NOVA IMS Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

# Cibersegurança e Inteligência Artificial

## Como garantir a segurança de um Sistema de Informação

Vânia Filipa Moreira Queirós de Oliveira

Dissertação como requisito parcial para obtenção do grau de Mestre em Gestão de Informação

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

**NOVA Information Management School**
**Instituto Superior de Estatística e Gestão de Informação**
Universidade Nova de Lisboa

# CIBERSEGURANÇA E INTELIGÊNCIA ARTIFICIAL

por

Vânia de Oliveira

Dissertação como requisito parcial para a obtenção do grau de Mestre em Gestão de Informação, Especialização em Gestão de Sistemas e Tecnologias de Informação

**Orientador:** Professor Doutor Vítor Duarte dos Santos

Março 2021

# RESUMO

Com o avanço da tecnologia e da difusão de informação que tem ocorrido ao longo dos anos, os Sistemas de Informação tendem a ser cada vez mais tecnológicos e independentes da utilização humana. Tal tem sido possível com a ajuda da Inteligência Artificial, pois permite definir um comportamento a aplicar ao sistema, através do estudo e da compreensão do cérebro humano, sendo possível replicar as suas ações e reações.

No entanto, nem tudo é positivo, na medida em que ao serem construídos com base no comportamento humano, os Sistemas de Informação tornam-se alvos fáceis de atacar, dando origem a ataques cibernéticos com repercussões graves e que podem dar origem a situações de caos.

Consequentemente, torna-se fundamental proteger os Sistemas de Informação destes ataques, definindo possíveis barreiras nos sistemas, melhorando a Inteligência Artificial colocada nos mesmos e é crucial entender quais as vulnerabilidades da Inteligência Artificial.

Assim, será inicialmente realizada uma revisão detalhada da Cibersegurança e da Inteligência Artificial, permitindo compreender melhor as suas características, bem como as suas áreas.

O objetivo desta investigação será propor uma solução que irá ajudar as organizações a diminuir as suas vulnerabilidades. A intenção será evitar ataques cibernéticos, bem como minimizar os seus danos. Será construído um referencial, onde serão cruzadas as áreas da Cibersegurança com as áreas da Inteligência Artificial. Para tal, serão analisados vários artigos académicos, que permitirão comprovar o que será apresentado.

# PALAVRAS-CHAVE

Cibersegurança; Inteligência Artificial; Sistemas de Informação; Tecnologias de Informação; Vulnerabilidade;

# ABSTRACT

With the progress in technology and the dissemination of information that has been occurring over the last years, Information Systems tend to be more technological and independents of human beings. Artificial Intelligence has been a valuable helper, allows us to determine the behaviour we want to implement on the system, based on the study and comprehension of the human brain, let us replicate actions and reactions.

However, not everything is excellent once they were built considering human behaviour, Information Systems become an easy target for hackers, which is a reason why, nowadays, we have a lot of cyber-attacks with consequences that cause situations of chaos.

Consequently, becomes necessary to protect Information Systems against these potential attacks, define new barriers, improve the Artificial Intelligence that exists on the systems and is also extremely crucial to understand which are the weakness of Artificial Intelligence.

Thus, a detailed revision about Cybersecurity and Artificial Intelligence will be initially carried out, allowing a better understanding of their features, as well as the areas.

The main goal of this investigation will be proposing a solution that will help organizations reduce their vulnerabilities. The intention will be to avoid cyberattacks, as well as minimize the damage caused. Will be developed a referential where will be crossed Cybersecurity areas with the Artificial Intelligence areas. Thus, several academic articles will be analyzed, which will help to prove what will be presented.



# KEYWORDS

Cybersecurity; Artificial Intelligence; Information Systems; Information Technologies; Vulnerabilities.

# ÍNDICE

# ÍNDICE DE FIGURAS



# ÍNDICE DE TABELAS

# LISTA DE SIGLAS E ABREVIATURAS

| | |
|---|---|
| **A&R** | Automatização e Robótica |
| **AIS** | Artificial Immune System (Sistemas de Imunidade Artificial) |
| **ANN** | Artificial Neural Networks (Rede Neural Artificial) |
| **BP-PNN** | Back-Propagation Probabilistic Neural Network (Rede Neural Probabilística de Retropropagação) |
| **CPS** | Cyber-Physical System (Sistema Ciberfísico) |
| **DARE** | Data Analysis and Remediation Engine (Análise de Dados e Mecanismo de Correção) |
| **DoS** | Denial of Service (Ataque de negação) |
| **DDoS** | Distributed Denial of Service (Ataque de Negação Distribuído) |
| **DFA** | Differential Fault Analysis (Análise de Falha Diferencial) |
| **DL** | Deep Learning (Aprendizagem Profunda) |
| **dNIDS** | Distributed Network based Intrusion Detection System (Rede Distribuída baseada em Sistema de Deteção de Intrusão) |
| **DNN** | Deep Neural Network (Rede Neural Profunda) |
| **DSR** | Design Science Research |
| **DPA** | Differential Power Analysis (Análise de Potência Diferencial) |
| **FL** | Fuzzy Logic (Lógica Difusa) |
| **HIDS** | Host Intrusion Detection System (Sistema de Deteção de Intrusão de Hospedeiros) |
| **IA** | Inteligência Artificial |
| **IDS** | Intrusion Detection Systems (Sistema de Deteção de Intrusão) |
| **IDPS** | Intrusion Detection Prevention Systems (Sistema de Prevenção de Deteção de Intrusão) |
| **IEC** | International Electrotechnical Commission (Comissão Eletrotécnica Internacional) |
| **IoT** | Internet of Things (Internet das Coisas) |
| **IPS** | Intrusion Prevention Systems (Sistema de Prevenção de Intrusões) |
| **ISMS** | Information Security Management Systems (Sistemas de Gestão de Segurança da Informação) |
| **ISO** | International Organization for Standardization (Organização Internacional para Padronização) |
| **KNN** | K-Nearest Neighbor |

| | |
|---|---|
| **LR** | Logistic Regression (Regressão Logistíca) |
| **LSTM** | Long Short-Term Memory (Memória Longa de Curto Prazo) |
| **ML** | Machine Learning (Aprendizagem da Máquina) |
| **MV** | Machine Vision (Visão da Máquina) |
| **NB** | Naïve Bayes |
| **NFV** | Network Function Virtualization (Virtualização da Função da Rede) |
| **NID** | Network Intrusion Detection (Rede de Deteção de Intrusão) |
| **NLP** | Natural Language Processing (Processamento de Linguagem Natural) |
| **NIDS** | Network based Intrusion Detection Systems (Rede baseada em Sistemas de Deteção de Intrusão) |
| **NN** | Neural Network (Rede Neural) |
| **NSF** | Network Security Function (Rede baseada em Sistemas de Deteção de Intrusão) |
| **PCA** | Principal Component Analysis (Análise Componente Principal) |
| **PE** | Portable Executable (Executável Portátil) |
| **PRISMA** | Preferred Reporting Items for Systematic Reviews and Meta- Analyses (Itens de Relatório Preferidos para Revisões Sistemáticas e Meta-análises) |
| **ROC** | Receiver Operator Characteristic (Característica de Operador Recetor) |
| **RF** | Random Forest (Floresta Aleatória) |
| **RNN** | Recurrent Neural Network (Redes Neurais Recorrentes) |
| **ROS** | Robot Operating System (Sistema Operacional Robótico) |
| **SI** | Sistema de Informação |
| **SVC** | Support Vector Classification (Classificador de Vetores de Suporte) |
| **SVM** | Support Vector Machine (Máquina de Vetores de Suporte) |
| **SVR** | Support Vector Regression (Regressão de Vetores de Suporte) |
| **TF-IDF** | Term Frequency and Inverse Document Frequency (Frequência de Termo e Frequência de Documento Inverso) |
| **TI** | Tecnologias de Informação |
| **URL** | Uniform Resource Locators (Localizador Uniforme de Recursos) |
| **WBC** | White-Box Criptography (Criptografia Caixa Branca) |

(*) Nota: As siglas foram utilizadas, na sua maioria, de acordo com o significado em inglês para uma melhor compreensão. Apenas as siglas com tradução definida para português é que foram adaptadas para tal.

# 1. INTRODUÇÃO

## 1.1. CONTEXTUALIZAÇÃO

Atualmente, a tecnologia tem um papel crucial e fundamental no quotidiano. As pessoas estão ligadas entre si nas mais variadas redes de informação, já que possuem computadores pessoais, redes domésticas ou telefones ligados entre si ou à internet, o que lhes permite alcançar qualquer parte do mundo (Zúquete, 2018).

Devido à facilidade existente nos dias de hoje, tornou-se fundamental armazenar e gerir a informação de uma forma diferente, já não é possível guardar tudo apenas em formato de papel. Assim, os Sistemas de Informação passaram a ser utilizados com maior frequência (Estrela, 2014).

Atualmente, várias empresas utilizam os Sistemas de Informação para gerir o seu negócio, de forma a facilitar a interação com os clientes e/ou com os fornecedores, bem como, para aumentar a competitividade no mercado. Além disso, muitas empresas atualmente funcionam de acordo com o que o seu Sistema de Informação permite, isto é, o negócio foi adaptado ao que o Sistema de Informação permitia, não às necessidades da organização (Laudon & Laudon, 2014).

Os governos utilizam também estes sistemas para satisfazerem as necessidades dos cidadãos. Individualmente as pessoas acabam também por utilizar Sistemas de Informação diariamente e não têm conhecimento relativamente a isso, nomeadamente quando acedem a uma rede social ou a uma plataforma para poderem realizar compras online (Laudon & Laudon, 2014).

Os Sistemas de Informação podem ser definidos como sendo um conjunto de elementos, isto é, dados, atividades ou recursos (Rainer Jr. & Cegielski, 2007), utilizados para recolher, processar, armazenar e distribuir informação (Laudon & Laudon, 2014).

Podemos considerar que um Sistema de Informação se concentra em 3 atividades principais: a entrada, o processamento e a saída. Estas atividades são fundamentais para produzir a informação que uma organização e os seus respetivos membros necessitam para poderem controlar operações, tomar decisões, analisar problemas e criar produtos. A entrada capta e recolhe dados dentro da organização, enquanto o processamento trata a informação proveniente da empresa e, por sua vez, a saída devolve a informação tratada à parte interessada na mesma (Laudon & Laudon, 2014).

A maioria dos Sistemas de Informação são computorizados, no entanto, nem todos o são, daí o termo Sistema de Informação ser associado sinonimamente como Sistema de Informação baseado em computador, pois envolvem uma variedade de Tecnologias de Informação. Os seus componentes básicos são: hardware, software, bases de dados, rede, procedimentos e pessoas (Boell & Cecez-Kecmanovic, 2015).

O hardware envolve os processadores, monitores, teclados, impressoras, isto é, tudo o que aceita dados e informação, processa a mesma e permite ao utilizador observar. O software envolve um programa ou vários, que permitem que o hardware processe os dados. Por sua vez, as bases de dados permitem armazenar a informação em tabelas, que podem ser relacionadas entre si, através da utilização de chaves primárias, ou não relacionadas. A rede envolve toda a ligação entre o sistema, isto é, os utilizadores necessitam de estar ligados entre si, portanto, precisam de ter ligações

via *wireless* ou *wireline*. Os procedimentos são conjuntos de instruções que permitem combinar os componentes referidos anteriormente, de forma a processar a informação e obter o resultado desejado. Por fim, as pessoas são as que utilizam o hardware e o software e que pretendem obter o resultado (Rainer Jr. & Cegielski, 2007).

Os Sistemas de Informação podem ser utilizados para apoiar no momento de tomar uma decisão numa organização, bem como para coordenar e controlar, além de serem extremamente úteis para os funcionários no momento de analisarem problemas e procurar possíveis soluções para aplicar aos mesmos (Laudon & Laudon, 2014).

Estes têm-se tornado cada vez mais robustos, inteligentes e os seres humanos mais dependentes deles (Brynielsson, Franke, Adnan Tariq, & Varga, 2016), assim sendo, é importante garantir que os mesmos são seguros e confiáveis (Laudon & Laudon, 2014).

Devido a armazenarem cada vez mais informação confidencial e que deve ser mantida em sigilo para entidades externas às organizações (O'Brien & George, 2011), os Sistemas de Informação tornam-se alvo do interesse dos criminosos do universo tecnológico, os chamados *hackers*. Estes são conhecidos por acederem aos sistemas de forma ilegal e por norma, venderem a informação a concorrentes da organização que foi atacada (Fang, et al., 2016).

Como tal, é importante garantir que um Sistema de Informação cumpre as normas de segurança requeridas para se manter longe de possíveis ataques cibernéticos. Garantindo que os mesmos não são vulneráveis, sendo que a vulnerabilidade é uma característica de um sistema que o torna sensível a certos ataques, na medida em que um ataque bem-sucedido é um risco ou uma ameaça (Zúquete, 2018).

Por conseguinte, a Cibersegurança é importante, uma vez que um ataque bem-sucedido pode originar uma situação muito dispendiosa para uma organização, podendo arruinar a sua reputação, o seu negócio e, ainda, o seu poder económico, uma vez que nem sempre é possível detetar o momento exato em que um ataque ocorre. Além de que os ataques têm se tornado cada vez mais sofisticados. Sendo que a Cibersegurança é uma medida usada para proteger um computador ou um sistema de computadores contra um acesso não autorizado ou um ataque (Von Solms & Van Niekerk, 2013).

Um “aliado” útil para ajudar a construir um Sistema de Informação mais seguro é a Inteligência Artificial (Desouza, Dawson, & Chenok, 2019). Pode ser definida como uma disciplina que se centra no “estudo e construção de entidades artificiais com capacidades cognitivas semelhantes às dos seres humanos” (Costa & Simões, 2011), como a aprendizagem e a resolução de problemas.

Além disso, um Sistema de Informação com Inteligência Artificial permite simular ataques de forma a conhecer as vulnerabilidades do sistema e torná-lo mais forte (Costa & Simões, 2011).

## 1.2. IDENTIFICAÇÃO DO PROBLEMA

A utilização da Inteligência Artificial nas mais variadas áreas de computação, está nos dias de hoje vulgarizada entre as diferentes áreas, onde existe uma promessa de grandes avanços através da utilização da Inteligência Artificial e da Cibersegurança.

Contudo, não é claro para um gestor de redes de computação e de recursos informáticos, qual a melhor estratégia para tirar proveito de forma eficaz destas tecnologias.

Assim sendo, este projeto tem o intuito de ajudar a perceber de que forma será possível tornar os Sistemas de Informação mais seguros e menos vulneráveis a possíveis ataques cibernéticos, recorrendo a técnicas de Inteligência Artificial.

## 1.3. RELEVÂNCIA E IMPORTÂNCIA DO ESTUDO

Devido à importância dos Sistemas de Informação nos dias de hoje, as pessoas tornaram-se dependentes dos mesmos, visto que muitos permitem que as tarefas do dia-a-dia sejam realizadas em um menor espaço de tempo ou até mesmo que o utilizador não necessite de se preocupar com as mesmas, uma vez que funcionam sozinhas.

Portanto, o presente estudo visa clarificar o papel da Inteligência Artificial na Cibersegurança e produzir um "guião" de aplicabilidade destas tecnologias na prevenção, deteção e reação da Cibersegurança.

Tal estudo poderá ser utilizado pelos responsáveis da segurança de dados na definição de estratégias e projetos de implementação de segurança, mas também poderá servir de base a outros estudos sobre a utilização das tecnologias mais modernas nesta área.

É necessário garantir que o utilizador confia na execução e no comportamento do Sistema de Informação que utiliza. Além disso, é importante assegurar que o mesmo é seguro e que vai ser resistente a possíveis ataques. É de extrema viabilidade salvaguardar que as possíveis fragilidades do sistema se encontram protegidas e fora do alvo de possíveis malfeitores.

Assim sendo, é importante definir uma proteção para estes sistemas, para tal, devemos analisar as fraquezas apresentadas por estes sistemas, verificar ataques passados e procurar uma solução para uma melhoria dos mesmos, tornando-os sistemas inabaláveis e fora do alcance dos *hackers*.

O resultado deste estudo poderá ser muito importante para as organizações que utilizam Sistemas de Informação e colocam nos mesmo um grande poder relativamente aos seus negócios, isto é, organizações que dependem a grande escala do bom funcionamento do seu Sistema de Informação. Se por alguma razão o Sistema de Informação falha, irá provocar o caos dentro da organização ou até mesmo poderá levar à exposição de dados confidenciais perante terceiros, podendo apresentar uma repercussão enorme perante a sociedade, deixando a organização numa situação bastante complicada.

Além disso, ajudará a compreender e a conhecer as vulnerabilidades existentes nos Sistemas de Informação, bem como a melhoria que pode ser realizada nos mesmos de forma a torná-los mais

resistentes. Contudo, irá ser apenas efetuada uma análise teórica do que se pretende provar, irão ser analisados ataques anteriores, de forma a ajudar a compreender as fraquezas dos sistemas e de que forma os ataques ocorreram.

## 1.4. OBJETIVO DO ESTUDO

O objetivo principal deste estudo é propor uma abordagem que possa ajudar a proteger os Sistemas de Informação de ataques cibernéticos utilizando Inteligência Artificial.

De forma a se encontrarem possíveis soluções, será construído um referencial e definido um processo capaz de relacionar as técnicas de Inteligência Artificial com os problemas de segurança existentes.

Assim, cruzando ambos será possível determinar de que modo a Inteligência Artificial conseguirá colmatar os problemas encontrados ao nível de Cibersegurança. Após a criação do referencial o mesmo será validado, para garantir que o referencial está de acordo com a informação recolhida.

O intuito será:

- Construir um referencial, onde serão cruzadas as técnicas de Inteligência Artificial a aplicar na Cibersegurança, de modo a determinar qual a melhor técnica de Inteligência Artificial para cada situação de falha de segurança e ajudar os gestores das áreas de Tecnologias de Informação da organização a garantir a segurança dos seus Sistemas de Informação, assim foi definida a hipótese:

  **H1**: Será possível construir um referencial de aplicabilidade das técnicas de Inteligência Artificial na Segurança Informática que possa ser facilmente utilizado pelos profissionais investigacionais na área?

- Construir um processo que ajuda a identificar qual a melhor abordagem na utilização de Inteligência Artificial para garantir uma maior segurança nos Sistemas de Informação, que possa ser utilizado pelos responsáveis de Tecnologias de Informação nas organizações, assim foi definida a hipótese:

  **H2**: Será possível construir um processo capaz de definir a melhor estratégia de usar Inteligência Artificial na Segurança Informática que possa ser facilmente utilizado pelos profissionais investigacionais na área?

O resultado obtido deverá variar em ambas as hipóteses e irá permitir aos gestores de Tecnologias de Informação de uma organização protegerem os seus Sistemas de Informação, tendo como base qual o possível ataque de segurança que podem sofrer e cruzar os mesmos com as diferentes áreas da Inteligência Artificial.

Para garantir um melhor entendimento e uma análise mais detalhada e consciente do problema, será necessário realizar um estudo primário dos conceitos, Cibersegurança e Inteligência Artificial, como se relacionam entre si e de que forma podem complementar-se.

## 1.5. ESTRUTURA

Este documento estará dividido em 5 capítulos, representando os diferentes passos seguidos para se obter o resultado. A estrutura seguida será a seguinte:

- **Introdução** – Contextualização do tema abordado ao longo do documento, apresentação dos diferentes conceitos a estudar, objetivo do estudo e qual o caminho a seguir para obter o resultado pretendido.

- **Revisão de Literatura** – Descrição detalhada dos conceitos fundamentais para a realização deste estudo. A Inteligência Artificial e a Cibersegurança serão os conceitos abordados neste capítulo. Existirá ainda uma secção referente à Revisão Sistemática de Literatura onde serão selecionadas diferentes palavras-chave que irão devolver diferentes artigos. Estes serão analisados e os artigos relevantes selecionados, isto é, que abordem a utilização da Inteligência Artificial em situações onde possam ocorrer ciberataques.

- **Metodologia** – Apresentação e detalhe do método abordado na realização do presente estudo, isto é, qual a estratégia que será seguida de forma a implementar o estudo.

- **Proposta** – Análise realizada após a conclusão da Revisão Sistemática da Literatura, onde será construído um referencial, na qual se irão cruzar as técnicas de Inteligência Artificial com as técnicas de Cibersegurança. Será ainda incluída uma secção para as entrevistas com intervenientes que têm experiência na área da Inteligência Artificial e/ou da Cibersegurança, para verificar a viabilidade do referencial construído. No final deste capítulo, será apresentada a discussão efetuada após a validação das entrevistas realizadas.

- **Conclusão –** Análise do resultado obtido nos pontos anteriores e resumo do que foi possível obter com o presente estudo, bem como, quais foram as limitações encontradas no processo de desenvolvimento do estudo e o que se pretende realizar de forma a resolver no futuro as limitações encontradas.

# 2. REVISÃO DE LITERATURA

Neste capítulo, irão ser apresentadas as bases teóricas que serão utilizadas para o desenvolvimento deste projeto.

A Revisão de Literatura tem como objetivo identificar quais as áreas da Cibersegurança que podem ser suportadas por técnicas de Inteligência Artificial. Assim sendo, a revisão de literatura foi organizada em duas partes.

Na primeira parte, foram investigados os principais conceitos, bem como as áreas associadas à Cibersegurança e à Inteligência Artificial.

Na segunda parte, tendo como base de pesquisa os conceitos investigados na secção anterior, foi realizada uma Revisão Sistemática de Literatura, seguindo a metodologia PRISMA onde foram definidas as seguintes questões de pesquisa, para o qual pretendemos obter resposta:

1. Quais as técnicas de Inteligência Artificial com impacto na Cibersegurança?
2. Como pode ser utilizada a Inteligência Artificial na prática da Cibersegurança?
3. Como uma técnica de Inteligência Artificial permitiu a resolução de um problema de Cibersegurança?

## 2.1. CONCEITOS

### 2.1.1. Cibersegurança

A Cibersegurança consiste na proteção e defesa de sistemas, redes e programas no ciberespaço contra possíveis ataques maliciosos, nomeadamente através da aplicação de normas, regulamentos, encriptação, de forma a se conseguir evitar possíveis danos, sejam estes a nível de hardware ou de software (Miller & CISSP, 2016).

Pode ser também considerada como um conjunto de ferramentas, políticas, conceitos de segurança, guias, abordagens de gestão de risco, melhores práticas, tecnologias que podem ser utilizadas para proteger o ciberespaço, organizações e utilizadores. Computadores, infraestruturas, serviços de sistemas de telecomunicações também devem ser protegidos, uma vez que guardam informação relevante no ciberespaço (Von Solms & Van Niekerk, 2013).

A Cibersegurança é utilizada para garantir determinadas propriedades de segurança (Gourisetti, Mylrea, & Patangia, 2020), para se evitarem possíveis riscos no ciberespaço, tal como, integridade, disponibilidade e confidencialidade (Von Solms & Van Niekerk, 2013).

Contudo, a Cibersegurança não se foca apenas na proteção do ciberespaço, mas também na proteção do que funciona no ciberespaço e em qualquer um dos seus ativos que possam ter uma relação direta ou indireta com o ciberespaço (Von Solms & Van Niekerk, 2013).

O conceito de ciberespaço aplica-se ao domínio global existente dentro de um ambiente que gera uma coleção de informação, isto é, dentro de uma rede de Sistemas de Informação, como a internet, rede de telecomunicações, sistemas computacionais, processadores e controladores (Shiode, 2000).

A Cibersegurança é por vezes alvo de atividades ilícitas, tal como, o acesso a informação que se encontra reservada e é confidencial a entidades externas de determinada organização e que por conseguinte, não devem ser tornadas públicas, estas são guardadas em sistemas computacionais ou em redes de informação. Efetuar alterações à informação, isto é, alterar ou apagar informação à qual não se tem autorização para modificar. Por vezes, ocorre uma utilização exagerada ou abusiva dos recursos computacionais, ou seja, um recurso tem uma utilização exagerada relativamente a uma exploração acima do esperado e sofre de uma utilização abusiva quando é utilizado por quem não deve ter acesso ao recurso. (Zúquete, 2018).

A segurança nos sistemas computacionais é cada vez mais um problema simultaneamente técnico e social. É um problema técnico, no sentido em que as arquiteturas de hardware, os sistemas operativos e os protocolos aplicacionais são cada vez mais complexos, o que exige politicas de segurança mais complexas, que se tornam mais dificeis de pôr em prática e manter o seu bom funcionamento. Além disso, é um problema social, uma vez que a maioria dos utilizadores de Sistemas de Informação não possuem capacidades técnicas suficientes para saber lidar com os problemas de segurança (Zúquete, 2018).

Garantir que um sistema é seguro o bastante para sobreviver a um ataque cibernético, é algo com o qual as organizações necessitam de se preocupar nos dias de hoje, na medida em que um ataque bem-sucedido por parte de um *hacker*, ou seja, a pessoa que se infiltra num sistema sem permissão do dono do mesmo, pode originar uma situação bastante dispendiosa para a entidade que sofreu o dano, podendo arruinar a sua reputação, o seu negócio e o seu poder económico, pois nem sempre é possível detetar o momento exato em que um ataque ocorre, além de que os ataques têm se tornado cada vez mais sofisticados (Pande, 2017).

#### 2.1.1.1. Normas da Cibersegurança

A norma 27001 (ISO/IEC 27001:2013, 2013) foi definida pela ISO e pela IEC. Inclui um conjunto de normas que cooperam relativamente ao interesse de ambas as entidades. A norma foi adaptada posteriormente para Portugal, de forma a compreender as necessidades existentes no país.

Uma norma deve basear-se nos riscos existentes a nível de segurança, avaliando e identificando os mesmos, para se conseguir determinar como reduzir os seus efeitos e atingir uma melhoria contínua. Assim, deverão ser considerados os seguintes objetivos de controlo:

**Políticas de segurança da informação**

Políticas de segurança têm como objetivo proporcionar diretrizes para ajudarem as organizações a gerir a sua informação, tendo em conta as necessidades do seu negócio, com as leis do país, bem como regulamentações existentes. As mesmas deverão ser revistas consoante possíveis alterações que podem existir nos negócios e/ou leis do país em questão (ISO/IEC 27001:2013, 2013).

**Organização de segurança da informação**

Tem como objetivo estabelecer um modelo de apoio à gestão, de forma a controlar a implementação das normas dentro de uma organização, como tal, é necessário definir as responsabilidades. As funções com maior responsabilidade devem ser definidas inicialmente, não permitindo uma alteração mais tarde, deverão ser também envolvidas entidades e grupos competentes na área, além disso, a segurança deverá ainda ser definida na gestão do projeto (ISO/IEC 27001:2013, 2013).

Contudo, é importante também garantir a segurança no teletrabalho, bem como na utilização de dispositivos móveis, assim, deverão ser adotadas políticas e medidas de segurança para proteger a informação acedida em locais de teletrabalho e mitigar os riscos que advêm da utilização de equipamentos móveis (ISO/IEC 27001:2013, 2013).

**Segurança na gestão de recursos humanos**

Pretende garantir que os funcionários e prestadores de serviços compreendem o seu papel na organização e que são capazes de executar a sua função, como tal, deverão ser previamente verificadas as suas credenciais e referências, considerando as leis e regulamentos. Além disso, quando é efetuado um contrato, a função tanto dos funcionários quanto dos prestadores de serviços deverá ser descrita. O mesmo deve acontecer quando ocorre a cessação do contrato, deverá ficar claro quais as responsabilidades e deveres a cumprir. (ISO/IEC 27001:2013, 2013).

Por conseguinte, é importante garantir que as responsabilidades dos funcionários e dos prestadores de serviços são cumpridas, a gestão deverá validar o mesmo, tendo em consideração as políticas e procedimentos da organização. Os funcionários e prestadores de serviço deverão ter conhecimento de possíveis formações que ajudem a consciencializar relativamente à melhor forma de garantir a segurança da informação, bem como, quando ocorre uma atualização da política da organização. Se porventura, um funcionário ou prestador de serviço incumpre alguma regra, o mesmo deverá ser informado, através de uma ação disciplinar formal (ISO/IEC 27001:2013, 2013).

**Gestão de ativos**

O objetivo é identificar os ativos existentes numa organização e definir responsabilidades para garantir a proteção apropriada, deverá ser mantido um inventário com a informação referente aos ativos, os mesmos deverão ter um responsável associado. As regras de utilização dos ativos terão de ser identificadas, documentadas e implementadas para uma utilização aceitável da informação. Se os ativos estiveram na posse de alguma entidade externa, os mesmos deverão ser devolvidos após a cessação do contrato existente com a organização (ISO/IEC 27001:2013, 2013).

É importante garantir que a informação tem um nível de proteção adequado, considerando a importância que possui perante a organização, a informação deverá ser classificada tendo em conta os requisitos legais, valor, importância e sensibilidade caso ocorra uma divulgação não autorizada. A mesma terá de ser etiquetada e deverão ser desenvolvidos procedimentos para gestão de ativos utilizando o esquema de classificação adotado pela organização (ISO/IEC 27001:2013, 2013).

A divulgação não autorizada da informação, bem como a sua modificação ou eliminação deverá ser prevenida, devendo ser implementados procedimentos para gestão da informação, tendo em conta a classificação adotada pela organização. Se for efetuada uma eliminação a mesma terá de ser realizada de maneira segura, utilizando procedimentos formais (ISO/IEC 27001:2013, 2013).

**Controlo de acesso**

É necessário limitar o acesso à informação e aos recursos de processamento de informação, para tal, deverá ser realizada uma política de controlo de acesso, considerando os requisitos de negócio e de segurança da informação. Além disso, os utilizadores deverão ter atribuídos a si apenas os acessos para os quais têm autorização (ISO/IEC 27001:2013, 2013).

Deverá ser garantido que apenas os utilizadores com permissões conseguem aceder à informação e que utilizadores sem permissões não o conseguem fazer. Assim sendo, deverá ser criada uma estrutura de dados que permita definir as permissões e cancelamentos que deverão estar atribuídos aos utilizadores, contudo, deverão ser definidas restrições para a atribuição de privilégios, sendo apenas o responsável pelo ativo capaz de controlar as suas permissões. Quando ocorrer uma cessação de contrato, o acesso do funcionário deverá ser removido ou cancelado, deixando de ter acesso a qualquer tipo de informação (ISO/IEC 27001:2013, 2013).

No processo de autenticação, os utilizadores deverão ser responsáveis pela proteção da sua informação, no entanto, deverão cumprir as normas de autenticação definidas pela organização (ISO/IEC 27001:2013, 2013).

**Criptografia**

Garantir uma utilização adequada e eficaz da criptografia de forma a proteger a confidencialidade, integridade e autenticidade da informação, deverá ser criada uma política para gestão dos controlos criptográficos durante o seu ciclo de vida (ISO/IEC 27001:2013, 2013).

**Segurança física e ambiental**

Evitar que alguém sem autorização consiga aceder à informação, para evitar danos e interferências na informação e gestão da informação, será necessário criar perímetros de segurança para proteger as áreas que têm informação sensível ou critica, uma opção será definir controlos de entrada, garantindo que apenas quem tem autorização consegue aceder. Deverão ainda ser garantidas medidas de proteção para situações de catástrofes naturais garantindo que não se irá perder informação (ISO/IEC 27001:2013, 2013).

Evitar a perda, dano ou roubo de ativos, para tal, deverão colocar os equipamentos de forma a protegê-los, considerando todos os riscos existentes, nomeadamente as ameaças e os acessos não autorizados. Deverá ser garantido que os equipamentos se encontram seguros e estáveis para garantir a sua disponibilidade e integridade. Assim, não deverão ser retirados sem autorização prévia (ISO/IEC 27001:2013, 2013).

**Segurança de operações**

Os procedimentos de operação deverão ser documentados e disponibilizados a todos os utilizadores que necessitem de utilizar os recursos para processamento de informação. Qualquer alteração existente na organização, seja ao nível de processos de negócio, recursos e sistemas deverá ser controlada. A utilização de recursos deve ser monitorizada e ajustada para requisitos de capacidade futura, a fim de assegurar o necessário desempenho dos sistemas. Os diferentes ambientes existentes na organização deverão ser separados de forma a reduzir os riscos de acessos não autorizados ou alterações no ambiente de produção (ISO/IEC 27001:2013, 2013).

Garantir a proteção contra código malicioso, tal poderá ser feito através da implementação de controlos de prevenção e deteção. Garantir ainda que não irá ocorrer perda de dados, devendo frequentemente fazer cópias de segurança, para salvaguardar a informação (ISO/IEC 27001:2013, 2013).

Regularmente deverá haver uma revisão e manutenção dos registos de eventos que têm informação sobre atividades dos utilizadores, os mesmos devem ser protegidos contra alterações e acessos não autorizados (ISO/IEC 27001:2013, 2013).

Em ambiente de produção, devem ser implementados procedimentos para controlar a instalação de software. Gerir vulnerabilidades é algo necessário de ser feito, para evitar que a organização seja exposta às suas vulnerabilidades, além disso, devem ser definidas medidas para evitar riscos. Em situações de auditoria, em ambientes de produção, deverão ser definidos previamente os passos a seguir, de forma a minimizar as interrupções no processo de negócio (ISO/IEC 27001:2013, 2013).

**Segurança de comunicações**

As redes devem ser geridas para proteger a informação existente nos sistemas, nas aplicações devem ser definidos mecanismos de segurança e requisitos de gestão para todos os serviços de rede identificados, sendo eles internos ou externos. Utilizadores, serviços de informação e sistemas devem ser agregados em redes (ISO/IEC 27001:2013, 2013).

Devem ser definidas políticas para garantir a transferência de informação segura, a mesma deverá ser acordada entre a organização e as entidades externas, a informação deverá ser protegida no momento da transferência (ISO/IEC 27001:2013, 2013).

**Aquisição, desenvolvimento e manutenção de sistemas**

Os Sistemas de Informação devem garantir a segurança da informação ao longo do seu tempo de vida, os requisitos de segurança devem ser incluídos nos Sistemas de Informação e a informação deverá ser protegida contra atividades ilegais e não autorizadas (ISO/IEC 27001:2013, 2013).

Para desenvolver novo software, deverão ser previamente definidas regras e as mesmas deverão ser controladas através da utilização de processos formais para controlar as modificações efetuadas (ISO/IEC 27001:2013, 2013).

**Relações com fornecedores**

A organização tem de garantir a proteção dos ativos acessíveis aos fornecedores, mitigando os riscos que poderão ser atribuídos aos acessos dos fornecedores, devendo os mesmos ser acordados e documentados (ISO/IEC 27001:2013, 2013).

Os serviços disponibilizados pelos fornecedores deverão ser auditados, monitorizados e revistos de forma regular para garantir que todos os requisitos são definidos. Os requisitos deverão variar consoante o fornecedor e o tipo de informação que o mesmo gere (ISO/IEC 27001:2013, 2013).

**Gestão de incidentes de segurança da informação**

Gerir os incidentes de segurança, considerando os pontos fracos de segurança, estabelecendo procedimentos e responsabilidades na gestão dos incidentes (ISO/IEC 27001:2013, 2013).

**Aspetos de segurança de informação na gestão da continuidade do negócio**

A organização deve determinar os requisitos de segurança da informação e a continuidade da gestão em situações de adversidade, devendo ser estabelecida, documentada e gerida a continuidade da segurança da informação (ISO/IEC 27001:2013, 2013).

**Conformidade**

Evitar o incumprimento de regulamentos, contratos e obrigações legais com os requisitos de segurança, definindo previamente os requisitos que deverão ser considerados para o Sistema de Informação e para a organização (ISO/IEC 27001:2013, 2013).

### 2.1.1.2. Riscos da Cibersegurança

Existem várias áreas na Cibersegurança que devemos considerar enquanto riscos, como:

- **Intrusão** - Uma intrusão é qualquer conjunto de ações com o intuito de comprometer a integridade, a confidencialidade ou a disponibilidade de um recurso. Uma intrusão resulta da execução de um ou mais ataques aos sistemas que gerem esse recurso. Ataques esses que podem ou não provocar alterações permanentes na informação guardada nesses sistemas. Constitui um risco difícil de avaliar, uma vez que não necessita de envolver exatamente um dado, no entanto, concede acesso a algo que normalmente é negado ao intruso (Zúquete, 2018).
- **Acesso a informação reservada ou confidencial** - Os computadores armazenam informação, por conseguinte, todos os acessos que não são autorizados, são definidos como riscos (Zúquete, 2018).

- **Perda ou roubo de informação** – Abriga todas as situações onde a informação se perde ou é subtraída por indivíduos não autorizados, podendo passar inclusive para a sua posse (Zúquete, 2018).
- **Personificação** - Ocorre quando um individuo subverte um sistema de autenticação, fazendo-se passar por outra pessoa ou quando um comportamento definido para uma aplicação sofre uma alteração. Por vezes é usada como despiste (esconder a verdadeira identidade de uma máquina) ou apropriação (utilização de identidade alheia) (Zúquete, 2018).

#### 2.1.1.3. Ataques de Cibersegurança

Existem vários tipos de ciberataques que ocorrem nos dias de hoje e com os quais os utilizadores de Sistemas de Informação devem ter atenção e cuidado de forma a não serem alvos dos mesmos. Os ataques mais conhecidos são os Malware, scarewares, botnets e ataques de DoS.

**Malware**

O conceito de Malware ou software malicioso engloba todo e qualquer software que tenha sido alterado com o objetivo de danificar dispositivos, roubar informação e assumir controlo, seja a nível individual ou a nível organizacional. Existem vários tipos de Malware, como backdoors, spyware, cavalo de troia, vírus, entre outros. Por exemplo, o spyware é um tipo especial de Malware que é instalado no computador alvo, com ou sem a permissão do utilizador, é utilizado para adquirir informação confidencial e sensível (Pande, 2017).

Os scarewares são softwares que à primeira vista são inofensivos, no entanto, são bastante nocivos, na medida em que por vezes induzem o utilizador ao engano, indicando que um website é fiável, todavia, o mesmo tem Malware e infeta a máquina utilizada para o acesso ao website. Este software é por vezes vendido como sendo uma solução para proteger o utilizador de ataques, no entanto, o seu objetivo é roubar informação pessoal de quem o adquiriu (Landage & Wankhade, 2013).

**Ataques Denial of Service (DoS)**

Um ataque de DoS é um ataque que tem o objetivo principal de inativar uma máquina ou uma rede, para que esta se torne inacessível aos utilizadores, este é executado através do bloqueio do tráfego, nomeadamente, enchendo-o de pedidos ou então acionando uma falha no sistema (Carl, Kesidis, Brooks, & Rai, 2006).

**Phishing**

Phishing é um tipo de ciberataque que usa o e-mail como arma, o recetor da mensagem acredita na viabilidade do remetente, abre o e-mail, seleciona a hiperligação que normalmente é disponibilizada e a informação sobre o utilizador é obtida sem a sua permissão. Consegue obter informação como senhas ou dados do cartão de crédito (Jagatic, Johnson, Jakobsson, & Menczer, 2007).

**Ransomware**

Este tipo de ataque cibernético restringe o acesso ao sistema do computador ou aparelho que se pretende atacar e tentam pedir um resgate ao dono, de forma a libertar o acesso. O Ransomware consegue atingir uma máquina através da receção de um anexo enviado por e-mail ou através do browser, normalmente quando se visita uma página que já tenha sido infetada com o vírus (Pope, 2016).

**SQL Injection**

É um tipo de ciberataque na qual se aproveita falhas em sistemas ligados ou que interagem com bases de dados. O ataque é executado através de comandos SQL, onde o atacante insere uma instrução SQL personalizada dentro que uma *query* (Halfond, Viegas, & Orso, 2006).

**Cross-site Scripting**

É uma vulnerabilidade definida no sistema de um computador, presente normalmente em aplicações web que ativam ataques maliciosos ao inserirem scripts dentro de páginas web que são acedidas por outros utilizadores. Estes scripts permitem que os atacantes consigam escapar do controlo efetuado durante o acesso (Wassermann & Su, 2008).

**Credential Stuffing**

Este tipo de ciberataque consiste no roubo de credenciais de acesso, normalmente de utilizadores e endereços de e-mail, bem como as passwords correspondentes, estas são utilizadas posteriormente para obter acesso não autorizado em aplicações web (Pal, Daniel, Chatterjee, & Ristenpart, 2019).

### 2.1.2. Inteligência Artificial

Os seres humanos possuem uma capacidade cognitiva, tal como a aprendizagem e a resolução de problemas, que durante vários anos foi investigada através de diferentes estudos e experiências, de forma a se entender o seu funcionamento, isto é, como compreendemos, como previmos ou como manipulamos o que está à nossa volta. A Inteligência Artificial (IA), no entanto, não se limita apenas em entender, mas também em construir entidades inteligentes (Russell, S. J. , & Norvig, P., 2010).

A Inteligência Artificial começou a dar os primeiros passos depois da Segunda Guerra Mundial, por volta dos anos 40, contudo, só por volta de 1956 passou a ter maior relevância (Russell, S. J. , & Norvig, P., 2010). No entanto, só na última década é que se tornou possível avançar com descobertas nesta área, devido ao avanço que ocorreu nas ciências computacionais e de informação (Desouza, Dawson, & Chenok, 2019).

Existem quatro abordagens diferentes para a Inteligência Artificial que devem ser seguidas, nomeadamente, pensar humanamente, pensar racionalmente, agir humanamente e agir

racionalmente. É possível verificar que as abordagens se baseam em duas dimensões, o pensamento e o raciocínio (Russell, S. J. , & Norvig, P., 2010).

Também podemos considerar diferentes metáforas para compreender a Inteligência Artificial, sendo elas a computacional, conexionista e biológica. A computacional parte do princípio de olhar para a inteligência enquanto computação, isto é, os computadores e a mente humana têm mecanismos de perceção e atuação, para além dos cognitivos. A conexionista vê a inteligência como sendo uma propriedade emergente das interações de um número elevado de unidades elementares de processamento, nomeadamente em relação ao cérebro humano, mais especificamente a relação existente entre os neurónios. Já a metáfora biológica foca-se na maneira como as espécies evoluem, sendo que esta evolução se deve à seleção natural e promove a sobrevivência das espécies mais evoluídas e adaptadas. Um bom exemplo, é a possibilidade de camuflagem por parte de algumas espécies, permitindo-lhes esconderem-se dos seus predadores (Costa & Simões, 2011).

Engloba uma grande variedade de componentes, varia desde a forma como se aprende até à forma como se joga xadrez, podemos assim afirmar que a Inteligência Artificial é relevante em qualquer atividade do dia a dia. Foca-se essencialmente na utilização de algoritmos ou cálculos, por vezes baseados num objetivo. Os algoritmos são instruções ambíguas que um computador está apto a executar, por vezes estes algoritmos são adaptados a partir de outros já existentes, mas que não estão focados no que se pretende (Russell, S. J. , & Norvig, P., 2010).

Carros que conduzem sozinhos são uma possibilidade devido à Inteligência Artificial, o que permite ao sistema pilotar o veículo sem a interferência do condutor, através da visão do computador, reconhecimento da imagem e a aprendizagem profunda (Russell, S. J. , & Norvig, P., 2010).

Como dito anteriormente, o objetivo da Inteligência Artificial é construir entidades inteligentes, estas podem ser também denominadas como agentes. Um agente pode recolher informação do ambiente, para que possa atuar sobre o mesmo, de forma, a determinar qual a melhor decisão (Costa & Simões, 2011).

Agentes deverão possuir capacidades como a autonomia, flexibilidade e aprendizagem. Se um agente for autónomo, então o mesmo é capaz de tomar decisões sem a intervenção de outros agentes. Contudo, um agente não necessita apenas de esperar uma mudança no ambiente na qual se encontra inserido para poder devolver uma resposta, podendo assim possuir o seu próprio estado emocional ou personalidade (Costa & Simões, 2011).

Existem 5 tipos de agentes:

- **Agentes reativos** – são máquinas simples que se limitam a reagir aos estímulos que recebem, do ambiente, no entanto, poderão agir mesmo sem a receção de estímulos, apenas como resposta ao ambiente na qual se encontra. Podendo assim ser representados tendo como base a perceção e a partir dela descobrir-se a ação (Costa & Simões, 2011).
- **Agentes de procura** – estes agentes devem ser capazes de entender os estados existentes nas ações e construir com base nisso uma representação interna dos mesmos, para além disso, devem possuir a capacidade de agir sobre eles, tendo em consideração as regras de funcionamento dos sistemas em questão. Permitem a ocorrência de transições entre estados e têm a capacidade de reconhecer quando atingem um estado definido como final, isto é,

quando atingem o objetivo ou encontram a solução para o qual foram planeados. É também importante que o agente tenha presente uma estratégia que seja completa (possível de encontrar solução), discriminadora (se existirem várias soluções, a melhor será a escolhida) ou económica (encontra a solução no menor espaço de tempo, gastando a menor quantidade de memória possível) (Costa & Simões, 2011).

- **Agentes baseados em conhecimento** – estes agentes necessitam de ter conhecimento e raciocínio para aumentarem o seu desempenho, assim sendo, um agente tem de construir a sua imagem do mundo, como tal, é necessário saber representar o conhecimento e interagir para conseguir desenvolver o raciocínio, tendo assim a capacidade de decidir (Costa & Simões, 2011).
- **Agentes aprendizes** – aprender é uma característica dos seres inteligentes, de tal forma que a aprendizagem artificial se tornou numa área central da Inteligência Artificial. A aprendizagem artificial possui três objetivos principais: o desenvolvimento de teorias computacionais da aprendizagem, a implementação de sistemas com capacidade para aprender e, ainda, a análise teórica e desenvolvimento de algoritmos genéricos de aprendizagem. Um agente aprendiz baseia-se assim na perceção e na ação (agentes reativos), bem como na capacidade de decidir e aprender (Costa & Simões, 2011).
- **Agentes adaptativos** - utilizam algoritmos genéticos, estes são técnicas que permitem efetuar otimização, sendo úteis para a resolução de problemas, podem ser consideradas técnicas inteligentes, pois permitem trabalhar simultaneamente em soluções alternativas, sendo ferramentas poderosas quando aplicadas para resolver problemas. A ideia principal de um algoritmo genético é imitar o que a natureza faz através do seu processo iterativo, evolui o algoritmo, com base numa população e define uma solução padrão, adaptando as soluções aos problemas existentes (Costa & Simões, 2011).

No entanto, apesar de serem máquinas com uma inteligência superior à usual para este tipo de sistemas, não conseguem, por vezes, sobreviver a ataques de *hackers*, como aconteceu há uns anos, na qual um ataque afetou câmaras de videovigilância e gravadores de vídeo, o que torna pouco seguro a utilização deste tipo de equipamentos, na medida em que existe uma diminuição da credibilidade deste tipo de serviços.

#### 2.1.2.1. Utilização da Inteligência Artificial

A Inteligência Artificial é útil em várias áreas, nomeadamente:

- Indústria Financeira, para ajudar na recolha de dados que podem ser usados mais tarde para conselhos financeiros.
- Educação, no sistema de graduação, essencialmente na sua automação, bem como na verificação da performance dos estudantes.
- Saúde, ajuda na melhoria dos diagnósticos dados aos pacientes, na ajuda ao consumidor, na marcação de consultas ou no processo de pagamento.
- Negócios, automatizando as tarefas realizadas por humanos, mas que podem ter ajuda de robots, melhorando a satisfação do cliente.
- Casas inteligentes, ao nível de segurança e vigilância, tornando a casa mais automática.

### 2.1.2.2. Áreas da Inteligência Artificial

A Inteligência Artificial é composta por várias áreas de investigação, podendo ser considerada enquanto área uma característica que se pretende atribuir à máquina/sistema. A identificação destas áreas varia de autor para autor. Sendo comum, elencar no contexto da Cibersegurança as áreas que em seguida se apresentam (Pannu, 2015):

**Aprendizagem**

A aprendizagem reflete-se na melhoria da *performance* durante a execução de tarefas após serem realizadas observações sobre o mundo e o seu respetivo comportamento, adquirindo capacidade para exibir o que foi aprendido (Russell, S. J. , & Norvig, P., 2010).

**Razão/Raciocínio**

Os humanos têm a capacidade de saber coisas, bem como o que os pode ajudar a efetuar determinadas tarefas, assim o raciocínio consiste em aplicar esse comportamento a um sistema. Isto é, atribuir a uma máquina a capacidade de tomar decisões, prever e julgar (Russell, S. J. , & Norvig, P., 2010).

**Perceção**

A perceção é o processo de adquirir, interpretar, selecionar e organizar informação sensorial. Assim são utilizados sensores para medirem alguns aspetos existentes no ambiente onde se encontram, permitindo obter informação (Russell, S. J. , & Norvig, P., 2010).

**Inteligência Linguística**

A inteligência linguística é uma habilidade que permite compreender, falar e escrever numa determinada linguagem, é importante para comunicações interpessoais (Russell, S. J. , & Norvig, P., 2010).

**Resolução de Problemas**

A resolução de problemas começa quando se encontra um problema, permitindo começar a planear os passos a seguir para a sua possível solução. Inclui também o processo de decisão, isto é, definir qual o melhor caminho a seguir para chegar de maneira mais eficaz a uma determinada solução (Russell, S. J. , & Norvig, P., 2010).

**Visão**

A visão tem o intuito de interpretar automaticamente imagens digitais genéricas de situações arbitrárias no ambiente em que se encontra. A interpretação deve permitir ao sistema detetar

eficientemente inferências nos desenhos, tomar decisões e executar tarefas de forma racional (Chella, Frixione, & Gaglioa, 1995).

**Jogos**

Os jogos são interações realizadas entre um ou mais jogadores, onde cada jogador executa diferentes movimentos, efetuando passos de acordo com o que foi definido nas regras (McBurney & Parsons, 2002). Contudo, os jogos também englobam cenários de pesquisa e de decisão, na qual é efetuada uma previsão relativamente aos cenários de jogo possíveis, calcula qual a probabilidade ótima e qual a melhor jogada a executar (Russell, S. J. , & Norvig, P., 2010).

Existem vários algoritmos que são aplicados na inteligência dos jogos, por exemplo, o algoritmo minimax e o alpha-beta pruning. O algoritmo do minimax obtém o valor minimax do estado atual, utiliza computação recursiva a partir dos valores de cada estado anterior, este algoritmo funciona de forma semelhante a uma estrutura em árvore. O algoritmo alpha-beta pruning aplica a norma do algoritmo minimax, retorna o mesmo valor que seria retornado em caso da execução do algoritmo minimax, no entanto, retira ramificações que não influenciam na decisão final (Russell, S. J. , & Norvig, P., 2010).

**Reconhecimento de padrões e Processamento de Imagem**

O reconhecimento de padrões é um processo complexo que requer uma variedade de técnicas, para garantir a transformação de dados em informação possível de ser reconhecida. Vários métodos foram propostos, podendo aplicar o reconhecimento não só às imagens como a várias outras situações do mundo (Daisheng, 1998).

O processamento de imagem engloba a geração, gravação e transmissão de uma imagem. Uma imagem visual é rica em informação, a receção e análise faz parte da rotina dos seres humanos. As técnicas de processamento de imagem são utilizadas atualmente por um vasto número de aplicações, que apesar de não relacionadas, partilham entre si a necessidade de métodos capazes de analisar e interpretar o que veem (Daisheng, 1998).

O reconhecimento de padrões está focado na descrição e classificação de entidades, enquanto o processamento de imagem está focado na qualidade e nas medidas de imagens de objetos, assim são ambos complemento um do outro (Daisheng, 1998).

### 2.1.2.3. Técnicas de Inteligência Artificial

Alan Turing foi um dos pioneiros na investigação da Inteligência Artificial e elaborou um teste que é atualmente conhecido como Teste de Turing. Este teste foi desenhado para providenciar uma definição satisfatória sobre a inteligência (Russell, S. J. , & Norvig, P., 2010).

Um computador passa no teste de inteligência, se a pessoa que efetua o interrogatório, não conseguir definir se quem responde é um individuo ou uma máquina (Russell, S. J. , & Norvig, P., 2010).

Em seguida, apresentam-se algumas técnicas que permitem que uma máquina possua inteligência:

**Machine Learning (ML)**

O Machine Learning consiste na aquisição de conhecimento automático por parte das máquinas, sem a necessidade explícita de ser programado, a aprendizagem centra-se em factos conhecidos. Além disso, explora o estudo e construção de algoritmos que podem aprender e efetuar previsões sobre os dados (Ongsulee, 2017).

Normalmente os métodos ou técnicas de Machine Learning são classificados em 4 categorias:

- Aprendizagem Supervisionada (Supervised Learning) – geralmente utilizada em ambientes conhecidos e familiares, com conhecimento prévio relativamente às características do mesmo (Bkassiny, Li, & Jayaweera, 2013).
- Aprendizagem Não Supervisionada (Unsupervised Learning) – as instâncias de dados não se encontram classificadas, um exemplo destes algoritmos é o *clustering* (conjunto de técnicas para prospeção de dados) (Haq, et al., 2015).
- Aprendizagem Semi-supervisionada (Semi-Supervised Learning) – é uma mistura da Aprendizagem Supervisionada e da Não Supervisionada, isto é, os dados podem estar ou não qualificados. Pode ser utilizado em métodos como classificação, regressão e previsão. É útil quando o custo relacionado com a qualificação é demasiado elevado para permitir um processo totalmente qualificado (Ongsulee, 2017).
- Aprendizagem por Reforço (Reinforcement Learning) – ocorre uma interação entre o computador e o ambiente de forma a alcançar determinado objetivo, por exemplo, perguntar a um utilizador uma determinada classificação para uma instância que pode fazer parte de um conjunto de instâncias não identificadas (Haq, et al., 2015).

O Machine Learning engloba também vários métodos, nomeadamente:

- Artificial Immune Systems (AIS) - técnica de Machine Learning, que foi inspirada nos princípios e processos de um sistema imune, os algoritmos são modelados considerando características de aprendizagem e memória para ajudar na resolução de problemas.
- Redes Neurais - área desenvolvida entre os anos 1950 e 1970, foi quando surgiu o primeiro entusiasmo por máquinas capazes de pensar. É um tipo de Machine Learning composto por unidades interligadas, como os neurónios, que processam informações, sendo inspiradas no funcionamento do cérebro humano. Estão organizadas em camadas, contendo vários nós interligados de forma a conterem funções de ativação. Além disso, as redes contêm padrões que permitem a comunicação entre as várias camadas existentes (Haq, et al., 2015).
- Deep Learning - utiliza redes neurais com várias camadas de unidades de processamento, utilizando o avanço tecnológico que tem existido ao longo dos anos. Inclui reconhecimento de imagens e de fala. Uma das suas mais-valias é o facto de permitir a substituição de recursos artesanais por algoritmos eficientes de aprendizagem não supervisionada ou semi-supervisionada, bem como a extração hierárquica de recursos (Ongsulee, 2017).

- Árvores de Decisão - representa a função que recebe um vetor de parâmetro de entrada e retorna uma decisão. Os valores de entrada e de saída podem ser discretos e contínuos. A decisão é alcançada após a execução de vários testes (Russell, S. J. , & Norvig, P., 2010).

**Natural Language Processing (NLP)**

O Natural Language Processing (NLP) é uma área de pesquisa e aplicabilidade que explora a forma como os sistemas podem ser utilizados para manipular e entender os textos ou discursos para as diferentes situações (Chowdhury, 2005).

Nos anos 80, os sistemas de NLP baseavam-se em regras manuscritas, no entanto, no final da década de 80 ocorreu uma evolução, baseada no Machine Learning para ajudar no processamento da linguagem (Kulkarni & Shivananda, 2019).

Os NLP provaram já ser úteis para certas atividades, nomeadamente na redução do tempo necessário em ensaios clínicos e identificação de possíveis reações adversas (Kreimeyer, et al., 2017).

**Machine Vision**

As máquinas podem capturar informação visual e analisá-la, utilizando câmaras, desde analógicas até às digitais. Após a captura, o resultado é passado para o computador (Flores-Fuentes, et al., 2014).

Existem dois aspetos importantes a ter em conta, a capacidade da máquina para compreender impulsos fracos e o intervalo no qual a máquina pode distinguir os objetos. A Machine Vision é utilizada para a identificação de assinaturas, reconhecimento de padrões, análises médicas, entre outras (Flores-Fuentes, et al., 2014).

Além disso, permite duplicar as habilidades da visão humana, através do entendimento eletrónico sobre uma imagem com elevada dimensão, melhorando o requisito de armazenamento, bem como a melhoria do tempo de processamento comparativamente com outros algoritmos (Flores-Fuentes, et al., 2014).

**Automatização e Robótica**

O objetivo da automação é utilizar máquinas para realizar tarefas monótonas e repetitivas, o que vai permitir melhorar a produtividade, diminuir custos e obter resultados mais eficientes. A Automatização ajuda a prevenir fraudes (Bahrin, Othman, Azli, & Talib, 2016).

A Robótica é programada para realizar um grande volume de tarefas repetitivas, que podem ser adaptadas em diferentes circunstâncias (Bahrin, Othman, Azli, & Talib, 2016).

**Fuzzy Logic**

Os sistemas Fuzzy Logic, são métodos de raciocínio, que propõem cálculos matemáticos para traduzir o conhecimento humano relativamente aos processos reais. Sendo uma forma de manipular o conhecimento prático com algum nível de incerteza (Vieira, Dias, & Mota, 2004).

O mecanismo de inferência dos sistemas Fuzzy Logic consiste em 3 etapas, a primeira consiste no mapeamento, utilizando uma função com os valores numéricos da entrada. Na segunda etapa, o sistema Fuzzy Logic processa as regras de acordo com a robustez da entrada. Na terceira etapa, os valores resultantes são novamente transformados em valores numéricos (Vieira, Dias, & Mota, 2004).

Este processo, torna possível a utilização das categorias em representação de palavras e ideias abstratas do ser humano na descrição das decisões tomadas (Vieira, Dias, & Mota, 2004).

## 2.2. Revisão Sistemática de Literatura sobre Inteligência Artificial na Cibersegurança

Tendo em conta o objetivo deste presente estudo, que se baseia na análise dos conceitos de Cibersegurança e Inteligência Artificial, com maior foco na forma como ambos se relacionam, para garantir que um Sistema de Informação se torna mais seguro e menos vulnerável.

Os artigos foram procurados tanto em português quanto em inglês. Foram selecionados artigos com data superior a janeiro de 2015. Os artigos procurados incluíam as palavras-chave definidas no título ou no resumo/sinopse. As palavras-chave definidas foram consideradas devido à relevância que tinham para a pesquisa. Para se identificarem os artigos a analisar foi utilizado o método PRISMA (Moher, Liberati, Tetzlaff, & Altman, 2015).

Assim, foram definidas palavras-chave para a Inteligência Artificial: “Inteligência Artificial”, “Machine Learning”, “Deep Learning”, “Artificial Immune Systems”, “Automatização”, “Robótica”, “Fuzzy Logic”, “Natural Language Processing” e “Machine Vision”. E palavras-chave para a Cibersegurança: “Cibersegurança”, “Segurança Informática”, “Phishing”, “Malware”, “Ransomware”, “Ataques de Denial of Service” e “Criptografia”. O processo de pesquisa consistiu na junção das áreas da Inteligência Artificial com a Cibersegurança, na qual foi efetuada uma combinação de palavras-chave, de forma a se relacionarem entre si. Um exemplo de pesquisa realizada é o seguinte: [“Inteligência Artificial” e “Cibersegurança”].

A pesquisa da literatura foi efetuada em plataformas com artigos académicos, nomeadamente, “Scimago”, “Google Académico”, “Research Gate” e “Science Direct”. A pesquisa foi efetuada em abril de 2020.

Foram considerados artigos que abordam a utilização de técnicas de Inteligência Artificial para a resolução de problemas de Cibersegurança. Os artigos apresentam por parte dos autores uma explicação detalhada da maneira como foi resolvido o problema de segurança. Enquanto critério de exclusão, foram retirados os artigos que apenas se focavam ou na Inteligência Artificial ou na Cibersegurança e artigos que eram apenas revisão de outros.

Na primeira fase, foi iniciado o processo de seleção dos artigos devolvidos nas pesquisas efetuadas, utilizando as palavras-chave definidas anteriormente. A pesquisa efetuada relacionou as palavras-chave definidas para a Inteligência Artificial com as definidas para a Cibersegurança, na qual ambas foram cruzadas de forma combinada. Além disso, foi aplicado um filtro para não devolver artigos anteriores a janeiro de 2015.

Após ser efetuado o cálculo para determinar o número total de artigos devolvidos nas diferentes plataformas onde a pesquisa foi efetuada. Foram removidos os duplicados, os que não apresentaram uma sinopse explicativa, bem como os ficheiros que não estavam em formato PDF ou resultados que não se encontrassem categorizados como artigos.

De seguida, foram excluídos os artigos que não correspondiam aos parâmetros definidos anteriormente. Para tal, foi efetuado um cálculo para comparar a diferença entre o total de artigos obtidos na primeira fase, com os artigos obtidos após a pesquisa com a aplicação dos parâmetros definidos.

Para a realização da primeira etapa, isto é, a fase de identificação de registos devolvidos na execução da pesquisa utilizando as palavras-chave, foi obtido um total de 935194 artigos/registos.

Na imagem (*Figura 1*) é apresentado o fluxo PRISMA, na qual são apresentadas as diferentes fases do processo de seleção dos artigos, bem como, o número de artigos que foram satisfazendo os critérios de pesquisa e o número de artigos que foram sendo desconsiderados até a obtenção dos artigos a se considerar.

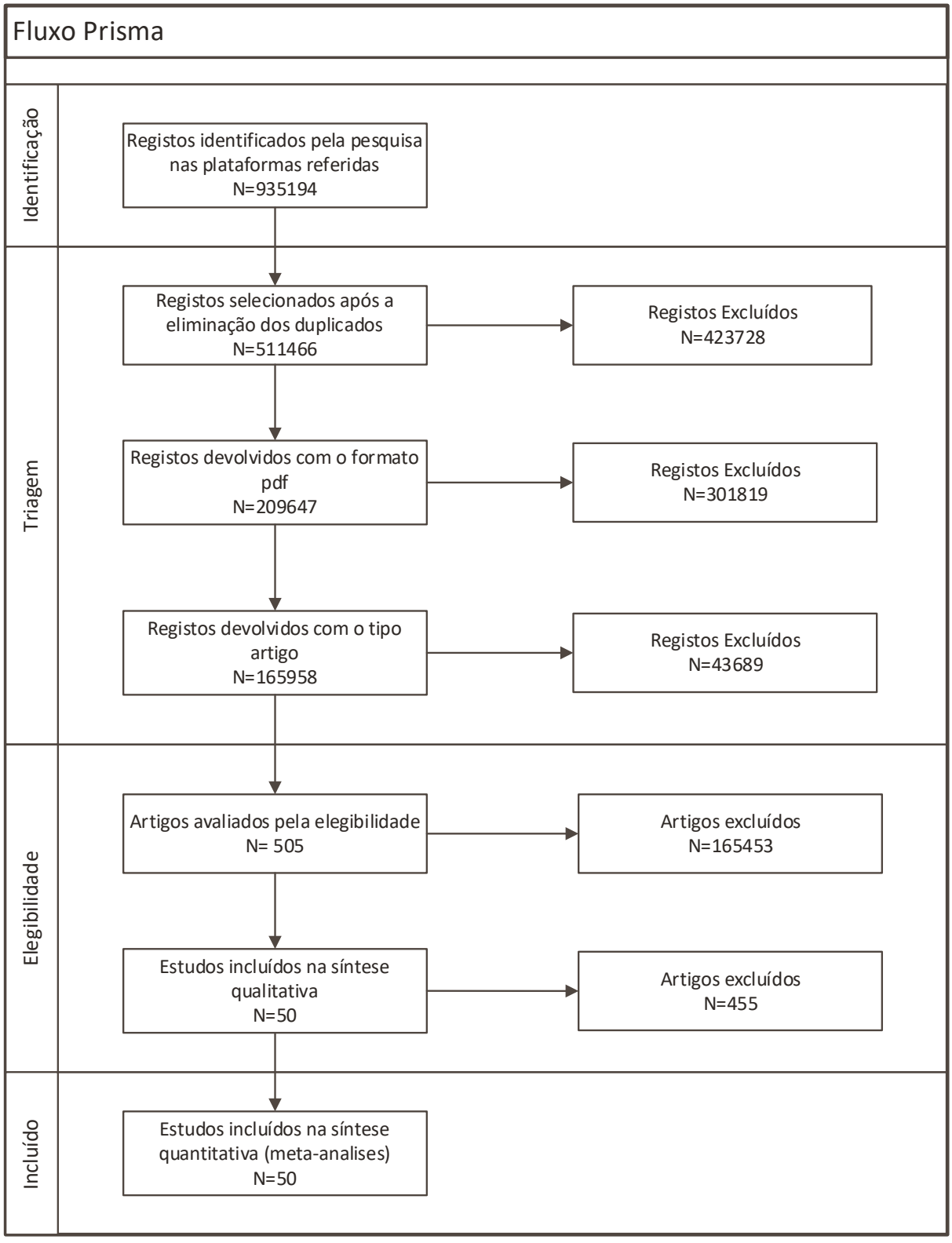


Figura 1- Fluxo Prisma

Assim, na fase de Triagem, foram retirados artigos que se encontravam duplicados, ficando um total de 511466 artigos. De seguida, foram removidos artigos sem resumo e que não se encontravam em formato PDF, onde se passou a ter um total de 209647 registos à pesquisa efetuada.

Para finalizar a fase de Triagem, foi aplicado um filtro para devolver apenas os registos definidos como artigo publicados em revistas académicas, ficando um total de 165958 artigos, na qual foram excluídos 43689 artigos.

Na fase de Elegibilidade, o número de artigos a definir como elegíveis era elevado, eram 165958 artigos. Assim, foram verificados os 165958 artigos selecionados, no qual foi conferido se correspondiam ao que se pretendia analisar no presente estudo, isto é, se utilizavam a Inteligência Artificial para evitar possíveis ataques de Cibersegurança, se utilizavam Inteligência Artificial para provocar ataques cibernéticos ou se apenas mencionavam separadamente ambas as áreas. Para tal, foram incluídas validações na pesquisa para funcionarem como filtro, de forma a garantir que os artigos abordavam apenas a utilização da Inteligência Artificial para combater ataques cibernéticos, tendo assim sido obtido um total de 505 artigos definidos como elegíveis, ou seja, artigos aptos para serem analisados.

Os 505 artigos, apurados como elegíveis, foram todos analisados, para garantir que utilizavam a Inteligência Artificial como ferramenta para ajudar a evitar ataques cibernéticos e os seus possíveis danos. Dessa análise foram selecionados 50 artigos, pois foi determinado que apresentavam o conteúdo necessário para corroborar o que se pretende obter com este presente estudo.

Os artigos incluídos na análise quantitativa / meta-análise foram os seguintes:

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| 1 | Artificial Intelligence in Cybersecurity | Janeiro 2017 | Nadine Wirkuttis, Hadas Klein | A Cibersegurança é sem dúvida a disciplina que mais poderia beneficiar com a introdução da IA. Quando os sistemas de segurança convencionais são lentos e insuficientes, as técnicas de Inteligência Artificial podem melhorar o seu desempenho ao nível de segurança e fornecer uma melhor proteção contra o número crescente de ameaças cibernéticas sofisticadas.<br>Além das grandes oportunidades atribuídas à IA na Cibersegurança, a sua utilização tem riscos e preocupações justificados. Para aumentar ainda mais a maturidade da Cibersegurança, é necessária uma visão holística do ambiente cibernético das organizações, na qual a IA é combinada com o discernimento humano, uma vez que nem as pessoas, nem a IA sozinhas comprovaram por si só o sucesso. Portanto, o uso responsável das técnicas de IA será essencial para mitigar ainda mais os riscos e as preocupações relacionados. |
| 2 | Artificial Intelligence based Network Intrusion Detection with hyper-parameter optimization tuning on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing | Abril 2019 | V. Kanimozhi, T. Prem Jacob | A Inteligência Artificial é uma das tecnologias emergentes mais recentes, que permite à máquina imitar o comportamento humano. A componente mais importante utilizada para detetar ataques cibernéticos ou atividades maliciosas é o IDS. Além disso, desempenha um papel vital na deteção de intrusões e é considerada a melhor maneira de adaptar e construir IDS.<br>Nos dias modernos, os algoritmos de NN têm emergido como uma nova técnica de IA, que pode ser aplicada a problemas que ocorrem em tempo real. O sistema proposto pretende detetar uma classificação de ataque de *botnet*, estes representam uma ameaça séria para o setor financeiro e para os serviços bancários. Este sistema é criado através da aplicação de IA num conjunto real de dados de defesa cibernética. |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | O sistema de ANN permite um excelente desempenho de pontuação de precisão de 99,97%, uma área média sob a curva ROC de 0,999 e a taxa média de falso positivo é um mero valor de 0,03. O sistema proposto para deteção de intrusões baseado em Artificial para classificação de ataques de *botnet* é poderoso, mais eficaz e preciso. Pode ser aplicado à análise de tráfego da rede convencional, do sistema ciber-físico e da rede em tempo real. |
| 3 | Mobile Malware attacks: Review, taxonomy & future directions | Março 2019 | Attia Qamar, Ahmad Karim, Victor Chang | Um aumento generalizado na taxa de adoção de smartphones com sistema operacional Android é observado nos últimos anos. O ambiente popular e atraente do Android não chamou apenas à atenção dos utilizadores, mas aumentou também as preocupações com a segurança. Como resultado, a deteção de Malware do Android é um dos tópicos urgentes no domínio da segurança móvel. Este documento fornece uma revisão abrangente dos ataques de Malware de última geração, vulnerabilidades, técnicas de deteção e soluções de segurança durante o período de 2013–2019 que visavam principalmente a plataforma Android. São apresentadas várias taxonomias bem organizadas e detalhadas que revelam abordagens de deteção de Malware móvel com base nas técnicas de análise, plataforma de trabalho, aquisição de dados, impacto operacional, resultados obtidos e as componentes de Inteligência Artificial envolvidas. Outra taxonomia composta pelo vetor de ataque de Malware móvel é apresentada para examinar grupos ameaças e lacunas para localizar o impacto malicioso generalizado nas comunidades.<br>Além disso, serão discutidos e classificados os esforços de análise forense na perspetiva de deteção de Malware móvel. Do ponto de vista do intruso, serão comparadas várias técnicas de evasão usadas com destaque pelos autores de Malware para impedir os esforços de deteção. Finalmente, as futuras direções de trabalho são apresentadas como diretrizes para a indústria, de forma a ajudá-los a reduzir ou mesmo evitar o impacto prejudicial desses esforços incómodos. |
| 4 | Symbiotic Artificial Intelligence and its challenges in cybersecurity and Malware research | Dezembro 2018 | Anthony N. Merrill | A Inteligência Artificial é um projeto de aplicações de "hardware" ou de software que interagem e agem racionalmente dentro do seu ambiente para completar uma tarefa especificada. A taxa de adoção e a importância da IA aumentou consideravelmente em muitas indústrias. Tanto, a Cibersegurança, a computação forense e pesquisa de Malware exigem modelos de IA altamente disponíveis, precisos e confiáveis. Na indústria da Cibersegurança, esta começou a empurrar os casos de uso para a IA em muitas aplicações, como a análise do comportamento de entidade do utilizador, deteção de intrusão e sistemas de prevenção, e análise estática e dinâmica. O objetivo deste estudo é identificar quais os diferentes algoritmos de IA utilizados na Cibersegurança, como são implementados e as suas vulnerabilidades, desvantagens e outros riscos relacionados com a utilização de IA.<br>Estas aplicações estão repletas de problemas que podem interromper o sistema e as dependências, levando a uma |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | perda de confiança. As aplicações que utilizam IA para criar alertas geralmente enviam grandes volumes de falsos positivos. Ataques de perda de dados e envenenamento de dados são os principais criminosos da IA mal treinada. Os *hackers* começam a identificar os algoritmos de IA utilizados para que possam encontrar maneiras de manipular o sistema e permanecer ocultos. As organizações para pesquisa que utilizam IA, devem contratar engenheiros de pesquisa para ajudar a adotar e desenvolver a tecnologia corretamente para o problema que tentam resolver. Se possível, com dados confidenciais, treinar um modelo de IA de forma isolada reduz o risco de manipulação de um invasor externo. Ao trabalhar com dados de Cibersegurança, as organizações podem obter ajuda de organizações terceira como a MITRE para a análise adequada dos dados. |
| 5 | Applications of Artificial Intelligence (AI) to Network Security | Março 2018 | Alberto Perez Veiga | Os ataques às redes estão se tornando mais complexos e sofisticados a cada dia. Além dos chamados *script-kiddies* e *hackers* inexperientes, há uma infinidade de invasores profissionais que pretendem obter lucros sérios infiltrando-se em redes corporativas. Tanto os governos hostis, grandes corporações ou máfias têm aumentado constantemente os seus recursos e as habilidades no crime cibernético para espiar, roubar ou causar danos de forma eficaz. Com a capacidade e os recursos dos *hackers* a crescer, as abordagens tradicionais de segurança de rede parecem começar a atingir os seus limites e é reconhecida a necessidade de uma abordagem mais inteligente para a deteção de ameaças.<br>Este artigo fornece uma introdução sobre a necessidade da evolução das técnicas de Cibersegurança e como a Inteligência Artificial (IA) pode ser aplicada para ajudar a resolver alguns dos problemas. Fornece também uma visão geral de alto nível de algumas técnicas de segurança de rede de IA de última geração, para concluir a análise do que é o futuro previsível da aplicação de IA para segurança de rede. |
| 6 | Cyber security meets artificial intelligence: a survey | Setembro 2018 | Jian-Hua LI | Existe uma ampla gama de interseções interdisciplinares entre a Cibersegurança e a Inteligência Artificial (IA). Por um lado, as tecnologias de IA, como o Deep Learning, podem ser introduzidas na Cibersegurança na construção de modelos inteligentes para implementar a classificação de Malware e deteção de intrusão e deteção de ameaças de inteligência. Por outro lado, os modelos de IA irão enfrentar várias ameaças cibernéticas, que irão atrapalhar a amostra, a aprendizagem e as decisões. Assim, os modelos de IA precisam de tecnologias específicas de defesa e proteção da Cibersegurança para combater a Machine Learning adversária, preservar a privacidade na ML, proteger a aprendizagem federada, etc. Com base nos dois aspetos acima, foi realizada a interseção da IA e da Cibersegurança. Primeiro, resumimos os esforços de pesquisa existentes no combate de ataques cibernéticos utilizando a IA, incluindo a adoção de métodos tradicionais de ML e soluções de Deep Learning. Em seguida, analisamos os contra-ataques que a própria IA pode sofrer, dissecamos as suas características e classificamos os métodos de |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | defesa correspondentes. Finalmente, a partir dos aspetos da construção de uma rede neural criptografada e da realização de DL federada e segura, explicamos a pesquisa existente sobre como construir um sistema de IA seguro. |
| 7 | The Benefits of Artificial Intelligence in Cybersecurity | Janeiro 2019 | Ricardo Calderon | As ameaças cibernéticas aumentaram amplamente na última década. Os criminosos cibernéticos tornaram-se mais sofisticados. Os controlos de segurança atuais não são suficientes para defender as redes do número de criminosos cibernéticos altamente qualificados. Estes aprenderam como contornar as ferramentas mais sofisticadas, como IDPS e os *botnets* são quase invisíveis para as ferramentas atuais. Felizmente, a aplicação de Inteligência Artificial pode aumentar a taxa de deteção de sistemas IDPS, e as técnicas de Machine Learning são capazes de extrair dados para detetar as fontes de *botnets*. No entanto, a implementação da IA pode trazer outros riscos, e os especialistas em Cibersegurança precisam de encontrar um equilíbrio entre risco e benefícios. |
| 8 | Trusting artificial intelligence in cybersecurity is a double-edged sword | Dezembro 2019 | Mariarosaria Taddeo, Tom McCutcheon, Luciano Floridi | As aplicações de Inteligência Artificial para tarefas de Cibersegurança atraem uma maior atenção dos setores público e privado. As últimas estratégias nacionais de Cibersegurança e defesa de vários governos mencionam explicitamente os recursos de IA. Ao mesmo tempo, iniciativas para definir novos padrões e procedimentos de certificação para obter a confiança dos utilizadores na IA surgem em escala global. No entanto, a confiança na IA para entregar tarefas de Cibersegurança é uma espada de dois gumes: pode melhorar substancialmente as práticas de Cibersegurança, mas também pode facilitar novas formas de ataques às aplicações de IA, que podem representar graves ameaças à segurança. Argumentamos que a confiança na IA para Cibersegurança é injustificada e que, para reduzir os riscos de segurança, é necessária alguma forma de controle para garantir uma implantação de IA "confiável" para a Cibersegurança. |
| 9 | Applying Artificial Intelligence Techniques to Prevent Cyber Assaults | Novembro 2017 | Amaan Anwar, Syed Imtiyaz Hassan | A Cibersegurança é ostensivamente a disciplina que mais poderia lucrar com a introdução da Inteligência Artificial. É difícil fazer um software para a defesa contra poderosos ataques aos sistemas. Pode ser curado com a aplicação de técnicas de Inteligência Artificial. Onde os sistemas de segurança convencionais podem ser lentos e deficientes, as técnicas de Inteligência Artificial podem aprimorar a sua execução geral de segurança e oferecer melhor proteção contra um número crescente de ameaças cibernéticas complexas. Além das grandes oportunidades atribuídas à IA na Cibersegurança, a sua utilização legitimou riscos e preocupações. Para promover o incremento do desenvolvimento da Cibersegurança, é necessária uma perspetiva holística do ambiente cibernético de associações, no qual a IA se deve consolidar com o conhecimento humano, uma vez que nem os indivíduos, nem a IA comprovaram o sucesso geral nesta esfera. Dessa forma, a utilização socialmente consciente de técnicas de IA será necessária para mitigar ainda mais os riscos e preocupações relacionados. |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| 10 | Benefit from AI in Cybersecurity | Outubro 2019 | Ljubomir Lazić | Este artigo pretende explicar o papel da IA na Cibersegurança e propõe recomendações de como as organizações beneficiam com a IA na Cibersegurança. O Machine Learning, um componente da IA, aplica os dados existentes para melhorar constantemente as suas funções e estratégias. Aprende e entende o comportamento normal do utilizador e pode identificar até mesmo a menor variação desse padrão. Mas, além de reunir informações para detetar e identificar ameaças, a IA pode usar estes dados para melhorar as suas próprias funções e estratégias. Neste artigo, pesquisamos as técnicas de ofuscação e desofuscação existentes e que atualmente são aplicadas às aplicações Android, em seguida, sugerimos a plataforma de desofuscação baseada em Low-Level Virtual Machine para realizar o processo de desofuscação de forma eficiente. Além disso, é investigada a solução AndrODet, um sistema de aprendizagem "online" para detetar três tipos comuns de técnicas de ofuscação em aplicações Android, conhecidas como renome do identificador, criptografia e ofuscação de fluxo de controlo. |
| 11 | Cyber defense using Artificial Intelligence | Novembro 2016 | A. Anitha, Girish Paul, Savera Kumari | Os ataques cibernéticos são uma preocupação muito comum atualmente. Se um indivíduo não tiver um sistema de segurança adequado, as informações podem ser facilmente roubadas. Uma das causas mais comuns de ataques cibernéticos deve-se ao invasor. Assim, é aprimorado o processo de segurança, evitando a intrusão nos vários níveis das camadas de rede do sistema, para tal, é utilizada a Inteligência Artificial. O objetivo principal é criar um programa que se possa defender de vários ataques de rede e deteção de intrusão. O principal desta experiência é desenvolver uma estrutura, na qual uma grande variedade de processos possa ser mapeada. Um modelo de software é desenvolvido para representar, capturar e aprender o comportamento da consciência cibernética de um processo de computador em relação às várias ameaças. |
| 12 | Phishing URL Detection: A Machine Learning and Web Mining-based Approach | 2015 | Bhagyashree E. Sananse, Tanuja K. Sarode | O aumento da sofisticação dos criminosos cibernéticos levou à proliferação de ataques de Phishing. A expansão contínua da World Wide Web levou à rápida disseminação de Phishing, Malware e spam. É proposta uma abordagem baseada nos recursos para classificar URL, na categoria de Phishing ou não Phishing. O uso de uma variedade de recursos de URL é feito ao estudar a anatomia de URL. Para a classificação de URL, foram usados dois algoritmos diferentes. O algoritmo de ML, Random Forest é usado para construir um classificador eficiente que decidirá se um determinado URL é Phishing ou não. Além disso, um novo esquema foi proposto para detetar URL de Phishing, explorando o conteúdo disponível publicamente nas URL. |
| 13 | Using Deep Learning Techniques for Network Intrusion Detection | Dezembro 2019 | Sara Al-Emadi, Aisha Al-Mohannadi, Felwa Al-Senaid | Nos últimos anos, houve um aumento significativo nos ataques de intrusão de rede, o que desperta uma grande preocupação nos aspetos de privacidade e segurança. Devido ao avanço da tecnologia, os ataques de Cibersegurança tornaram-se muito complexos, de tal modo que os sistemas de deteção atuais não são suficientes para resolver esse problema. Portanto, a |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | implementação de um NIDS inteligente e eficaz seria crucial para resolver este problema. Neste artigo, são utilizadas técnicas de Deep Learning, nomeadamente CNN e RNN para criar um sistema de deteção inteligente capaz de detetar diferentes intrusões na rede. Adicionalmente, avaliamos o desempenho da solução proposta utilizando diferentes matrizes de avaliação e apresentamos uma comparação entre os resultados da solução proposta para encontrar o melhor modelo para o sistema de deteção de intrusão na rede. |
| 14 | Website Phishing Detection using Heuristic Based Approach | Maio 2016 | Jaydeep Solanki, Rupesh G. Vaishnav | A Internet tornou-se uma parte útil da nossa vida, pois fazemos quase todas as nossas atividades sociais e financeiras "online". Hoje, todas as pessoas dependem fortemente da Internet e de atividades "online", como o *homebanking*, entre outros. O Phishing é uma forma de ameaça na web, o atacante cria uma réplica do site original e tenta ilegalmente obter as informações pessoais da vítima, como nome de utilizador, palavra-passe, detalhes do cartão de crédito e usá-las em benefício próprio. Um utilizador não regular não consegue identificar se o site é Phishing ou legítimo. Não existe uma solução única para impedir esta atividade fraudulenta. É proposto um modelo que identifica o site de Phishing. O sistema primeiro extrai os recursos que diferenciam se o site é Phishing ou legítimo. Em seguida, aplicamos estes recursos às técnicas de Machine Learning para identificar sites com Phishing ou legítimos. Desta forma, ajudará a sociedade. |
| 15 | Phishing Websites Detection Using Machine Learning Based Classification Techniques | Novembro 2016 | Mazharul Islam, Nihad Karim Chowdhury | Os sites de Phishing que esperam obter dados confidenciais das vítimas, fazem-no desviando-as para navegar numa página falsa que se assemelha a uma página de um site honesto, são atos criminosos através da Internet e uma das preocupações em várias áreas. A deteção de sites de Phishing é realmente um problema imprevisível e elementar, incluindo vários componentes e critérios que não são estáveis. Por conta da última e tendo em conta as ambiguidades na organização de sites devido aos procedimentos inteligentes que os programadores estão a utilizar, algumas estratégias proativas afiadas podem ser úteis e podem ser utilizadas ferramentas poderosas, por exemplo, fuzzy, sistema neurais e métodos de Data Mining, são mecanismos de sucesso para distinguir sites de Phishing. Foram aplicados diferentes algoritmos de classificação baseados em Machine Learning, incluindo Naïve Bayes, Support Vector Machine, Neural Network, Random Forest, classificador preguiçoso IBK e árvore de decisão (J48). |
| 16 | Feature selection and Machine Learning classification for Malware detection | Outubro 2015 | Ban Mohammed Khammas, Alireza Monemi, Joseph Stephen Bassi, Ismahani Ismail, Sulaiman Mohd Nor, | O Malware é um problema de segurança de computador que se pode transformar para evitar os métodos tradicionais de deteção com base na correspondência de assinatura conhecida. Como as novas variantes de Malware contêm padrões semelhantes aos do Malware observado, as técnicas de Machine Learning podem ser usadas para identificar novos Malwares. Apresenta um estudo comparativo de vários métodos de seleção de recursos com quatro classificadores diferentes de Machine Learning no contexto de deteção estática de Malware com base na análise de n-gramas. O resultado |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | Muhammad Nadzir Marsono | mostra que o uso da seleção de recursos do PCA e da classificação de SVM fornece a melhor precisão de classificação utilizando um número mínimo de recursos. |
| 17 | A Survey of Adversarial Machine Learning in Cyber Warfare | Julho 2018 | Vasisht Duddu | A natureza mutável da guerra viu uma mudança de paradigma da guerra convencional para a assimétrica e sem contato, como a guerra de informação e a cibernética. A dependência excessiva de tecnologias de informação e comunicação, infraestruturas de *cloud*, análise de Big Data, Data Mining e automação na tomada de decisão representam graves ameaças para os negócios e para a economia em ambientes adversários. O Machine Learning adversário é uma área de pesquisa de rápido crescimento que estuda o projeto de algoritmos de ML que são robustos em ambientes adversários. Este artigo apresenta um levantamento abrangente desta área emergente e as várias técnicas de modelagem do adversário. Exploramos os modelos de ameaça para sistemas de ML e descrevemos as várias técnicas para os atacar e para os defender. Apresentamos questões de privacidade nestes modelos e descrevemos um caso de teste de guerra cibernética para testar a eficácia das várias estratégias de defesa de ataque e concluímos com alguns problemas em aberto nesta área de pesquisa. |
| 18 | Malware Detection using Sub-Signatures and Machine Learning Technology | Abril 2018 | Ban Khammas | O Malware é uma grande preocupação de segurança do computador, pois muitos sistemas estão ligados à Internet. O número de Malware aumentou ao longo dos anos e novos Malwares surgiram, na qual as novas variantes são capazes de escapar da deteção convencional do sistema por ofuscações. Um dos métodos promissores usados para detetar Malware são as técnicas de Machine Learning (ML).<br>Este trabalho apresenta um sistema de deteção de Malware estático que utiliza técnicas de n-gram e ML, utiliza sub-assinaturas de Malware conhecidas para reduzir grandes espaços de procura de recursos, que são criados devido aos métodos de extração de recursos de n-gram. O espaço dos recursos afeta diretamente o desempenho e a precisão de deteção dos classificadores de ML de Malware. A análise dos vários métodos de seleção de recursos para minimizar o número de recursos e a análise de vários classificadores de ML são também apresentados para melhorar a precisão da deteção de Malware. Os resultados mostram que a análise de n-gram com recursos de sub-assinatura do Snort usando ML fornece boa precisão de deteção de Malware de mais de 99,78% e zero FPR quando recursos de 4 gramas são usados para a maioria dos classificadores de ML verificados. |
| 19 | Intrusion Detection Using Machine Learning: A Comparison Study | Fevereiro 2018 | Saroj Kr. Biswas | Com o avanço da Internet ao longo dos anos, o número de ataques pela Internet também aumentou. Um poderoso sistema de deteção de intrusão (IDS) é necessário para garantir a segurança de uma rede. O objetivo do IDS é monitorizar os processos que prevalecem numa rede e analisá-los quanto a indícios de possíveis desvios. Alguns estudos foram realizados neste campo. É proposto um IDS utilizando o ML para a rede com uma boa união de técnica de seleção de recursos e classificador, através do estudo das combinações da |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | maioria das técnicas e classificadores populares de seleção de recursos. Um conjunto de recursos significativos é selecionado do conjunto original de recursos utilizando técnicas de seleção. Em seguida, o conjunto dos recursos significativos é utilizado para treinar diferentes classificadores para criar o IDS. A validação cruzada é realizada no conjunto de dados NSL-KDD para encontrar os resultados. É finalmente observado que o classificador K-NN apresenta um melhor desempenho do que outros e, entre os métodos de seleção de características, o método de seleção de características baseado na taxa de benefício de informação é melhor. |
| 20 | Application of distributed computing and machine learning technologies to cybersecurity | Novembro 2018 | Hamza Attak, Marc Combalia, Georgios Gardikis, Bernat Gastón, Ludovic Jacquin, Dimitris Katsianis, Antonis Litke, Nikolaos Papadakis, Dimitris Papado-poulos, Antonio Pastor, Marc Roig, Olga Segou | SHIELD é um sistema de Cibersegurança distribuído que aproveita a NFV para implementar dinamicamente a NSF virtual. As funções de segurança enviam dados de monitorização do movimento de rede para um armazenamento de Big Data. O DARE executa módulos de análise de segurança, além dos módulos de dados de monitorização para detetar ameaças. A análise de segurança aproveita os algoritmos de ML para detetar anomalias e classificar as ameaças. Este artigo apresenta os diferentes algoritmos de ML e detalha os resultados obtidos e o rumo do projeto, quanto à sua implementação, incluindo as capacidades de negócio para a solução de Cibersegurança. |
| 21 | Machine Learning for Cybersecurity: Network-based Botnet Detection Using Time-Limited Flows | Setembro 2017 | Stephanie Ding | Os *botnets* são conjuntos de anfitriões ligados e infetados por Malware que podem ser controlados por um atacante remoto e são uma das ameaças proeminentes na Cibersegurança, pois podem ser utilizados para uma ampla variedade de finalidades, como ataques de DoS, spam ou bitcoin. É proposto um método de deteção de duas fases, utilizando técnicas de ML supervisionadas e não supervisionadas para distinguir entre o tráfego da rede *botnet* e não *botnet*. Na primeira fase, são examinados os registos de fluxo de rede criados em intervalos de tempo limitados, que fornecem um resumo conciso, mas parcial, do perfil de tráfego de rede completo e são classificados os fluxos como maliciosos ou benignos, com base num conjunto de recursos extraídos. Na segunda fase, é executado o *clustering* em anfitriões internos envolvidos em comunicações maliciosas previamente identificadas para determinar quais os anfitriões com maior probabilidade de estar infetados por *botnet*. Utilizando os conjuntos de dados existentes, irá ser demonstrada a viabilidade do método e implementado um sistema de deteção em tempo real que agrega os resultados de vários classificadores para identificar os anfitriões infetados. |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| 22 | Machine Learning and Deep Learning methods for Cybersecurity | Abril 2019 | Prof. Nanda M. B., Parinitha B. S | A deteção e prevenção de intrusões na rede é uma grande preocupação. Os métodos de Machine Learning e Deep Learning detetam intrusões na rede, prevendo o risco com a ajuda do tratamento dos dados. Vários métodos de ML e DL têm sido propostos ao longo dos anos, os quais se mostram mais precisos quando comparados a outros sistemas de deteção de intrusão na rede. Este artigo de pesquisa fornece uma breve introdução sobre vários algoritmos de Machine Learning e Deep Learning. |
| 23 | Ransomware, Threat and Detection Techniques: A Review | Fevereiro 2019 | SH Kok, Azween Abdullah, NZ Jhanjhi, Mahadevan Supramaniam | A popularidade do Ransomware criou um ecossistema único de cibercriminosos. Portanto, os objetivos deste artigo são fornecer um melhor entendimento da ameaça do Ransomware e discutir as técnicas de deteção utilizadas recentemente. Um ataque de Ransomware bem-sucedido tem uma implicação financeira direta, que é alimentada por vários facilitadores maduros, tal como tecnologia de criptografia, moeda cibernética e acessibilidade. A criptografia é eficaz e quase inquebrável. A moeda cibernética é anónima e pode evitar o rastreio. O código de Ransomware é facilmente obtido, o que permite uma entrada fácil. A combinação destas opções proporciona um caminho atraente para os cibercriminosos, produzindo assim cibercriminosos especializados. Em relação às técnicas de deteção, constatou-se que o Machine Learning (ML) através de algoritmos de regressão, foi a técnica mais utilizada pelos investigadores de ransomware. No entanto, nenhum dos investigadores produziu qualquer modelo para proteção contra possíveis ataques de ransomware. Esta pesquisa destaca a necessidade de uma solução utilizando o algoritmo ML para o mecanismo de deteção. |
| 24 | Deep Learning for Ransomware Detection | Março 2017 | Aragorn Tsengy, YunChun Chen, YiHsiang Kao, TsungNan Lin | O Ransomware é Malware que se instala de forma secreta no computador ou no smartphone da vítima, na qual executa um ataque de criptovirologia e exige o pagamento de um resgate para restaurá-lo. Os ransomwares foram a ameaça mais séria no ano de 2016 e a situação tende a piorar. Devido à alta recompensa para Ransomwares, mais e mais famílias de Ransomware aparecem, o que dificulta a sua deteção. Existem diferentes assinaturas ou comportamentos entre diferentes famílias, ou versões de Ransomwares. Seria maravilhoso se houvesse uma maneira de detetar potenciais ameaças de Ransomware. Neste artigo, será utilizado o método de Deep Learning para detetar Ransomwares. Inicialmente, apresentaremos como legendamos os dados com diferentes comportamentos e quais os recursos escolhidos. Posteriormente, será apresentado o modelo para detetar vários Ransomwares e evitar que os dados da vítima sejam criptografados. A avaliação experimental demonstra que o modelo de Deep Learning pode detetar os Ransomwares mais recentes em redes de alta velocidade em tempo útil. |
| 25 | From Artificial Intelligence to Security: Back and Forth | Maio 2018 | Islam Faisal | A Inteligência Artificial e, especialmente, o Machine Learning difundiram-se em todos os campos das ciências da computação. A engenharia da Segurança é um dos campos na qual as aplicações de IA e ML invadiram os paradigmas tradicionais baseados em regras. As |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | aplicações de ML em segurança incluem deteção de intrusão, antivírus e deteção de ataques de DoS. Recentemente, tem havido interesse em usar o Machine Learning para calcular o tráfego criptografado para facultar a segurança na rede, como inspeção profunda dos pacotes. A relação entre a segurança e o ML é bilateral. Conceitos de segurança e privacidade são usados para proteger os dados dos utilizadores quando alimentados para um modelo de ML, bem como para proteger os segredos comerciais dos modelos de ML. Pela atual onda de interesse pelo ML e pela descentralização, espera-se que essa relação bilateral continue crescendo cada vez mais. |
| 26 | Survey of Deep Learning Based Intrusion Detection Systems for Cyber Security | Maio 2019 | Riyaz Jamadar, Shreyas Ingale, Anuj Panhalkar, Anup Kakade, Mohit Shinde | Devido ao rápido crescimento da Internet e à crescente complexidade dos ataques cibernéticos, a necessidade da Cibersegurança aumentou. Ocorre uma extensa pesquisa para o desenvolvimento de soluções eficientes e económicas para a deteção de intrusão. As pesquisas mais recentes aproveitam o Machine Learning e os algoritmos de Deep Learning para fornecer as soluções mais eficientes. Neste artigo, é realizado um relatório de pesquisa que descreve alguns dos trabalhos de literatura realizados para o desenvolvimento de sistemas de deteção de intrusão de rede. A limitação do trabalho existente atualmente, bem como as suas vantagens, é discutida com o âmbito e a direção das pesquisas realizadas. O objetivo principal desta pesquisa é apresentar a um investigador, o que já foi realizado nestes campos de pesquisa. |
| 27 | Machine Learning Methods For Cyber Security | 2018 | Rashmi Deep, Dr. Vinay Goyal | Com o uso crescente da Internet pelas pessoas, a proteção de dados confidenciais tornou-se algo que devemos ter em consideração e é um problema muito grave. Porque na era atual, o termo Cibersegurança é ouvido por todas as pessoas e relativamente a isso, as condições de Cibersegurança não são muito boas. Assim, para contornar estes problemas, introduzimos os métodos de Machine Learning na análise da rede para a Deteção de Intrusão. Como a segurança dos dados é bastante importante, o método de Machine Learning é apresentado. Nesta revisão, discutimos como o ML pode ser usado para compreender a ameaça à segurança. |
| 28 | Deep Learning Based-Phishing Attack Detection | Setembro 2019 | K. Sumathi, V. Sujatha | Devido ao rápido desenvolvimento das tecnologias de comunicação e das redes globais, muitas atividades da vida humana diária, tal como o *homebanking*, as redes sociais, o e-commerce, entre outros, são transferidos para o ciberespaço. A infraestrutura anónima, aberta e não controlada da Internet permite uma excelente plataforma para ataques cibernéticos. O Phishing é um dos ataques cibernéticos na qual os invasores abrem sites fraudulentos semelhantes aos sites populares e legais para roubar informações confidenciais do utilizador. As técnicas de Machine Learning, como o J48, o SVM, o LR, o NB e o ANN, foram utilizadas para detetar os ataques de Phishing. Porém, obter dados de tratamento com uma boa qualidade é um dos maiores problemas do Machine Learning. Portanto, um método de Deep Learning chamado Deep Neural Network (DNN) é introduzido para detetar os URL de Phishing. Inicialmente, um extrator de |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | recursos é utilizado para construir um vetor de recursos de 30 dimensões baseados em URL, HTML e domínio. Estes recursos são fornecidos como entrada para o classificador DNN para a deteção de um ataque de Phishing. Consiste numa camada de entrada, várias camadas escondidas e uma camada de saída. As várias camadas escondidas no DNN tentam aprender os recursos de alto nível de forma incremental. Finalmente, o DNN devolve um valor de probabilidade que representa os URL de Phishing e URL legais. Ao utilizar o DNN, a exatidão, precisão e recuperação da deteção de ataque de Phishing são melhoradas. |
| 29 | Machine Learning Based Phishing E-mail detection | Março 2018 | Nidhin A. Unnithan, Harikrishnan N.B., Akarsh S., Vinayakumar R., Soman K.P. | A deteção de um e-mail de Phishing é uma ameaça significativa atualmente. A taxa de criação de Phishing aumenta a cada dia. É tempo de implementar um sistema de autoaprendizagem que permita a deteção e prevenção de e-mails de Phishing de forma eficiente. Este trabalho propõe um sistema que utiliza a matriz de documentos de termos enquanto mecanismo de engenharia de recursos e técnicas clássicas de Machine Learning para detetar e-mails de Phishing e os legais. O sistema incorpora também o conhecimento de domínio e recursos lexicais como parte do mecanismo de engenharia de recursos. A eficiência do sistema é comparada ao utilizar diferentes técnicas clássicas de Machine Learning. Com base na precisão, propomos o melhor modelo para resolver o problema formulado de forma eficiente. |
| 30 | An Artificial Immune System and Neural Network to Improve the Detection Rate in Intrusion Detection System | Fevereiro 2016 | Pallvi Dehariya | O uso de AIS na deteção de intrusão é um conceito atraente por duas razões. Em primeiro lugar, o sistema imunitário humano fornece ao corpo humano um alto nível de proteção contra patógenos invasores, de forma robusta, auto-organizada e distribuída. As ameaças e intrusões nos sistemas de TI, podem ser comparadas às doenças humanas com a diferença de que o corpo humano tem uma forma eficaz de lidar com elas, o que ainda precisa de ser aplicados nos sistemas de TI. O sistema de deteção de intrusão proposto irá utilizar o conceito do AIS, este é um modelo de computação com inspiração biológica promissor. Os conceitos de AIS podem ser aplicados para melhorar a eficácia do IDS. |
| 31 | A Competent Approach for Type of Phishing Attack Detection Using Multi-Layer Neural Network | Janeiro 2017 | Bhawana Goyal, Meenakshi Bansal | Com a expansão das tecnologias contemporâneas e as redes de computadores globais em grande escala, os ataques à web têm aumentado devido à curiosidade emergente de pessoas e instituições legais em relação à internet. O Phishing é um ataque na Web executado por um atacante que utiliza tanto a engenharia social como a técnica. Na web são lançados a cada vez mais ataques que pretendem atingir os seus utilizadores com a intenção de adquirir informação sobre a sua identidade, senhas de acesso e detalhes de conta. Os métodos de classificação e deteção de ataques de Phishing são utilizados para a prevenção e análise aprofundada dos ataques. Neste artigo, o modelo proposto foi desenhado com a análise de características multidirecionais com a classificação da BP-PNN. O modelo proposto teve um melhor desempenho relativamente a precisão em todos |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | os domínios com base na deteção e classificação dos ataques. |
| 32 | An Effective Android Ransomware Detection Through Multi-Factor Feature Filtration and Recurrent Neural Network | Agosto 2019 | Iram Bibi, Adnan Akhunzada, Jahanzaib Malik, Ghufran Ahmed, Mohsin Raza | Com a diversidade crescente do Malware no Android, a eficácia dos mecanismos convencionais de defesa está em risco. Esta situação confirmou um interesse notável na melhoria da exatidão e da escalabilidade na deteção de Malware para smartphones. Neste estudo, propusemos um modelo de deteção de Malware baseado em Deep Learning que seja eficaz na deteção de Ransomware no ambiente Android, através da observação do algoritmo de Long Short-Term Memory. A seleção de recursos foi feita a utilizar 8 algoritmos de seleção diferentes. Os 19 recursos importantes são selecionados por um processo de votação por maioria, na qual é efetuada uma comparação dos resultados de todas as técnicas de filtragem de recursos. O algoritmo proposto é avaliado a utilizar um conjunto de dados de Malware do Android (CI-CAndMal2017) e parâmetros de desempenho padrão. O modelo proposto apresenta 97,08% de precisão na deteção. Com base num desempenho excecional, endossamos o algoritmo proposto para ser eficiente em Malware e análise forense. |
| 33 | Neural Networks and Deep Learning in Cyber Security | Março 2019 | Mihnea Horia Vrejoiu | Nos últimos anos, a tecnologia do Deep Learning utiliza vários modelos / arquiteturas de Deep Neural Network tornando-se o estado da arte em Machine Learning e Inteligência Artificial, as suas aplicações têm atingido melhores desempenhos que os humanos em mais e mais domínios. Enquanto as técnicas tradicionais de ML eram baseadas principalmente em certas fases da engenharia e da extração de recursos “artesanais”, a nova abordagem de DL executa automaticamente a etapa da extração de representações de recursos específicos diretamente das amostras de entrada. Essa habilidade intrínseca torna-a aplicável aos vários problemas com os quais a Cibersegurança tem lidado atualmente, como: deteção de intrusão, classificação e deteção de Malware, spam e Phishing, e análise binária. Neste artigo, pretende-se apresentar uma breve visão geral das ANN e alguns exemplos de soluções baseadas em Deep Learning para a Cibersegurança. |
| 34 | Distributed Network Intrusion Detection System: An Artificial Immune System Approach | Agosto 2016 | Obinna Igbe, Ihab Darwish, Tarek Saadawi | A deteção de intrusão é a identificação de uso não autorizado, mau uso e abuso de infraestruturas dos sistemas computacionais por utilizadores do sistema e utilizadores externos. Detetar uma intrusão numa rede distribuída de um segmento de rede externo, bem como de dentro, é um problema difícil. O NIDS deve analisar um largo volume de dados, sem colocar uma carga adicional significativa nos sistemas de monitorização e redes. Este artigo apresenta uma *framework* para um sistema de deteção de intrusão numa rede distribuída (dNIDS) baseado no conceito de AIS. Neste contexto, um mecanismo imunológico adaptável por métodos de ML supervisionado é proposto para classificar o tráfego da rede tanto em perfis normais e em perfis suspeitos. Experimentalmente, a abordagem distribui o NIDS entre todos os segmentos de rede ligados, permitindo que o NIDS em cada segmento identifique quais as ameaças |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | potenciais individualmente e permitindo a partilha de vetores de ameaças identificados entre os NIDS distribuídos em comunicação. É apresentada uma análise da técnica de distribuição dessas informações sobre vetores de ameaças. |
| 35 | Neural Network Implementation for Detection of Denial of Service Attacks | Fevereiro 2020 | Irina Topalova, Pavlinka Radoyska, Strahil Sokolov | Os ataques de DoS são considerados um grande risco porque podem facilmente interromper um serviço, um negócio ou processo educacional. Esses ataques são relativamente simples de conduzir, mesmo por um atacante não qualificado e causam perdas significativas. Por esta razão, é particularmente importante que esses ataques sejam detetados, reconhecidos e bloqueados a tempo. A maioria dos métodos e ferramentas avançados de proteção contra esses ataques baseia-se em monitorização e rastreio constante para detetar qualquer tráfego de IP suspeito. A aplicação destes métodos está associada a recursos computacionais adicionais e especialização, o que leva a uma subjetividade na avaliação da ameaça. Portanto, é necessário propor métodos para uma deteção adaptável e automatizada, bem como para o reconhecimento de ataques de DoS. Este estudo apresenta um método para deteção e reconhecimento automatizado de alguns dos ataques de DDoS, por um sistema adaptável e automatizado, baseado numa Rede Neural com várias camadas. É treinado tanto normalmente quanto com sinais que refletem diferentes condições de tráfego quando ocorrem ataques de DDoS. A Rede Neural é testada para reconhecer "sinais de base", representando as diferentes condições normais de tráfego e para detetar as situações anormais de tráfego. A pesquisa é conduzida para diferentes ataques de DDoS internos numa rede local real. Os resultados de precisão de reconhecimento obtido são representados e os benefícios alcançados são discutidos. |
| 36 | A Futuristic Approach: Incorporating Artificial Intelligence with Cyber Security | Dezembro 2018 | Shefali Nangia, Megha Malik, Deepak Chahal, Latika Kharb | Num mundo digital de IoT e dispositivos interligadas, a Cibersegurança tornou-se num motivo para preocupação. Os especialistas precisam de toda a ajuda para evitar ataques e falhas na segurança, bem como, de responder aos ataques. Como pode ser visto, não é possível criar um sistema de software baseado em lógica e bem ligados para lidar com ataques cibernéticos tão graves. E tem sido visto por um bom período, que se os procedimentos de Inteligência Artificial forem bem implementados, várias questões de segurança podem ser resolvidas com o progresso.<br>Este artigo fala sobre a Inteligência Artificial, que pode ser referida como a habilidade de uma máquina ou de um software para pensar e aprender. O conceito de IA baseia-se na ideia de construir máquinas capazes de pensar, agir e aprender tal como os humanos. E a Cibersegurança que se refere simplesmente às técnicas de proteção de computadores, redes, programas e dados contra acessos não autorizados ou ataques que pretendem explorar as vítimas.<br>Com o aumento o número de ataques cibernéticos, a Cibersegurança tem-se tornado numa grande preocupação. Então, aqui falamos de como a Inteligência |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | Artificial pode ajudar nisso, as várias técnicas de IA que podem ser implementadas para a Cibersegurança e qual é o seu futuro. |
| 37 | A Multi-Level Ransomware Detection Framework using Natural Language Processing and Machine Learning | Outubro 2019 | Subash Poudyal, Dipankar Dasgupta, Zahid Akhtar, Kishor Datta Gupta | Os ataques de Ransomware nos últimos anos provaram ser dispendioso devido aos danos e obstruções significativos causados em vários setores, tal como saúde, seguros, negócios e educação. Foram propostos vários métodos de deteção de Malware para descobrir diferentes famílias de Malware, mas o problema permaneceu sem solução por o Malware estar em constante evolução. Neste trabalho, propusemos uma estrutura de multinível de Big Data Mining combinada com as abordagens da engenharia reversa, NLP e ML. A estrutura analisa o Ransomware em diferentes níveis através de diferentes algoritmos de ML supervisionados. O Apache Spark foi aplicado para um processamento mais rápido de um grande conjunto de recursos gerados. O analisador PE e a ferramenta de descarga de objetos do sistema Linux foram utilizados para obter os dados brutos do Ransomware e os binários normais que foram processados posteriormente usando o processamento de NLP personalizado. As probabilidades de n-gram, frequência de termo e frequência inversa de documento (TF-IDF) foram usadas para gerar os conjuntos de recursos finais. As experiências foram realizadas com o valor N diferente do modelo de linguagem n-gram que demonstra que a precisão de deteção do Ransomware é inversamente proporcional ao valor de N. Entre os cinco classificadores supervisionados escolhidos, a regressão logística superou os outros com uma taxa de deteção de 98,59% para TF-IDFs em vários níveis combinados, o que é uma precisão aprimorada em comparação com os níveis individuais. |
| 38 | A Survey On Automatic Phishing Email Detection Using Natural Language Processing Techniques | Novembro 2019 | Anirban Mukherjee, Nimit Agarwal, Shubham Gupta | Na era global de hoje, conforme o mundo fica menor, tornando-nos todos ligados uns aos outros. Os e-mails são uma ferramenta essencial para realizar a interligação de máquinas e pessoas em todo o mundo. Com o advento das tecnologias de comunicação e o aumento da utilização da Internet em termos exponenciais nos últimos anos, os e-mails passaram a fazer parte do nosso dia a dia. Dito isto, infelizmente também deu origem a muitos golpes online, estes utilizam técnicas de Phishing para fazer com que utilizados desinformados disponibilizem as suas informações pessoais, o que pode levar a roubos significativos de dinheiro, usurpação de identidade, perda de credibilidade e muitas outras atividades maliciosas com consequências graves para os cidadãos da Internet. Estes problemas levam a que enfrentar o desafio de e-mails de Phishing seja uma prioridade. Portanto, neste artigo, examinamos a deteção de e-mails de Phishing utilizando técnicas de NLP. São investigados os métodos que ajudam a identificar esses e-mails maliciosos utilizando NLP, o que ajuda a detetar e classificar os e-mails potencialmente prejudiciais comparativamente aos que não causem nenhum dano potencial aos preciosos dados do utilizador. É verificada a implementação destes métodos para detetar os e-mails e qual o trabalho que pode ser |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | realizado para melhorar os cenários atuais e quais os possíveis trabalhos a realizar no futuro, porque conforme os atacantes continuam a evoluir e vão ficando mais sofisticados, precisamos de nos desenvolver e enfrentar o desafio também. |
| 39 | Catch me, Yes we can! - Pwning Social Engineers using Natural Language Processing Techniques in Real-Time | 2018 | Myeongsoo Kim, Changheon Song, Hyeji Kim, Deahyun Park, Yeeji Kwon, Eun Namkung, Ian G. Harris, Marcel Carlsson | Os ataques de engenharia social são uma das ameaças mais comuns à segurança e das menos defendidas atualmente. Apresentamos uma abordagem que analisa o conteúdo do ataque de forma a detetar declarações inadequadas, as quais indicam possíveis ataques de engenharia social. As pesquisas anteriores para a deteção de ataques de engenharia social dependem bastante da análise de vários meta-dados específicos para o vetor de ataque por email, incluindo informações de cabeçalho e links de URL. A abordagem seguida é nova em comparação com o trabalho anteriormente realizado porque se foca no texto em linguagem natural incluído no ataque, é efetuada uma análise semântica do texto para detetar intenções maliciosas. O foco na análise de texto torna a abordagem seguida aplicável na deteção de ataques de engenharia social utilizando vetores de ataque que não seja o e-mail, incluindo aplicações de mensagens de texto, aplicações de conversa e ataques por telefone/pessoalmente, estes foram convertidos em texto usando uma aplicação que traduz fala em texto. Para demonstrar a eficácia da abordagem seguida, foi realizada uma avaliação utilizando um largo conjunto de e-mails de Phishing. |
| 40 | Modified honey encryption scheme for encoding natural language message | Abril 2019 | Abiodun Esther Omolara, Aman Jantan | Os esquemas de criptografia convencionais são suscetíveis a ataques de força bruta. Isso ocorre porque os bytes codificam caracteres UTF8 (ou ASCII). Consequentemente, um adversário que impede um texto cifrado e que tenta descriptografar a mensagem por força bruta com uma chave incorreta pode filtrar algumas das combinações da mensagem descriptografada, através da observação de que algumas das sequências são uma combinação de caracteres que são distribuídos de forma não uniforme e não apresentam nenhum significado plausível. O esquema de criptografia Honey foi proposto para reduzir a vulnerabilidade existente na criptografia convencional, produzindo textos cifrados com uma aparência válida, que se encontram uniformemente distribuídos, mas, são falsos após a descriptografia com as chaves incorretas. No entanto, o esquema funciona apenas para senhas e PINS. A sua adaptação para suportar a codificação de mensagens em linguagem natural continua a ser um problema por resolver. As propostas existentes para estender o esquema para suportar a codificação das mensagens em linguagem natural revelam fragmentos do texto simples no texto cifrado, daí a sua suscetibilidade a estes ataques. Neste artigo, modificamos os esquemas de encriptação Honey para suportar a codificação de mensagens em linguagem natural utilizando as técnicas de NLP. A principal contribuição deste artigo, foi criar uma estrutura que permitisse que uma mensagem fosse codificada inteiramente em binário. Como resultado dessa estratégia, a maioria das cadeias binárias produz |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | mensagens sintaticamente corretas que serão geradas para enganar um atacante que tente descriptografar um texto cifrado utilizando chaves incorretas. |
| 41 | Natural Language Processing to Quantify Security Effort in the Software Development Lifecycle | 2015 | Constantine Aaron Cois, Rick Kazman | Abordar a segurança no ciclo de vida de desenvolvimento de software é uma preocupação sempre presente para os engenheiros de software e as organizações. Na perspetiva de gestão e monitorização, é difícil medir a quantidade de esforço colocado nas questões de segurança durante o desenvolvimento ativo e o sucesso dos esforços colocados no projeto e desenvolvimento relacionados com a segurança. Estes dados simplesmente não são registados. Se as medições confiáveis estivessem disponíveis, os líderes de projeto de software teriam uma ferramenta poderosa para avaliar qual o risco e informar a tomada de decisão. Isso permitiria aos gestores direcionar o desenvolvimento e os testes para garantir qual o nível desejado de segurança nos produtos de software, para proteger as suas organizações e os seus clientes. Para preencher essa necessidade e fornecer esses dados, propomos uma técnica para realizar a deteção de tópicos em dados disponíveis na maioria dos projetos de desenvolvimento de software: artefactos de texto para rastreio de problemas e sistemas de controlo de versão. Aplicamos técnicas de ML e NLP para criar classificadores capazes de detetar com precisão se um determinado pedaço de texto está relacionado com o tópico de segurança.<br>A realização de tal capacidade dará às equipas de software a capacidade de analisar os níveis atuais e anteriores do esforço aplicado à segurança, revelando qual o foco imediato do projeto e quais os impactos a longo prazo das tarefas de segurança. Foi validada a abordagem seguida por experiências em dados de grande escala do projeto de software de código aberto Chromium. Os resultados mostram que o esquema de classificação Naïve Bayes ao utilizar um espaço de recursos n-gram se torna uma abordagem apropriada e eficaz para a deteção automatizada de tópicos de fragmentos de texto de segurança de software, e que os dados de teste quando eficazes podem ser derivados de fontes de dados públicas sem a necessidade de intervenção manual. |
| 42 | Security for Biometrics Protection between Watermarking and Visual Cryptography | Março 2017 | P. Anitha, K. Narayana Rao, V. Rajasekhar, Ch. Hari Krishna | Este artigo apresenta uma nova arquitetura de segurança para proteger a integridade de imagens e modelos de íris utilizando a marca de água e a criptografia visual. A biometria da retina é considerada um dos métodos mais precisos e robustos na verificação de identidade. Oferece uma estrutura de proteção completa para a biometria das íris, que consiste em duas etapas: a primeira etapa é para a proteção da imagem da íris, enquanto a segunda é para o modelo da íris. Em primeiro lugar, para proteger a imagem da íris, um texto de marca de água que carrega informações pessoais é incorporado na região de frequência de banda média da imagem da íris utilizando um novo algoritmo de marca de água. Em segundo lugar, para a proteção do modelo da íris, o modelo binário da íris é dividido em dois blocos utilizando a criptografia visual, onde um dos blocos propostos é armazenado na |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | base de dados como sendo um formato simétrico criptografado utilizando o método de criptografia convencional como DES ou AES e o outro é mantido com o utilizador num cartão inteligente com a chave secreta de criptografia simétrica. Além disso, a função *hash* SHA-2 é utilizada para manter a integridade do modelo de íris armazenado na base de dados e no cartão inteligente. Como os modelos biométricos são armazenados na base de dados centralizada, devido às ameaças à segurança, o modelo biométrico pode ser modificado pelo atacante. Se o modelo biométrico for alterado, o utilizador autorizado não terá permissão para aceder ao recurso. A estrutura proposta disponibiliza à segurança de ambas as bases de dados contra possíveis ataques maliciosos. A combinação da biométrica e da criptografia visual é uma técnica de segurança da informação promissora que oferece uma forma eficiente de proteger o modelo biométrico. |
| 43 | A fast defect detection algorithm for glass tube based on ROI reduction | 2019 | Gabriele Antonio De Vitis, Pierfrancesco Foglia, Cosimo Antonio Prete | Neste trabalho, apresentamos um algoritmo para a deteção de defeitos em produtos de vidro, o que nos permite minimizar o tempo de processamento. A ideia principal baseia-se na redução do tamanho da área da imagem a investigar com a utilização das características das imagens de vidro. Os resultados de um conjunto de imagens de teste mostram que a solução proposta não compromete a qualidade da deteção e permite alcançar uma melhoria 7x superior em relação à solução existente em condições particulares, com a mesma precisão na deteção de defeitos. |
| 44 | The New Embedded ATM Security based on Machine Vision using MATLAB | Abril 2016 | R. Mahendran, C. Karthik | Um sistema de reconhecimento facial é uma aplicação computacional para identificar ou verificar automaticamente uma pessoa a partir de uma imagem digital, ou um fragmento de um vídeo. O artigo proposto usa a técnica de reconhecimento facial para verificações no sistema ATM. Para o reconhecimento de rosto, existem duas comparações. A primeira é a verificação, que é onde o sistema compara o determinado indivíduo com quem ele diz ser e devolve uma decisão de sim ou não. O próximo é a identificação, que é onde o sistema compara o determinado indivíduo com todos os outros indivíduos na base de dados e disponibiliza uma lista classificada de correspondências. A tecnologia de reconhecimento facial analisa a forma, o padrão e o posicionamento exclusivos das características faciais. O reconhecimento de rosto começa com uma imagem, na tentativa de encontrar uma pessoa na imagem usando o software MATLAB. A saída do resultado do MATLAB é dado ao microcontrolador, que verifica o rosto do utilizador e a senha. Se ambos forem combinados com os dados, então apenas o utilizador pode continuar a sua transação. |
| 45 | Optimation through Automation of Malware Update Process, Capable of Evading Anti-Malware Systems | Setembro 2016 | Daniel Soto Carabantes, Cristian Barría Huidobro, David Cordero Vidal | A implementação e manutenção de medidas de proteção contra Malware implicam uma elevada utilização de recursos. É o caso dos ISMS, cuja estrutura sugerida está descrita na norma ISO 27.001:2013. Nesta norma, o trabalho com Malware é contemplado para fins de teste de penetração, permitindo avaliar a resposta dos sistemas computacionais frente a estes eventos. O |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | documento aborda um dos métodos de utilização de Malwares existentes para esse fim: ofuscação de Malware criptografado, por meio da adição de código morto. Este método é avaliado em termos monetários e em tempo necessário, por simulações, para posteriormente avaliar essas métricas contra um modelo automatizado, testado por um software protótipo. A otimização desse processo através da automação proposta, resultou numa redução significativa do custo monetário e do tempo necessário. |
| 46 | Automation of White-Box Cryptography attacks in Android Applications | Dezembro 2018 | Víctor Sánchez Ballabriga | A criptografia é cada vez mais implementada em aplicações executadas em dispositivos abertos (tal como computadores portáteis, tablets ou smartphones). A natureza aberta destes sistemas torna o software extremamente vulnerável a ataques, uma vez que o atacante tem controlo total sobre a plataforma de execução e a implementação do software. Isto significa que um atacante pode analisar facilmente o código binário da aplicação e as páginas de memória correspondentes durante a execução. Portanto, a WBC é uma tecnologia essencial em qualquer estratégia de proteção de software. Na prática, o WBC é um algoritmo criptográfico, geralmente simétrico, que incorpora a chave e ofusca o processo, de forma a torná-lo mais difícil de reverter. Esta tecnologia permite realizar operações criptográficas sem revelar nenhuma parte das informações confidenciais, tal como a chave criptográfica. Sem isto, os atacantes poderiam facilmente roubar as chaves secretas da implementação binária, da memória ou intercetar as informações que levariam à revelação no momento da execução. Em paralelo, novos caminhos de ataque apareceram com esta tecnologia, ataques que no passado eram usados para atacar criptografia implementada no hardware, como DPA ou DFA encontraram os seus equivalentes ao nível do software. Estes ataques equivalentes são baseados na análise dos acessos à memória ou dos valores armazenados nas inscrições em cada ciclo do processador. Para recuperar estes vestígios, a instrumentação binária dinâmica e a emulação dos binários são as melhores abordagens. No entanto, para proteger o WBC implementado no software, normalmente outras contramedidas de segurança são aplicadas, como a ofuscação de código, o que pode inviabilizar a tarefa de localizar as funções que executam as tarefas de criptografia/descriptografia, tornando inviável o seu rastreio para recuperar rastos válidos para um ataque bem-sucedido. Este projeto tem dois objetivos principais. O primeiro é facilitar a procura pelas funções WBC em binários ofuscados, tornando-o mais fácil para os analistas de segurança que queiram testar a força das diferentes implementações WBC em produtos reais. Pesquisando e aprofundando as diferentes técnicas que poderiam ser utilizadas para identificar a impressão digital que esta criptografia pode deixar no binário que a utiliza, a sua identificação pode se tornar uma tarefa semiautomática. O segundo objetivo do projeto é implementar um emulador de um processador ARM com |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | capacidade para recolher as informações necessárias para os ataques mencionados anteriormente, gerando também as falhas necessárias na execução quando necessário. |
| 47 | Cybersecurity in Autonomous Systems: Evaluating the performance of hardening ROS | Junho 2016 | Francisco Javier Rodríguez Lera, Jesús Balsa, Fernando Casado, Camino Fernández, Francisco Martín Rico, Vicente Matellán | À medida que os sistemas robóticos se propagam, surge a Cibersegurança como principal preocupação. Atualmente, a maioria dos sistemas autónomos de pesquisa são construídos a utilizar a estrutura ROS, com outro software comercial. O ROS é uma estrutura distribuída na qual os nós existentes publicam informações, que outros vão consumir, mais tarde. Este modelo simplifica a comunicação de dados, mas representa uma grande ameaça porque um processo malicioso pode facilmente interferir nas comunicações, ler mensagens privadas ou até mesmo substituir os nós. Neste artigo, é proposto que as comunicações ROS sejam criptografadas. É também medido a forma como a criptografia afeta o seu desempenho. É utilizado o algoritmo de cifra 3DES e avaliado o desempenho do sistema, tanto do ponto de vista computacional quanto de comunicação. Os resultados preliminares mostram que cifras simétricas ao utilizar chaves privadas impõem atrasos significativos. |
| 48 | Security framework for industrial collaborative robotic cyber-physical systems | Março 2018 | Azfar Khalid, Pierre Kirisci, Zeashan Hameed Khan, Zied Ghrairi, Klaus-Dieter Thoben, Jürgen Pannek | O documento apresenta uma estrutura de segurança para a aplicação da colaboração do humano com o robot num contexto futurístico de num CPS industrial da indústria 4.0. Os elementos básicos e os requisitos funcionais de um CPS robótico colaborativo seguro são explicados e, em seguida, os modos de um ataque cibernético são discutidos no contexto do CPS colaborativo, enquanto uma estratégia de mecanismo de defesa é proposta para um sistema tão complexo. Os ciberataques são categorizados de acordo com a extensão da controlabilidade e os possíveis efeitos sobre o desempenho e a eficiência do CPS. O documento também descreve a gravidade e a categorização de tais ataques cibernéticos e o efeito causal na segurança do trabalhador humano durante a colaboração entre humanos e robots. Ataques em três dimensões de disponibilidade, autenticação e confidencialidade são propostos como base para um plano de mitigação consolidado. É proposta uma estrutura de segurança baseada numa estratégia de duas vertentes, na qual o impacto desta metodologia é demonstrado por um *benchmark* de tele-operação (NeCS-Car). A estratégia de mitigação inclui o aperfeiçoamento da segurança dos dados em importantes nós adaptadores e interligados, bem como o desenvolvimento de um módulo inteligente que emprega um conceito semelhante à monitorização e à reconfiguração do sistema. |
| 49 | An Efficient Approach Based on Neuro-Fuzzy for Phishing Detection | Abril 2016 | Luong Anh Tuan Nguyen, Huu Khuong Nguyen, Ba Lam To | Na era da Internet, o comércio online tem crescido rapidamente em vários campos. Como resultado, o crime cibernético aumenta constantemente. O Phishing é um novo crime que visa roubar informações do utilizador através de páginas da web falsas. A maioria dessas páginas da web de Phishing são semelhantes às páginas reais como a interface do site e do endereço de URL. Muitas técnicas têm sido propostas para detetar sites de |

| # | Título | Ano | Autor(es) | Descrição / Sinopse |
|---|---|---|---|---|
| | | | | Phishing, tal como a técnica baseada na lista negra, a técnica baseada na heurística, entre outros. No entanto, o número de vítimas tem aumentado devido à técnica de proteção ser ineficiente. Redes neurais e sistemas fuzzy podem ser combinados para unir as suas vantagens e tratar das suas vulnerabilidades individuais. Este artigo propõe um novo modelo neuro-fuzzy sem a utilização de conjuntos de regras para deteção de Phishing. Especificamente, a técnica proposta que calcula o valor das heurísticas a partir de funções de adesão. Em seguida, os pesos são treinados por Rede Neural com taxa de aprendizagem adaptável. A técnica proposta é avaliada por um conjunto de dados de 11.660 sites de Phishing e de 10.000 sites legítimos. Os resultados mostram que a técnica proposta consegue detetar mais de 99% de sites de Phishing. |
| 50 | A Neuro-Fuzzy System to Detect IPv6 Router Alert Option DoS Packets | Janeiro 2020 | Shubair Abdullah | A deteção de ataques de DoS que visam exclusivamente o router são um imperativo de segurança máximo na implementação das redes de IPv6. Os métodos de deteção de DoS visam aproveitar as vantagens dos recursos estatísticos do fluxo e das técnicas de Machine Learning. No entanto, o desempenho na deteção é altamente afetado pela qualidade do selecionador de recursos e pela confiabilidade dos conjuntos de dados nas informações do fluxo IPv6. Este artigo propõe um novo sistema de inferência neuro-fuzzy para resolver o problema de classificação dos pacotes nas redes IPv6, em situações decisivas num pequeno conjunto de dados supervisionados. O sistema proposto pode classificar os pacotes com opção de alerta do router IPv6 em DoS e normal, utilizando as forças do neuro-fuzzy para aumentar a precisão da classificação. Uma análise matemática da perspetiva da teoria dos conjuntos fuzzy é disponibilizada para expressar o benefício do desempenho do sistema proposto. Um teste de desempenho empírico é conduzido por um conjunto de dados abrangente dos pacotes IPv6 produzidos num ambiente supervisionado. O resultado mostra que o sistema proposto supera de maneira robusta alguns sistemas de última geração. |

Tabela 1 – Artigos para análise quantitativa

# 3. METODOLOGIA

A metodologia utilizada será *a* Design-Science Research (DSR), esta metodologia é bastante utilizada em projetos de Sistemas de Informação, nomeadamente em questões de desenho de processos de negócio e de soluções que permitem uma combinação de sínteses.

## 3.1. DESIGN-SCIENCE RESEARCH (DSR)

A Design-Science Research (DSR) é uma metodologia que apresenta um conhecimento científico e que permite obter um ponto de vista bastante analítico. Uma vez que o resultado consistirá na elaboração de um referencial, é de uma enorme mais valia utilizar esta metodologia, já que procura investigar os problemas de uma forma intuitiva e prática, contudo, é importante clarificar que o resultado não se centra na apresentação de um produto ou serviço, mas sim em algo conceptual, como, teorias de desenho, construções, modelos, princípios de desenho e regras tecnológicas (Gregor, S., & Hevner, A. R., 2013).

O método DSR deve ser utilizado e aplicado em análises de Sistemas de Informação, como referido anteriormente, sendo que os conhecimentos necessários para realizar uma pesquisa nos Sistemas de Informação envolvem paradigmas na ciência do comportamento e do desenho, a primeira aborda a pesquisa pelo desenvolvimento das teorias, enquanto a segunda aborda a pesquisa pelo desenvolvimento e avaliação do que foi criado para satisfazer o objetivo (Hevner, 2007).

Assim, podemos considerar o DSR como um método de pesquisa que envolve a criação e/ou melhoria de algo de uma forma inovadora, que advém de um problema especifico (Hevner, A. R., March, S. T., Park, J., & Ram, S., 2004), acaba por se enquadrar no principal objetivo desta tese, que se foca na garantia de segurança para os Sistemas de Informação que são usados no dia a dia, para que os mesmos não sofram danos devido a possíveis ameaças.

O objetivo é garantir a coerência e a credibilidade do resultado que se pretende demonstrar, realizando uma pesquisa científica consistente. A comunicação com entidades que lidam com Sistemas de Informação frágeis será um ponto crucial para se conseguir perceber de que forma é possível melhorar, além disso, é importante não deixar escapar nenhum detalhe (Hevner, A. R., March, S. T., Park, J., & Ram, S., 2004).

Deverão ser seguidos seis passos fundamentais que compõem a metodologia de pesquisa DSR, que são os seguintes:

- **Identificar o problema e a motivação**, neste passo o objetivo é identificar qual será o foco do estudo que pretendemos realizar e qual o seu valor, bem como os possíveis interessados na solução (Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S., 2007).
- **Definir os objetivos e a solução**, especificar os requisitos, tanto quantitativos quanto qualitativos, para que de uma maneira clara consigamos expor uma possível solução e como a mesma poderá ser implementada (Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S., 2007).
- **Desenho e desenvolvimento**, foca-se na procura do conhecimento para a construção do artefacto final, que pode ser alcançado através da separação dos problemas principais

encontrados durante a pesquisa em algo mais simples (Hevner, A. R., March, S. T., Park, J., & Ram, S., 2004).

- **Avaliação**, neste passo, o objetivo é validar a eficiência do resultado, a mesma deve-se focar em métodos que permitam determinar a viabilidade da solução final (Hevner, A. R., March, S. T., Park, J., & Ram, S., 2004).
- **Comunicação**, o último passo é referente à apresentação da solução final para o mundo, permitindo assim determinar se o resultado será aceite no mercado, além disso, ao dar a conhecer o resultado é importante clarificar como se chegou ao seu valor final (Hevner, A. R., March, S. T., Park, J., & Ram, S., 2004).

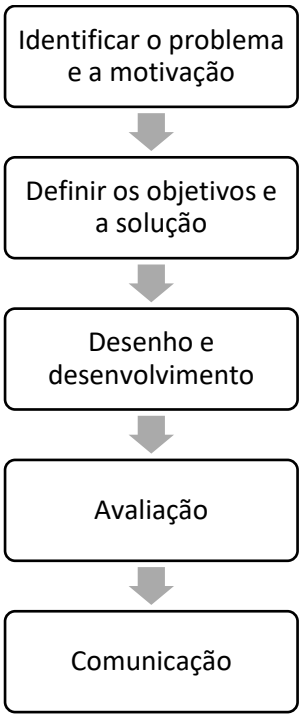


Figura 2 – Passos da Metodologia DSR

## 3.2. ESTRATÉGIA DE INVESTIGAÇÃO

Primeiramente foi necessário verificar quais os tópicos que se pretendia abordar neste presente estudo, diante disso, foram definidos os conceitos, Inteligência Artificial e Cibersegurança. De seguida, foram analisadas as possíveis relações entre os conceitos, bem como possíveis problemas para os quais se pretendia procurar uma solução.

Assim, foi definido como objetivo deste estudo verificar a possibilidade de cruzar as diferentes técnicas de Inteligência Artificial com as diferentes técnicas de Cibersegurança, de forma a determinar qual a melhor maneira de proteger os Sistemas de Informação, deixando-os menos

vulneráveis e mais resistentes. A solução para a resolução deste problema irá ser apresentada por um referencial onde ambas as técnicas serão cruzadas e onde se determinará como proceder nas variadas situações.

Para se obter a solução, foi previamente efetuada uma análise dos conceitos, onde foram abordadas as diferentes técnicas de Inteligência Artificial e de Cibersegurança. Foi de seguida, efetuada uma revisão sistemática da literatura para determinar quais os artigos mais importantes para corroborar o que se pretende provar.

Com base na informação obtida na revisão sistemática, será apresentada uma proposta onde se mostram as conclusões a que foi possível chegar, bem como, um referencial representativo das conclusões obtidas.

O resultado obtido irá ser apresentado a responsáveis com conhecimento sobre ambos os conceitos, que conseguem garantir que o que foi apresentado é efetivamente válido e pode ser usado por organizações, de forma, a que estas consigam resolver os problemas existentes ao nível de segurança.

# 4. PROPOSTA

## 4.1. PRESSUPOSTOS

Após a execução da revisão sistemática de literatura foi possível obter as seguintes assunções:

- As áreas da Cibersegurança onde são frequentemente utilizados algoritmos de Inteligência Artificial são:
    - Deteção de Intrusão
    - Malware
    - Ransomware
    - DoS (Denial of Service)
    - Phishing
- As técnicas de Inteligência Artificial consideradas promissoras e que são atualmente foco de investigação para a Cibersegurança são:
    - Machine Learning
    - NLP (Natural Language Processing)
    - Automatização & Robótica
    - Fuzzy Logic
    - Vision Machine
- As áreas da Cibersegurança onde são utilizados algoritmos de Machine Learning são:
    - Deteção de Intrusão
    - Malware
    - Ransomware
    - DoS (Denial of Service)
    - Phishing
- As áreas da Cibersegurança onde são utilizados algoritmos de NLP (Natural Language Processing) são:
    - Deteção de Intrusão
    - Ransomware
    - Phishing
- As áreas da Cibersegurança onde são utilizados algoritmos de Automatização & Robótica são:
    - Malware
- As áreas da Cibersegurança onde são utilizados algoritmos de Fuzzy Logic são:
    - DoS (Denial of Service)
    - Phishing
- As áreas da Cibersegurança onde são utilizados algoritmos de Machine Vision:
    - Deteção de Intrusão

## 4.2. MODELO / RECOMENDAÇÕES INTELIGÊNCIA ARTIFICIAL NA CIBERSEGURANÇA

Com base nos pressupostos obtidos no ponto anterior (4.1), foi criado um referencial onde foram inseridas as técnicas de Inteligência Artificial e as áreas de Cibersegurança encontradas nos artigos selecionados.

| | | | Áreas de Cibersegurança | | | | |
|---|---|---|---|---|---|---|---|
| | | | **Deteção de Intrusão** | **Malware** | **Ransomware** | **DoS** | **Phishing** |
| **Técnicas da IA** | **Machine Learning** | **AIS** | (1) | | | | |
| | | **Deep Learning** | (2) | | (9) | | (14) |
| | | **ANN** | (3) | | | | |
| | | **Machine Learning** | (4) | (7) | (10) | (12) | (15) |
| | **NLP** | - | (5) | | (11) | | (16) |
| | **A&R** | - | | (8) | | | |
| | **FL** | - | | | | (13) | (17) |
| | **VM** | - | (6) | | | | |

Tabela 2 – Referencial das técnicas de Inteligência Artificial e Cibersegurança

Após a criação do referencial, foi realizado um cruzamento entre as técnicas de Inteligência Artificial e as áreas da Cibersegurança, isto é, foi verificado quais as áreas que apresentavam uma possível relação.

Abaixo é detalhado de que forma ambas as técnicas se relacionam entre si:

(1) Artificial Immune Systems e Deteção de Intrusão:

Os Artificial Immune Systems (AIS) podem ser utilizados para resolver questões relacionadas com a intrusão, nomeadamente para ajudarem os sistemas preparados para a deteção de intrusões, conhecidos como Intrusion Detection Systems (IDS) [34].

Os IDS são úteis para identificar a atividade maliciosa existente na rede, tal como, o uso indevido ou não autorizado, contudo, têm a sua capacidade limitada no que toca à deteção de ataques distribuídos e coordenados [34][30].

Atualmente várias redes estão protegidas contra o tráfico malicioso utilizando infraestrutura de multicamada também conhecida como Network Intrusion Detection Systems (NIDS) [34].

Os AIS exploram as características dos sistemas imunológicos, nomeadamente a capacidade de aprender e memorizar, de forma a possibilitar a resolução de diversos problemas [34].

Além disso, são uma excelente abordagem para a resolução de problemas, ao nível de segurança, existentes nas redes de computadores, uma vez que estas são semelhantes ao corpo humano. Um sistema digital imune pode ser criado pelas redes para combater intrusões, tal como as células imunes combatem patógenos [34].

Como tal, a sua aplicação consiste na utilização de AIS nos IDS, para ajudar na resolução de possíveis problemas de segurança e facilitar a intercomunicação entre as redes de deteção de intrusões, permitindo uma maior proteção [30][34].
Artigos utilizados para esta relação do referencial: [34] (Igbe, Darwish, & Saadawi, 2016), [30] (Dehariya, 2016)

(2) Deep Learning e Deteção de Intrusão:

O Machine Learning desempenha um papel importante contra as ameaças ao nível de Cibersegurança, por ser bastante poderoso, depende bastante da extração de recursos, no entanto, para se detetarem ocorrências de Malware é necessário compilar vários recursos manualmente, o que limita a eficiência e a precisão na deteção da ameaça [6].
Assim, surge a necessidade de utilizar o Deep Learning, este algoritmo apresenta uma maior complexidade comparativamente ao Machine Learning, não necessita que seja efetuada uma extração de recursos, consegue detetar correlações não lineares escondidas nos dados, suporta qualquer tipo de novos ficheiros e permite detetar ataques desconhecidos, tornando-se assim uma vantagem ao nível de segurança. Tal como o Machine Learning, o Deep Learning é baseado na aprendizagem e é utilizado em sistemas autónomos, devido às vantagens que apresenta ao nível de otimização, discriminação e previsão [6][7][26][33].
Deep Belief Network (DBN) é um modelo probabilístico, constituído por várias camadas, utilizado para o processamento de dados proposto por Geoffrey Hinton, que é aplicado com sucesso em áreas como o reconhecimento de discurso e de objetos, sendo capaz de processar um largo volume de dados, tornando-o assim efetivo para complementar o Deep Learning enquanto ferramenta para detetar possíveis intrusões [6] [7][22][33].
Os Intrusion Detection System (IDS) foram desenvolvidos com o objetivo de introduzir mecanismos que ajudam a diferenciar atividades benignas de malignas, envolvendo segurança tanto ao nível de rede quanto de hospedeiro [13][33].
Os Network Intrusion Detection System (NIDS) são utilizados em situações onde podem ocorrer problemas ao nível de rede, uma vez que oferecem uma segurança superior às tecnologias tradicionais de defesa, como as *firewalls*, assim ajudam os administradores de rede a detetar ataques, vulnerabilidades e brechas existentes na rede da organização [13][26][33].
Os Host Intrusion Detection System (HIDS) são utilizados no sistema hospedeiro para observar ao longo da operação do sistema, bem como para verificar o seu estado a fim de detetar instalações e acessos não autorizados, vai ainda verificar o estado da RAM e dos ficheiros alocados no sistema [33].
Os Intrusion Detection Prevention Systems (IDPS) são utilizados para detetar intrusões, na qual após a possível deteção de uma intrusão, os administradores autorizados recebem e-mails de alerta. Além de detetar intrusões, os IDPS também as impedem quando um intruso tenta aceder à rede [7][33].
Estes sistemas de prevenção de intrusões podem gerar problemas no momento da sua configuração, na medida em que não existe um modelo padrão a seguir, devido a isso podem gerar um elevado número de falsos positivos [7].
A tecnologia IDPS depende de dois sistemas diferentes, um que engloba a deteção de intrusões que se encontram registadas numa base de dados e outro que vai analisar os pacotes, recuperar as suas assinaturas da rede e comparar com o existente na base de

dados. No entanto, só funciona perante ameaças de intrusões previamente conhecidas. Contudo, existe ainda outro sistema que vai analisar o comportamento na rede, sendo mais eficiente que os sistemas referidos anteriormente [7].

A utilização de Deep Learning em IDS, na deteção e resistência a intrusões, demonstra uma maior precisão, apresentando assim melhores resultados, melhorando o desempenho e utilizando o modelo DBN para ajudar a detetar ameaças. O IDS vai analisar a rede e enviar os dados produzidos para o DBN, que vai criar um conjunto de dados para iniciar o processo, de forma a determinar se é uma ameaça ou um falso positivo [6][7][26].

Artigos utilizados para esta relação do referencial: [6] (Li, 2018), [7] (Calderon, 2019), [13] (Al-Emadi, Al-Mohannadi, & Al-Senaid, 2019),[22] (B & S, 2019), [26] (Jamadar, Ingale, Panhalkar, Kakade, & Shinde, 2019), [33] (Vrejoiu, 2019)

(3) Artificial Neural Network e Deteção de Intrusão:

Os Artificial Neural Network (ANN) são um modelo estatístico de aprendizagem que imita a estrutura de funcionamento do cérebro humano, nomeadamente as redes neurais biológicas naturais, que ajudam a aprender e a resolver problemas, especialmente em ambientes onde as regras e os algoritmos para resolução de problemas são difíceis de expressar ou são desconhecidos [1][2].

Podem ser utilizados para compreender as atividades de rede anteriores e ataques, para prevenir ataques futuros, isto é possível pela capacidade existente nos ANN para aprenderem, podendo assim ser treinados para identificar padrões automáticos através da utilização de dados transferidos ao longo da rede. Podem também ser utilizados para avaliar a informação passada no cabeçalho existente nos pacotes de dados da rede, permitindo assim uma aprendizagem dos padrões de forma completa [1].

Os Intrusion Detection Prevention Systems (IDPS) tanto são configurações de software como de hardware, são utilizados para proteger um sistema ou redes inteiras e já demonstraram ao longo do tempo ser uma ferramenta útil para a Cibersegurança [1].

Os IDPS têm dois objetivos principais, a abordagem de mau uso e a abordagem de deteção de anomalias. A abordagem de mau uso identifica atividade maliciosa através da definição de padrões de comportamento incomum da rede e/ou do sistema. A abordagem de deteção de anomalias é baseada na definição de padrões de comportamento comum da rede e/ou do sistema [1][2].

As características de um IDPS permitem uma otimização de desempenho, máxima proteção e minimização do erro. No entanto, têm a sua capacidade limitada no que toca à deteção de atividade maliciosa, demonstram uma falha ao nível de escalabilidade, resiliência e automatismo [1].

Quando os ANN são integrados nos IDPS, se for detetada alguma irregularidade na informação a mesma é classificada como maliciosa e é rejeitada. Enquanto os IDPS trabalham essencialmente contra intrusões conhecidas, a abordagem dos ANN está protegida contra instâncias de intrusões que são desconhecidas até então [1].

Esta abordagem, protege com sucesso o sistema contra possíveis intrusões, provando assim a sua viabilidade, na medida em que torna o IDPS mais robusto, adaptável e preciso. A utilização de ANN nos IDPS não limita a sua utilização e pode ser inclusive utilizada em qualquer sistema que monitoriza atividades de redes, além de permitir que sejam detetados ataques em tempo real [1][2].

Artigos utilizados para esta relação do referencial: [1] (Wirkuttis & Klein, 2017), [2] (Kanimozhi & Jacob, 2019)

(4) Machine Learning e Deteção de Intrusão:

Devido ao largo volume de dados e informação, é importante garantir a segurança da rede, como tal os Intrusion Detection System (IDS) são uma mais-valia a explorar [19].

As deteções de intrusões têm como principal objetivo evitar intrusões, controlam os processos predominantes num sistema ou rede e analisam-nos de forma a detetar qualquer anomalia, tal como possíveis incumprimentos de políticas de segurança. Existem dois métodos diferentes de deteção de intrusões: má utilização e anomalia, a má utilização tem como objetivo determinar assinaturas de ataques nos recursos controlados, enquanto a anomalia depende do conhecimento do comportamento normal, permitindo a verificação de possíveis desvios [11][19][25].

Os algoritmos de Machine Learning focam-se essencialmente na classificação e regressão, podem ser utilizados para a deteção de anomalias, uma vez que são preparados e posteriormente aplicados em parâmetros de entrada usualmente invisíveis pelo processo normal de deteção de intrusões. O Machine Learning engloba algoritmos, como o Support Vector Machine (SVM) e o K-Nearest Neighbor (KNN) [11][19][22][25][27].

O SVM é um algoritmo robusto, preciso e poderoso, incluí o Support Vector Classification (SVC) e o Support Vector Regression (SVR). O SVM é baseado no conceito de limites de decisão, no qual as instâncias têm diferentes valores nas classes de dois grupos. No processo de classificação, os vetores mapeados na entrada coincidem com uma das classes e as posições com outra. Em situações em que ambos os pontos não são separáveis, então o SVM vai utilizar funções apropriadas para realizar o mapeamento. O K-Nearest Neighbor é um algoritmo de aprendizagem preguiçoso, isto é, não paramétrico, que prevê a classificação e medição de uma nova amostra, utilizando um conjunto de dados no qual cada ponto é separado em diversas classes, é maioritariamente utilizado para classificações e regressões, sendo que o seu classificador é baseado na função de distância que mede a diferença entre duas instâncias [22][27].

Utilizando os algoritmos de Machine Learning e começando com uma pré-preparação dos dados, seguido de uma conversão e otimização conseguimos classificar se existe ou não uma possível intrusão [22][27].

Artigos utilizados para esta relação do referencial: [11] (Anitha, Girish, & Kumari),[19] (Biswas, 2018), [22] (B & S, 2019), [27] (Deep & Goyal, 2018), [25] (Faisal, 2018)

(5) Natural Language Processing e Deteção de Intrusão:

O Natural Language Processing (NLP) captura a relação entre as palavras e a forma como foram sentenciadas, tem vários métodos que quando aplicados ajudam a derivar os conjuntos de recursos dos dados de texto não estruturados. Para usar as técnicas do NLP na deteção informação de segurança específica dentro de documentos de texto completos é necessário analisar o esquema. As técnicas são utilizadas por pesquisadores para analisar artefactos no desenvolvimento de software, focando-se no compromisso das mensagens e no defeito dos relatórios [40][41].

Utilizando esta informação conseguimos criar medidas relativamente ao processo do software e a sua qualidade. Podemos ainda automatizar o processo de extrair a semântica da

informação significativa existente no sistema de rastreio de questões, tendo a informação uma elevada precisão e alta recordação [40][41].
Artigos utilizados para esta relação do referencial:[40] (Omolara & Jantan, 2019),[41] (Cois & Kazman, 2015)

(6) Machine Vision e Deteção de Intrusão:
O reconhecimento facial é uma característica do Machine Vision, é uma aplicação computacional para identificar automaticamente ou verificar uma pessoa através de uma imagem digital, a tecnologia de reconhecimento facial analisa o formato único, o padrão e a posição dos recursos faciais [44].
É utilizado para iniciar sessão em alguns sistemas e garantir a autenticidade da pessoa que executou a ação [44].
Os sistemas foram criados para desenvolverem a deteção facial e o rastreio. Assim, é necessário utilizar mapeamentos para as características comportamentais e relacionar com as características biométricas e fisiológicas. Os aspetos comportamentais do sistema relacionam a atitude entre as diferentes expressões enquanto propriedade base. A expressão facial baseada no reconhecimento facial é mais eficiente com algoritmos genéticos invariantes, permitindo um elevado rácio de reconhecimento evitando assim falsos acessos [42][44].
Artigos utilizados para esta relação do referencial: [42] (Anitha, Rao, Rajasekhar, & Krishna, 2017),[44] (Mahendran & Karthik, 2016)

(7) Machine Learning e Malware:
O Malware é considerado qualquer software com intenções maliciosas, isto é, todo o software com capacidade para explorar vulnerabilidades do sistema operativo e de aplicações computacionais. Pode ser usado para obter controlo e para bloquear funcionalidades, por exemplo. Apresenta um crescimento contínuo e implementa diversos métodos de ameaça aos utilizadores, além disso, tem se tornado cada vez mais complexo e imprevisível [3][18][25].
O Malware pode espalhar-se rapidamente pelas redes sem qualquer intervenção do utilizador, os programadores de Malware são capazes de gerar novas versões, utilizando *kits* disponíveis na internet, contudo, atacando apenas um utilizador é possível propagar por outros utilizadores. As diferentes variantes contêm funcionalidades básicas e semelhantes herdadas de versões anteriores [4][16][18]. A deteção de Malware é um problema pelo facto do Malware estar sempre em constante evolução, conseguindo desviar-se de possíveis atenções, assim sendo, é necessário desenvolver novas técnicas de forma a prevenir novas categorias de Malware [16][18].
As técnicas de Machine Learning desempenham um papel fundamental no desenvolvimento de sistemas inteligentes, pois conseguem distinguir entre o que é definido como malicioso e o que é benigno [3][4][5][36].
A abordagem da Machine Learning é aplicada de forma efetiva em situações de Malware, nomeadamente na sua deteção e identificação, ajudando assim na diminuição dos ataques por Malware de que os sistemas são alvo [3][10][18]. A utilização de Machine Learning geralmente foca-se em encontrar possíveis ligações em dados observados e minimização de relações, como tal, provou-se capaz de detetar possíveis variações de Malware, além de que

consegue compreender o comportamento do Malware e analisar como o mesmo evolui [5][10][16][25].

Artigos utilizados para esta relação do referencial: [3] (Qamar, Karim, & Chang, 2019), [4] (Merrill & Capstone, 2018) ,[5] (Veiga, 2018),[10] (Lazic, 2019),[16] (Khammas, 2018), [18] (Khammas, 2018), [25] (Faisal, 2018),[36] (Nangia, Malik, Chahal, & Kharb, 2018)

(8) Automatização/Robótica e Malware:

O Malware é considerado qualquer software que atinge o objetivo de um atacante, normalmente com o intuito de causar dano a quem sofreu o ataque [45].

Sistemas autónomos propagam não só no mundo virtual ou em filmes de ficção científica, mas também no mundo real [47][48].

Os sistemas têm 3 vulnerabilidades básicas, disponibilidade, confidencialidade e integridade. Como tal, surgiu a necessidade de utilizar *robots* para combater estas vulnerabilidades, os *robots* disponibilizam informação para os utilizadores, bem como a representação do seu comportamento de forma a completar as tarefas necessárias [47][48].

A utilização da Robótica e Automação para evitar Malware tem a utilidade de combater comportamento malicioso e evitar manipulações que possam ocorrer no sistema [47].

Artigos utilizados para esta relação do referencial: [45] (Carabantes, Huidobro, & Vidal, 2016),[47] (Lera, et al., 2016), [48] (Khalid, et al., 2018)

(9) Deep Learning e Ransomware:

O Ransomware consiste no congelamento do sistema da vítima, usualmente os ataques têm início na encriptação de ficheiros importantes, como fotografias e documentos. Após a realização do ataque é feito um pedido de resgate para que a vítima consiga recuperar o que lhe roubaram. Um ataque de Ransomware pode ocorrer de várias formas, através de downloads em páginas maliciosas, anexos em e-mails de spam, entre outros. A partir do momento em que se acede ao sistema do utilizador, o Ransomware começa a agir, conseguindo obter informação acerca do utilizador, bem como informação sobre o sistema a que acedeu [24][32].

Se um Ransomware apresentar uma mudança de comportamento ou se forem utilizados pacotes para o camuflar, então não é mais visto até ocorrer uma nova atualização de software. Apenas a utilização de software de antivírus não é suficiente para defender o sistema de novos ataques, além de que não protegem o utilizador de um possível ataque, pois não conseguem determinar o momento em que o ataque tem início [24].

O Deep Learning tem sido usado na deteção de Ransomware, pois a sua técnica permite que se aprenda a representação abstrata de dados, como imagens e discursos, é também eficiente. Como tal, de forma intuitiva e prática, é possível criar um classificador que permite detetar possíveis ataques de Ransomware. No entanto, também pode ser efetuada uma ligação à informação, permitindo extrair uma melhor representação dos dados [24][32].

A utilização de técnicas de Deep Learning na deteção de ataques de Ransomware mostra-se uma mais-valia, uma vez que é possível detetar através da análise dos dados a viabilidade de determinado objeto que pretende aceder ao sistema [32].

Artigos utilizados para esta relação do referencial: [24] (Tseng, Chen, Kao, & Lin, 2017), [32] (Bibi, Akhunzada, Malik, Ahmed, & Raza, 2019)

(10)Machine Learning e Ransomware:

O Ransomware impede as vítimas de aceder aos próprios dados até que seja efetuado o pagamento do resgate pedido, como tal, tem uma implicação financeira, acabando por promover um ecossistema de cibercrime. Ransomware as a Service (RaaS) é um serviço que permite a aquisição fácil de códigos de Ransomware por determinado preço (uma compra ou o lucro da partilha de um programa). Ataques de Ransomware têm vários passos, começa pela criação do Ransomware, seguido da distribuição ou disseminação no sistema da vítima, após o acesso ao sistema é efetuado um controlo. Depois de se realizar o controlo, é efetuada uma pesquisa pelos ficheiros importantes, após os mesmos serem encontrados são encriptados e é realizado o processo do pedido de resgate [23].

O Machine Learning (ML) envolve a aprendizagem de padrões em dados de forma a criar um modelo, este irá permitir prever o resultado quando adicionados novos dados. Os algoritmos de ML são vantajosos uma vez que conseguem prever de forma precisa o resultado com dados de treino adequados, estes dados devem variar com distribuição balanceada de resultados a ser previstos [23].

A deteção de Ransomware utilizando ML é possível com um modelo híbrido de regressão e um algoritmo baseado em regras. O algoritmo de regressão permite construir o modelo de previsão baseado na relação entre prognósticos e resultados, enquanto o algoritmo baseado em regras irá gerar um conjunto de regras para o modelo de previsão. Esta combinação permite a criação de um modelo robusto que segue um conjunto de regras, que serão fortalecidas com a formulação do relacionamento com o algoritmo de regressão [23].

Artigos utilizados para esta relação do referencial: [23] (Kok, Abdullah, Jhanjhi, & Supramaniam, 2019)

(11)Natural Language Processing e Ransomware:

O Ransomware utiliza algoritmos criptográficos para encriptar e bloquear o sistema, comunica com o comando e com o servidor de controlo. Forçando os utilizadores a pagar para ter de volta os ficheiros originais, contudo, a recuperação nem sempre é garantida. O Ransomware tenta tirar vantagens das vulnerabilidades existentes no sistema [37].

O Natural Language Processing tem vários modelos que se têm provado úteis para controlar recomendações do sistema, tal como, classificação de texto, reconhecimento de discurso [37].

A utilização de Natural Language Processing para combater ataques de Ransomware mostra-se eficiente, pois permite que sejam verificados comportamentos diferentes no sistema e ajuda no seu combate [37].

Artigos utilizados para esta relação do referencial: [37] (Poudyal, Dasgupta, Akhtar, & Gupta, 2019)

(12)Machine Learning e Denial of Service:

Ataques de Distributed Denial of Service (DDoS) são lançados através de computadores geridos remotamente, bem organizados e amplamente distribuídos, têm como objetivo deixar exaustos recursos da internet, de forma a deixar determinados serviços inacessíveis, fazendo os utilizadores enviarem um largo número de pedidos inválidos. Os Ataques de DDoS podem ser definidos como um ataque de múltiplas camadas, pois operam sobre mecanismos na camada de rede e na camada de dados [25][35].

Os *botnets* são utilizados para lançar ataques de Denial Of Service (DoS), um *botnet* é uma rede de computadores e outros dispositivos, normalmente referidos como *bots* que se encontram ligados entre si através de um Malware. A deteção de *botnets* é alcançada através da configuração de *honeypots*, estes são mecanismos que contêm dados e são parte de uma rede, contudo, são sistemas isolados desenhados para detetar tentativas de intrusão na rede. Além disso, são desenvolvidas ainda assinaturas específicas para os diferentes *botnets*, para efetuar uma defesa em caso de ocorrência de ataques futuros do mesmo tipo [21].

As abordagens com Inteligência Artificial têm aumentado de forma bem-sucedida para a defesa perante ataques de DDoS, utilizando as principais técnicas de Machine Learning como a teoria de decisão Bayesian, Random Forest, tecnologias multivariadas, discriminação linear [21][35].

As Neural Networks (NN) têm utilidade na aprendizagem paralela e na tomada de decisão, são conhecidas pela velocidade da operação. São utilizadas para a aprendizagem de reconhecimento de padrões, além de que suportam tanto instalação de hardware quanto de software. Maioritariamente utilizadas para executar a deteção e prevenção de intrusões [9].

Os Expert Systems estão presentes em diferentes formulários de pequenos sistemas para resolver problemas complexos, tornando o sistema poderoso. Inclui o conhecimento básico, na qual o conhecimento especializado é armazenado num domínio aplicacional particular. São úteis para definir quais os recursos a utilizar, quando os mesmos devem ser limitados e é poderoso na deteção de intrusões na rede [9].

O Machine Learning envolve estratégias computacionais para procurar por novo conhecimento e aptidões, também apresenta melhores abordagens para compor conhecimento já existente. A variação do problema de aprendizagem depende da sua complexidade, que vai desde a aprendizagem paramétrica até aos formulários de aprendizagem simbólica [9][21][25].

A grande vantagem da utilização do algoritmo de Machine Learning na deteção de ataques de DDoS centra-se na flexibilidade existente no Machine Learning, alteram o seu desempenho com base nos dados recolhidos [21][35].

Artigos utilizados para esta relação do referencial: [7] (Calderon, 2019), [9] (Anwar & Hassan, 2017), [21] (Ding & Bunn, 2017), [25] (Faisal, 2018), [35] (Topalova, Radoyska, & Sokolov, 2020)

(13)Fuzzy Logic e Denial of Service:

Um ataque Denial of Service (DoS) é um ataque malicioso que tem como objetivo não disponibilizar os recursos de rede para aplicações legítimas, utilizando as suas vulnerabilidades [50].

Os sistemas Fuzzy cujo objetivo se centra numa aproximação e não numa solução exata. Um sistema Neuro-Fuzzy produz um alcance de valor verdadeiro entre 0 e 1, deixando o veredito final para os profissionais de rede. As funções Fuzzy podem ser utilizados para explicar o resultado, o valor 0 demonstra a normalidade e o 1 demonstra ser efetivamente um ataque de DoS [50].

Assim, é possível verificar que utilizando sistemas Fuzzy conseguimos de forma eficiente detetar possíveis ataques de DoS [50].

Artigos utilizados para esta relação do referencial: [50] (Abdullah, 2020)

(14)Deep Learning e Phishing:

O Phishing é uma forma de ciberataque que utiliza sites contrafeitos para roubar informações sensíveis aos utilizadores, como o número do cartão de crédito, dados para início de sessão, entre outros. Apresenta um elevado risco para instituições que gerem o seu negócio online, pois diminuem a confiança do consumidor, além do elevado dano financeiro que ocorre com a realização de um ataque de Phishing [28][33].

Um ataque de Phishing pode ser disseminado através da utilização da imagem de uma empresa, fazendo parecer que é algo real, com o objetivo de adquirir credenciais de início de sessão ou pode ser um ficheiro enviado em anexo que o utilizador descarrega e ataca o sistema, permitindo assim roubar qualquer informação do utilizador [28].

Deep Neural Network (DNN) é uma técnica de Deep Learning que consegue aprender sobre recursos com múltiplos níveis de abstração, o que permite ao sistema aprender funções complexas de mapeamento, para através dos dados da entrada obter os dados de saída, diretamente dos dados, sem depender de recursos humanos. Este algoritmo suporta um largo volume de dados. É utilizado para classificar os URL enquanto URL's de Phishing ou legítimos [28][33].

A utilização do Deep Learning para detetar e combater possíveis ataques de Phishing é eficiente pela sua elevada precisão, é ainda possível utilizar diferentes algoritmos de Deep Learning para tornarem a defesa do sistema mais forte, visto que os algoritmos se complementam [28][33].

Artigos utilizados para esta relação do referencial: [28] (Sumathi & Sujatha, 2019), [33] (Vrejoiu, 2019)

(15)Machine Learning e Phishing:

O Phishing é um processo fraudulento que ocorre quando um site malicioso age como sendo o site oficial, permitindo assim roubar dados confidenciais e informação do utilizador que acede ao site, tal como palavras-passe ou informações bancárias [14][15][31].

Faz utilização de mensagens falsas que apresentam uma aparência verdadeira e honesta, levando a que a pessoa que acede ao conteúdo acredite na sua viabilidade e veracidade, por isso é uma área de enorme preocupação ao nível da Cibersegurança, principalmente pelo largo volume de dados a que acedem e ao dinheiro que podem roubar [15].

Um URL de Phishing é criado com uma intenção maliciosa, quem efetuar o download do Malware pode realizar um ataque de Phishing ou manipular os resultados de pesquisa [12].

Foi proposta uma abordagem, com utilização de Machine Learning, para a classificação de URL, definindo os que são Phishing e os que não são, tendo como base detalhes do URL. Os URL de Phishing são classificados como positivos e os de não Phishing como negativos [12][14].

Para esta classificação são utilizados algoritmos de Machine Learning: Random Forest, Content-based, Naive Bayes, J48, Support Vector Machine, Neural Net e Probabilistic Neural Network [12][15][31].

O Random Forest é um dos algoritmos mais eficientes de Machine Learning para construir protótipos com dados de treino, que consiste em pares de valores de recursos e referências a classes. O algoritmo de Content-based foca-se na importância de distinguir quais os sites são Phishing e os que não são [12]. Naive Bayes é um algoritmo de probabilidades, usado para

classificações pela sua simplicidade, além disso, o volume de dados que necessita é menor. O J48 é um algoritmo que constrói árvores de decisão para um conjunto de dados, olha para os resultados que são devolvidos por uma parte dos dados, lida tanto com atributos contínuos como discretos, preparando os dados com estimativas sobre propriedades perdidas, bem como qualidades com despesas variadas. Support Vector Machine é um algoritmo baseado no conceito de planos de decisão, no qual separa um conjunto de objetos com diferentes classes de membro. Neural Net é um algoritmo que é organizado como um arranjo de unidades interligadas, estas interligações são utilizadas para enviar sinais de um neurónio para o seguinte, tendo pesos para facilitar o transporte entre neurónios, tem assim uma capacidade elevada para aprender mapeamentos complexos. [15]. O Probabilistic Neural Network é uma Rede Neural que funciona de forma semelhante à rede Bayesian e à análise discriminante do *kernel* Fisher [31].

Após a análise ao nível de precisão de ambos os algoritmos, foi verificado que o Random Forest é o mais eficaz na deteção de URL's de Phishing, a amostra que foi utilizada para esta conclusão foi de 11055 websites, tendo cada amostra 31 atributos.

Artigos utilizados para esta relação do referencial: [12] (Sananse & Sarode, 2015), [14] (Solanki & Vaishnav, 2016), [15] (Islam & Chowdhury, 2016), [29] (Unnithan, NB, S, R, & KP, 2018), [31] (Goyal & Bansal, 2017)

(16) Natural Language Processing e Phishing:

O Phishing é um esforço desonesto para adquirir informação sensível e crucial, tal como utilizadores pessoais, palavras-passe e detalhes do cartão do crédito e, por vezes, dinheiro. A razão principal para o acontecimento de fraudes de Phishing é a falta de consciência dos utilizadores, assim não é fácil prevenir a larga perda de recursos como dados, informação sensível e dinheiro. Tem ao longo dos anos mostrado ser um ataque efetivo e perigoso [38][39].

Os Phishers, pessoas que executam um ataque de Phishing, usam diferentes técnicas para iniciar diferentes ataques de Phishing, o método principal é a utilização de e-mail, podem modificar facilmente, de forma a utilizar qualquer método para alcançar as suas vítimas [38]. As técnicas de Natural Language Processing (NLP) são aplicadas para analisar cada sentença e identificar os papéis semânticos das palavras essenciais no predicado. O algoritmo utilizado analisa a sentença no e-mail, um de cada vez e identifica se o e-mail é fiável ou não, é ainda utilizada uma verificação para os links existentes no e-mail [38].

Utilizando o NLP para evitar ataques de Phishing, podemos detetar os diferentes contextos e conteúdos num golpe de e-mail padrão, o qual ajuda facilmente na identificação. No decorrer dos anos, os golpes de Phishing têm aumentado, tornou-se assim fundamental encontrar métodos que ajudem na sua deteção [38][39].

Artigos utilizados para esta relação do referencial: [38] (Mukherjee, Agarwal, & Gupta, 2019),[39] (Kim, Song, Kim, & Park, 2018)

(17) Fuzzy Logic e Phishing:

O Phishing é um novo crime que tem como objetivo roubar informação aos utilizadores, utilizando páginas de Web falsas, a maioria dessas páginas parecem-se com as páginas reais em questão de interface e de URL [49].

Um sistema de inferência Fuzzy incluí a regra de IF-THEN do sistema Fuzzy na forma de obter conhecimentos de especialistas, permitindo lidar com problemas imprecisos e vagos. Tem sido largamente utilizado em várias aplicações para otimização, controlo e identificação de sistema [49]. Sistemas Fuzzy por norma não aprendem nem se ajustam a eles próprios, como tal, surgiu a ideia de aos sistemas Fuzzy aplicarem o comportamento das Redes Neurais, uma vez que estes conseguem aprender o ambiente, a auto-organização e adaptar-se. Assim surgiu o sistema Neuro-Fuzzy [49].
A partir do sistema Neuro-Fuzzy foi desenvolvido um modelo que é constituído por quatro diferentes camadas, sendo elas, a primeira, que é considerada a camada de entrada e contem 6 nós. A segunda camada tem 12 nós, no qual o valor de cada nó é um valor fuzzy e é calculado através de uma função. A terceira camada contém 2 nós, o que representa o sim e o que representa não. A última camada é a camada de saída. [49].
O modelo referido anteriormente pode ser aplicado ao paradigma do Phishing, permitindo identificar se o ficheiro ou endereço acedido é Phishing, ou se é legitimo. O sim indica ser Phishing e o não indica que não é Phishing, permitindo obter assim um rácio de deteção de ataques de Phishing elevado [49].
Artigos utilizados para esta relação do referencial: [49] (Nguyen, Nguyen, & To, 2016)

## 4.3. VALIDAÇÃO

Após a elaboração do referencial, de forma a proceder à sua validação foram realizadas quatro entrevistas a pessoas com experiência na área da Cibersegurança e da Inteligência Artificial.

Foram colocadas três perguntas a cada entrevistado:

1. O referencial proposto é útil para ajudar a combater possíveis ataques cibernéticos?
2. Concorda com o referencial? Tem algum complemento a acrescentar?
3. Existe algum ponto de melhoria?

Abaixo foram transcritas as respostas dos entrevistados agregadas por pergunta. Em anexo encontram-se as transcrições completas de todas entrevistas com o detalhe das conversas realizadas.

**Entrevista ao Professor José Barateiro:**

Pergunta: O referencial proposto é útil para ajudar a combater possíveis ataques cibernéticos?

Resposta: Imaginemos que eu estava responsável pela segurança de uma determinada organização, ia usar o mapa e no que é que ele me poderia ajudar. É uma boa pergunta, que não sei bem como responder. Isto porque, vai um bocado abaixo do que normalmente é um nível de decisão, quem tipicamente está com esta responsabilidade sabe o tipo de riscos que corre, e os sistemas de IDS e IPS já usam IA para saber o tipo de ataque. Assim, as técnicas usadas não serão um problema. Já existem no mercado ofertas que permitem garantir a segurança, e na ótica do pessoal de segurança irão aplicar esse software. Uma organização, na área do banco ou seguros, eles não vão desenvolver

ML para sistemas de intrusão, eles vão comprar algo já pronto. Daí atualmente já se encontrarem ofertas disponíveis online e atualizadas constantemente, sempre a correr os algoritmos e a alimentar dados, para garantir que qualquer evolução é detetada. Se o meu chapéu na organização fosse esse, não precisaria de ir a esse nível de detalhe, ou seja, sabia que teria de ter um bom sistema de deteção de intrusão. Se eu me abstrair das soluções já existentes, tendo de desenvolver algoritmos, de certa forma, sim ajuda, porque me permite saber quais as técnicas de IA que posso usar para cada situação de segurança.

Pergunta: Concorda com o referencial? Tem algum complemento a acrescentar?

Resposta: Sim, o referencial poderia ser usado, se eu pretender criar uma solução de raiz, devido a possibilitar cruzar as soluções de IA com os possíveis ataques. Faz-me sentido as relações apresentadas. Em termos de comentário geral, a matriz em si, faz-me sentido.

Pergunta: Existe algum ponto de melhoria?

Resposta: Alterar as áreas no referencial nunca é uma solução, porque existem várias áreas, desde que as que apresente tenham justificação não me parece mal, porque a Cibersegurança é uma área que nunca mais acaba. Outra coisa, é pensar nas vulnerabilidades existentes, podem estar em diferentes níveis, não só ao nível de software e hardware, mas também ao nível de pessoas e essas ainda não as conseguimos controlar. No entanto, há aí um que é demasiado evidente, que é a ligação dos AIS com a deteção de intrusões, que são considerados uma subparte dos sistemas de Intrusão. Vendo isso, como muito tradicional, tem Phishing, Malware, eu adicionaria spoofing.

**Entrevista a Daniel Soares:**

Pergunta: O referencial proposto é útil para ajudar a combater possíveis ataques cibernéticos?

Resposta: Eu acho que com este referencial se consegue dar uma direção de qual o caminho a seguir. Acho que sim que é útil porque de uma forma objetiva ajuda a encaminhar o decisor**.**

Pergunta: Concorda com o referencial? Tem algum complemento a acrescentar?

Resposta: Sim, acho que tendo exemplos práticos poderia ser mais útil. Também irá ser útil para uma pessoa que não esteja tão dentro do departamento de TI. Com este referencial são apresentados vários pontos de decisão, está a ser facilitada essa decisão ao utilizador. O que quis dizer é que um diretor de Segurança pode usar o referencial como informação técnica para ele próprio usar, mas também pode ser útil para ele apresentar a outro departamento, como o de Gestão, para o mesmo validar e poder dar uma aprovação na implementação da solução. Com o referencial, é mais fácil fundamentar o porquê da decisão tomada.

Pergunta: Existe algum ponto de melhoria?

Resposta: Apenas a questão do Machine Learning, de acrescentar ou outros, ou um asterisco, para facilitar o processo de análise. Relativamente às áreas consideradas, parecem-me estar as fundamentais e principais.

**Entrevista a Anónimo:**

Pergunta: O referencial proposto é útil para ajudar a combater possíveis ataques cibernéticos?

Resposta: Sim, penso que sim. Se for alguém dentro da organização que sabe à partida qual a fragilidade existente, conseguirá com base nisso cruzar com as técnicas de Inteligência Artificial.

Pergunta: Concorda com o referencial? Tem algum complemento a acrescentar?

Resposta: Sim, concordo com o referencial. Contudo, não aborda todos os tipos de ciberataques existentes, existindo ainda mais tipos de ataques e que começam a aparecer mais tipos de ciberataques. Por exemplo, ataques por Data Manipulation. É um tipo de cibercrime em crescimento, contudo, poderá ser interessante falar sobre ele. Este tipo de ataque centra-se na falsificação de documentos oficiais ou na alteração dos dados de uma organização, onde a informação é depois tornada pública pelo atacante e tem a intenção de destabilizar e denegrir a imagem das organizações.

Pergunta: Existe algum ponto de melhoria?

Resposta: Penso que tirando o que disse na pergunta anterior, relativamente à inclusão de outras áreas da Cibersegurança, não vejo como se pode efetuar alterações, acredito que será um modelo que deverá estar em constante evolução devido à movimentação que existe ao nível tecnológico. Vão sempre surgindo novas informações que poderão ser úteis para garantir uma maior viabilidade. Tirando isso, penso que será útil para um técnico de TI ter um apoio quando for tornar o seu serviço mais seguro.

**Entrevista a Anónimo:**

Pergunta: O referencial proposto é útil para ajudar a combater possíveis ataques cibernéticos?

Resposta: Penso que será útil sim. Contudo, terei de analisar melhor.

Pergunta: Concorda com o referencial? Tem algum complemento a acrescentar?

Resposta: Concordo, no entanto, precisarei de ler e analisar com um maior foco.

Pergunta: Existe algum ponto de melhoria?

Resposta: As áreas de segurança abordadas são efetivamente as mais importantes atualmente, os ataques de DoS estão na ordem do dia e são dos que mais ocorrem nos dias de hoje, os ataques de Malware incluem os vírus e estão também na ordem do dia. Os ataques de Ransomware estão também na ordem do dia e são os mais frequentes de acontecer em Portugal, foram inclusive ataques desses efetuados nos sistemas hospitalares e os mesmos foram resolvidos sem a utilização de Inteligência Artificial e de uma forma bastante rudimentar, tendo sido apenas reposto os backups anteriores. Portanto, considero uma boa abordagem e que as 5 áreas da Cibersegurança são as de importante foco, visto que as outras áreas também existentes não têm vulnerabilidades detetadas atualmente, portanto, não é necessário a existência de um foco neste momento.

## 4.4. DISCUSSÃO

Após a realização das entrevistas foi efetuada uma análise às respostas dadas pelos entrevistados relativamente ao referencial que lhes foi apresentado. As entrevistas ajudaram a verificar a utilidade do referencial, contando com a perspetiva de participantes com conhecimento nas áreas de estudo deste projeto. Os entrevistados foram participativos, apresentaram de forma clara a sua opinião, deram soluções e oportunidades de melhoria.

Foram apresentados alguns pontos fortes, nomeadamente, o facto de poder ajudar as organizações a encontrarem uma solução de implementação para conseguirem combater os ataques cibernéticos e as suas vulnerabilidades, tornando os seus sistemas mais fortes. Além disso, foi referida a utilidade de se poder apresentar o referencial a áreas da organização que não pertencem ao departamento de Tecnologias de Informação, pois poderão ter alguma dificuldade na compreensão de uma linguagem técnica no momento de apresentação da solução que se pretende implementar, tendo um descritivo funcional conseguem ter uma melhor perceção.

No entanto, foram referidos alguns pontos que poderão ser considerados menos positivos, particularmente, o facto do modelo não abordar mais áreas da Cibersegurança, bem como o facto do modelo poder sofrer alterações ao longo do tempo devido ao aparecimento de novas áreas da Cibersegurança. Referiu-se ainda que já existem modelos fabricados no mercado para combater possíveis ciberataques e que estes não exigem investimento por parte da organização. Todavia, o objetivo do referencial é ajudar as organizações a criarem algo à sua medida, possuindo um Sistema de Informação único e com uma capacidade de defesa específica para as suas vulnerabilidades.

Além disso, foram indicados pontos de melhoria, na qual, foi sugerida a alteração da legenda nas diferentes áreas do Machine Learning, onde se apresenta novamente Machine Learning deveria ser substituído por outros. No entanto, como são pontos que se centram, essencialmente, no Machine Learning, apesar de referir diferentes algoritmos, não poderão ser definidos enquanto subáreas, mas sim como complementos.

Foi ainda sugerido, como ponto de melhoria, a inclusão de novas áreas da Cibersegurança, como o *spoofing*, por ser uma área que tem estado a crescer atualmente. Porém, por ser uma área recente,

ainda não existem estudos aprofundados sobre a mesma, nem de que forma a utilização de Inteligência Artificial pode ser útil para combater este tipo de ataque.

Assim, foi possível concluir que o referencial será útil para as organizações, nomeadamente, para as ajudar a tomar decisões, pois conseguirão ter um ponto de partida e uma base estruturada para a solução que pretendem implementar. Para além disso, poderá ser útil para vários departamentos da organização, ou seja, poderá ser utilizado pelo departamento de Tecnologias de Informação na definição das características da solução a desenvolver ou pelo departamento de gestão para ter conhecimento relativamente aos gastos necessários na implementação da solução.

# 5. CONCLUSÃO

## 5.1. SÍNTESE

Com este trabalho foi possível fomentar o conhecimento quer ao nível de Cibersegurança, quer ao nível da Inteligência Artificial, através do conhecimento mais detalhado das técnicas de cada uma das áreas.

No início do presente estudo, foram definidos quais os objetivos, sendo o principal deles propor uma abordagem na qual se pudesse ajudar os Sistemas de Informação a combater possíveis ataques a nível de segurança, na qual se utilizaria Inteligência Artificial.

Assim, numa primeira fase foram estudadas as áreas da Cibersegurança e da Inteligência Artificial, permitindo assim conhecer ao detalhe as suas características e possíveis valências, bem como, os seus pontos fracos.

O foco da Cibersegurança centra-se na proteção e defesa de um sistema no ciberespaço, através da utilização de ferramentas, políticas e outras, para ajudar no combate de possíveis ameaças, contudo, nem sempre está presente o lado bom, havendo assim muitos atos ilícitos e ilegais que têm sempre uma vertente maliciosa, isto é, têm sempre o objetivo de prejudicar uma terceira parte.

Deste modo, o que se pretendeu foi encontrar formas de minimizar os danos causados com estes ataques, sendo então útil a utilização da Inteligência Artificial, esta centra-se na aprendizagem das máquinas tendo como base o comportamento cognitivo do ser humano.

Como tal, foram definidas quais as técnicas da Inteligência Artificial que apresentam uma mais-valia e quais as áreas da Cibersegurança que atualmente geram um maior problema. Foi assim, efetuada uma seleção de artigos científicos que cruzam ambas as áreas e apresentam já provas dadas da sua eficiência e eficácia. A partir dessa seleção, foi criado um referencial do que foi possível extrair da análise efetuada.

Após a construção do referencial e do detalhe relativo à forma como ambas as técnicas se relacionam e conseguem depreciar o dano causado por um ataque cibernético, o mesmo foi validado. A validação foi efetuada por 4 entrevistados, com experiência tanto na área da Segurança/Cibersegurança quanto na área da Inteligência Artificial, permitindo assim ter uma perspetiva relativamente à utilidade do referencial.

Posteriormente à validação efetuada, foi possível validar a utilidade do referencial e assim ter uma salvaguarda de que os profissionais de Tecnologias de Informação vão conseguir ter uma ajuda na construção de Sistemas de Informação mais seguros e que ajudarão na redução dos danos de ataques cibernéticos.

Posto isto, pode-se dizer que o principal objetivo deste estudo foi bem conseguido, pois foi possível apresentar e descrever algo útil para os gestores de Tecnologias de Informação, que ajudará no momento de decidirem o que pretendem para melhorar o seu Sistema de Informação.

## 5.2. LIMITAÇÕES ENCONTRADAS

Uma das limitações encontradas no processo de realização do presente estudo centrou-se no facto de apenas ter sido possível validar o referencial com 4 intervenientes, deste modo foi apenas validado de forma teórica. Posto isto, não foi possível realizar a validação prática com gestores de Tecnologias de Informação, não tendo sido assim obtida uma opinião por parte dos responsáveis e possíveis utilizadores do referencial construído.

Foi também referido pelos entrevistados a possibilidade de inclusão de novas áreas da Cibersegurança, pois existem outras ameaças para as organizações a nível de segurança que não foram referidas no presente estudo. Devido a ser uma área em constante evolução, existem ainda muitas áreas por estudar e compreender, assim o estudo foi restringido às áreas da Cibersegurança com maior conhecimento e na qual existem várias fontes de informação que permitem garantir a viabilidade do que se afirma.

Também a área da Inteligência Artificial é muito ampla, como tal, neste presente estudo foram apenas mencionadas as principais áreas e as que demonstraram maior utilidade para ajudar a combater os ataques cibernéticos.

De forma a permitir um melhor entendimento e compreensão na leitura das siglas a maioria foi mantida de acordo com o seu significado no inglês. Por exemplo, o Machine Learning poderia ter sido traduzido para aprendizagem da máquina, no entanto, poderia levar a alguma dificuldade de entendimento, assim, manteve-se em inglês, bem com o seu acrónimo. Esta lógica foi aplicada a outros acrónimos e expressões. Outras palavras, como *honeypots* que poderiam ser traduzidas para português como potes de mel, no entanto, perderiam em parte o seu significado, porque também utilizamos esta palavra enquanto estrangeirismo.

## 5.3. TRABALHO FUTURO

Numa próxima fase, pretende-se apresentar o referencial construído a gestores de Tecnologias de Informação, de forma a darem o seu *feedback* e permitindo obter a confirmação de que o referencial é útil e que ajudará as organizações a apresentarem soluções próprias para os seus Sistemas de Informação e que irá minimizar os danos causados por atividades maliciosas.

Além disso, será também efetuada uma análise às áreas da Cibersegurança e da Inteligência Artificial, para validar se faz sentido incluir novas áreas no referencial, devido à constante evolução poderão já existir estudos mais completos sobre outras áreas.

# 6. REFERÊNCIAS


Abdullah, S. (2020). A Neruo-Fuzzy System to Detect IPv6 Router Alert Option DoS Packets. *The International Arab Journal of Information Technology*.

Al-Emadi, S., Al-Mohannadi, A., & Al-Senaid, F. (2019). Using Deep Learning Techniques for Network Intrusion Detection.

Anitha, A., Girish, P., & Kumari, S. (n.d.). Cyber Defense Using Artificial Intelligence.

Anitha, P., Rao, K. N., Rajasekhar, V., & Krishna, C. H. (2017). Security for Biometrics Protection between Watermarking and Visual Cryptography.

Anwar, A., & Hassan, S. I. (2017). Applying Artificial Intelligence Techniques to Prevent Cyber Assaults. *International Journal of Computational Intelligence Research*.

Apple. (2019, Junho 7). *SIRI*. Retrieved from Apple: https://www.apple.com/siri/

Attak, H., Combalia, M., Gardikis, G., Gastón, B., Jacquin, L., Katsianis, D., . . . Segou, O. (2018). *Application of distributed computing and machine learning technologies to cybersecurity.*

Azfar, K., Kirisci, P., Khan, Z. H., Ghrairi, Z., Thoben, K.-D., & Pannek, J. (2018). Security framework for industrial collaborative robotic cyber-physical systems.

B, N. M., & S, P. B. (2019). Machine Learning and Deep Learning methods for Cybersecurity. *International Research Journal of Engineering and Technology (IRJET)*.

Bahrin, M., Othman, M., Azli, N., & Talib, M. (2016). Industry 4.0: A review on industrial automation and robotic. *Jurnal Teknologi, 78(6-13)*, 137-143. Penerbit UTM Press.

Bax, M. P. (2015). Design science: filosofia da pesquisa em ciência. *Ciência da Informação v.42 nº 2*.

Bibi, I., Akhunzada, A., Malik, J., Ahmed, G., & Raza, M. (2019). An Effective Android Ransomware Detection Through Multi-Factor Feature Filtration and Recurrent Neural Network. *2019 UK/China Emerging Technologies, UCET 2019.* Institute of Electrical and Electronics Engineers Inc.

Biswas, S. (2018). Intrusion Detection Using Machine Learning: A Comparison Study. *International Journal of Pure and Applied Mathematics, 118*(19), 101-114.

Bkassiny, M., Li, Y., & Jayaweera, S. (2013). A survey on machine-learning techniques in cognitive radios. *IEEE Communications Surveys and Tutorials, 15*(3), 1136-1159.

Boell, S., & Cecez-Kecmanovic, D. (2015). What is an information system? *Proceedings of the Annual Hawaii International Conference on System Sciences. 2015-March*, pp. 4959-4968. IEEE Computer Society.

Brynielsson, J., Franke, U., Adnan Tariq, M., & Varga, S. (2016). Using cyber defense exercises to obtain additional data for attacker profiling. *IEEE International Conference on Intelligence*

*and Security Informatics: Cybersecurity and Big Data, ISI 2016* (pp. 37-42). Institute of Electrical and Electronics Engineers Inc.

Butler, B., & Gray, P. (2006, 6). Reliability, mindfulness, and information systems. *MIS Quarterly: Management Information Systems, 30*(2), 211-224.

Calderon, R. (2019). The Benefits of Artificial Intelligence in Cybersecurity. *Economic Crime Forensics Capstones. 36.*

Carabantes, D. S., Huidobro, C. B., & Vidal, D. C. (2016). Optimation through Automation Of Malware Update Process, Capable of Evading Anti-Malware Systems. *Research in Computer Science*.

Carl, G., Kesidis, G., Brooks, R. R., & Rai, S. (2006). Denial-of-Service AttackDetection Techniques.

Chella, A., Frixione, M., & Gaglioa, S. (1995, Junho). A cognitive architecture for artificial vision. *A cognitive architecture for artificial vision*.

Chowdhury, G. (2005, 1 31). Natural language processing. *Annual Review of Information Science and Technology, 37*(1), 51-89.

Cisco. (2019, Junho 1). *What Is Cybersecurity?* Retrieved from Cisco: https://www.cisco.com/c/en/us/products/security/what-is-cybersecurity.html

Cois, C. A., & Kazman, R. (2015). Natural Language Processing to Quantify Security Effort in the Software Development Lifecycle.

Costa, E., & Simões, A. (2011). *Inteligência Artificial Fundamentos e Aplicações (3º Edição).* Lisboa: FCA.

Daisheng, L. (1998). *Pattern Recognition and Image Processing.* Woodhead Publishing.

Daniel M. A da Silva, D. F. (2016). Abordagem utilizando o Design Science Research para o Desenvolvimento de Sistema Colaborativo Assistivo. *Revista de Informática Aplicada*, 75.

Deep, R., & Goyal, V. (2018). Machine Learning Methods for Cyber Security. *Pramana Research Journal, Volume 8*.

Dehariya, P. (2016). An Artificial Immune System and Neural Network to I mprove the Detection Rate in Intrusion Detection System. 4.

Desouza, K., Dawson, G., & Chenok, D. (2019). Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector. *Business Horizons*.

Ding, S., & Bunn, J. (2017). *Machine Learning for Cybersecurity: Network-based Botnet Detection Using Time-Limited Flows.*

E., B., & K., T. (2015, 8 18). Phishing URL Detection: A Machine Learning and Web Mining-based Approach. *International Journal of Computer Applications, 123*(13), 46-50.

Estrela, S. C. (2014). A gestão da informação na tomada de decisão das PME da região centro : um estudo exploratório e de multicasos no âmbito da Ciência da Informação. *http://hdl.handle.net/10316/25956*.

Faisal, I. (2018). *From Artificial Intelligence to Security: Back and Forth.*

Fang, Z., Zhao, X., Wei, Q., Chen, G., Zhang, Y., Xing, C., . . . Chen, H. (2016). Exploring key hackers and cybersecurity threats in Chinese hacker communities. *IEEE International Conference on Intelligence and Security Informatics: Cybersecurity and Big Data, ISI 2016* (pp. 13-18). Institute of Electrical and Electronics Engineers Inc.

Flores-Fuentes, W., Rodríguez-Quiñonez, J., Hernandez-Balbuena, D., Rivas-López, M., Sergiyenko, O., Gonzalez-Navarro, F., & Rivera-Castillo, J. (2014). Machine vision supported by artificial intelligence. *IEEE International Symposium on Industrial Electronics* (pp. 1949-1954). Institute of Electrical and Electronics Engineers Inc.

Gourisetti, S., Mylrea, M., & Patangia, H. (2020, 4 1). Cybersecurity vulnerability mitigation framework through empirical paradigm: Enhanced prioritized gap analysis. *Future Generation Computer Systems, 105*, 410-431.

Goyal, B., & Bansal, M. (2017). A Competent Approach for Type of Phishing Attack Detection Using Multi-Layer Neural Network. *International Journal of Advanced Engineering Research and Science, 4*(1), 210-215.

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly Vol. 37 No. 2*, 337-355.

Halfond, W. G., Viegas, J., & Orso, A. (2006). A Classification of SQL Injection Attacks.

Haq, N., Onik, A., Avishek, M., Hridoy, K., Rafni, M., Shah, F., & Farid, D. (2015). *Application of Machine Learning Approaches in Intrusion Detection System: A Survey.*

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian journal of information systems v.19 nº2*.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information System Research. *MIS Quarterly Vol. 28 No. 1*, 75-105.

Horia Vrejoiu, M. (2019). Neural Networks and Deep Learning in Cyber Security.

Igbe, O., Darwish, I., & Saadawi, T. (2016). *Distributed Network Intrusion Detection System: An Artificial Immune System Approach.*

Islam, M., & Chowdhury, N. (2016). Phishing Websites Detection Using Machine Learning Based Classification Techniques.

*ISO/IEC 27001:2013*. (2013, 10 01). Retrieved from ISO - ISO/IEC 27001:2013 - Information technology — Security techniques — Information security management systems — Requirements: https://www.iso.org/standard/54534.html

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social Phishing. *Communication of the ACM*.

Jamadar, R., Ingale, S., Panhalkar, A., Kakade, A., & Shinde, M. (2019). Survey of Deep Learning Based Intrusion Detection Systems for Cyber Security. *IJRAR19K1034 International Journal of Research and Analytical Reviews*.

Jie, Y., Choo, K., Li, M., Chen, L., & Guo, C. (2019). Tradeoff gain and loss optimization against man-in-the-middle attacks based on game theoretic model. *Future Generation Computer Systems, 101*, 169-179.

Kanimozhi, V., & Jacob, T. (2019). Artificial Intelligence based Network Intrusion Detection with hyper-parameter optimization tuning on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing. *ICT Express, 5*(3), 211-214.

Kasabov, N. (2000). *Future directions for intelligent systems and information sciences : the future of speech and image technologies, brain computers, WWW, and bioinformatics.* Physica-Verlag.

Khalid, A., Kirisci, P., Khan, Z. H., Ghrairi, Z., Thoben, K.-D., & Pannek, J. (2018). Security framework for industrial collaborative robotic cyber-physical systems.

Khammas, B. (2018, 9 1). Malware Detection using Sub-Signatures and Machine Learning Technique. *Journal of Information Security Research, 9*(3), 96.

Kim, M., Song, C., Kim, H., & Park, D. (2018). Catch me, Yes we can! - Pwning Social Engineers using Natural Language Processing Techniques in Real-Time.

Kok, S., Abdullah, A., Jhanjhi, N., & Supramaniam, M. (2019). *Ransomware, Threat and Detection Techniques: A Review*.

Kreimeyer, K., Foster, M., Pandey, A., Arya, N., Halford, G., Jones, S., . . . Botsis, T. (2017, 9 1). Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. *Journal of Biomedical Informatics, 73*, 14-29. Academic Press Inc.

Kulkarni, A., & Shivananda, A. (2019). *Natural Language Processing Recipes: Unlocking Text Data with Machine.*

Landage, J., & Wankhade, M. P. (2013). Malware and Malware Detection Techniques: A Survey. *International Journal of Engineering Research & Technology*.

Laudon, C. K., & Laudon, P. J. (2014). *Management Information Systems.* United States of America: Pearson.

Laurence, A. (2019, agosto 22). *The Impact of Artificial Intelligence on Cyber Security*. Retrieved from CPO Magazine: https://www.cpomagazine.com/cyber-security/the-impact-of-artificial-intelligence-on-cyber-security/

Lawrence A. Gordon, Martin P. Loeb. (2005). *Managing Cybersecurity Resources: A Cost-Benefit Analysis.* McGraw-Hill Education.

Lazic, L. (2019). Benefit From AI in Cybersecurity. *The 11th International Conference on Business Information Security, 1*(October), 1-8.

Lera, F. J., Balsa, J., Casado, F., Fernández, C., Rico, F. M., & Matellán, V. (2016). Cybersecurity in Autonomous Systems: Evaluating the performance of hardening ROS.

Li, J.-h. (2018). Cyber security meets artificial intelligence: a survey.

Loaiza, F., Birdwell, J., Kennedy, G., & Visser, D. (2019). Utility of Artificial Intelligence and Machine Learning in Cybersecurity. (June), 66.

Mahendran, R., & Karthik, C. (2016). The New Embedded ATM Security based on Machine Vision using MATLAB.

McBurney, P., & Parsons, S. (2002). Dialogue Games in Multi-Agent Systems. *Dialogue Games in Multi-Agent Systems*.

Merrill, A., & Capstone, A. (2018). *SYMBIOTIC ARTIFICIAL INTELLIGENCE AND ITS CHALLENGES IN CYBERSECURITY AND MALWARE RESEARCH.*

Miller, L. C., & CISSP. (2016). *Cybersecurity For Dummies, Palo Alto Networks 2nd Edition.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2015). Principais itens para relatar Revisões sistemáticas e Meta-análises: A recomendação PRISMA. 335-342.

Mukherjee, A., Agarwal, N., & Gupta, S. (2019). A SURVEY ON AUTOMATIC PHISHING EMAIL DETECTION USING NATURAL LANGUAGE PROCESSING TECHNIQUES. *International Research Journal of Engineering and Technology (IRJET)*.

Nangia, S., Malik, M., Chahal, D., & Kharb, L. (2018). A Futuristic Approach: Incorporating Artificial Intelligence with Cyber Security. *International Journal of Research In Engineering, Science and Management*.

Nguyen, L. A., Nguyen, H. J., & To, B. L. (2016). An Efficient Approach Based on Neuro-Fuzzy for Phishing Detection. *Journal of Automation and Control Engineering*.

Nidhin A, U., Harikrishnan, N., Akarsh, S., Vinayakumar, R., & Soman, K. (2018). Machine Learning Based Phishing E-mail detection.

Nilsson, N. J. (1980). *Principles of Artificial Intelligence.* Morgan Kaufmann Publishers, Inc.

O'Brien, J. A., & George, M. M. (2011). *Management Information Systems, 10e.* New York: McGraw-Hill Irwin.

Omolara, A. E., & Jantan, A. (2019). Modify honey encription schema for encoding natural language processing message.

Ongsulee, P. (2017). Artificial Intelligence, Machine Learning and. *2017 Fifteenth International Conference on ICT and Knowledge Engineering*.

Pal, B., Daniel, T., Chatterjee, R., & Ristenpart, T. (2019). Beyond Credential Stuffing: Password Similarity Models using Neural Networks.

Panch, T., Szolovits, P., & Atun, R. (2018, October 21). Artificial intelligence, machine learning and health systems.

Pande, J. (2017). *Introduction To Cyber Security.* Uttarakhand Open University, Haldwani: Uttarakhand Open University.

Pannu, A. (2015). Artificial Intelligence and its Application in Different Areas. *International Journal of Engineering and Innovative Technology (IJEIT)*.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 45-78.

Pope, J. (2016). Ransomware: Minimizing the Risks. *Innovation in Clinical Neuroscience*.

Poudyal, S., Dasgupta, D., Akhtar, Z., & Gupta, K. D. (2019). A Multi-Level Ransomware Detection Framework using Natural Language Processing and Machine Learning.

Qamar, A., Karim, A., & Chang, V. (2019). Mobile Malware attacks: Review, taxonomy & future directions. *Future Generation Computer Systems, 97*, 887-909.

Rainer Jr., R. K., & Cegielski, C. G. (2007). *Introduction to Information Systems.* United States of America: John Wiley & Sons, Inc.

Roman V. Yampolskiy, M. S. Spellchecker. (2016). Artificial Intelligence Safety and Cybersecurity: a Timeline of AI Failures. *Cornell University*, 12.

Russell, S. J. , & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach 3rd Edition.* Upper Saddle River, New Jersey 07458.: Pearson Education, Inc.

Sananse, B. E., & Sarode, T. K. (2015). Phishing URL Detection: A Machine Learning and Web Mining-based Approach. *International Journal of Computer Applications*.

Selma Dilek, Hüseyin Çakır, Mustafa Aydın. (2015). Applications of Artificial Intelligence Techniques to Combating Cyber Crimes: A Review. *International Journal of Artificial Intelligence & Applications (IJAIA) Vol.6 No. 1*, 19.

Shiode, N. (2000). Urban Planning, Information Technology, and Cyberspace. *Journal of Urban Technology, 7*(2), 105-126.

Solanki, J., & Vaishnav, R. (2016). Website Phishing Detection using Heuristic Based Approach. *International Research Journal of Engineering and Technology*.

Sumathi, K., & Sujatha, V. (2019, 9 1). Deep learning based-Phishing attack detection. *International Journal of Recent Technology and Engineering, 8*(3), 8428-8432.

Taddeo, M. (2019). Three Ethical Challenges of Applications of Artificial Intelligence in Cybersecurity. *Minds and Machines, 29*(2), 187-191.

Taddeo, M., McCutcheon, T., & Floridi, L. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. *Nature Machine Intelligence, 1*(12), 557-560.

Teknologi, J., Khammas, M., Monemi, A., Stephen Bassi, J., Ismail, I., Mohd Nor, S., & Marsono, M. (2015). FEATURE SELECTION AND MACHINE LEARNING CLASSIFICATION FOR MALWARE DETECTION. *77*, 2180-3722.

Topalova, I., Radoyska, P., & Sokolov, S. (2020). Neural Network Implementation for Detection of Denial of Service Attacks. *Journal of Engineering Science and Technology Review*.

Tseng, A., Chen, Y., Kao, Y., & Lin, T. (2017). *Deep Learning for Ransomware Detection.*

UK, I. G. (2019, Junho 1). *What is Cyber Security?* Retrieved from IT Governance UK: https://www.itgovernance.co.uk/what-is-cybersecurity

Unnithan, N. A., NB, H., S, A., R, V., & KP, S. (2018). Machine Learning Based Phishing E-mail detection.

Veiga, A. P. (2018). Applications of Artificial Intelligence (AI) to Network Security.

Vieira, J., Dias, F. M., & Mota, A. (2004). Neuro-Fuzzy Systems: A Survey.

Von Solms, R., & Van Niekerk, J. (2013). From information security to cyber security. *Computers and Security, 38*, 97-102.

Vrejoiu, M. H. (2019). Neural Networks and Deep Learning in Cyber Security. *Romanian Cyber Security Journal*.

Wassermann, G., & Su, Z. (2008). Static Detection of Cross-Site Scripting Vulnerabilities.

Wirkuttis, N., & Klein, H. (2017). *Artificial Intelligence in Cybersecurity.*

Wu, D., Ren, A., Zhang, W., Fan, F., Liu, P., Fu, X., & Terpenny, J. (2018). Cybersecurity for digital manufacturing. *Journal of Manufacturing Systems, 48*, 3-12.

Yadav, D., Arora, M., Tiwari, K., & Ghosh, J. (2018, 11 1). Detection and identification of camouflaged targets using hyperspectral and LiDAR data. *Defence Science Journal, 68*(6), 540-546.

Zúquete, A. (2018). *Segurança em Redes Informáticas, 5º Edição.* Lisboa: FCA.

# 7. ANEXOS

**Entrevista ao Professor José Barateiro (Completa):**

Autor: Boa tarde professor, agradeço desde já a sua presença. Podemos assim começar.

Entrevistado: Boa tarde. Sim, podemos começar

Autor: A tese centra-se basicamente na Cibersegurança e na Inteligência Artificial, em que a ideia principal foi criar um referencial onde fosse possível cruzar as técnicas de Inteligência Artificial com possíveis ataques a nível de segurança. E com essas técnicas de IA diminuir o risco de um ciberataque. Assim sendo, foi criado o referencial, onde é possível ver as técnicas de IA e as áreas de Cibersegurança. Foi efetuado um mapeamento, com base num largo volume de artigos, cerca de 50 artigos científicos.

Entrevistado: Os números existentes nos artigos são referências a artigos ou é outra coisa?

Autor: Estes números são referências aos artigos e o que foi possível extrair de cada artigo, nomeadamente características das áreas da IA e da Cibersegurança e de que forma as características de IA conseguiam combater os ciberataques. Existe ainda referência aos artigos usados em cada tópico, permitindo assim comprovar que a informação descrita teve algum fundamento.

Entrevistado: Certo.

Autor: Assim sendo, tenho algumas questões que lhe gostaria de colocar**. Olhando para este referencial, acha que poderá ser útil para as organizações?** Ajudando a combater os possíveis ameaças de segurança.

Entrevistado: Isso é uma boa pergunta. Tendo apenas o referencial como base?

Autor: Sim, tendo apenas o referencial como base.

Entrevistado: **A pergunta é: imaginemos que eu estava responsável pela segurança de uma determinada organização, ia usar o mapa e no que é que ele me poderia ajudar. É uma boa pergunta, que não sei bem como responder. Isto porque, vai um bocado abaixo do que normalmente é um nível de decisão, quem tipicamente está com esta responsabilidade sabe o tipo de riscos que corre, e os sistemas de IDS e IPS já usam IA para saber o tipo de ataque. Assim, as técnicas usadas não serão um problema.**

Autor: A ideia no caso, é aplicar as técnicas de IA, sabendo os riscos e conhecendo as vantagens saber como aplicar.

Entrevistado: Estou a entender, no caso, será fazer algo por medida, não usar software que já considere essas técnicas.

Autor: Exato, será algo bastante específico para cada Sistema de Informação.

Entrevistado: **Sim. Porque já existem no mercado ofertas que permitem garantir a segurança, e na ótica do pessoal de segurança irão aplicar esse software.**

Autor: A ideia é criar mesmo algo específico para cada organização, devido a cada organização ter as suas necessidades, garantindo uma melhor eficiência e eficácia.

Entrevistado: **Não sei se me estou a fazer entender, isto é, uma organização, na área do banco ou seguros, eles não vão desenvolver ML para sistemas de intrusão, eles vão comprar algo já pronto. Daí atualmente já se encontrarem ofertas disponíveis online e atualizadas constantemente, sempre a correr os algoritmos e a alimentar dados, para garantir que qualquer evolução é detetada. Se o meu chapéu na organização fosse esse, não precisaria de ir a esse nível de detalhe, ou seja, sabia que teria que ter um bom sistema de deteção de intrusão.**

Autor: Estou a entender, então sendo assim, se eu lhe perguntar se **concorda com o referencial e se tem alguma coisa a acrescentar, o que tem a dizer?**

Entrevistado: Essa já é uma pergunta um pouco diferente e mesmo a anterior eu posso responder de uma forma diferente, já pode efetuar as diferentes análises.

Autor: Certo.

Entrevistado**: Se eu me abstrair das soluções já existentes, tendo de desenvolver algoritmos, de certa forma, sim ajuda, porque me permite saber quais as técnicas de IA que posso usar para cada situação de segurança.** Mais umas perguntas, na deteção de intrusões, estamos a falar apenas de intrusão tecnológica ou física?

Autor: Intrusão tecnológica

Entrevistado: Não foi considerada a física?

Autor: Não foram encontrados artigos sobre intrusão física, mas é uma boa pergunta, porque também pode haver intrusão física.

Entrevistado: É engraçado como usam técnicas de visão, câmaras de controlo. Já agora, só um complemento, a Cibersegurança deve ser vista de duas formas, quando usamos meios tecnológicos para atacar algo ou quando a parte tecnológica é o alvo do ataque, ou seja, quando um servidor é roubado, é também um problema de Cibersegurança**. Sim, o referencial poderia ser usado, se eu pretender criar uma solução de raiz, devido a possibilitar cruzar as soluções de IA com os possíveis ataques. Faz-me sentido as relações apresentadas. Em termos de comentário geral, a matriz em si, faz-me sentido**. Como é que a Vânia chegou ao resultado das áreas de Cibersegurança apresentadas?

Autor: Com base na pesquisa prévia que foi realizada e que foram vistas como as áreas de maior risco.

Entrevistado: Certo, existe também a questão do Social Engineering que é uma área emergente.

Autor: Sim, é uma área que está a crescer mais atualmente.

Entrevistado: O que apresenta são técnicas mais correntes, portanto não me choca nada do que é apresentado.

Autor: Sim, a pesquisa foi efetuada tendo como base o modelo PRISMA, no qual foi sendo diminuindo o volume de artigos obtidos.

Entrevistado: E com base nisso, foram obtidas as técnicas de IA e técnicas de Cibersegurança apresentadas?

Autor: Sim, tudo com base dos artigos retornados.

Entrevistado: Certo.

Autor: Assim sendo, tenho uma última pergunta**. Considera algum ponto de melhoria?** Como estava a falar da Social Engineering.

Entrevistado: **Alterar as áreas no referencial nunca é uma solução, porque existem várias áreas, desde que as que apresente tenham justificação não me parece mal, porque a Cibersegurança é uma área que nunca mais acaba. Outra coisa, é pensar nas vulnerabilidades existentes, podem estar em diferentes níveis, não só ao nível de software e hardware, mas também ao nível de pessoas e essas ainda não as conseguimos controlar. No entanto, há aí um que é demasiado evidente, que é a ligação dos AIS com a deteção de intrusões, que são considerados uma subparte dos sistemas de Intrusão.**

Autor: E tem mais alguma coisa a acrescentar? Alguma coisa que ache relevante?

Entrevistado: **Vendo isso, como muito tradicional, tem Phishing, Malware, eu adicionaria spoofing**.

Autor: Sim, conheço essa técnica.

Entrevistado: Também pode referenciar o sniffing.

Autor: Relativamente a esse não encontrei nada.

Entrevistado: O conceito de sniffing é essencialmente como o nome indica, é ir cheirar, ir buscar pacotes que não nos pertencem. E aí também existem algumas técnicas que detetam tráfego que não é normal, permitindo detetar a existência de pessoas a fazer sniffing. Outra coisa, nos sistemas de deteção de intrusão, pode considerar também, os IDS e os IPS.

Autor: Sim, estão ambos incluídos no grupo de deteção de intrusão, apenas não ocorre uma separação entre ambos no referencial.

Entrevistado: Sim, porque é importante referir os dois, porque vão usar a mesma técnica de IA, mas têm finalidade diferente.

Autor: Exato. Não tenho mais nada a acrescentar.

Entrevistado: Sim, tem artigos bastante recentes, que tenho alguma curiosidade de ver.

Autor: Sim, são artigos entre 2015 e este ano, tenho alguns do ano corrente.

Entrevistado: O trabalho que está a fazer é muito sobre revisão de literatura, ler artigos e representar no referencial.

Autor: Sim, foi maioritariamente isso, como a ideia principal era cruzar técnicas e garantir a fiabilidade dos resultados apresentados, foi a melhor forma encontrada para isso.

Entrevistado: Sim, foi uma boa abordagem. No entanto, se for as pontas todas nunca vai acabar, porque existe muita coisa. Em resumo, acho que deveria acrescentar spoofing.

Autor: Certo. Obrigada.

Entrevistado: Bom trabalho

Autor: Obrigada pela disponibilidade

Entrevistado: Ora essa. Até uma próxima.

**Entrevista a Daniel Soares (Completa):**

Autor: Boa tarde, obrigada por aceitares participar desta entrevista. Podemos assim começar.

Entrevistado: Boa tarde. Não tens de agradecer. Sim, podemos começar.

Autor: A ideia principal da tese, centrou-se na realização de um referencial que cruza as áreas da Inteligência Artificial com as áreas da Cibersegurança.

Entrevistado: Certo.

Autor: Após uma análise efetuada a diferentes artigos académicos, foi possível construir o referencial, que permite o cruzamento das áreas de Cibersegurança com a Inteligência Artificial, tendo como fundamento ajudar elementos das organizações a encontrarem forma de diminuir o impacto causado por possíveis ciberataques. Assim sendo, de um lado do referencial estão as técnicas de Inteligência Artificial, como o Machine Learning, o Natural Language Processing e o Fuzzy Logic, e do outro as técnicas de Cibersegurança, como o Malware, Phishing e a Deteção de Intrusões.

Entrevistado: Sim, parece-me bem. O Vision Machine pode ser utilizado para ler padrões, utilizado em paralelo com o OCR, vai permitir identificar num documento que queiras analisar, na qual tens que dar um código, ele vai ler o código e faz a análise.

Autor: Exato, assim sendo vou começar com as perguntas. **Olhando para o referencial, parece-te útil para ajudar as entidades responsáveis nas organizações conseguirem criar um sistema com Inteligência Artificial para ajudar a combater possíveis ataques cibernéticos?**

Entrevistado: **Eu acho que com este referencial se consegue dar uma direção de qual o caminho a seguir. Contudo, gostaria de saber se vão existir exemplos para cada situação?**

Autor: Certo, sim. Cada número existente no referencial aponta para um cruzamento entre técnicas de IA com as técnicas da Cibersegurança, funcionando como uma legenda, isto é, há um detalhe do que é cada técnica e como podemos relacionar cada uma delas de forma a evitar um ciberataque.

Entrevistado: Está bem, era isso que queria saber. Outra coisa, na organização das técnicas de Machine Learning, o Machine Learning aparece lá como subárea.

Autor: Sim, porque o Machine Learning pode ser aplicado para evitar um ciberataque, além disso, tem vários algoritmos, que podem ser utilizados também, a ideia de colocar o Machine Learning novamente, foi para explicar que não era nenhum dos outros algoritmos referidos anteriormente.

Entrevistado: Poderia ficar Machine Learning / Outros, iria tornar-se mais claro. **Voltando à pergunta inicial, acho que sim que é útil porque de uma forma objetiva ajuda a encaminhar o decisor.**

Autor: Ou seja, vai ajudar a pessoa a ter uma decisão mais completa.

Entrevistado: Sim.

Autor: Assim sendo, **concordas com o referencial?**

Entrevistado: **Sim, acho que tendo exemplos práticos poderia ser mais útil. Também irá ser útil para uma pessoa que não esteja tão dentro do departamento de TI.**

Autor: Como assim?

Entrevistado: Com este referencial são apresentados vários pontos de decisão, está a ser facilitada essa decisão ao utilizador.

Autor: Certo, mas porque dizes que será mais útil para alguém que está fora do departamento de TI e não quem pertence mesmo ao departamento de IT?

Entrevistado: **O que quis dizer é que um diretor de Segurança pode usar o referencial como informação técnica para ele próprio usar, mas também pode ser útil para ele apresentar a outro departamento, como o de Gestão, para o mesmo validar e poder dar uma aprovação na implementação da solução. Com o referencial, é mais fácil fundamentar o porquê da decisão tomada.**

Autor: Entendi. Do teu ponto de vista, **achas que há alguma coisa que possa ser melhorada no referencial?**

Entrevistado: **Apenas a questão do Machine Learning, de acrescentar ou outros, ou um asterisco, para facilitar o processo de análise. Relativamente às áreas consideradas, parecem-me estar as fundamentais e principais.**

Autor: Sim, há mais áreas, no entanto, ainda não estão validadas cientificamente, o que torna a validação da informação não tão fiável. Mais alguma coisa que aches relevante acrescentar?

Entrevistado: Não, penso que seja apenas isso.

Autor: Obrigada por teres aceitado o convite para participar.

Entrevistado: De nada e boa sorte.

Autor: Obrigada.

**Entrevista a Anónimo (Completa):**

Autor: Boa tarde, agradeço a sua disponibilidade para a entrevista.

Entrevistado: Boa tarde, não é preciso agradecer.

Autor: Assim sendo podemos começar, o tema da tese centra-se na Inteligência Artificial e na Cibersegurança, tendo como foco principal a proteção dos Sistemas de Informação contra possíveis ciberataques, para tal, irá ser utilizada a Inteligência Artificial. Foi assim construído um referencial onde se realizou o cruzamento entre as técnicas de Inteligência Artificial e as técnicas de Cibersegurança.

Entrevistado: Certo, como foram obtidas essas relações?

Autor: Foi efetuada uma análise de diversos artigos académicos e científicos com base num conjunto de palavras-chave definidas. Utilizou-se o método PRISMA para ir reduzindo o nº de artigos, até se ficar com um nº mais reduzido e que continha efetivamente mais informação relativamente ao que se pretendia.

Entrevistado: Parece-me uma boa abordagem.

Autor: Ainda bem. Tenho algumas questões que gostaria de fazer. **Acha que o referencial proposto é útil para ajudar a combater possíveis ataques cibernéticos**?

Entrevistado: **Sim, penso que sim. Se for alguém dentro da organização que sabe à partida qual a fragilidade existente, conseguirá com base nisso cruzar com as técnicas de Inteligência Artificial.**

Autor: Certo. A próxima pergunta que tenho para fazer é se **concorda com o referencial? E se tem algo complementar a acrescentar?**

Entrevistado: **Sim, concordo com o referencial. Contudo, não aborda todos os tipos de ciberataques existentes, existindo ainda mais tipos de ataques e que começam a aparecer mais tipos de ciberataques.**

Autor: Consegue dar exemplos, desses tipos de ataques?

Entrevistado: Claro, **por exemplo, ataques por Data Manipulation. É um tipo de cibercrime em crescimento, contudo, poderá ser interessante falar sobre ele. Este tipo de ataque centra-se na falsificação de documentos oficiais ou na alteração dos dados de uma organização, onde a informação é depois tornada pública pelo atacante e tem a intenção de destabilizar e denegrir a imagem das organizações.**

Autor: Certo, no entanto, são áreas na qual ainda não existe grande conteúdo científico e académico que consiga garantir que o que for dito é verdadeiro.

Entrevistado: Sim, ainda é uma área em crescimento.

Autor: Certo, tenho apenas mais uma questão. Do seu ponto de vista, existe **algum ponto de melhoria?**

Entrevistado: **Penso que tirando o que disse na pergunta anterior, relativamente à inclusão de outras áreas da Cibersegurança, não vejo como se pode efetuar alterações, acredito que será um modelo que deverá estar em constante evolução devido à movimentação que existe a nível tecnológico. Vão sempre surgindo novas informações que poderão ser úteis para garantir uma maior viabilidade. Tirando isso, penso que será útil para um técnico de TI ter um apoio quando for tornar o seu serviço mais seguro.**

Autor: Certo, então parece-lhe uma boa proposta.

Entrevistado: Sim, penso que terá utilidade para um técnico ou responsável de TI quando tiver que tomar uma decisão e até terá o referencial como base para outras situações futuras.

Autor: Ainda bem. Tem mais alguma coisa que gostaria de acrescentar?

Entrevistado: Não, penso que não tenho mais nada a acrescentar.

Autor: Assim sendo, penso que terminamos. Obrigada pela sua disponibilidade e pela sua ajuda.

Entrevistado: Não tem que agradecer. Até à próxima e boa sorte.

Autor: Obrigada. Até à próxima

**Entrevista a Anónimo (Completa):**

Autor: Bom dia, agradeço desde já a sua disponibilidade para conversarmos.

Entrevistado: Bom dia, obrigada eu pelo convite.

Autor: Podemos assim começar, o tema da tese foca-se na Inteligência Artificial e na Cibersegurança, tendo como foco proteger os Sistemas de Informação de possíveis ciberataques, utilizando a Inteligência Artificial para tal. Assim sendo, foi elaborado um referencial onde é possível verificar o cruzamento entre as técnicas de Inteligência Artificial e as técnicas de Cibersegurança.

Entrevistado: O referencial parece-me bastante interessante, a numeração representada no referencial tem algum significado?

Autor: A numeração indica depois o detalhe obtido através da análise realizada a um diferente conjunto de artigos, que permite assim provar a veracidade do que se apresenta no referencial.

Entrevistado: Parece-me bem.

Autor: Sendo assim, tenho algumas questões que lhe gostaria de fazer, nomeadamente, o **referencial proposto é útil para ajudar a combater possíveis ataques cibernéticos**?

Entrevistado: **Penso que será útil sim. Contudo, terei de analisar melhor**.

Autor: Certo, posso-lhe enviar depois um documento com o referencial e o detalhe da análise efetuada.

Entrevistado: Pode ser, analiso e digo-lhe o que me parece, de uma forma mais completa.

Autor: A próxima pergunta que lhe gostaria de colocar é se **concorda com o referencial? Tem algum complemento a acrescentar?**

Entrevistado: **Concordo, no entanto, precisarei de ler e analisar com um maior foco**.

Autor: Ficarei a aguardar a sua análise, após o envio do documento com a informação do referencial. Assim sendo, tenho uma última pergunta que lhe gostaria de colocar, existe **algum ponto de melhoria?**

Entrevistado: **As áreas de segurança abordadas são efetivamente as mais importantes atualmente, os ataques de DoS estão na ordem do dia e são dos que mais ocorrem nos dias de hoje, os ataques de Malware incluem os vírus e estão também na ordem do dia. Os ataques de Ransomware estão também na ordem do dia e são os mais frequentes de acontecer em Portugal, foram inclusive ataques desses efetuados nos sistemas hospitalares e os mesmos foram resolvidos sem a utilização de Inteligência Artificial e de uma forma bastante rudimentar, tendo sido apenas reposto os backups anteriores.**

Autor: Certo, não sendo de todo a melhor abordagem a ser seguida, no entanto, resolveu o problema.

Entrevistado: **Portanto, considero uma boa abordagem e que as 5 áreas da Cibersegurança são as de importante foco, visto que as outras áreas também existentes não têm vulnerabilidades detetadas atualmente, portanto, não é necessário a existência de um foco neste momento.**

Autor: Correto, tem mais alguma coisa que gostaria de acrescentar e que ache relevante de considerar na tese?

Entrevistado: Por agora não.

Autor: Então, de seguida envio-lhe o documento com o referencial e espero as suas notas.

Entrevistado: Combinado.

Autor: Obrigada pela sua disponibilidade e até uma próxima.

Entrevistado: Obrigada e até à próxima.