-se às alterações no comportamento do alvo (BRUNDAGE, et al., 2018 p. 25).

Uma possível aplicação com resultados interessantes é a aplicação de ML aos dados de utilização de uma organização. O domínio de segurança gera enormes quantidades de dados, provenientes de *logs*, sensores de rede e SIEM's, bem como de sistemas de gestão de acessos e sistemas de gestão de privilégios que indicam quais as ações que são permitidas a cada utilizador. Coletivamente, esta massa de dados pode fornecer as pistas contextuais necessárias para identificar e isolar as ameaças. Este é precisamente o tipo de processamento em que a ML pode ser um fator diferenciador. Aplicada corretamente, a ML pode fornecer o contexto que necessitamos para produzir um modelo padrão ou de normal funcionamento da organização, permitindo gerar alertas baseados em discrepâncias entre a ação do utilizador e a ação determinada expectável pelo algoritmo. Este modelo pode ser transposto para dezenas de aplicações diferentes, dependendo do tipo de dados e objetivos do sistema (The Cylance Data Science Team, 2017).

Esta abordagem é sem dúvida interessante e algumas empresas, como a *Darktrace*[49] *e Tanium*[50] estão a avançar com produtos baseados em modelos produzidos com recurso aos dados do seu cliente.

Outras duas aplicações comerciais de IA vocacionadas para a Cibersegurança, estudadas no âmbito do trabalho de campo – são o HP *Sure Sense* da *Hewlett Packard* e a Plataforma *X-Force Exchange* da IBM.

O HP *Sure Sense* procura solucionar um problema bastante conhecido, o *malware* ou *malicious software,* que inclui todas as formas de *software* malicioso, tais como: vírus, cavalos de troia, *spyware*, *adware* e *ransomware*. *Softwares* que quando introduzidos num sistema, interrompem ou bloqueiam a sua capacidade de operar, recolhem informações que podem ser sensíveis e que resultam na perda de privacidade. Podem também obter acesso não autorizado a recursos do sistema para a disseminação do vírus, para integrar *botnets*, para minerar *bitcoin* ou para outros fins abusivos.

A abordagem tradicional dos antivírus (AV) para detetar *malware* consiste na comparação de uma assinatura digital – designada *hash* – extraída do ficheiro suspeito com

[49] www.darktrace.com/en/
[50] www.tanium.com/

as assinaturas de amostras de *malware* conhecido. De acordo com o AV Test[51], mais de 350.000 novos tipos de *malware* são criados todos os dias. Este modelo de comparação de assinaturas obriga à criação de vastas bases de dados e atualizações permanentes. Para além disso, as formas avançadas de *malware* de hoje-em-dia podem mudar dinamicamente a sua assinatura digital para evadir a sua deteção, recorrendo a permutação de código, alteração do registo ou introdução de código aleatório.

Estes fatores culminam num desafio sério para a Cibersegurança. Empresas como a HP[52], *Cylance*[53] e *Patternx*[54] começaram a desenvolver *software* de IA com o objetivo de detetar, investigar, classificar e mitigar os tipos mais avançados de *malware* de forma preventiva e em tempo real.

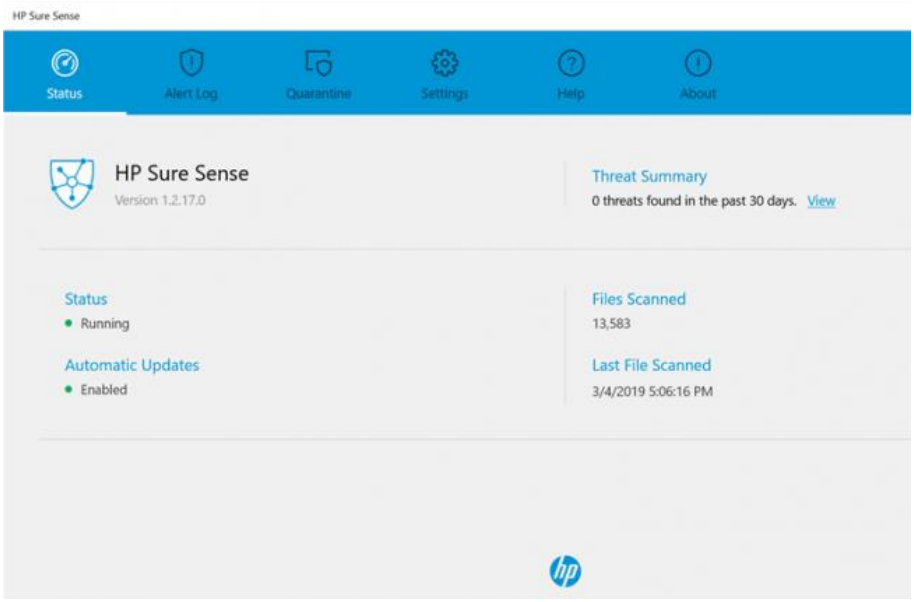

Figura 4 Agente HP Sure Sense

O HP *Sure Sense*[55] (Figura 4)[56] recorre a um modelo da IA baseado em DL para prevenir e detetar *malware* em tempo real, incluindo aqueles cuja assinatura é desconhecida.

Recorre a um modelo (Figura 5)[57] de deteção previamente treinado com dados de centenas de milhões de ficheiros, classificados como seguros ou suspeitos. Durante o processo de treino, os algoritmos definem as características ou atributos que diferenciam um ficheiro seguro de um malicioso. O resultado do treino é modelo de previsão de IA com funções semelhantes a um classificador, como vimos no capítulo anterior, e que pode ser distribuído e integrado noutras aplicações (SOUTO, E10).

---

[51] www.av-test.org/en/statistics/malware/
[52] https://press.ext.hp.com/us/en/press-releases/2019/hp-transforms-pc-security-with-AI-driven-hp-sure-sense.html
[53] www.cylance.com
[54] www.patternex.com
[55] Desenvolvido em parceria com a empresa DeepInstict https://www.deepinstinct.com/
[56] Fonte: https://h20195.www2.hp.com/v2/GetDocument.aspx?docname=4AA7-6875ENW
[57] Fonte: https://info.deepinstinct.com/datasheet_windows

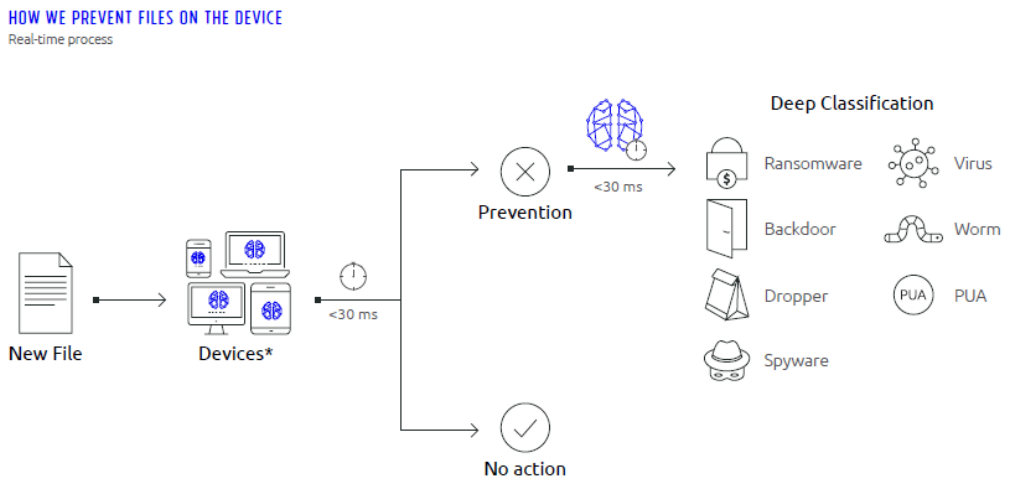

Figura 5 Modelo de Classificação

Depois de distribuído, qualquer novo ficheiro é verificado pelo agente e recebe uma pontuação. A pontuação representa o nível de maliciosidade do ficheiro. Caso esse valor ultrapassar o patamar estabelecido para um ficheiro seguro, o agente impede a execução do mesmo. Para além de classificar se é seguro ou não, também classifica em tempo real o tipo de ameaça ou família de *malware* (SOUTO, E10).

A *DeepInstict* refere que o seu agente tem uma taxa de prevenção de 100% e 0 falsos positivos num teste realizado pela SELabs um laboratório certificado para testes de *software* de Cibersegurança (SE Labs, 2018). No entanto não foi possível confirmar este facto com outras fontes.

A vantagem destes novos AV está na capacidade de resposta a ameaças zero-days graças a um modelo de classificação robusto e devidamente treinado e independente de assinaturas.

Miguel Souto salienta que estes sistemas de defesa não devem de substituir por completo os sistemas tradicionais, devem sim ser integrados numa *framework multi-layer*, garantindo assim uma arquitetura mais eficiente no combate (SOUTO, E10).

Outra aplicação, o IBM *X-Force Exchange*. A IBM apresenta desenvolvimentos na área de IA: os seus supercomputadores – como o *DeepBlue* ou o *Summit*[58] – e pelo IBM *Watson*[59], que tem diversas aplicações: saúde, finanças, assessoria jurídica, retalho, entre outras.

O IBM *X-Force Exchange* é uma ferramenta de partilha de informação sobre ameaças à Cibersegurança, comumente designada *threat intelligence*. Esta pode ser definida como a recolha de indícios sobre ciberataques, seja o seu contexto, mecanismos

[58] De 11.4 gigaFLOPS ou mil milhões do DeepBlue para os 1.88 exaFLOPS ou triliões do *Summit* https://www.olcf.ornl.gov/2018/06/08/genomics-code-exceeds-exaops-on-summit-supercomputer/
[59] https://www.ibm.com/watson

do ataque, ações recomendadas ou outras informações necessárias para aumentar o nível de conhecimento da ameaça em apoio aos elementos dos *Security Operations Center* (SOC)[60]. A *threat intelligence* pode materializar-se em dados estruturados e dados não estruturados. Os dados estruturados geralmente incluem informações como o nome dado ao *hack*, a assinatura digital, dados de *logs* e endereços IP associados. Os dados não estruturados podem incluir elementos como uma descrição das etapas do ataque, notas do analista, comentários partilhados, informações históricas, artigos de investigação, blogs, boletins de segurança e ainda conselhos de resolução (COOMBS, 2018).

A informação partilhada nesta plataforma é gerada pelo *Watson for Cybersecurity*[61]. Esta aplicação permite por um lado receber dados estruturados de fontes – indicadores de comprometimento, relatórios de eventos ou de incidentes, e por outro lado, dados não estruturados – como relatórios de investigação, Blogs de segurança, *websites,* entre outras (RIBEIRO, E11).

A aplicação da IA foi vocacionada para usar algoritmos de NLP de forma a consumir todos esses dados, processá-los e produzir informação relevante para investigações de incidentes de Cibersegurança. Na sua essência, esta ferramenta é como um especialista em segurança que recolhe informação permanentemente como um *webcrawler* com memória perfeita (RIBEIRO, E11).

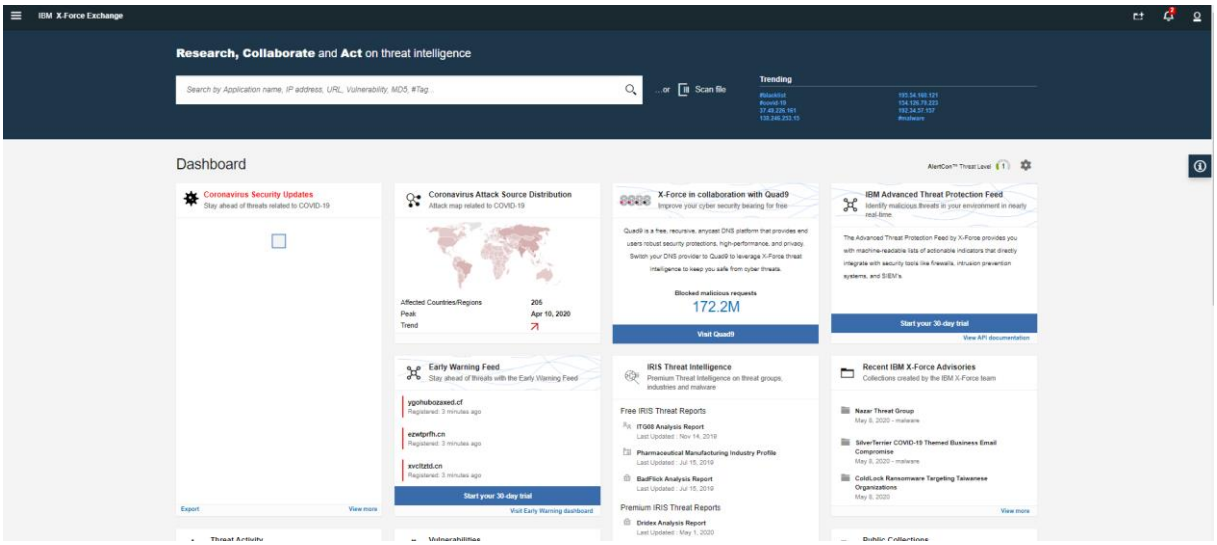

Figura 6 Portal X-Force Exchange

Todo este conhecimento é partilhado pelo portal (Figura 6)[62] ou integrando em diversas ferramentas de segurança através da sua API[63]. No caso da IBM, permite a

[60] O SOC é célula dentro da organização que é responsável pela monitorização e melhoramento da postura de segurança dessa organização. Operacionalmente visa evitar, detetar, analisar, responder, reportar e recuperar de incidentes de Cibersegurança, apoiando-se em tecnologia, processos e procedimentos bem definidos.
[61] www.ibm.com/security/artificial-intelligence
[62] Fonte: https://exchange.xforce.ibmcloud.com/
[63] *Application programming interface* ou Interface de Programação de Aplicações permite partilha de funcionalidades ou serviços de determinada aplicação de forma pública ou privada. Um exemplo, caso

integração do *X-Force Exchange* na SIEM nativa da IBM para controlo de eventos de segurança – *QRadar*.

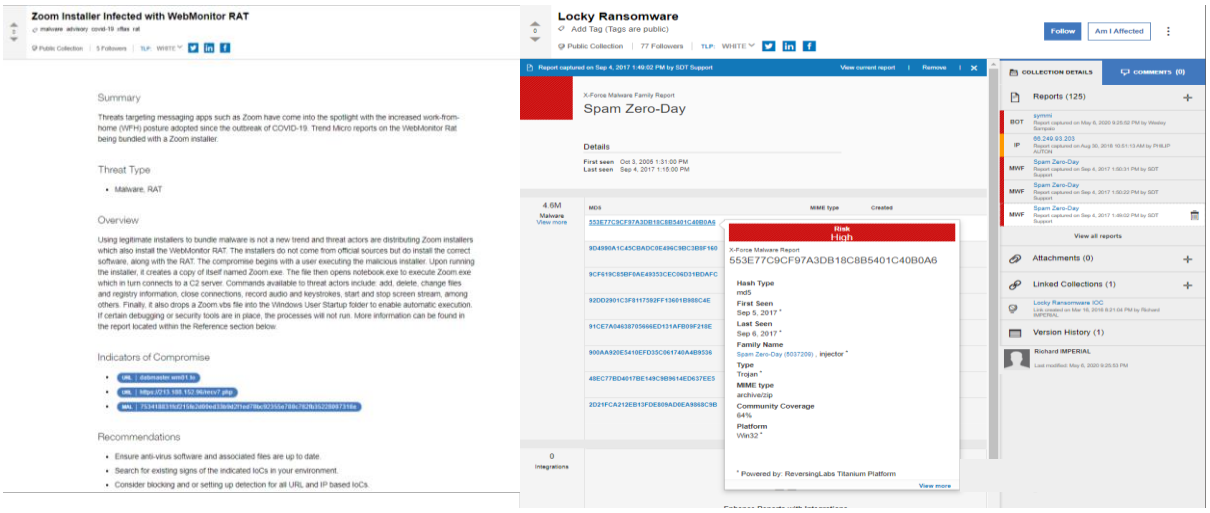

Figura 7 Exemplo de Relatório de Informação

Na Figura 7 verificamos a informação geralmente disponibilizada nos relatórios da plataforma: um sumário da vulnerabilidade ou ameaça, uma descrição detalhada, indicadores de comprometimento, recomendações e até assinaturas digitais do *malware* em causa que poderá ser utilizado para a sua deteção pelos AV.

Para Rui Ribeiro a automação de processos defensivos é um fator diferenciador dos sistemas de IA aplicados à Cibersegurança. Desta forma, é possível reduzir o tempo de resposta a incidentes e acionar os mecanismos de continuidade de negócio de forma mais célere. A automação das etapas dos processos tem de ser feita com supervisão humana e de forma metódica, para evitar que estes processos se tornem em vulnerabilidades (RIBEIRO, E11).

Os sistemas de segurança da informação dispõem as suas defesas em profundidade, as ferramentas apresentadas complementam os sistemas atuais. Carecem de integração com os sistemas em utilização para obtermos os ganhos de rentabilidade proporcionado pela IA.

Estes exemplos representativos dos desenvolvimentos na área de Cibersegurança com recurso a IA perspetivam um progresso contínuo, uma vez que o problema do cibercrime continuará também ele a evoluir. Estas ferramentas contribuem para a segurança das organizações, por aumentarem a proteção dos sistemas de informações através da deteção de *malware* mais eficazmente e por facilitarem a disseminação de conhecimento entre elementos que lidam diariamente com as ameaças do ciberespaço.

pretenda apresentar a localização de um restaurante no seu site poderia usar a API da GoogleMaps para gerar uma janela interativa com a localização do restaurante.

3.2. Policiamento Preditivo

A capacidade de prevenir ou impedir a ocorrência de um crime de forma antecipada é algo que pode alterar o rumo da sociedade ou até mesmo da humanidade. As implicações sociais, demográficas e económicas de uma tecnologia que fosse capaz de prever com precisão a hora, local e autor de qualquer crime são de difícil compreensão. No entanto, devem ser estudadas e acauteladas, uma vez que já existem tecnologias que se podem considerar precursores desta realidade. O atual desenvolvimento desta tecnologia levanta questões éticas, de transparência e responsabilização, pois a sua aplicação influencia diretamente os DLG dos cidadãos.

O PP pode ser definido pela aplicação de técnicas analíticas - principalmente técnicas quantitativas - para prevenir a ocorrência de crimes, identificar potenciais vítimas ou criminosos através de previsões estatísticas (Rand Corporation, 2013 p. 2). Complementando e segundo Ratcliffe: "*O policiamento preditivo, é o uso de dados históricos para criar uma previsão espácio-temporal de áreas de criminalidade ou pontos críticos, que servirão de base para a tomada de decisão relativa à alocação de recursos, na expectativa de que os efetivos policiais estarão no momento e no local das ocorrências criminais*." (RATCLIFF, 2016 p. 151).

A análise de dados estatísticos e geoespaciais para fins de prevenção criminal não é uma técnica recente, existindo registos que remontam a 1829 com a criação de um mapa com informação sobre crimes passados, educação e situação socioeconómica dos residentes, desenvolvido por Adriano Balbi, um geógrafo italiano, e Andre-Micher Guerry, um matemático francês (KOBAYASHI, 2020 p. 12). Nos últimos anos, foi possível observar um aumento do interesse, da procura e do investimento em ferramentas analíticas que utilizam grandes conjuntos de dados, ou *big data*, para fazer previsões em apoio à prevenção criminal. A utilização destas ferramentas torna a polícia mais dependente da infraestrutura da tecnologia de informação para recolher, analisar, armazenar e partilhar esses conjuntos de dados. Outro fator que aumenta esta dependência centra-se no facto das FS não terem capacidade de serem autossustentáveis nesta matéria, recorrendo necessariamente a parcerias com universidades ou soluções comerciais (Rand Corporation, 2013 p. 2).

Este facto verifica-se pelo número limitado de cientistas de dados, programadores e gestores de projeto com conhecimentos nesta área dentro das FS. Aliás, há uma escassez internacional deste tipo de recursos após a absorção dos mesmos pelas grandes empresas tecnológicas. Tal limitação deverá ser colmatada com parcerias e programas de desenvolvimento conjunto com a comunidade académica.

Apesar destas limitações, as aplicações de PP surgiram de diversos contextos com o objetivo de aumentar eficácia e eficiência das FS, procurando passar de uma postura reativa para uma atitude proativa e potenciando os seus recursos limitados, como explica Pedro Domingos: *"Ao prever tendências de crime e ao concentrar patrulhas estrategicamente onde*

é mais provável que estas sejam necessárias, bem como tomando outras medidas preventivas, a força policial de uma cidade pode fazer na prática o trabalho de uma outra muito maior." (DOMINGOS, 2017 p. 45).

Segundo o relatório da Rand: "*O objetivo destes métodos é desenvolver estratégias eficazes que impeçam o crime ou tornem os esforços de investigação mais eficazes. No entanto, deve ser entendido em todos os níveis que a aplicação de métodos preditivos de policiamento não é equivalente a encontrar uma bola de cristal. Para que uma estratégia de policiamento seja considerada eficaz, deve produzir resultados tangíveis."* Concluímos assim que a capacidade de fazer previsões deve ser apenas considerada como parte da solução. Cabe às FS adotar a melhor tática policial para intervir na situação ou área referenciada pela ferramenta.

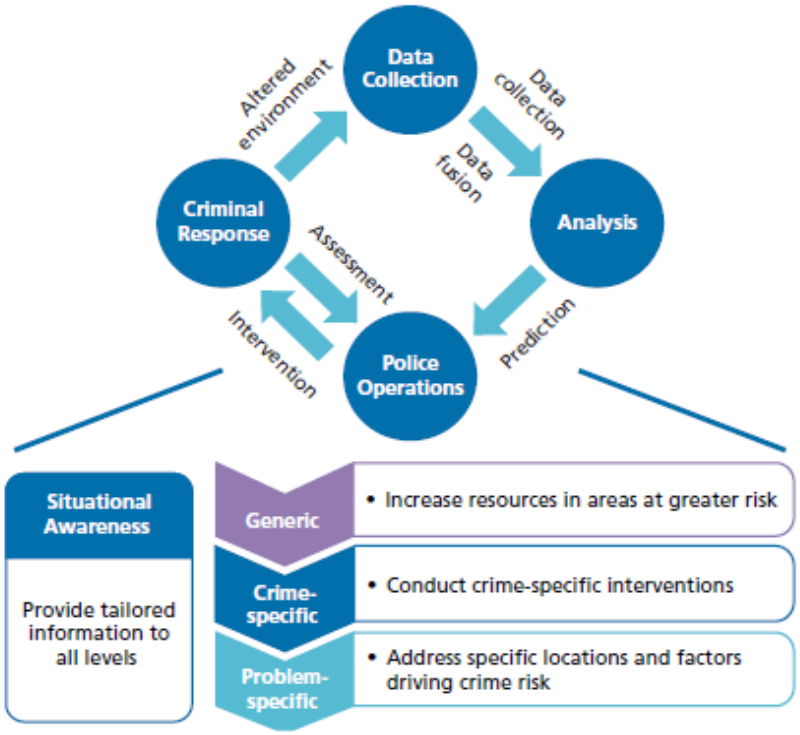

Figura 8 Processo de Gestão do Policiamento Preditivo

O processo de gestão do PP (Figura 8)[64] pode ser representado por um ciclo de quatro etapas. A primeira etapa envolve a recolha e fusão de dados criminais[65], incidentes e infratores para a segunda etapa, a análise. Nesta segunda fase, os dados são processados pelo algoritmo preditivo da aplicação podendo introduzir ajustes no modelo previamente treinado.

[64] Fonte: (Rand Corporation, 2013 p. xviii)
[65] Dependendo da aplicação em causa, podem ser utilizados dados específicos da comunidade em causa, concretamente infratores conhecidos que vivem na comunidade, indicadores socioeconômicos, época do ano, clima ou eventos específicos da comunidade entre outros (FERGUSON, 2019 p. 496).

O terceiro passo é atuar com base na predição efetuada, empregando a tática policial mais adequada à situação. Estas ações podem ser: genéricas, pelo aumento dos recursos alocados (ex.: reforço de patrulhamento), específicas em relação ao crime (ex.: reforço de vigilâncias) ou vocacionadas para o problema (pela atuação sobre os fatores motivadores do crime). Independentemente do tipo de intervenção em causa, é fundamental que seja disponibilizada informação suficiente para resolver a situação de forma eficaz, ressalvando a importância de complementar o conhecimento situacional de quem reage às situações. Idealmente, estas intervenções devem reduzir a criminalidade ou ajudar na resolução de problemas, o que constitui o quarto passo, a resposta criminal. Esta alteração do ambiente espoleta um novo ciclo de gestão do PP. A avaliação da aplicação de um modelo preditivo deve ser uma constante. Inicialmente, deve-se garantir que os procedimentos são executados em conformidade com objetivo do programa, sendo que a longo prazo deve ser avaliada a resposta criminal pela alteração das tendências e dados iniciais do modelo preditivo (Rand Corporation, 2013 p. xviii).

Dois exemplos de programas de PP, que se proporcionam como bons objetos de estudo, designadamente o *PredPol* e o *Hunchlab*. Estes dois produtos foram desenvolvidos nos EUA e categorizam-se como plataformas de gestão de patrulhamento e de comando e controlo (C2) [66]. A escolha justifica-se por serem os produtos mais empregues pelas FS, em especial nos EUA e no Reino Unido.

O *PredPol*, o nome mais reconhecido na área, resultou de um projeto de investigação entre o Departamento de Polícia de Los Angeles (LAPD) e a Universidade da Califórnia, Los Angeles (UCLA). O diretor da LAPD, William Bratton[67], procurava rentabilizar a COMPSTAT[68] para além da sua função de estatística histórica. O objetivo era entender se esses dados poderiam fornecer recomendações prospetivas sobre onde e quando os crimes poderiam ocorrer. Ser capaz de antecipar quais os locais e horários mais propícios para a ocorrência de crimes poderia permitir que se destacassem agentes, que pela sua presença e conduta, ajudassem a prevenir esses crimes.

O grupo de trabalho incluiu matemáticos e investigadores comportamentais da UCLA e da Universidade de Santa Clara, que avaliaram diversos tipos de dados e modelos de previsão comportamental. Juntaram-se também à equipa analistas criminais e agentes da LAPD e do Departamento de Polícia de Santa Cruz (Califórnia). Finalmente, determinaram que os três

[66] C2 - conjunto de capacidades e processos organizacionais que permitem a alocação de recursos humanos, físicos e de informação para resolver problemas e cumprir os objetivos de uma organização (VASSILIOU, et al., 2015).
[67] Também conhecido pela aplicação da teoria das janelas partidas (FAGAN, et al., 2001).
[68] Derivado de *Computer Statistics*, é um sistema multifacetado de gestão de operações policiais. É uma ferramenta de analise do crime e seus efeitos na comunidade, e ao mesmo tempo, e apoia a organização na implementação das melhores práticas na gestão de recursos humanos e gestão do risco.

pontos de dados mais objetivos recolhidos pelos departamentos de polícia são: o tipo de crime, o local do crime e a data e hora do crime (PredPol, 2020).

A política de utilização de dados da *PredPol* rejeita a utilização de dados pessoais, demográficos, étnicos ou socioeconômicos, reduzindo assim a possibilidade de violações da privacidade ou dos DLG observados noutros modelos de PP. Esta abordagem minimalista limita as variáveis (e o ruído) na análise dos dados e evita o reforço de práticas discriminatórias, características de alguns modelos preditivos, como veremos no capítulo seguinte. A fórmula (Equação 1) que representa o algoritmo da *PredPol* está disponível no seu *website*[69].

$$\frac{\partial A}{\partial t} = B + \frac{nD}{4} + \frac{nD}{4}\nabla^2 A - \omega A + \theta\omega\delta$$

Equação 1- Algoritmo patenteado PredPol

Para as FS, as previsões baseadas em dados são operacionalizadas por meio de mapas gerados por computador, assinalando as áreas com probabilidade elevada de ocorrência de crimes.

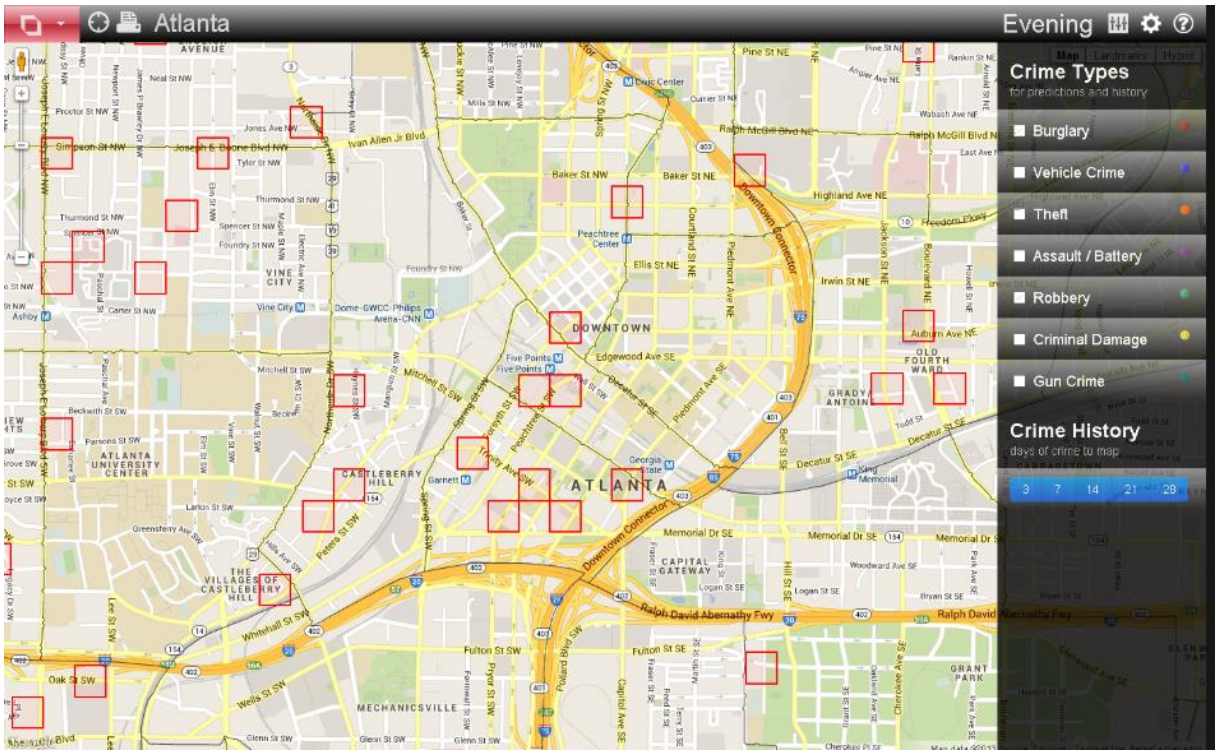

Figura 9 Dashboard do PredPol

Estes mapas (Figura 9)[70], acedidos através de um computador ou *smartphone,* tendem a prever áreas precisas (com 150 m^2) e representam um risco elevado de um crime em particular. Em geral, cada turno é aconselhado a patrulhar estas áreas sempre que possível. A estratégia é aumentar a presença policial em determinados locais e em determinados

[69] www.predpol.com/technology/
[70] Fonte: www.wabe.org/concerns-arise-over-new-predictive-policing-program

horários previstos. O *PredPol* pode cronometrar o tempo despendido em patrulhamento numa área com recurso a GPS e permite também acompanhar as patrulhas por uma sala de situação.

A teoria subjacente tem por premissa que o patrulhamento de visibilidade em áreas de maior probabilidade de ocorrência de delitos deterá a sua ocorrência por dissuasão (FERGUSON, 2019 p. 493).

Um relatório independente de controlo randomizado avaliou a capacidade do *PredPol* comparativamente às capacidades de analistas criminais de três departamentos da LAPD e um da Polícia de Kent, no Reino Unido. O estudo concluiu que a ferramenta consegue prever entre 1.6 a 2.5 vezes mais crimes que o analista humano, correspondendo a uma redução ao fim de seis meses nos crimes violentos em Kent de 6%, de crimes contra a propriedade de 12% e de assaltos de 25% em Los Angeles (PredPol, 2019).

A *PredPol* define-se como uma ferramenta que "recorrendo a aritmética avançada e dados relativos a crimes em tempo real, avalia os crimes de ontem no contexto de todos os crimes que ocorrem durante um longo horizonte de tempo e amplo campo espacial para calcular as probabilidades precisas de onde e quando o crime ocorrerá hoje. As FS podem fazer uso destas informações para dissuadir a ocorrência de crimes nessas áreas levando a uma redução geral da criminalidade." (PredPol, 2019).

Também no mesmo documento, a *PredPol* afirma que não recorre a perfis criminais, nem utiliza qualquer informação sobre indivíduos ou populações e suas características. Os padrões inerentes aos próprios crimes fornecem informações suficientes para prever onde e quando os crimes ocorrerão no futuro. O PP interrompe as causas situacionais a curto prazo do crime. Não resolve a criminalidade ou a propensão para os indivíduos cometerem crimes. O PP não substitui as estratégias de participação política e comunitária necessárias para a integração na sociedade das pessoas mais suscetíveis de incorrerem na prática de crimes (PredPol, 2019).

Outra aplicação é a *HunchLab, que começou a ser* desenvolvido pela empresa *Azavea* em 1998 com um protótipo por Rober Cheetham em colaboração com o Departamento de Polícia de Filadélfia (PPD) nos EUA. Era um sistema de alerta precoce, ou *early warning system,* com base nos dados e estatísticas. Usado para detetar picos localizados de atividade criminal, compará-los com dados históricos e gerar informações para as FS (CHEETHAM, 2019).

Em 2013, desenvolveram uma nova abordagem estatística para prever crimes. Tiveram a colaboração de académicos como Jerry Ratcliffe[71] e Ralph Taylor[72] da *Temple University* de Filadelfia*,* EUA na construção do modelo para as tendências criminais de médio a longo prazo

[71] Investigador (TU) responsável pelo estudo e implementação de diversos programas de POI.
[72] Investigador do modelo de policiamento comunitário.

com base em indicadores demográficos. E ainda com Joel Caplan e Leslie Kennedy da *Rutgers University* de Nova Jérsia, EUA – para automatizar o processo de modulação da relação risco-terreno, *Risk Terrain Modeling.* Este é um modelo explicativo da emergência de ocorrências criminais em localizações específicas, com base na natureza dessas localidades e proximidade com outros locais – como bares, paragens de autocarros e casas devolutas.

Posteriormente, foi desenvolvido o *HunchLab 2.0* para agregar as capacidades de um sistema que recorre a várias teorias sobre a criminalidade e gera um modelo unificado de predição da ocorrência de crimes – uma matriz de risco sobre o terreno. O objetivo é a aplicação dessa matriz como ferramenta de apoio ao C2, com sugestões de formas de atuação sob critérios estabelecidos pelas FS. A equipa que desenvolveu o *HunchLab* tinha uma visão específica para o projeto: desenhar uma ferramenta de gestão de patrulhamento que minimizasse a ocorrência de dano[73]. O algoritmo conjuga os seguintes modelos de predição criminal: Índices criminais, semelhantes a mapas de pontos quentes; *contágio*, relativo à propagação de eventos recentes; modulação da relação risco-terreno, proximidade e densidade geográfica de pontos de interesse; teoria da atividade de rotina, que integra as ações dos criminosos, forças da ordem e potenciais alvos; eficácia coletiva, com utilização de indicadores socioeconómicos, ciclos temporais, relativos à sazonalidade, à hora do dia, semana e mês, eventos recorrentes, férias, épocas desportivas; e por fim as condições climatéricas, como a temperatura, precipitação, etc. Acreditam que a inclusão de dados não criminais torna o sistema mais robusto e menos suscetível de gerar enviesamentos ou preconceitos no modelo.

Esta abordagem é utilizada para criar um modelo para cada tipo de crime individualmente, que serve de base para calcular o dano prevenível. Esta pode inclusive estar alinhada com o impacto social do crime (Figura 10)[74].

Figura 10 Modelos de Crime

[73] Traduzido do original *harm,* que se pode traduzir também em prejuízo, ofensa, perigo ou malefício segundo o dicionário Linguee, 2020

[74] Fonte: (HEFFNER, 2017)

Os locais apresentados como sugestão de alocação de meios são selecionados com base na procura de uma distribuição justa. Evitando-se a saturação e sobrepoliciamento de zonas quentes que poderão originar roturas na ligação entre as FS e a comunidade. O *HunchLab* minimiza este problema com a aplicação de uma série de fatores aleatórios como apenas apresentar as áreas entre a 3ª e 5ª posições de nível de risco de ocorrência de determinado crime. Como ferramenta de C2, a aplicação sugere uma forma de atuação, designada por missão, que melhor se adequa ao crime específico e que pode ser configurável pela FS. A aplicação promove diversas táticas que são vocacionadas para o policiamento de proximidade, como por exemplo a distribuição panfletos informativos e aconselhamento dos proprietários de pequeno comércio. Cada missão tem uma duração aconselhada de 15 minutos que pode ser cronometrada com recurso à aplicação. A ferramenta pode ser utilizada a partir de um smartphone (Figura 11)[75] ou de um computador (HEFFNER, 2017).

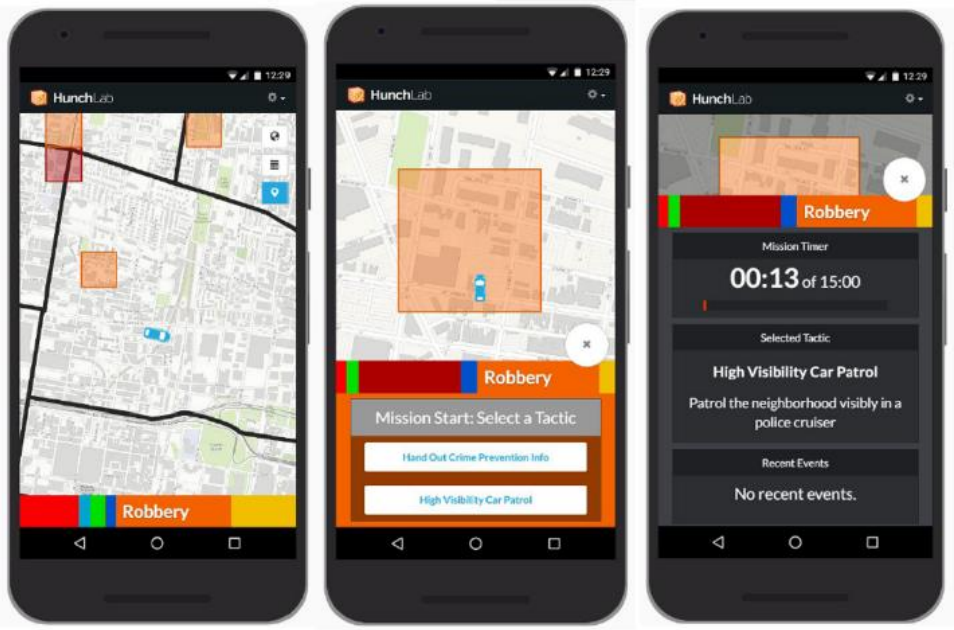

Figura 11 Aplicação HunchLab para smartphone

Em resumo, o *HunchLab* visa encontrar formas mais eficientes de alocar recursos, direcionando as patrulhas para locais onde previsivelmente seriam mais eficazes na dissuasão de ocorrências criminais e por sua vez reduzir os níveis totais de criminalidade. Ao fornecer mais contexto relativamente ao principal tipo de crime previsível de acontecer naquele local, permite aumentar o conhecimento situacional e identificar oportunidades para interagir com a comunidade (HEFFNER, 2017).

A 3 de outubro de 2018, a empresa *Shotspotter* adquiriu a ferramenta da *HunchLab* e continua o seu processo de integração com o seu próprio programa direcionado para a

[75] Fonte: (HEFFNER, 2017)

deteção e alerta de disparos provenientes de armas de fogo[76]. A informação apresentada e o racional de desenvolvimento da aplicação poderão alterar-se com esta aquisição.

Relativamente a resultados, ainda não foi avançado nenhum estudo compreensivo de avaliação da eficácia da ferramenta, estando a decorrer estudos nos departamentos que a utilizam. Avaliam a sua eficácia e os resultados da integração de novos dados no sistema de treino.

É importante analisar as implicações éticas da utilização destas ferramentas e ou outras ferramentas com resultados questionáveis e possíveis perpetuações de preconceitos enraizados nos seus algoritmos.

[76]https://www.shotspotter.com/press-releases/shotspotter-announces-acquisition-of-hunchlab-to-springboard-into-ai-driven-analysis-and-predictive-policing/

3.3. Videovigilância com recurso ao Reconhecimento Facial

"A IA será o caminho para uma democracia mais perfeita ou para uma ditadura mais insidiosa? A eterna vigília acabou de começar." (DOMINGOS, 2017 p. 313). A questão apresentada revela o grande dilema atual da IA: representa um potencial incrível para resolver problemas e otimizar soluções, mas caso a sua aplicação seja segundo os valores e princípios errados, poderá constituir o primeiro passo para concretizar uma profecia Orwelliana. E não poderia ser mais relevante no que se refere à videovigilância com recurso a reconhecimento facial. Esta tecnologia está a ser aplicada nas mais variadas áreas da sociedade, não sendo uma ferramenta exclusiva das FS. Está a ser diversificada em aplicações de verificação de assiduidade, atenção e produtividade nas escolas[77] e nos locais de trabalho[78], nas lojas de retalho[79] com reconhecimento do cliente ou pagamentos automatizados, entre muitos outros que estão em desenvolvimento, prevendo-se que o mercado do reconhecimento facial cresça dos 3.2 mil milhões de USD em 2019 para os 7 mil milhões de USD até 2024[80] – motivado pela procura dos grandes gigantes tecnológicos ocidentais como a *Google, Apple, Facebook*, *Amazon*, e *Microsoft* e os asiáticos *China Mobile*, *Tencent, Alibaba, Baidu e Xiaomi*.

A capacidade de reconhecer uma face nos humanos desenvolve-se desde os primeiros dias de vida, onde recém-nascidos conseguem imitar expressões faciais, e matura ao longo da vida, considerando-se que os mecanismos neurológicos necessários estão presentes aos cinco anos de idade (JEFFERY, et al., 2011).

Esta capacidade de reconhecimento facial humana é aplicada naturalmente no exercício do patrulhamento pelas FS, no reconhecimento presencial ou por meio de fotografia, filme ou gravação que esteja previsto no CPP [81] como meio de prova para julgamento.

Consultando a base de dados governamental para os contratos públicos, deparamo-nos com a realidade portuguesa. As ferramentas de reconhecimento facial automatizado começaram a chegar à administração pública[82] a 21 de dezembro de 2009 quando foi celebrado um contrato para aquisição de licenças de *VBWatcher server e client*, para a criação de uma solução piloto de reconhecimento facial por parte dos Serviço de Estrangeiros e Fronteiras, onde os dados relativos ao contrato não foram publicados. O reconhecimento facial é utilizado no controlo das entradas e saídas nas fronteiras internacionais nos aeroportos, nomeadamente nos quiosques automáticos de verificação do passaporte, com

[77] https://wonderfulengineering.com/chinese-schools-use-facial-recognition-to-make-sure-students-are-paying-attention/
[78] https://www.findd.io/
[79] https://sightcorp.com/knowledge-base/facial-recognition-in-retail/
[80] https://www.marketsandmarkets.com/PressReleases/facial-recognition.asp
[81] Cf. Artigo 147.º do CPP
[82] Com o Concurso Público N.º 235/19/DCP/GACD *Software* de reconhecimento facial e deteção de vida da Agência para a Modernização Administrativa

vista a verificar a correspondência da pessoa que se apresenta no controlo ao existente na base de dados e no documento apresentado.

A 23 de dezembro de 2016, a entidade Turismo de Portugal também adquiriu 110 licenças de *software* para efetuar o reconhecimento facial através dos sistemas de CCTV de onze Casinos de Portugal[83] . Mais recentemente, a Agência para a Modernização Administrativa lançou o Concurso Público N.º 235/19/DCP/GACD para adquirir *software* de reconhecimento facial e deteção de vida, para permitir a autenticação com recurso a reconhecimento facial da chave móvel digital. Também é possível observar que os municípios de Lousã, Lisboa, Castelo de Paiva e Ovar adquiriram serviços de reconhecimento facial aplicado ao controlo de assiduidade dos funcionários municipais, em diferentes fases de implementação.

No âmbito da segurança, a PSP avançou com um projeto de implementação de sistemas de videovigilância com recurso a reconhecimento facial e à analítica de vídeo em Leiria[84] e Portimão [85], que mereceram parecer negativo da CNPD "*atendendo especialmente que a utilização de um sistema de videovigilância* […] *representa um elevado risco para a privacidade dos cidadão."*

A proteção legal do cidadão face à utilização desta tecnologia varia consideravelmente dependendo da área geográfica[86] e da finalidade do sistema, destacando-se o estudo[87] que revela que 11 polícias de países europeus usam o reconhecimento facial. Por outro lado, nos EUA, diversos estados já baniram ou aplicaram moratórias[88], assim como Marrocos, Bélgica e Luxemburgo. A China afigura-se como líder mundial no desenvolvimento da tecnologia (LEE, 2018 p. 37), sendo obrigatório o registo por reconhecimento facial para a aquisição de cartões SIM[89].

Em Portugal, o reconhecimento facial para identificação de uma pessoa para efeitos de prevenção, investigação, deteção ou repressão de infrações penais ou execução de sanções penais é considerado uma forma de tratamento de dados biométricos definidos como dados sensíveis de acordo com o n. º1 do artigo 9.º da Lei n.º 59/2019 que transpôs a Diretiva EU 2016/680.

A utilização desta tecnologia em Portugal afigura-se legal, por existir uma lei que permite o tratamento de dados – a Lei n.º 1/2005 que regula a videovigilância pelas forças de segurança em locais públicos de utilização comum. Pode ser utilizada unicamente para os

[83] https://www.publico.pt/2019/05/06/economia/noticia/turismo-gastou-338-mil-euro-reconhecimento-facial-nao-funciona-1871551
[84] https://www.cnpd.pt/home/decisoes/Par/PAR_2019_92.pdf
[85] https://www.cnpd.pt/home/decisoes/Par/PAR_2019_93.pdf
[86] https://surfshark.com/facial-recognition-map
[87] https://algorithmwatch.org/en/story/face-recognition-police-europe/
[88] https://epic.org/state-policy/facialrecognition/
[89] www.theguardian.com/world/2019/dec/02/china-brings-in-mandatory-facial-recognition-for-mobile-phone-users

fins previstos na lei[90], supondo sempre a avaliação prévia do impacto sobre a proteção de dados[91], a garantia da segurança dos sistemas de informação assim como direitos reforçados para os titulares dos dados. O fundamento legal para a utilização deste tipo de tecnologias obriga à apreciação[92] da CNPD: da segurança do tratamento dos dados recolhidos; das medidas especiais de segurança a implementar – adequadas para garantir os controlos de entrada nas instalações, dos suportes de dados, da inserção, da utilização, de acesso, da transmissão, da introdução e do transporte de dados –, bem como a verificação do cumprimento do dever de informação.

Para Rui Pereira devem ser estabelecidas regras claras e equilibradas – o que se pode ou não pode fazer, quais as finalidades, quais as entidades competentes e entidades fiscalizadoras (PEREIRA, E12). Depreende-se que para ser legal, não basta estar previsto na lei, tem de respeitar um ecossistema de governação constituído por todas as partes interessadas, que salvaguardem os DLG dos cidadãos.

Comparativamente a Portugal, o panorama internacional demarca-se pela velocidade e magnitude da evolução destas tecnologias. Na China estavam previstas cerca de 626 M de câmaras até ao final de 2020[93]. Este número astronómico de câmaras gera um volume ainda maior de dados, sendo que é humanamente impossível observar individualmente cada *feed* de vídeo. Para tal, foram desenvolvidos diversos *softwares* que permitem realizar essa tarefa autonomamente[94].

O reconhecimento facial é uma forma de reconhecimento biométrico[95], pois a face possui atributos que podem ser medidos em número suficiente para serem distintivos, que se traduzem num identificador único possível de ser verificado de forma eficiente. Este identificador único tem a capacidade de ser convertido em formato digital por forma a permitir o seu armazenamento e pesquisa (WOODWARD, et al., 2003 p. 1).

A tecnologia de reconhecimento facial é usada para executar várias funções, sendo as principais a comparação de uma pessoa desconhecida com uma base de dados de pessoas conhecidas – designado um-para-muitos –, e a verificação de identidade – designado um-para-um (EDMOND, et al., 2009).

---

[90] Cfr o n. º1 do artigo 2.º da Lei n. º1/2005
[91] Cfr o Art.º 29.º do RGPD
[92] Cfr o n. º2 do artigo 3.º da Lei n. º1/2005
[93] https://www.comparitech.com/vpn-privacy/the-worlds-most-surveilled-cities/
[94] https://youtu.be/aE1kA0Jy0Xg?t=182
[95] Referente à medição de um atributo físico do corpo humano. Vários destes atributos podem ser usados como identificadores biométricos como por exemplo: o tom de pele, a cor e forma dos olhos, forma das mãos, vasos sanguíneos, características da face, traços comportamentais como o andar ou modulação da voz, e até características biodinâmicas como a pressão, padrão e velocidade com que se escreve com teclados de computador (CLARKE, 1999).

O processo de reconhecimento de um-para-muitos (Figura 12)[96] geralmente consiste em duas fases, a fase de pré-processamento seguida da fase de correspondência.

Figura 12 Processo de Correspondência no reconhecimento facial

Durante o pré-processamento, a foto de uma pessoa conhecida é posta à escala, alinhada e processada pelo *software* de reconhecimento facial com diversos métodos possíveis de digitalização. As feições faciais são quantificadas e mapeadas sobre diferentes máscaras representativas da estrutura facial e podem ser guardadas sobre diversos formatos como a representação da face um indivíduo, uma impressão facial (Figura 13)[97] (RICANEK, et al., 2012).

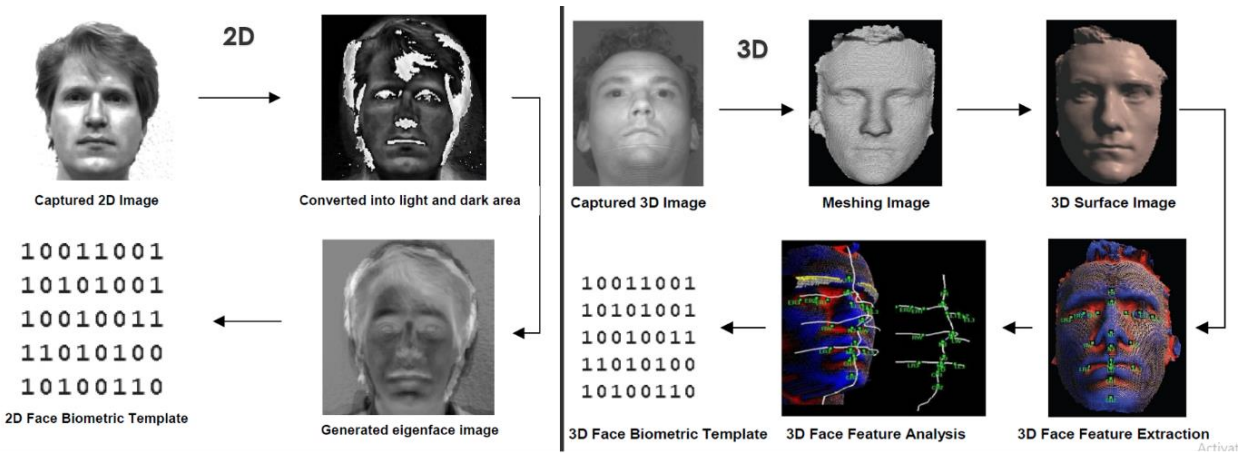

Figura 13 Digitalização Facial

Esta impressão facial é armazenada com informações biográficas do indivíduo na base de dados de pessoas conhecidas. Durante a fase da correspondência, uma fotografia ou fotograma[98] da pessoa a identificar é processada pelo *software* de reconhecimento facial, e é deduzida uma impressão facial, que é comparada com as impressões faciais presentes na

[96] Fonte: https://news.softpedia.com/news/fbi-has-access-to-412-million-face-recognition-images-even-of-foreign-citizens-505298.shtml
[97] Fonte: FingerTec Face Recognition Technology White Paper
[98] Nos casos de utilização de vídeos, designa-se um fotograma a cada uma das imagens que compõe o vídeo.

base de dados de pessoas conhecidas. Para tal, o *software* de reconhecimento facial usa um algoritmo para comparar as impressões faciais que apresentam um valor de similaridade entre as várias correspondências (GARVIE, et al., 2016 p. 9). Se o *software* determinar que este valor de similaridade é superior a um valor estipulado, elas serão apresentadas como uma correspondência provável. Dependendo do *software*, são identificadas uma ou mais correspondências prováveis que idealmente são validadas por um ser humano (INTRONA, et al., 2010).

Um exemplo de *software* de reconhecimento facial é o *Hitachi Live Face Matching* (LFM), integrado na plataforma *Smart Spaces and Video Intelligence* (SSVI) da *Hitachi Vantara*, e visualizado na *Hitachi Visualization Suite* (HVS).

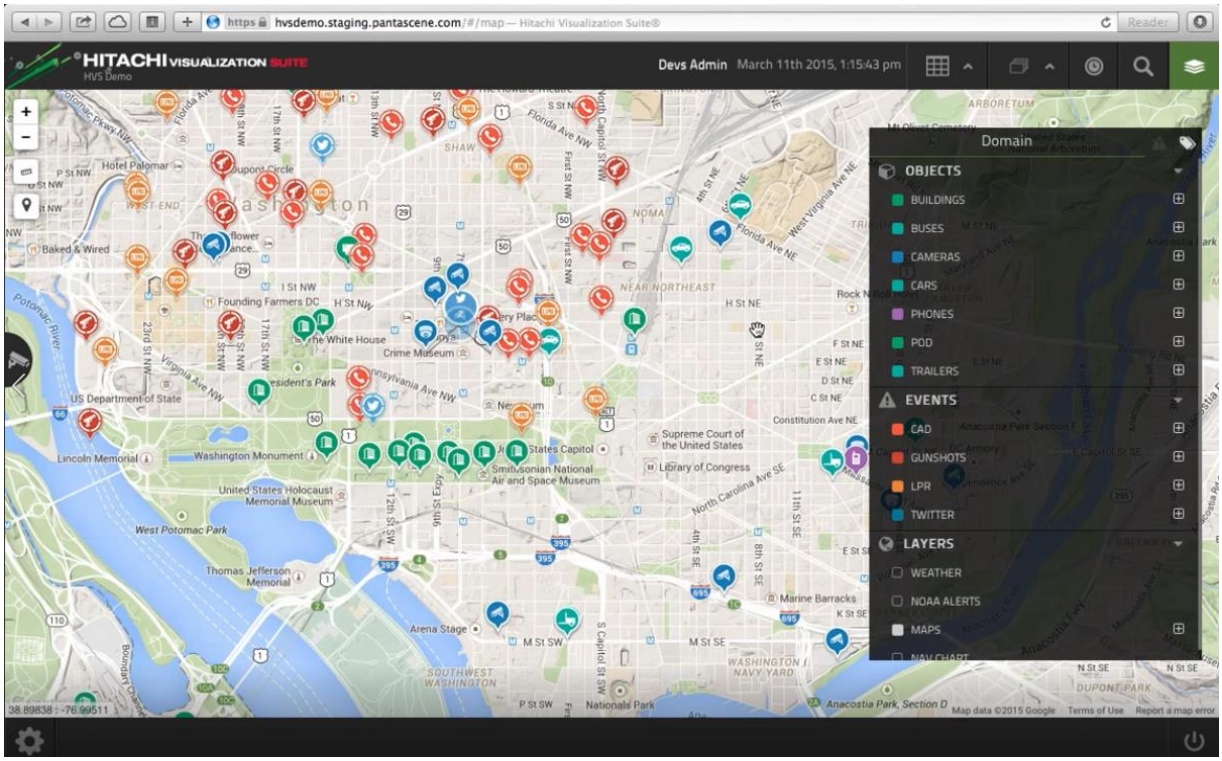

Figura 14 Portal do SSVI

Esta ferramenta apresenta-se como uma plataforma de gestão de cidades inteligentes (Figura 144)[99], que engloba zonas urbanas comerciais e industriais que utilizam o vídeo, IoT, ferramentas analíticas e tecnologias de IA para produzir conhecimento para pessoas, edifícios e máquinas de forma a aumentar a eficiência e qualidade de vida (Hitachi Vantara, 2020).

A demonstração do produto revela uma visão do futuro, próximo ou afastado, dependendo do desenvolvimento atual das cidades no que diz respeito à videovigilância, IoT ou sensorização. O SSVI trata-se de um Centro de Gestão de Cidades Inteligentes, que congrega diversas camadas de informação numa única plataforma de representação

[99] Fonte: (Hitachi Vantara, 2020)

geoespacial e temporal[100] e que pode integrar diversos módulos, entre eles a capacidade de reconhecimento facial. Os módulos integrados na plataforma ingerem dados provenientes dos *feeds* de vídeo do sistema, ou de qualquer outra câmara de IP[101]que seja integrada nele, do sistema de emergência 112, dos leitores automáticos de matrículas, das redes sociais e de outros sensores (Hitachi Vantara, 2020).

A LFM (Figura 15)[102] aplica um processo de extração de até seis imagens vídeo de ângulos diferentes da mesma pessoa, para criar uma impressão facial mais completa com capacidade para processar até 60 faces por segundo num único servidor com os dados provenientes de quatro câmaras.

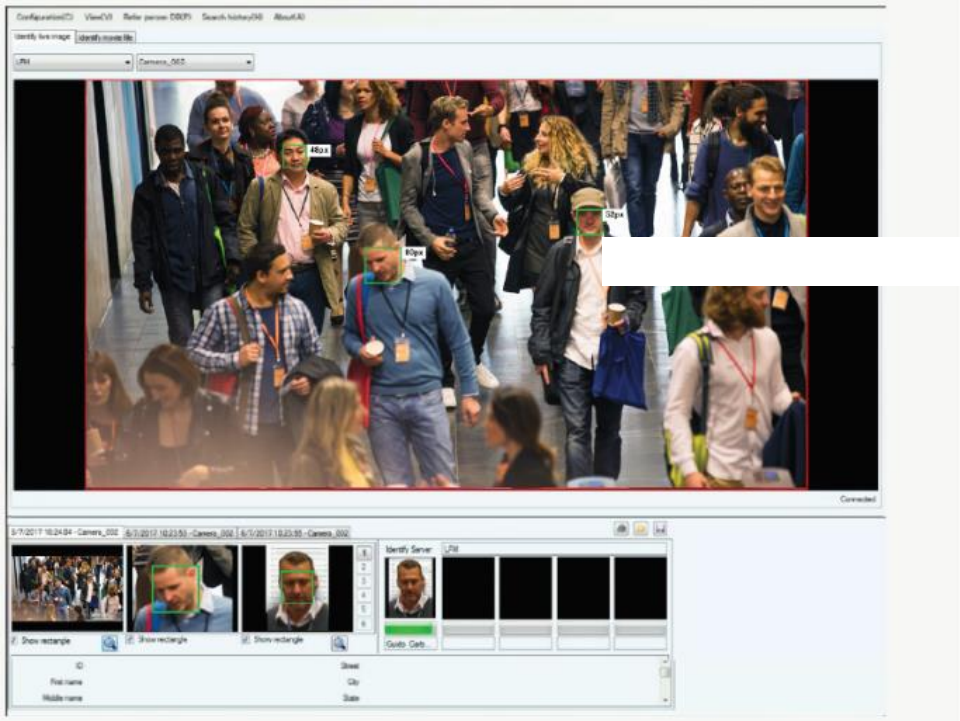

Figura 15 Correspondência de Impressão Facial com LFM

Para alem disso, integra diversas capacidades analíticas que permitem: a deteção e alerta de ameaças; a procura de suspeitos ou criminosos conhecidos (Figura 16); a procura pessoas desaparecidas; o alertas de bagagens abandonadas; sistemas de alerta por vedações virtuais; verificação de identidade entre outras.

Quando incorporada no HVS, permite também visualizar a atividade num determinado local, analisar o trânsito, contar veículos e as suas matrículas, detetar intrusões, detetar objetos como armas ou barreiras e até uma ferramenta para analisar e gerir os parques de estacionamento (Hitachi Vantara, 2019).

[100] Permite aceder a eventos no passado visualizando a sua georreferenciação e que camaras integradas na rede captaram imagens do evento
[101] Com a capacidade de transmitir os seus dados através de um protocolo de rede.
[102]Fonte: (Hitachi Vantara, 2019)

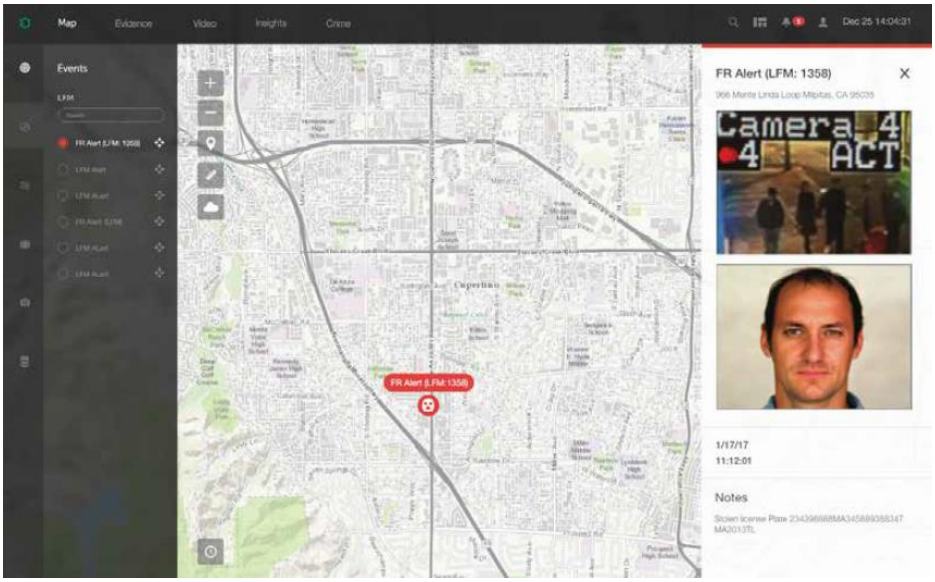

Figura 16 Alerta de deteção de Suspeito

O produto demonstra as capacidades que, do ponto de vista de segurança, são fatores de multiplicação de qualquer força. No entanto a documentação é relativamente opaca quanto aos dados, algoritmos, modelos e técnicas utilizadas no treino, e até a precisão do próprio produto não é revelada – há apenas referência a um resumo da implementação do sistema nas Ilhas Caimão, nas Bahamas e Antígua e Barbuda[103]. Mas não apresentaram qualquer métrica de avaliação da plataforma.

Como ponto positivo ressalva-se a procura de mitigar o impacto na privacidade do público. A *Hitachi Vantara* desenvolveu um filtro de privacidade[104] que anonimiza uma pessoa, revelando apenas a sua silhueta enquanto o *software* continua a aplicar as suas capacidades preditivas. A visualização da silhueta, acrescida do alerta dado pelo sistema, permite sinalizar uma situação na qual poderá ser relevante e necessário retirar o filtro, como uma ordem judicial, algo que é configurável para níveis superiores de autenticação de utilizadores. Todo o processo é auditável, na perspetiva de aumentar a responsabilização dos seus utilizadores desencorajando práticas pouco éticas.

O potencial do emprego do reconhecimento facial como ferramenta antidemocrática realça a necessidade de consciencialização, de garantias de transparência, de definição de critérios – necessidade, proporcionalidade, adequabilidade –, instrumentos de responsabilização e fiscalização do seu uso.

Caso se avance para a sua utilização desta tecnologia, com respeito pelos princípios e preocupações acima elencados, é expectável que as FS sejam mais eficazes e eficientes no

[103] https://social-innovation.hitachi/en/case_studies/security-centres-international-and-hitachi-visualization-success-story
[104] https://social-innovation.hitachi/en/case_studies/encrypted_personal_data

cumprimento das suas funções, uma vez que o processo de deteção de criminosos será apoiado pelas capacidades destes sistemas.

A transição digital é uma vontade do atual governo. Facto disso é a implementação em curso da estratégia "AI Portugal 2030", com o objetivo de democratizar a utilização da IA nos serviços públicos, pequenas e médias empresas e pelo cidadão, de forma a aumentar a qualidade de vida e competências digitais. Onde se incluem ainda iniciativas de Segurança Pública e Defesa, desenvolvendo a conectividade global das redes de polícia e segurança (incluindo IoT), com digitalização crescente dos meios de segurança e defesa.[105]

Em suma, existem diversas aplicações de IA, na área da Cibersegurança. Ferramentas, rápidas e eficazes em identificar *malware*, como o HP Sure Sense. Plataformas de partilha de *Threat Intel* como a IBM X-Force Exchange, com capacidade autónoma de identificação e agregação de conteúdos.

As Ferramentas de PP quantificam a relação risco-terreno para identificar as áreas mais suscetíveis de ocorrência de crimes, permitem a orientação de patrulhas para dissuadir potenciais criminosos. Ao mesmo tempo que contribuem para o desenvolvimento de programas de policiamento comunitário.

Terminámos a apresentação das ferramentas com uma plataforma de gestão de cidades inteligentes com capacidade de reconhecimento facial e outras ferramentas analíticas de vídeo. Esta foi apresentada na sede da Warpcom[106], que comercializa o produto em Portugal.

[105] www.incode2030.gov.pt/ai-portugal-2030
[106] https://warpcom.com/

Capítulo 4 – Discussão

Após uma breve apresentação das formas de aprendizagem que possibilitam a IA, e de exemplos de aplicações e funcionalidades de sistemas de IA. É importante refletir sobre:

QD2 - Que ingerências poderão ocorrer na esfera dos direitos, liberdades e garantias com o recurso a sistemas de IA pelas FS?

Percebemos que, apesar de vivermos a infância da IA, já existem diversos sistemas que apresentam um grau avançado de maturidade e que suscitam preocupação do foro ético. Exemplos disso são a classificação de cidadãos, reconhecimento facial, sistemas de monitorização de comunicações, sistemas de armas letais autónomas. E também preocupações a longo prazo com desenvolvimento da IA geral (SMUHA, Nathalie; Comissão Europeia, 2019 p. 43).

Os casos mais mediáticos são os sistemas de atribuição de crédito autónomos, que podem afastar potenciais credores com uma justificação baseada na análise de dados de diversas fontes, como compras online, amigos em redes sociais ou espaços de diversão frequentados. A China está a implementar um sistema de classificação para cidadãos através do *Citizen Score Card*, que representa o valor de um cidadão individual sob uma perspetiva governamental. Este sistema de classificação tem a capacidade de avaliar as diversas ações e rotinas diárias de uma pessoa, as suas amizades, dívidas, atitudes antissociais, os seus padrões de compras ou de deslocação e serve de base para tomar decisões que visam limitar as liberdades individuais dos cidadãos. Por exemplo: a limitação da deslocação entre zonas administrativas dentro da China e territórios internacionais; acesso facilitado a crédito bancário; possibilidade de compra de habitação em bairros de renome; acesso à primeira classe nos transportes públicos; entre diversas outras possibilidades que estão em desenvolvimento (STORM, 2015).

Outro exemplo de práticas de ética questionável pelas FS chinesas é a utilização integrada de um banco de dados nacional de ADN, vigilância eletrónica e reconhecimento facial, para controlar a população minoritária de uigures muçulmanos na província ocidental de Xinjiang.[107]

Analisando o recurso a IA para a Cibersegurança pelas FS, este apresenta impactos reduzidos nos DLG dos cidadãos uma vez que são ferramentas defensivas.

Por outro lado, é previsível que a IA seja usada, por atores estatais ou não-estatais, para fins ofensivos ou maliciosos. Existem duas categorias distintas de uso malicioso de IA: ataques com recurso a IA – que incluem técnicas baseadas em IA destinadas a melhorar a eficácia dos ataques tradicionais; e ataques direcionados à IA – com foco em subverter os

[107] https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html

sistemas de IA existentes para alterar as suas capacidades. Alguns casos prováveis de uso são: a criação de tipos de ataques mais sofisticados – *malware* desenhado por IA com a capacidade de se transformar e de se adaptar ao ambiente infetado; a Engenharia social avançada – com ataques massivos automatizados de *spearphishing* com informações altamente personalizadas sobre a vítima; a criação e gestão autónoma de contas falsas nas redes sociais – com interação humana simulada por sistemas de IA; os Modelos generativos profundos – para criar dados falsos para ofuscação e envenenamento de conjuntos de dados de treino dos modelo de IA; a descoberta de senhas com algoritmos de IA, etc. (ENISA, 2020).

A falta de regulação da utilização de IA para Cibersegurança e Ciberdefesa pode originar uma proliferação de ciberarmas. O ciclo é conhecido, as autoridades desenvolvem capacidades defensivas e melhoram o estado de segurança. Os cibercriminosos adaptam-se e procuram novas maneiras de continuar as suas práticas criminosas. É possível antever uma escalada no número de ciberataques pela automação de processos pelos cibercriminosos. Desta forma vai aumentar o número e a complexidade destes mesmos ataques. É importante que Estados e Organizações Internacionais (ONU, NATO, UE) estabeleçam regras e princípios, à luz do direito nacional e internacional, para legitimar e fundamentar as aplicações desta tecnologia.

É urgente criar os mecanismos de auditoria e fiscalização internacionais da utilização destas ferramentas, por forma a garantir o cumprimento das regras estabelecidas ou, em caso de violação das mesmas, aplicar medidas políticas ou económicas sancionatórias, de forma semelhante ao procedimento com violações dos direitos humanos ou ao tratado de não proliferação de armas nucleares (TADDEO, et al., 2018).

Relativamente aos problemas associados a programas de PP, estes podem ser direcionados para duas tipologias de crimes: os crimes violentos, que incluem o homicídio, fogo posto, roubo e agressão, geralmente denunciados; ou crimes de pequena delinquência, onde se inclui a venda e consumo de pequenas quantidades de estupefacientes, condução sem habilitação legal, dano simples, furto simples, ou injúria (MALHEIROS, et al., 2007). Muitos destes crimes de pequena delinquência não seriam registados ou denunciados caso não fossem presenciados pelo patrulhamento direcionado.

Esta pequena criminalidade é endêmica de muitos bairros pobres. Para Cathy O'Neil[108] é infeliz que as ferramentas sejam dirigidas para este tipo criminalidade. Incluí estes dados no modelo ameaça distorcer as suas predições. Ao alimentar o modelo preditivo com os dados desta pequena criminalidade, poderão ser alocadas mais patrulhas para esses locais por serem avaliados como mais propensos à ocorrência de crimes. Estas patrulhas podem iniciar

[108] Escritora de um livro de leitura obrigatória para perceber o impacto que os algoritmos estão a ter nas sociedades mais avançadas a nível de IA. Caracteriza certos modelos em uso como *Weapons of Math Destruction*, armas amde destruição matemática, e que previsivelmente também se vai verificar em Portugal.

um ciclo de *feedback,* alimentando o modelo com ainda mais informação destas áreas e resultar num sistema que promove o policiamento excessivo e a marginalização da sociedade (O'NIEL, 2016 p. 75).

Os sistemas de PP têm potencialidades que devem ser exploradas e que podem ajudar as FS. No entanto, as preocupações apresentadas são legítimas e há exemplos comprovados de discriminação nestes sistemas. O desenvolvimento desta tecnologia, combinado com a regulamentação do seu desenvolvimento e uso, pode minimizar os impactos negativos assinalados ao longo do trabalho. Salienta-se a importância da necessidade da avaliação obrigatória dos impactos destes programas quando utilizados em países europeus. Esta garantia está prevista na Diretiva EU 2016/680 e em Portugal na Lei n.º 59/2019, designada avaliação do impacto sobre a proteção de dados.

O reconhecimento facial é uma tecnologia eficiente e pouco invasiva, mas carece de regulamentação e precisão nos algoritmos utilizados, uma vez que pode gerar preconceitos raciais e consequências sociais graves. Um estudo comprovou um dos primeiros casos de discriminação algorítmica relativo ao reconhecimento facial. Este estudo avaliou a eficácia de três produtos comerciais de reconhecimento facial, onde se incluíram a Microsoft, a IBM e a Face++, e revelou as seguintes conclusões: todos os classificadores têm melhor desempenho em rostos masculinos do que femininos (diferença de 8. 1% a 20.6% na taxa de erro); todos os classificadores têm melhor desempenho em faces mais claras do que faces mais escuras (diferença de 11. 8% a 19.2% na taxa de erro); todos os classificadores apresentam pior desempenho em rostos femininos mais escuros (diferença de 20.8% a 34.7 % na taxa de erro); os classificadores da Microsoft e IBM apresentam melhor desempenho em rostos masculinos mais claros (taxas de erro de 0.0% e 0.3%, respetivamente); o classificador Face ++ têm melhor desempenho em rostos masculinos mais escuros (taxa de erro 0.7%); e a diferença máxima na taxa de erro entre os melhores e os piores grupos classificados é 34, 4%. As conclusões do estudo revelam que os algoritmos podem apresentar diferentes taxas de desempenho consoante o sexo ou tom de pele, confirmando um dos primeiros casos de preconceito algorítmico (Buolamwini, et al., 2018).

A principal causa apresentada na obtenção destes resultados é utilização de dados de treino pouco representativos e limitados. Assim, os algoritmos não tinham exemplos suficientes para treinar as suas capacidades nos tipos de faces menos comuns. Recentemente, diversos investigadores publicaram bases de dados de treino mais representativas da população que, quando utilizadas, ajudam os algoritmos a obter melhores resultados no reconhecimento facial em faces distintas.

Considerando que já foi detida a primeira pessoa inocente devido a uma identificação incorreta por um programa de reconhecimento facial[109], e por causa da crescente preocupação com utilização desta tecnologia de forma indiscriminada para o controlo de pessoas em manifestações[110], diversas empresas como a IBM, a Amazon e a Microsoft pararam temporariamente ou permanentemente o desenvolvimento desta tecnologia.

QD3 - Que medidas devem ser tomadas para salvaguardar que o desenvolvimento e a utilização de sistemas de IA pelas FS é feito de forma responsável, ética e segura?

A crescente preocupação com os impactos das diversas aplicações de IA na sociedade e a perspetiva do surgimento de IAG despertam a necessidade de uma resposta a esta questão complexa e que está a ser estudada por diversas instituições.

Miguel Souto acredita que temos de trabalhar em vários planos: no plano das leis, no plano dos reguladores, no plano da ética e no próprio plano tecnológico desenvolvendo, sistemas que regulem a atividade da AI (SOUTO, E10).

Esta abordagem holística é consensual entre os entrevistados e investigadores. As potencialidades desta tecnologia obrigam à adoção de medidas concretas para controlar o seu desenvolvimento.

José Fontes considera fundamental a implementação de sistemas de controlo, monitorização e fiscalização. Considera também importante que o sistema jurídico sancione as situações conhecidas de violações dos direitos (FONTES, E1). A implementação desses sistemas pressupõe um processo contínuo de verificação do respeito dos DLG dos cidadãos e mecanismos para punir o seu uso malicioso.

Rui Pereira admite ser inevitável a criação de uma agência que tutele a IA, devido à dimensão que esta ganhará. Tal agência, à semelhança da Agência Internacional de Energia Atómica, deverá operar num plano supra-estatal na ordem jurídica internacional. A imprevisibilidade do desenvolvimento tecnológico obriga à definição de critérios morais e jurídicos para conter os perigos que dela advêm (PEREIRA, E12). A recetividade dos Estados para aderirem a esta agência poderá variar conforme o estado de desenvolvimento dos seus projetos de IA e a sua estratégia política.

Para definir os princípios éticos referentes à IA foram avançadas diversas propostas. Um estudo (JOBIN, et al., 2019) identificou 84 documentos de diferentes organizações sobre os princípios ético de IA. Existem iniciativas que visam responder à questão da utilização da IA pela sociedade em geral e não unicamente pelas FS. No entanto considera-se que as mesmas enquadram as atividades das FS e são por isso aplicáveis.

[109] https://www.reuters.com/article/us-michigan-facial-recognition/us-activists-decry-first-known-wrongful-arrest-blamed-on-face-recognition-idUSKBN23V1KJ
[110] https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25996&LangID=E

Destacamos o "Enquadramento Unificado dos 5 Princípios para a IA na Sociedade"[111] – que visa compilar, sistematizar e simplificar a panóplia de princípios avançados pelas principais entidades desta área. Também as Orientações éticas para uma IA de confiança do Grupo de Peritos de alto nível sobre a IA da Comissão Europeia[112] constituem um excelente ponto de partida para o desenvolvimento e avaliação das tecnologias a curto prazo, enquanto se maturam normas, regulamentos e leis sobre a IA a médio e a longo prazo.

A apresentação destes documentos afigura-se importante por serem pioneiros e indicarem um caminho a seguir para o desenvolvimento ético da IA. Este caminho deve ser compatível com a democracia, de modo a permitir maximizar os benefícios destes sistemas, mitigando simultaneamente os seus riscos.

4.1. Enquadramento Unificado dos 5 Princípios para a IA na Sociedade

Este documento elaborado por Luciano Floridi[113] e Josh Cowls pode servir como enquadramento para o processo de definição de melhores práticas, padrões técnicos, regulamentos ou legislação relativa ao recurso à IA nos diversos setores públicos, indústrias e jurisdições específicas. Neste contexto, o presente enquadramento pode desempenhar um papel habilitador – sustentando do ponto de vista ético a aplicação de IA restrita ou geral a objetivos governamentais –, ou um papel limitador – exaltando a necessidade de regulamentar as tecnologias de IA no contexto de cibercrime e ciberguerra.

Os diversos documentos analisados apresentam 47 princípios que, através processo de análise comparativa, foram destilados em apenas 5. Surgem princípios anteriormente conhecidos da área da bioética: a beneficência, a prevenção de danos[114], a autonomia e a justiça, acrescidos de um princípio exclusivo da IA, a explicabilidade.

A beneficência orienta o desenvolvimento da IA com o objetivo de promover o bem comum da humanidade, preservar a dignidade humana, o bem-estar e a sustentabilidade do planeta.

A prevenção de danos visa limitar qualquer aplicação de IA que seja prejudicial, provoque danos ou traga consequências negativas do uso da IA. Este princípio incorpora também a preservação da privacidade e segurança, pela responsabilização de quem desenvolve a IA. Visando evitar: a escalada de violência autónoma, o melhoramento autónomo recursivo e a perda do controlo sobre as capacidades da IA.

O princípio da autonomia está relacionado com o compromisso entre a capacidade de decisão que retemos e aquela que delegamos para os sistemas de IA. Visa salientar os riscos da perda de autonomia e a limitação da liberdade do ser humano, resultante da transferência

[111] https://hdsr.mitpress.mit.edu/pub/l0jsh9d1/release/6
[112] https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60435
[113] http://www.philosophyofinformation.net/about/
[114] Pode ser traduzido da expressão "*Do no harm"* referente à minimização do dano causado.

de decisão para os sistemas. Poderemos ceder esta capacidade de decidir onde se julgue que seja proveitosa, por razões de eficácia e eficiência, retendo, porém, a possibilidade de a reverter a qualquer momento. Certas capacidades autónomas devem ser limitadas à partida, obrigando à intervenção humana, como a possibilidade de enganar, causar dano, ferir e matar.

O desenvolvimento de uma IA justa consiste na promoção da equidade, da diversidade, da solidariedade, da prosperidade e da partilha dos seus benefícios. E em simultâneo, reduzir a desigualdade, o preconceito, a discriminação e a estigmatização contra pessoas ou grupos. Associados à justiça, surgem os conceitos de legalidade, adequabilidade e proporcionalidade entre o emprego dos meios e os fins que se pretendem atingir. As decisões e implicações da IA devem poder ser contestáveis e deve existir a possibilidade de recorrer das mesmas.

A explicabilidade da IA é um conceito que abrange a necessidade da sua compreensão, transparência, inteligibilidade, interpretabilidade e responsabilização. Estes conceitos materializam as preocupações relativas a algo novo na IA – a sua forma de funcionamento ultrapassa a compreensão da generalidade das pessoas. O problema da "caixa negra" apenas fica resolvido quando se consegue dar resposta a duas perguntas fundamentais: "Como funciona?" e "Quem é responsável pela forma como funciona?".

A explicabilidade é considerada uma pedra basilar do desenvolvimento da IA por complementar os restantes princípios. Para que seja benéfica, temos de conhecer e compreender as formas, os efeitos e as consequências, positivas e negativas, que advêm do seu uso na sociedade. Para que não seja prejudicial, temos de garantir que opera de forma segura e limitada aos objetivos determinados por via da transparência. Para que promova e não restrinja a autonomia humana, devemos ser informados dos motivos e fundamentos das decisões autónomas. E para que a IA seja justa, devemos saber quem responsabilizar no caso de se verificar um incidente grave ou a violação dos DLG de um cidadão.

4.2. Orientações Éticas para uma IA de Confiança

Este documento merece destaque por ser uma das primeiras abordagens pragmáticas sobre o desenvolvimento de IA, apresentando-se como um quadro de referência que procura materializar os princípios e preocupações da sociedade. Foi elaborado pelo Grupo de Peritos de Alto Nível sobre a Inteligência Artificial[115] – um grupo de peritos independente criado pela Comissão Europeia para aconselhamento na área da IA. Elenca os fatores essenciais para que qualquer entidade que esteja a desenvolver um produto que recorra a IA consiga fazê-lo com garantias mínimas de conformidade.

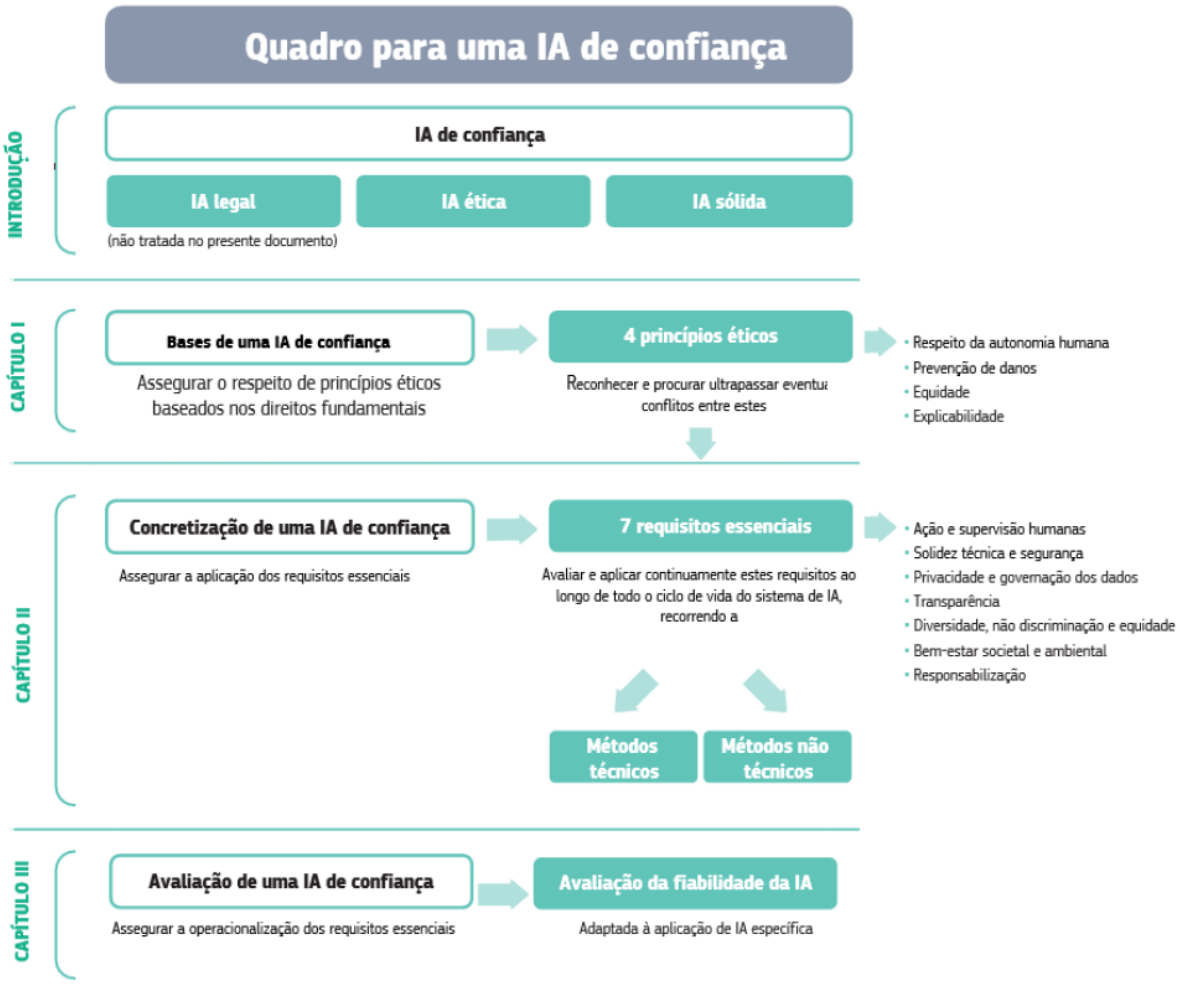

Figura 17 Quadro para uma IA de confiança

O documento (Figura 17)[116] divide-se em três partes: introdução, desenvolvimento e conclusão. Na parte do desenvolvimento são apresentados três capítulos que constituem o enquadramento: o primeiro capítulo estabelece as bases de uma IA de confiança; o segundo capítulo apresenta os requisitos e métodos para a concretização de uma IA de confiança; o terceiro capítulo estabelece uma lista de verificação para avaliar uma IA de confiança. Por fim,

[115] https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence
[116] Fonte: (EU High-Level Expert Group on Aritificial Intelligence, 2019)

é apresentado um resumo das oportunidades e preocupações críticas suscitadas pela IA e as conclusões do grupo de trabalho.

Para este grupo de peritos, a utilização de IA deve manter três pressupostos ao longo do seu ciclo de vida: deve ser Legal – cumprindo toda a legislação e regulamentação aplicáveis e respeitar os princípios e os valores internacionalmente reconhecidos; deve ser Ética – garantindo a observância de princípios e valores éticos; e deve ser Sólida - tanto do ponto de vista técnico como do ponto de vista social.

O conceito de legalidade pode evoluir pela existência de diversos processos legislativos sobre a IA em curso. Principalmente pela necessidade de adensar o ordenamento jurídico para a garantia de confiança nas aplicações de IA, quando apresentem riscos para os DLG dos cidadãos (Comissão Europeia, 2020).

O grupo de trabalho considerou a IA ética aquela que respeita os direitos fundamentais consagrados nos diversos tratados internacionais[117] e que garanta os princípios da autonomia humana, a prevenção de danos, a equidade e a explicabilidade. Cada um destes princípios está alinhado com as definições apresentadas na síntese do documento anterior. Além disso, estes autores ressalvam a necessidade de estudo das possíveis consequências do agravamento das assimetrias de poder e informação nas relações entre estado e cidadão, empregadores e colaboradores, empresas e consumidores, que resultam da aplicação da IA.

A concretização de uma IA de confiança obriga ao envolvimento de todas as partes interessadas: os seus criadores, os responsáveis pela sua implementação, os utilizadores finais e ainda a sociedade em geral, em todo o ciclo de vida dos sistemas de IA.

Os sete requisitos que materializam estes princípios consistem em aspetos sistémicos, individuais e sociais: a ação e supervisão humanas; a solidez técnica e segurança; a privacidade e governação dos dados; a transparência; a diversidade a não discriminação e equidade; o bem-estar social e ambiental; e a responsabilização. Estes requisitos podem ser aplicados através de métodos técnicos e não técnicos ao longo de todo ciclo de vida do sistema, de forma contínua e dinâmica. Os métodos técnicos podem ser baseados em arquiteturas estabelecidas, segurança e privacidade por desenho, métodos explicativos, testes, validação e indicadores de qualidade de serviço. Os métodos não técnicos deverão ser desenvolvidos e melhorados continuamente e consistem em códigos de conduta, regulamentação, normalização, certificação de conformidade, responsabilização por meio de quadros de governação, educação e sensibilização, diversidade, equipas de desenvolvimento inclusivas, participação das partes interessadas e diálogo social. O capítulo 3 apresenta uma

[117] Que, de forma não exaustiva, abrangem o respeito da dignidade humana, a liberdade do indivíduo, o respeito da democracia, da justiça e do estado de direito, a igualdade, a não discriminação e a solidariedade. (EU High-Level Expert Group on Aritificial Intelligence, 2019)

lista de avaliação[118] não exaustiva, desenvolvida para ser aplicada, na conceção, implementação e utilização do sistema de IA, figurando-se um processo contínuo de verificação de conformidade, atualização e adaptação da própria lista, ao sistema de IA em causa.

O desenvolvimento de quadros e normas para as aplicações de IA específicas para os sistemas de IA utilizados de FS são uma prioridade de diversos organismos internacionais, como a OSCE (POLIS, 2019) e a UE (Comissão Europeia, 2020).

No caso do PP, os fatores transparência e auditabilidade são fundamentais na utilização destas ferramentas pelas FS. Qualquer cidadão tem o direito de perceber quais os critérios utilizados para se chegar a uma determinada decisão (RIBEIRO, E11).

Os cidadãos devem ser informados se esta tecnologia é utilizada e de que forma são afetados pela mesma. Tendencialmente, os algoritmos, métodos de treino e dados utilizados devem ser abertos ao público. Pelo menos, deverá ser considerado um compromisso de abertura a grupos de fiscalização para não comprometer a sua finalidade. Apenas com controlo externo, aliado ao controlo interno, se poderá ganhar confiança na tecnologia e perceber os seus resultados. Em Portugal, cada FS tem o seu órgão inspetivo e existe a Inspeção Geral da Administração Interna que tutela todas as FS. São órgãos fundamentais para garantir o respeito pelos DLG dos cidadãos na utilização destas tecnologias.

Para Tiago Lopes será fundamental assegurar o equilíbrio entre os DLG dos cidadãos e assegurar o nível de prevenção para o estado de segurança que se pretende alcançar (LOPES, E8). Neste caso, a necessidade do equilíbrio aplica-se ao PP e à capacidade de reconhecimento facial. Este equilíbrio pretendido deve ser definido pela sociedade e respeitado pelas FS.

Pedro Verdelho considera ser necessário haver uma legislação própria para legitimar o recurso ao reconhecimento facial. Esta tecnologia acarreta a violação da privacidade da pessoa, logo, deve estar prevista na lei e apenas ser permitida nesses parâmetros (VERDELHO, E14). Pode fazer-se uma equiparação entre o reconhecimento facial e as escutas telefónicas. São formas invasivas da privacidade para fins de segurança e investigação criminal. Uma forma de controlo possível do seu uso é a obrigatoriedade de autorização judicial. A intervenção judicial legitima o processo e garante um controlo rigoroso da sua utilização. Consideramos que esta legislação será um passo importante e complementar ao RGPD na salvaguarda dos DLG - implementando um catálogo de crimes, especialmente graves, para os quais seja permitido o reconhecimento facial, assim como os procedimentos a utilizar.

[118] https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

De uma forma geral, a IA tem de ser regulamentada, pois a atuação de diversos sistemas autónomos pode resultar em consequências para terceiros. O direito tem de prever a atividade da AI na eventualidade de existirem consequências erradas ou em caso de acidente. Só assim será possível determinar quem é o responsável pelo mesmo (VERDELHO, E14). Neste caso, o grau de responsabilização irá depender de vários fatores, como o grau de autonomia do agente e a intervenção humana no processo. É importante referir que quem cria e comercializa produtos deve ser também abrangido por este princípio.

Sónia Pereira denota a importância da garantia de intervenção humana, assim como a demarcação de produtos gerados apenas por processos autónomos (PEREIRA, E4). A demarcação facilita a identificação de processos onde não existe qualquer mediação humana, e esta informação pode ser importante na escolha de serviços ou aplicações. A garantia de intervenção humana está prevista no RGPD, existindo o direito de não ser sujeito a decisões individuais automatizadas, incluindo definição de perfis.

O mediatismo dado ao tema, muitas vezes por representações alarmistas da IA, espoletou a necessária discussão pública com o envolvimento de organismos internacionais e regionais, da qual resultam várias iniciativas de regulamentação, assim como instituições reconhecidas, como a ISO e o IEEE, que estão a preparar quadros para o desenvolvimento de IA. Todo este esforço deverá ser contínuo e prevêem-se rápidos avanços devido ao impacto e potenciais consequências do uso da IA.

Tomámos consciência dos efeitos que o recurso a sistemas de IA pode produzir no cidadão e na sociedade, demonstrando uma manifesta necessidade do desenvolvimento destas tecnologias de forma mais consciente, ética e transparente, criando normas e quadros de referência, verificação e certificação de conformidade, e possivelmente agências ou organismos que tutelem estas tecnologias.

Capítulo 5 – Conclusão

Com o presente trabalho investigámos o surgimento da IA como vetor de segurança e as implicações que advêm da sua utilização na sociedade, demonstrámos algumas das suas capacidades atuais, observámos alguns exemplos de impactos negativos que decorrem da sua utilização e ainda apresentámos abordagens para os mitigar.

Para responder à questão de partida:

Qual é o impacto da aplicação dos sistemas de IA pelas FS na Segurança, nos dados, na liberdade e privacidade?

Identificámos questões derivadas que permitiram analisar fatores relevantes de forma mais detalhada e objetiva.

QD1 - Que sistemas de IA existem para apoiar as FS?

Existem atualmente diversos sistemas de IA em utilização e um leque ainda maior em desenvolvimento – como o recurso a IA para análise e tradução de voz para texto de telecomunicações, análise de texto para produção de informações, agentes virtuais para recolha de depoimentos ou receção de queixas, veículos autónomos terrestres e aéreos para patrulhamento de fronteiras, identificação de fraude fiscal, identificação de publicações de conteúdo proibido nas redes sociais, entre outros. Neste trabalho, focámos a nossa atenção em três áreas: a Cibersegurança, o Policiamento Preditivo e a Videovigilância com recurso a Reconhecimento Facial.

A Cibersegurança afigura-se como uma área profícua para o surgimento de aplicações de IA. Este facto está relacionado com a partilha do espaço digital, onde grande parte dos dados necessários para alimentar estes sistemas são produzidos pelos próprios dispositivos, computadores, routers e outros sensores.

Apresentámos o *HP Sure Sense,* que se caracteriza por ser um antivírus de nova geração. Este recorre a um modelo de deteção previamente treinado, onde o algoritmo define as características ou atributos que diferenciam um ficheiro seguro de um malicioso, para criar um modelo de previsão do nível de maliciosidade do ficheiro. Quando um documento é aberto, o mesmo é avaliado e gerado um valor e, caso esse valor se encontre acima do nível estabelecido para um ficheiro seguro, o agente decide impedir a execução do mesmo. Para além de classificar se é malicioso ou não, também classifica em tempo real o tipo de ameaça ou família de *malware.*

Uma outra ferramenta apresentada, a IBM *X-Force Exchange,* tem como objetivo a partilha de informação sobre ameaças à Cibersegurança, quer seja o seu contexto, mecanismos do ataque, ações recomendadas ou outras informações necessárias para aumentar o nível de conhecimento da ameaça, como apoio aos elementos dos SOC. Esta informação é gerada pelo *Watson for Cybersecurity* e partilhada num portal, podendo ser integrada em diversas ferramentas de segurança. Nos diversos relatórios, podemos verificar

o sumário da vulnerabilidade ou ameaça, uma descrição detalhada, indicadores de comprometimento, recomendações e até assinaturas digitais do *malware* em causa que poderão ser utilizados para a sua deteção pelos AV.

No presente trabalho apresentámos dois exemplos de programas de PP, o *PredPol* e o *Hunchlab*, que se categorizam como plataformas de gestão de patrulhamento e de C2. As previsões destes sistemas são operacionalizadas por meio de mapas, assinalando as áreas com probabilidade elevada de ocorrência de crimes, permitindo o direcionamento de patrulhas e a adoção de táticas adequadas para dissuadir potenciais criminosos.

A última aplicação de IA estudada é a videovigilância com recurso ao reconhecimento facial, geralmente incluída em programas de analítica de vídeo, onde apresentámos a plataforma da *Hitachi Vantara*. O reconhecimento facial pode ser usado para verificar a identidade de um indivíduo conhecido ou para identificar uma pessoa que conste numa base de dados. A grande vantagem destes sistemas é a possibilidade de serem aplicados a diversas fontes de vídeo em simultâneo, aumentando assim a eficácia dos sistemas de videovigilância atuais. Para além da identificação, os programas de analítica de vídeo também conseguem localizar pessoas pelo vestuário que utilizam, identificar objetos suspeitos abandonados e identificar quem os largou, alertar quando detetada uma intrusão ou quando alguém se encontra numa zona proibida, e ainda identificar veículos pela leitura automática de matrículas ou pela descrição do mesmo. Estas capacidades afiguram-se como valiosas para o desempenho das missões das FS, principalmente no combate ao terrorismo e à criminalidade violenta e grave.

Apesar de estas tecnologias já existirem no mercado e em utilização há vários anos a nível internacional, poucos são os países que desenvolveram enquadramento jurídico para as mesmas. A necessidade de previsão das consequências da utilização destes sistemas suscitou a seguinte questão derivada:

QD2 - Que ingerências poderão ocorrer na esfera dos direitos, liberdades e garantias com o recurso a sistemas de IA pelas FS.

O desenvolvimento e a utilização de sistemas de IA com supervisão limitada têm originado diversas situações de conflito e violação de DLG dos seus utilizadores ou alvos. As principais questões apresentadas relativas à área da Cibersegurança prendem-se com a antecipação de uma escalada do emprego de meios coercivos digitais e uma previsível corrida às ciberarmas. À medida que a segurança aumenta, os ciberciminosos e os Estados vão desenvolver mais capacidades ofensivas, prevendo-se o recurso à IA para aumentar a sua eficácia e eficiência. O sucesso da resposta a um ciberataque depende também da rapidez das ações tomadas e, sabendo que a IA reduz este tempo de resposta, é expectável que o emprego de contramedidas seja feito de forma cada vez mais autónoma. O ambiente de guerra difuso que caracteriza o ciberespaço permite que seja fácil atravessar a linha que

separa as ações defensivas das ofensivas, podendo originar retaliações e efeitos no mundo físico, à distância de algumas linhas de código. Esta indefinição propicia que seja difícil a aplicação de regras e princípios que legitimem, à luz do direito nacional e internacional, as intervenções de todos os agentes neste meio.

Quanto ao recurso do PP, as principais questões levantadas são relativas à falta de estudos que comprovem a sua eficácia e efeitos concretos na forma de policiamento levado a cabo, combinado com a falta de transparência e explicabilidade das sugestões apresentadas pelos programas. A falta de resposta do porquê, ou que fatores levam a direcionar as patrulhas para determinada área numa determinada janela temporal, levantam condicionantes à utilização deste tipo de programas. Outro fator fundamental prende-se com a importância dos dados históricos criminais que alimentam os sistemas, uma vez que, por regra, dados recolhidos de uma forma que não seja representativa da realidade e apenas representem uma margem da criminalidade vão originar sistemas que reforçam o combate a essa fração de crimes, podendo replicar preconceitos ou a discriminação já existente na sociedade por meio de sobrepoliciamento de determinadas zonas.

Por fim, a utilização de sistemas de reconhecimento facial afigura-se como a tecnologia que mais impacto terá nos DLG dos cidadãos, pois quando é empregue para fins contrários à democracia, e conjugada com outros sistemas de controlo e repressão social, poderá ter consequências imprevisíveis na sociedade. Nomeadamente a limitação de diversos direitos fundamentais, como o direito à privacidade, liberdade de expressão, reunião e manifestação. O respeito pela lei é um princípio fundamental de qualquer sociedade civilizada, mas a perfeita conformidade com o que é moralmente aceitável depende de quem define os termos e as consequências do seu desrespeito. O pior cenário hipotético da utilização desta tecnologia é a perseguição de indivíduos pela sua origem racial ou étnica, opiniões políticas, convicções religiosas ou filosóficas ou a filiação sindical, orientação sexual ou debilidades físicas. O sentimento de que se é constantemente observado provoca, por norma, a alteração de comportamentos individuais para se conformar com as normas estabelecidas, traduzindo-se num arrefecimento social e na manutenção de regimes autocráticos.

Para além dos riscos associados à sua utilização, a própria tecnologia apresenta limitações de caráter técnico, demonstrando graus de eficácia variável conforme o sexo, idade, tom de pele e adereços que a pessoa a ser identificada apresente.

Para compreender de que forma os riscos apresentados podem ser mitigados e garantir que a utilização desta tecnologia seja compatível com os princípios fundamentais, foi formulada a seguinte questão derivada:

QD3 - Que medidas devem ser tomadas para salvaguardar que o desenvolvimento e a utilização de sistemas de IA pelas FS é feito de forma responsável, ética e segura?

A consciencialização da evolução e alcance das capacidades dos sistemas de IA nas diversas áreas da sociedade, e em particular nas FS, espoletaram um processo de discussão nacional e internacional, de forma a garantir que a sua utilização seja benéfica para humanidade e para cada cidadão a nível individual.

Das diversas iniciativas avançadas, apresentámos o Enquadramento Unificado dos 5 Princípios para a IA na Sociedade, que estabelece os princípios fundamentais que devem pautar o desenvolvimento e utilização dos sistemas de IA, sendo eles a beneficência, a prevenção de danos, a autonomia, a justiça, e a explicabilidade.

Seguidamente apresentámos as Orientações Éticas para uma IA de Confiança, que estabelece um quadro de referência que procura materializar os princípios e preocupações da sociedade, elencando à partida fatores essenciais para desenvolver um sistema de IA com garantias mínimas de conformidade. O desenvolvimento destes sistemas tem obrigatoriamente de cumprir os pressupostos de ser legal, ético e sólido. Têm de respeitar os princípios éticos da autonomia humana, da prevenção de danos, da equidade e da explicabilidade – bases para uma IA de confiança.

Relativamente à cibersegurança, é importante implementar a segurança e privacidade por desenho. Fomentar uma cultura proteção de dados e o respeito pela privacidade. Devemos regulamentar do uso de IA para fins defensivos e ofensivos em especial as contramedidas autónomas. Por outro lado, devemos desenvolver a capacidade de cooperação internacional e promover a Ciberdiplomacia da EU, para mitigar ameaças internacionais, através de ações conjuntas e pela aplicação de sanções.

O reconhecimento facial e o PP consideram-se aplicações críticas de IA quanto aos direitos fundamentais. Esta classificação obriga à avaliação do impacto sobre a proteção de dados. Uma garantia que se encontra prevista em leis nacionais e europeias: na Diretiva EU 2016/680 e, em Portugal, na Lei n.º 59/2019.

Quanto ao Policiamento preditivo Quanto ao Policiamento Preditivo, devemos assegurar que os dados utilizados no treino dos modelos preditivos são representativos e livres de preconceitos e os algoritmos usados devem ser abertos e auditáveis. Os programas de PP devem ser vocacionados para os crimes mais violentos ou graves, devendo ser utilizados de forma metódica e com supervisão. A implementação destes sistemas deve ser faseada, priviligiando projetos piloto controlados com avaliação e transparência em todas as etapas. Em suma, Benéfico e Explicável.

No caso específico do reconhecimento facial, é defendida a criação de uma lei autónoma que a regule, complementar ao RGPD. Esta legislação própria deve prever e legitimar o recurso ao reconhecimento facial e estabelçer a forma, os princípios e limites da

sua utilização. Deve prever um catálogo de crimes, especialmente graves, para os quais seja permitido recorrer ao reconhecimento facial. Defende-se também a obrigatoriedade de intervenção judicial que legitima o processo e garante um controlo rigoroso à sua utilização.

Para além das iniciativas anteriores, devem também ser estabelecidos códigos de conduta, regulamentos e normas que orientem a utilização destes sistemas para os casos específicos de aplicação, assim como a avaliação de conformidade com normas e certificados que sejam desenvolvidos futuramente. Para que se garantam todos os fatores anteriormente referidos, devem ser criados órgãos reguladores e de inspeção, ou capacitar os órgãos existentes para a fiscalização deste tipo de sistemas.

Para culminar o resultado de investigação e concluir o presente esforço de pesquisa, respondemos à questão de partida:

Qual é o impacto da aplicação dos sistemas de IA pelas FS na Segurança, nos dados, na liberdade e privacidade?

Em potência, o impacto pode ser altamente disruptivo para a sociedade, vejamos o caso da sua utilização desenfreada de IA na República Popular da China. De ponto de vista académico, a utilização de IA para fins de segurança não implica a criação de um estado orwelliano, mas pode pôr em causa diversos direitos fundamentais como a liberdade de expressão, manifestação e à privacidade.

Presume-se que o impacto na sociedade ocidental será mitigado internacionalmente pelas salvaguardas implementadas pelas entidades supranacionais como a ONU e a UE. Atualmente em Portugal, recai sobre a CNPD a responsabilidade de fiscalizar as garantias de proteção previstas no RGPD e na Diretiva EU 2018/680.

A IA irá apoiar as FS no cumprimento das suas missões, na cibersegurança, no Policiamento Preditivo e na Videovigilância com Recurso ao Reconhecimento Facial.

O reconhecimento facial é útil no combate ao crime organizado e ao terrorismo por multiplicar a capacidade de vigilância das áreas guardadas. No entanto, deve ser criada uma lei que a regule. Esta deve definir quais os crimes aplicáveis, prossupostos e fundamentos que o justifique e ainda a obrigatoriedade de intervenção judicial.

O tratamento dos dados recolhidos para fins de prevenção criminal deve garantir a segurança e privacidade. Apenas desta forma podem ser salvaguardados os Direitos, Liberdades e Garantias dos Cidadãos.

Para concluir, devemos assegurar que a utilização desta tecnologia é benéfica para humanidade e para cada cidadão a nível individual, devemos fomentar a discussão e consciencialização sobre a temática, para que tomemos decisões mais informadas.

IA encerra um enorme potencial transformativo, mas assim como qualquer tecnologia outra, existe a possibilidade de a IA ser empregue para fins contrários aos princípios democráticos. Assim, devemos permanecer vigilantes para identificar os primeiros sinais de

ingerência nos DLG dos cidadãos e perspetivar caminhos, soluções ou travões para mitigar os riscos que esta tecnologia acarreta, fazer uso de forma segura, consciente e ética, para beneficiar da onda de desenvolvimento movida pela IA.

Limitações e Investigações futuras

No decorrer da elaboração deste trabalho de investigação foi necessário gerir duas limitações distintas:

A permanente discussão, produção de informação e a tomada de posições por parte de organismos internacionais como a UE relativas ao tema da IA, afigurando-se como um alvo de estudo em movimento, reafirmando o descrito na Introdução, nomeadamente no Problema de investigação e formulação de hipóteses.

A especificidade do tema da IA e a abrangência do capítulo II, levou a que a Revisão da Literatura consumisse mais tempo do que o inicialmente previsto e que cada assunto fosse um percurso de descoberta contínua, até alcançar o conhecimento sólido dos conceitos discutidos.

Ao longo do trabalho identificámos a necessidade de limitar o escopo a uma amostra de área de aplicação de sistemas de IA. No entanto, consideramos de elevada pertinência orientar investigações e discussões futuras para diversos temas que não foram devidamente aprofundados no presente trabalho. Dessas questões, salientam-se os riscos associados ao surgimento de uma IA Geral, nomeadamente na área da sua segurança, controlo, definição e manutenção de objetivos ou finalidades. Quanto ao desenvolvimento da IA restrita, deve-se continuar a investigar novas aplicações em diversas áreas da sociedade, incluindo a segurança, onde se pretende um desenvolvimento e evolução dos processos por forma a ganhar eficácia, eficiência e rentabilidade. Devemos sempre continuar a avaliar os efeitos da utilização desta tecnologia, nomeadamente a sua eficácia, segurança e os respeito pelos princípios éticos e democráticos, avaliando os impactos na sociedade em geral e nos utilizadores em particular. Deve-se discutir e avaliar a necessidade de edificar e desenvolver organismos, instituições, regulamentação e modelos numa perspetiva a longo prazo. São passos fundamentais para antecipar os efeitos disruptivos das capacidades que a IA encerra. Por outro lado, deve-se procurar reforçar a investigação e a implementação de medidas a curto prazo, para garantir que os sistemas em produção atualmente não incorrem em violações graves nos DLG dos cidadãos e que se adequam gradualmente aos padrões estabelecidos futuramente.

Bibliografia

ALLEN, Greg and CHAN, Taniel. 2017. *Artificial Intelligence and National Security.* Cambridge : s.n., 2017. https://www.belfercenter.org/sites/default/files/files/publication/AI%20NatSec%20-%20final.pdf.

ALMEIDA, Paulo. 2013. Políticas de Segurança Nacional. [book auth.] PAULO ALMEIDA. *Como tornar Portugal um País Seguro.* 2013.

ANTONAKAKIS, Manos , APRIL, Tim and BAILEY, Michael. 2017. *Understanding the Mirai Botnet.* Vancouver, CA : s.n., 2017. ISBN 978-1-931971-40-9.

ANYOHA, Rockwell. 2017. The History of Artificial Intelligence. *Special Edition on Artificial Intelligence.* [Online] Harvard University, 28 08 2017. http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/.

AR. 1976. Constituição da República Portuguesa. *Decreto de aprovação da Constituição.* Lisboa : Assembleia da República, 10 04 1976.

ASHWORTH, Andre. 2007. *Security, Terrorism and the Value of Human Rights.* s.l. : B. J. Goold and L., 2007.

BARLOW, Mike and FELL, Gregory. Not All Data Is Created Equal. *Balancing Risk and Reward in a Data-Driven Economy.* O'Reilly Media, Inc. 9781492049937.

BRUNDAGE, Miles, et al. 2018. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.* [Online] Fevereiro 2018. https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf.

BUOLAMWINI, Joy and GEBRU, Timnit. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Conference on Fairness, Accountability, and Transparency.* 2018.

CAETANO, Marcello. 1977. *Princípios Fundamentais do Direito Administrativo.* Rio de Janeiro : s.n., 1977.

CAIR. Artificial Intelligence and Robotics. *Centre for Artificial Intelligence and Robotics.* [Online] http://www.unicri.it/topics/ai_robotics/.

CAMBRIDGE. 2020. [Online] Cambridge Dictionary, 2020. https://dictionary.cambridge.org/dictionary/english/fake-news.

CANOTILHO, José and MACHADO, Jónatas. 2003. *Reality Shows e Liberdade de Programação.* Coimbra : Coimbra Editora, 2003.

CBS 60 Minutes. 2020. How Russian intelligence officers interfered in the 2016 election. *CBS NEWS.* [Online] 20 Agosto 2020. https://www.cbsnews.com/news/russian-hackers-2016-election-democratic-congressional-campaign-committee-60-minutes-2020-08-23/.

CE. 1950. Convenção Europeia dos Direitos do Homem. Roma : Conselho da Europa, 04 11 1950.

CHEETHAM, Robert. 2019. Why We Sold HunchLab. *Azavea.* [Online] 23 Janeiro 2019. https://www.azavea.com/blog/2019/01/23/why-we-sold-hunchlab/.

CLAPPER, James R. 2013. Worldwide Threat Assessment of the US Intelligence Community. [Online] 12 Março 2013. https://www.dni.gov/files/documents/Intelligence%20Reports/2013%20ATA%20SFR%20for%20SSCI%2012%20Mar%202013.pdf.

CLARKE, Roger. 1999. *Introduction to Dataveillance and Information Privacy and Definitions of Terms.* 1999. www.rogerclarke.com/DV/Intro.html.

CLEMENTE, Pedro. 2013. Prevenção e Segurança: Políticas e Estratégias. [book auth.] P. P Almeida. *Como tornar Portugal um país seguro: segurança nacional e prevenção da criminalidade.* Lisboa : bnomics, 2013.

CNCS. 2019. Centro Nacional de Cibersegurança. [Online] 2019. https://www.cncs.gov.pt/a-internet-das-coisas-iot-internet-of-things/.

COATS, Daniel R. 2019. Worldwide Threat Assessment of the US Intelligence Community. [Online] 29 01 2019. https://www.dni.gov/files/ODNI/documents/2019-ATA-SFR---SSCI.pdf.

Comissão Europeia. 2020. *White Paper on Artificial Intelligence - A European aproach to excellence and trust.* Bruxelas : COM, 2020.

—. 2020. *White Paper on Artificial Intelligence: a European approach to excellence and trust.* Bruxelas : s.n., 2020. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

Conselho da União Europeia. 2007. Regulamento (CE) N.o 1183/2007 Do Conselho. *Regime comunitário de controlo das exportações de produtos e tecnologias de dupla utilização.* 2007. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:278:0001:0240:PT:PDF.

COOMBS, Ted. 2018. *Artificial Intelligence & Cybersecurity.* Hoboken : John Wiley & Sons, Inc, 2018. 78-1-119-50825-0.

CYLANCE. 2018. Classifying AI-Driven EDR Capabilities. *http://www.cylance.com/.* [Online] 2018. https://s7d2.scene7.com/is/content/cylance/prod/cylance-web/en-us/resources/knowledge-center/resource-library/infographics/EDRInfographicComparisonChart.pdf.

DIXON, William and EAGAN, Nicole. 2019. https://www.weforum.org. *3 ways AI will change the nature of cyber attacks.* [Online] 2019. https://www.weforum.org/agenda/2019/06/ai-is-powering-a-new-generation-of-cyberattack-its-also-our-best-defence/.

DOMINGOS, Pedro. 2017. *A Revolução do Algoritmo Mestre.* Lisboa : Manuscrito, 2017. 978-989-8871-18-3.

EDMOND, Gary., et al. 2009. *Law's looking glass: Expert identification evidence derived from photographic and video images.* s.l. : Current Issues in Criminal Justice 20, 2009.

EEAS, European External Action Service. 2016. *A Global Strategy for the European Union's Foreign and Security Policy.* [Online] 2016. https://eeas.europa.eu/sites/eeas/files/eugs_review_web_0.pdf.

ENISA. 2020. *AI Cybersecurity Challenges, Threat Landscape for Artificial intelligence.* Attiki, Greece : European Union Agency for Network and Information Security, 2020. 978-92-9204-462-6.

—. 2015. *Definition of Cybersecurity – Gaps and overlaps in standardisation.* Heraklion : European Union Agency for Network and Information Security, 2015. 978-92-9204-155-7.

EU High-Level Expert Group on Aritificial Intelligence. 2019. *Ethic Guidelines for Trustworthy AI.* Brussels : European Comission, 2019.

EU High-Level Expert Group on Artificial Intelligence. 2019. *A definition of AI: Main capabilities and scientific disciplines.* Brussels : European Commission, 2019.

EUROPOL. 2019. *First report of the observatory function on encryption.* [Online] 11 01 2019. https://www.europol.europa.eu/sites/default/files/documents/final_report_of_the_observatory_function.pdf.

—. 2017. *Crime in the age of Technology – EUROPOL's Serious and Organised Crime Threat Assessment.* 2017.

—. 2020. Sobre a EUROPOL. [Online] 2020. https://www.europol.europa.eu/pt/about-europol.

FAGAN, Jeffrey and DAVIES, Garth. 2001. *Street Stops and Broken Windows: Terry, Race, and Disorder in New York City.* 2001. 28 Fordham Urb. L.J. 457 2000-2001.

FERGUSON, Andrew. 2019. Predictive Policing Theory. [book auth.] Tamara Lave and Eric j Miller. *The Cambridge Handbook of Policing in the United States.* Washington : Cambridge Univercity Press, 2019.

FERNANDES, Filipe Reina. 2013. *A Cibersegurança e as Estruturas Críticas: A GNR.* Academia Militar. Lisboa : s.n., 2013.

FERREIRA, Ricardo and CUNHA, António. 2013. *Captura, análise e identificação de malware: caso de estudo.* 2013.

FESTAS, David. 2004. O Direito à Reserva da Intimidade da Vida Privada do Trabalhador no Código do Trabalho. *REVISTA da Ordem dos Advogados.* 2004.

FRANKLIN, Daniel and FLORIDI, Luciano. 2017. *Megatech-Technology in 2050.* s.l. : Clube do Autor, 2017. 978-989-724-393-6.

FRIEDMAN, Jerome. 2002. Stochastic gradient boosting. *Computational Statistics & Data Analysis.* s.l. : Elsevier, 2002.

GARVIE, Clare., BEDOYA, Alvaro and FRANKLE, Jonathan. 2016. *The prepetual line-up: Unregulated police face recognition in America.* Georgetown : s.n., 2016.

GOUVEIA, Jorge Bacelar. 2018. *Direito da Segurança: Cidadania, Soberania e Cosmopolitismo.* Coimbra : s.n., 2018.

GOV. 1966. Código Civil. *Decreto-Lei n.º 47344 - Diário do Governo n.º 274/1966, Série I de 1966-11-25.* Lisboa : s.n., 1966.

HEFFNER, Jeremy. 2017. HunchLab. *A Citizen's Guide to HunchLab.* 2017. http://robertbrauneis.net/algorithms/HunchLabACitizensGuide.pdf.

HEIKKILÄ, Juha. EU Artificial Intelligence. [Online] https://ec.europa.eu/digital-single-market/en/artificial-intelligence.

Hitachi Vantara. 2019. Live Face Matching. *Hitachi Vantara.* [Online] 2019. https://www.hitachivantara.com/en-us/pdfd/datasheet/live-face-matching-datasheet.pdf.

—. 2020. Smart Spaces. *Hitachi Vantara.* [Online] 2020. https://www.hitachivantara.com/en-us/solutions/smart-spaces.html.

Homeland Security. 2018. *Phishing Don't be fooled!* [Online] 2018. https://www.dhs.gov/sites/default/files/publications/2018_AEP_Vulnerabilities_of_Healthcare_IT_Systems.pdf.

IEP, Institute for Economics & Peace. 2019. *Global Peace Index 2019: Measuring Peace in a Complex World.* Sydney : s.n., 2019.

INTERPOL & UNICRI. 2019. *Artificial Intelligence and Robotics for Law Enforcement.* 2019. https://www.europarl.europa.eu/cmsdata/196207/UNICRI%20-%20Artificial%20intelligence%20and%20robotics%20for%20law%20enforcement.pdf.

INTRONA, Lucas. and NISSENBAUM, Helen. 2010. *Facial recognition technology: A survey of policy and implementation issues.* Lancaster : The Department of Organisation, Work and Technology, Lancaster University, 2010.

ITU, International Telecommunication Union. United Nations Activities on Artificial Intelligence (AI). [Online] https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-UNACT-2018-1-PDF-E.pdf.

JABUR, Gilberto Haddad. 2000. *Liberdade de Pensamento e Direito à Vida Privada: conflitos entre Direitos da Personalidade.* São Paulo : Revista dos Tribunais, 2000.

JEFFERY, Linda and RHODES, Gillian. 2011. Insights into the development of face recognition mechanisms revealed by face aftereffects. *British Journal of Psychology.* 102, 2011, 11/09/2011.

JOBIN, Anna, IENCA, Marcello and VAYENA, Effy. 2019. *Artificial Intelligence: the global landscape of ethics guidelines.* Zurich : Health Ethics & Policy Lab, 2019. https://arxiv.org/ftp/arxiv/papers/1906/1906.11668.pdf.

KOBAYASHI, Audrey. 2020. *Encycopledia of Human Geography.* Kingston : Elsevier, 2020. 9870081022955.

LANGSTON, Jennifer. 2019. Microsoft. *Innovation Stories.* [Online] 21 3 2019. https://news.microsoft.com/innovation-stories/hello-data-dna-storage/.

LEE, Kai-Fu. 2018. *As Superpotências da Inteligência Artificial: A China, Silicon Valley e a Nova Ordem Mundial.* Lisboa : Relógio de Água, 2018. 9789896419349.

—. 2018. Utopia, Dystopia, and the Real AI Crisis. [Online] 2018. http://storage.googleapis.com/aisp-assets/pdf/Utopia-Dystopia-and-the-Real-AI-Crisis.pdf?mtime=20180926132928.

MALHEIROS, Jorge, et al. 2007. *Espaços e Expressões de Conflito e Tensão entre Autóctones, Minorias Migrantes e Não Migrantes.* Lisboa : Observatório da Imigração, 2007. 97898980000293.

MOLEIRINHO, Pedro. 2018. A importância dos modelos preditivos na área da segurança.. Entre riscos e equilíbrios instáveis. [book auth.] Teresa RODRIGUES and Marco PAINHO. *Modelos Preditivos & Segurança Pública.* Porto : Fronteira do Caos, 2018.

MOREIRA, Vital, GOMES, Carla and NEVES, Ana. 2012. *Compreender os Direitos Humanos.* Graz : s.n., 2012.

NAS, Sjoera, TERRA, Floor and BAEHRING, Jill. 2019. *Data protection impact assessment on the processing of diagnostic data - Office 365 Online and mobile Office apps.* Ministry of Justice and Securit. The Hague : s.n., 2019.

NEGREIRO, Mar and MADIEGA, Tambiama. 2019. *europarl.europa.eu.* [Online] Junho 2019. http://www.europarl.europa.eu/RegData/etudes/BRIE/2019/633171/EPRS_BRI(2019)633171_EN.pdf. PE 633.171.

New York Times. 2019. Russia Is Targeting Europe's Elections. So Are Far-Right Copycats. [Online] 12 05 2019. https://www.nytimes.com/2019/05/12/world/europe/russian-propaganda-influence-campaign-european-elections-far-right.html.

NIELSEN, Didrik. 2016. Tree Boosting With XGBoost. *Why Does XGBoost Win "Every" Machine Learning Competition?* Trondheim : Norwegian University of Science and Technology, 2016.

OECD. 2019. *Artificial Intelligence in Society.* Paris : OECD Publishing, 2019. ISBN 978-92-64-54519-9.

OECD, Legal Instruments. Artificial intelligence. *OECD.* [Online] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

OLIVEIRA, Arlindo. 2019. *Inteligência Artificial.* Lisboa : Fundação Francisco Manuel dos Santos, 2019. 9789898863316.

O'NIEL, Cathy. 2016. *Weapons of math destruction: how big data increases inequality and threatens democracy.* s.l. : Crown, 2016. 9780553418811.

ONU. 1948. Declaração Universal dos Direitos Humanos. s.l. : Organização das Nações Unidas, 10 12 1948.

—. 1966. Pacto Internacional sobre os Direitos Civis e Políticos. 16 12 1966.

OpenAI. 2019. OpenAI. [Online] 2019. https://openai.com/blog/gpt-2-6-month-follow-up/.

—. 2020. OpenAI. [Online] 2020. https://openai.com/about/.

ORWELL, George. *1984.* Lisboa : Antígona.

OSCE. 2017. *OSCE Guidebook, Intelliogence-Led Policing.* Viena : s.n., 2017. 978-3-903128-04-0.

Panda Security. 2017. *Ransomware from the Crysis/Dharma family Report.* Madrid : s.n., 2017. https://www.pandasecurity.com/mediacenter/src/uploads/2017/11/Ransomware_Crysis-Dharma-en.pdf.

PEREIRA, Rui. 2018. [book auth.] TERESA RODRIGUES and Marco PAINHO. *Modelos Preditivos & Segurança Pública.* s.l. : Fronteira do Caos, 2018.

PIRES, Lucas and FREITAS, Almendra. 2014. *Direito à Privacidade no Âmbito da Sociedade da Informação: reflexões em torno da questão nos inícios do século XXI.* FACULDADE DE DIREITO, UNIVERSIDADE DE COIMBRA. Coimbra : s.n., 2014.

POLIS. 2019. 2019 Annual Police Experts Meeting on "Artificial Intelligence and Law Enforcement: An Ally or an Adversary?". [Online] 2019. https://polis.osce.org/2019APEM.

POOLE, David, MACKWORTH, Alan and GOEBEL, Randy. 1998. *Computational Intelligence: A Logical Approach.* Nova York : Oxford University Press, 1998. 9780195102703.

PredPol. 2019. From Theory to Practical Deployment. *PredPol.* [Online] 2019. https://info.predpol.com/cs/c/?cta_guid=258f00e6-8e29-4288-9ef2-1851928151aa&placement_guid=4a23f8ec-41ff-4a9b-b5c8-93f1dc4322fc&portal_id=3362003&canon=http%3A%2F%2Finfo.predpol.com%2Fthanks-white-paper-of-science-testing-of-predictive-policing&redirect_.

—. 2020. Overview. *PredPol.* [Online] PredPol, 2020. https://www.predpol.com/about/.

RACONTEUR. 2019. *A Day in Data.* [Online] Raconteur, 2019. http://rcnt.eu/un8bg.

Rand Corporation. 2013. *Predictive Policing The role of Crime Forecasting in Law Enforcement Operations.* 2013. 978-0-8330-8148-3.

RATCLIFF, Jeff. 2016. *Intelligence-Led Policing.* Nova Yorke : Rotledg, 2016.

RÊGO, Helena. 2013. Segurança e prevenção do terrorismo: algumas notas. [book auth.] P. P. Almeida. *Como tornar Portugal um país seguro: segurança nacional e prevenção da.* Lisboa : bnomics, 2013.

RICANEK, Karl. and BOEHNEN, Chris. 2012. *Facial analytics: From big data to law enforcement.* s.l. : IEEE Computer Society, 2012.

RUSSEL, Stuart and NORVIG, Peter. 2009. *Artificial Intelligence: A Modern Approach.* London : Pearson, 2009. http://aima.cs.berkeley.edu/.

SCHAFFERS, Hans, KOMNINOS, Nicos and PALLOT, Marc. 2012. *Smart Cities as Innovation Ecosystems sustained by the Future Internet.* 2012. hal-00769635.

SE Labs. 2018. *On-Demand Malware Detection Certification.* Buckinghamshire : s.n., 2018. https://info.deepinstinct.com/whitepaper-se-labs-certificate.

SILVA, Nuno. 2010. *Cidadania e Segurança: Uma Análise Prospectiva.* Lisboa : s.n., 2010.

SMUHA, Nathalie; Comissão Europeia. 2019. Orientações éticas para uma IA de confiança. Bruxelas : Grupo de peritos de alto nível sobre a inteligência artificial, 2019. Vol. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60435.

SNOWDEN, Edward. 2019. *Vigilância Massiva Registo Permanente.* s.l. : Grupo Planeta, 2019. 978-989-777-312-9.

SOMERS, James. 2013. The man who would teach machines to think. *The Atlantic.* November, 2013, https://www.theatlantic.com/magazine/archive/2013/11/the-man-who-would-teach-machines-to-think/309529/.

STORM, Darlene. 2015. Orwellian Citizen Score, China's credit score system, is a warning for Americans. *Computerworld.* 2015. Vols. https://www.computerworld.com/article/2990203/security/aclu-orwellian-citizen-score-chinas-credit-score-system-is-a-warning-for-americans.html.

TADDEO, Mariorosaria and FLORIDI, Luciano. 2018. Regulate artificial intelligence to avert cyber arms race. *Nature.* [Online] 16 04 2018. https://www.nature.com/articles/d41586-018-04602-6.

TEIXEIRA, Nuno Severiano. 2002. *Contributos para a Política de Segurança Interna.* Lisboa : Ministério da Administração Interna, 2002.

The Cylance Data Science Team. 2017. *Introduction to artificial intelligence for security.* Irvine : The Cylance Data Science Team, 2017. 978-0-9980169-0-0.

The Guardian. 2019. [Online] 14 02 2019. https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction.

—. 2018. https://www.theguardian.com/world/2018/jan/10/russian-influence-brexit-vote-detailed-us-senate-report. [Online] 10 01 2018. https://www.theguardian.com/world/2018/jan/10/russian-influence-brexit-vote-detailed-us-senate-report.

—. 2019. The Guardian. [Online] 16 09 2019. https://www.theguardian.com/world/2019/sep/16/saudi-arabia-oil-attacks-everything-you-need-to-know.

—. 2019. The Guardian. [Online] 28 Janeiro 2019. https://www.theguardian.com/world/2019/jan/28/spanish-drug-smugglers-who-used-drones-to-spy-on-police-are-arrested.

—. 2018. The Guardian. [Online] 26 outubro 2018. https://www.theguardian.com/uk-news/2018/oct/26/seven-jailed-over-plot-fly-drones-drugs-uk-prisons.

THIES, Justus, et al. 2016. niessnerlab. *Visual Computing Group.* [Online] 2016. http://niessnerlab.org/papers/2016/1facetoface/thies2016face.pdf.

TONIN, Matej. Artificial Intelligence: Implications for NATO's Armed Forces. [Online] https://www.nato-pa.int/download-file?filename=sites/default/files/2019-04/088%20STCTTS%2019%20E%20-%20ARTIFICIAL%20INTELLIGENCE%20-%20DRAFT%20REPORT%20TONIN.pdf.

TURING, Alan. 1950. Computing machinery and intelligence. *Parsing the Turing Test.* Dordrech : Springer, 1950.

UW. 2006. *History of AI.* Washington : University of Washington, 2006. https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf.

VASCONCELOS, Álvaro. 2015. Debate: Terror em França/Uma questão europeia: a torre de Babel e os seus inimigos. 2015.

VASSILIOU, Marius, ALBERTS, David and AGRE, Jonathan. 2015. *C2 Re-Envisioned: the Future of the Enterprise.* Nova York : CRC Press, 2015. 9781466595804.

VIANA, Vítor Daniel. 2003. O conceito de segurança alargada e o seu impacto nas missões e organização das Forças Armadas. s.l. : Instituto de Altos Estudos Militares, 2003.

VISUAL CAPITALIST. 2019. Why Big Data Keeps Getting Bigger. [Online] 16 07 2019. https://www.visualcapitalist.com/big-data-keeps-getting-bigger/.

WELCH, Larry D. 2011. Cyberspace – The Fifth Operational Domain. *https://www.ida.org.* [Online] 2011. https://www.ida.org/-/media/corporate/files/publications/researchnotes/rn2011/2011-cyberspace---the-fifth-operational-domain.ashx?la=en&hash=4DCB5CA98848C3DDFE3B82067FD94768.

WOODWARD, John., et al. 2003. *Biometrics: A look at facial recognition.* Santa Mónica : RAND, 2003.

WTTC, World Travel & Tourism Council. 2019. www.wttc.org. *www.wttc.org.* [Online] 29 03 2019. [Cited: 12 08 2019.] https://www.wttc.org/about/media-centre/press-releases/press-releases/2019/1-in-every-5-euros-in-portugal-comes-from-tourism/.

YING, Tan, GODDARD, Steve and PÉREZ, Lance. 2008. *A Prototype Architecture for Cyber-Physical Systems.* Lincoln : ACM Sigbed Review, 2008.

ZEDNER, Lucia. 2003. *"Too much security?".* s.l. : International Journal of the Sociology Law, 2003. Vol. 31.

—. 2009. *Security, Routledge, Key Ideas in Criminology.* 2009.

ANEXO – I Entidades Entrevistadas

Código	Título	Nome	Organização
E1	Doutor	José Fontes	Academia Militar
E2	Doutora	Sofia Casimiro	Academia Militar
E3	Tenente Coronel	Rogério Raposo	Centro Nacional de Cibersegurança
E4	Doutora	Sonia Sousa Pereira	Europol
E5	Doutor	João Nuno Ferreira	FCT - FCCN
E6	Coronel	João Nortadas	GNR - Direção de Comunicação e Sistemas de Informação
E7	Tenente Coronel	João Nunes	GNR - Direção de Comunicação e Sistemas de Informação
E8	Major	Tiago Lopes	GNR - Direção de Investigação Criminal
E9	Major	Alfaro Pereira	GNR - Escola da Guarda
E10	Engenheiro	Miguel Souto	HP - Portugal
E11	Doutor	Rui Barata Ribeiro	IBM
E12	Professor	Rui Pereira	ISCPSI
E13	Mestre	António Nunes	OSCOT
E14	Procurador	Pedro Verdelho	PGR - Cibercrime
E15	Procurador	Rui Batista	PGR - Coordenação dos Sistemas de Informação
E16	Comissário	Rui Costa	PSP - Departamento de Investigação Criminal
E17	Inspetor-Chefe	Rogério Bravo	PJ - UNC3T
E18	Engenheiro	Bruno Banha	Warpcom - Diretor de Solutions Design
E19	Major	Silvestre Machado	Diretor de Segurança Auchan

ANEXO – II Guião de Entrevista

TÉCNICO LISBOA

Mestrado em Segurança da Informação e Direito do Ciberespaço

Segurança, Dados e a Inteligência Artificial

GUIÃO DE ENTREVISTA

Caraterização do(a) entrevistado(a):

Nome:

Cargo:

Organização:

Perguntas:

1. Considera o avanço da Inteligência Artificial (IA) como um dos principais desafios à segurança das sociedades modernas? Qual é a face do problema que lhe desperta mais preocupação?

2. Do ponto de vista da sua organização, considera importante desenvolver algum projeto/solução/programa com recurso a IA? Está algum em desenvolvimento ou operação? Que objetivos, funções ou tarefas desempenha?

3. Para que uma aplicação de IA seja eficaz, é necessário obter um grande volume de dados para o treino e ajustamento dos seus algoritmos. Considera que as Instituições Públicas, onde se incluem as Forças de Segurança, e o Sistema Judicial, devam ter acesso a dados públicos de forma generalizada sobre o cidadão comum (Dados de Saúde, Fiscais, Rendimentos, Biométricos) para fins de prevenção criminal? E dados privados (historial de pesquisas, compras, geo-localização, preferências pessoais)? E acesso a estes dados para investigação de crimes consumados?

4. As aplicações de reconhecimento facial permitem reconhecer pessoas desaparecidas, detetar comportamentos ou objetos estranhos e ainda alertar para crimes em curso. Capacidades que contribuem para uma sociedade mais segura. Por outro lado, permite que regimes autocráticos estejam a uns passos de concretizar uma profecia Orwelliana. Como conseguimos encontrar a virtude entre dois futuros possíveis?

5. Os princípios morais e éticos são pedras basilares de qualquer sociedade. As decisões tomadas sobre qualquer cidadão devem respeitar esses mesmos princípios. Que medidas devem ser tomadas para que garantir que as aplicações de IA são desenvolvidas de forma segura e livre de preconceitos?

6. A liberdade, a privacidade e a segurança parecem ser direitos inconciliáveis, que medidas devem ser tomadas para que a utilização das tecnologias que empregam Sistemas de IA, salvaguarde os Direitos, Liberdades e Garantias dos cidadãos? Que pressupostos, alcances e limites devem ser impostos às Forças de Segurança e ao Sistema Judicial na utilização de IA para combater o crime?