



**POLITECNICO**  
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE  
E DELL'INFORMAZIONE

# Named Entity Recognition Network for Cyber Risk Assessment in Healthcare Domain

TESI DI LAUREA MAGISTRALE IN  
COMPUTER SCIENCE AND ENGINEERING - INGEGNERIA IN-  
FORMATICA

Author: **Lorenzo Bancale**

Student ID: 10576176

Advisor: Prof. Zanero Stefano

Co-advisors: Antonio Falvo

Academic Year: 2022-23



# Abstract

Cybersecurity in the Healthcare sector has become an important asset that aims to protect patient personal information and prevent attacks on the technology infrastructure that can affect the patient's health and safety. For this purpose, cyber risk analysis must be performed to draw up a Risk Assessment document stating the company's risk level. Despite every business deals with cybersecurity, the healthcare sector has been one of the most affected by cyber attacks in the last ten years. This is because patients' clinical information is the most valuable data to steal and because an attack on a medical device (DM) can directly or indirectly affect the health and well-being of patients. Because of that, this sector is more inclined to evaluate the possibility of paying a ransom or coming to terms with criminals. This highlights the importance for a hospital or, in general, a healthcare company of having state-of-the-art technological facilities where it is impossible to have efficient and fast analysis methods to be aware of the risks and vulnerabilities. We proposed a model to speed up the process of drawing up a cyber-risk analysis with the help of Natural Language Processing (NLP) operations, exploiting Named Entity Recognition (NER) tasks in the healthcare domain and paying attention to remain strict at the standard provided by the National Institute of Standards and Technology (NIST) Framework. We took into account the use case of AOU Sant'Andrea of Rome, and we performed an AI-driven cyber-risk assessment that ended up with a risk index as an outcome to attach to a more general risk assessment document. Our model consists of an Artificial Neural Network (ANN) based on the SpaCy library and trained in the healthcare and cybersecurity domain to perform NER on MITRE's CVE annual reports. We then compared the found entities with the Installed Base (IB) of the hospital to produce a dynamic risk index through statistical methods and merged with a static one to produce a final risk grade: LOW, MEDIUM, or HIGH for each DM. The performances of our model are: F1 score of 0.93, precision score of 0.93, and recall score of 0.94.

**Keywords:** Cybersecurity, Healthcare, Natural Language Processing, Risk Assessment



## Abstract in lingua italiana

La cybersecurity nel settore sanitario è diventata un'importante risorsa per la protezione delle informazioni personali dei pazienti e per la prevenzione da attacchi all'infrastruttura tecnologica che possono ledere salute e sicurezza dei pazienti. A tal fine, deve essere eseguita un'analisi del rischio informatico per redigere un documento che ne indichi il livello dell'azienda. Nonostante ogni settore si occupi di cybersecurity, il settore sanitario è stato uno dei più colpiti dagli attacchi informatici negli ultimi dieci anni. Questo perché le informazioni cliniche dei pazienti sono i dati più preziosi da rubare e perché un attacco a un dispositivo medico (DM) può influenzare direttamente o indirettamente la salute e il benessere dei pazienti. Pertanto, questo settore è più incline a valutare la possibilità di pagare un riscatto o di venire a patti con i criminali. Ciò evidenzia l'importanza per un ospedale o, in generale per un'azienda sanitaria, di avere strutture tecnologiche all'avanguardia e avere metodi di analisi efficienti per essere consapevoli di rischi e vulnerabilità. Abbiamo quindi proposto un modello per velocizzare il processo di analisi del rischio informatico con l'aiuto del Natural Language Processing (NLP), sfruttando la Named Entity Recognition (NER) nel dominio sanitario e rispettando gli standard forniti dal Framework del National Institute of Standards and Technology (NIST). Abbiamo poi preso in considerazione il caso d'uso dell'AOU Sant'Andrea di Roma e abbiamo effettuato una valutazione del rischio informatico guidata dall'IA che ha prodotto un indice di rischio come risultato. Il nostro modello consiste in una Rete Neurale Artificiale (ANN) basata sulla libreria SpaCy e addestrata nel dominio sanitario e della cybersecurity per eseguire NER sui rapporti annuali CVE del MITRE. Abbiamo poi confrontato le entità trovate con l'Installato Base (IB) dell'ospedale per produrre, attraverso metodi statistici, un indice di rischio dinamico unito poi con uno statico per produrne uno finale: BASSO, MEDIO o ALTO per ogni DM. Le prestazioni del nostro modello sono: punteggio F1 di 0,93, precisione di 0,93 e punteggio di richiamo di 0,94.

**Parole chiave:** Sicurezza Informatica, Sanità, Elaborazione del Linguaggio Naturale, Valutazione del Rischio



# Contents

<b>Abstract</b>	<b>i</b>
<b>Abstract in lingua italiana</b>	<b>iii</b>
<b>Contents</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Althea Italia and AOU Sant' Andrea . . . . .	1
1.2 Purposes and structure of the internship . . . . .	3
<b>2 State of the art</b>	<b>5</b>
2.1 An overview of cybersecurity in healthcare sector . . . . .	5
2.2 Related Works . . . . .	7
<b>3 The Healthcare Domain</b>	<b>11</b>
3.1 Clinical Data . . . . .	11
3.2 Medical device . . . . .	13
3.3 Cyber attacks, threats and vulnerabilities in HT domain . . . . .	15
3.4 CyberSecurity's laws, regulations and frameworks in healthcare domain . .	19
3.4.1 CyberSecurity regulations in Italy for IT risk management . . . . .	27
<b>4 Natural Language Processing and Neural Networks</b>	<b>29</b>
4.1 NLP definition . . . . .	29
4.2 NLP tasks . . . . .	30
4.3 Named Entity Recognition . . . . .	32
4.4 Evaluation Metrics . . . . .	34
4.5 NER challenges . . . . .	36
4.6 NER tools: SpaCy . . . . .	37
4.7 Artificial Neural Network . . . . .	40
4.7.1 Working of ANN . . . . .	41

4.7.2	Learning process . . . . .	41
4.8	BERT: Deep Bidirectional Transformers for Language Processing . . . . .	44
4.8.1	Input Embeddings & Positional Encoding . . . . .	45
4.8.2	Multi-Head Attention and Add&Norm . . . . .	47
4.8.3	FeedForward and Add&Norm . . . . .	48
<b>5</b>	<b>Proposed method</b>	<b>49</b>
5.1	Data Requirements Identification . . . . .	51
5.2	Labels . . . . .	52
5.3	Dictionaries . . . . .	54
5.3.1	Scraping instruments . . . . .	55
5.4	Training and Test Set . . . . .	57
5.5	The model . . . . .	60
5.6	The parameters . . . . .	62
<b>6</b>	<b>Experimental evaluations</b>	<b>65</b>
6.1	Network output and evaluation . . . . .	65
6.2	Business intelligence . . . . .	73
6.3	Index of risk . . . . .	74
6.3.1	Static Index of Risk - IRs . . . . .	75
6.3.2	Dynamic Index of Risk - IRd . . . . .	82
6.4	Outcome of the use case . . . . .	84
<b>7</b>	<b>Conclusions and future works</b>	<b>89</b>
	<b>Bibliography</b>	<b>91</b>
	<b>List of Figures</b>	<b>99</b>
	<b>List of Tables</b>	<b>101</b>
	<b>Acknowledgements</b>	<b>103</b>



# Chapter 1

## Introduction

This chapter describes the company where i completed my internship, it highlights the objectives and how the whole project has been conducted.

### 1.1 Althea Italia and AOU Sant' Andrea

This work results from an internship at Althea Italia, Millepini (MI), a biomedical company that is an integral part of Althea Group.

Althea Group is an independent company that provides a multi-vendor service to manage all medical technologies. The company assists customers in designing their healthcare site, selecting the appropriate equipment, and providing support throughout the entire lifecycle of the technology, from maintenance to disposal.

The Group manages over 1 million medical devices in over 2,000 healthcare facilities in 12 countries. With 3 Centers of Excellence between Europe and the USA, and over 2000 employees, Althea stands as the leading independent operator in the management and multi-vendor maintenance of medical equipment for both public and private healthcare facilities.

The integration of the various business lines has strengthened the Group's ability to develop increasingly integrated and innovative solutions capable of supporting, with high-quality standards, the management processes of technologies within healthcare facilities worldwide.

The digitization of data and services offered by hospital facilities and healthcare compa-

nies has meant that IT security has become a real asset to be protected. Furthermore, the increasingly frequent and damaging attacks on the healthcare sector have required that all companies that gravitate around this sector have had to move toward greater awareness of cyber risk.

Therefore, Althea has also begun to integrate into its services those aimed at managing the IT risk to which the medical devices under management are subjected. This work is motivated by the need to improve the risk assessment process, which only recently began to exploit the state of the art of new AI technologies. The main objective is the prevention of future attacks through a risk analysis of all the basic installations in hospital facilities. The AOU Sant'Andrea hospital in Rome was considered a use case, with a machine and software fleet of around 10 thousand units.

The Sant'Andrea University Hospital is a highly specialized healthcare structure, housing the Faculty of Medicine and Psychology - Sapienza University of Rome. Opened in 2001, the hospital offers a full range of patient services, including emergency care, scheduled surgeries, labor services, diagnostic testing, and inpatient or outpatient care.

As a University Hospital, their mission is the continuous translation of research and scientific innovations to clinical practice, realizing the highest level of healthcare quality and proceeding towards the actuation of patient-sized Medicine.

Research and scientific innovation continuously improve patient care: as a center involved in Phase I clinical trials, the Sant'Andrea Hospital contributes to developing and delivering novel treatments for serious diseases.

The analysis that Althea offers through business intelligence processes, using efficient analysis and visualization software such as Qlik, allows hospital managers to view relevant information and trends relating to all the machinery in the hospital. Therefore, The aim is to support strategic decisions to prepare maintenance, sales, and purchase plans for medical devices.

## 1.2 Purposes and structure of the internship

Once the context is framed, the purpose of this project is to optimize the process of evaluation of cyber-threats and vulnerabilities of the technological infrastructure of the AOU Sant'Andrea hospital, exploiting the current state of the art of Deep Bidirectional Transformers (BERT).

Then, we performed Information Extraction (IE) from the daily reports of Common Vulnerabilities and Exposure (CVE), published on the MITRE Corporation's site, by performing Natural Language Processing (NLP) and, in particular, the task of Named Entity Recognition (NER).

We trained a model to perform NER that learns from many corpora and extracts entities belonging to the Healthcare and Cybersecurity domains. In order to do that, we needed to create a domain-specific dictionary that would be suited to our particular use case.

The model will have a double function: prevention and protection. Prevention will be allowed by Business Intelligence (BI) operations that will continually compare daily reports with the entire playground of medical devices (DM) present in the installed base (IB). Protection, indeed, will be allowed by a real-time comparison of internal alerts belonging to the hospital manager of medical devices with the dataset of all the recent reports.

At the end of these operations, an Index of cyber risk will be generated as output. It will affect the Index of Risk (IR) already present in the document of cyber risk assessment drawn up by experts. This index of risk will be the result of the combination of a static IR and a dynamic IR; the first will be computed in the playground of Sant' Andrea, while the second is the outcome of the comparison between the output of the neural network and the current active asset of the hospital.

The internship was carried out from March to November in person and, when possible, at home through the Microsoft Teams application. During the first quarter, I was introduced to the biomedical sector, and I got in touch with the healthcare domain, understanding the main actors and their roles. I also understood the importance of cybersecurity in this sector and the needs of Althea's customers.

Then, I started to study the literature to understand the current state of the art of AI in the healthcare domain, understanding the starting point of this work. The core of my internship has been collecting data and developing a neural network that could perform named entity recognition in healthcare and cyber domains.

Finally, in the last quarter, I worked on business intelligence to show the project's results. I used Qlik and realized dashboards that could be integrated into Althea's already deployed customer packets.

## Chapter 2

### State of the art

This chapter highlights the context of cybersecurity in the healthcare sector and shows the related works that we can find in the literature.

#### **2.1 An overview of cybersecurity in healthcare sector**

The role of Healthcare Technologies is to extend, save, and enhance people's lives through efficient management of patient's clinical data, easy interconnectivity, and a secure network and cloud system. Such technologies range from those used for the management and storage of electronic health records (EHRs), the devices used for monitoring the health of the patients, the increasingly popular wearable and all the family of Internet of Medical Things (IoMT), remote telemedicine tools used to provide healthcare services through innovative ICT technologies, to all the diagnostic equipment and machinery that can be found in a hospital. To this already extensive set of technologies, we have to add all the mobile devices and ad hoc applications that interface themselves with the healthcare network system. These technologies generate a vast amount of data that needs to be protected because they concern medical patients' information confidential to doctors and patients. The fundamental properties of clinical data are that they need to be up to date, usable anytime, and kept safe; this is necessary because one of the main assets in the Healthcare sector is patient safety, both from an IT side and an actual health side.

As technological innovation proceeds quickly in the direction of a digital and interconnected society, so does the cyber criminality that has become increasingly dangerous in the last 20 years, exploiting software vulnerabilities and developing new threats, from viruses to ransomware and phishing. Despite many sectors starting to move in the direction of cyber security awareness, the Healthcare sector is one of the most targeted and also one of the last to move in such a direction. Indeed, many are cyber attacks

on this sector and public health. The most famous is the Wannacry ransomware, 2017, which infected over 300'000 endpoints worldwide, encrypting the data and asking for a ransom. Another ransomware attack originated from a phishing email in the US at the Hollywood Presbyterian Medical Centre, which was shut down for ten days until a ransom of 17'000€ was paid. In the UK, a malware attack led to the shutdown of the IT systems and the cancellation of almost all planned operations and outpatient appointments for four days. The last relevant case in Italy was the attack on the ASL of Abruzzo in May; this attack paralyzed all the services and the operations of hospitals and ambulatories for days, threatening to publish all sensible data stored by the organizations if a ransom of 2 Million euros would not be paid in bitcoin.

Nevertheless, why the Healthcare sector is the most hit by cyber-offenders? The reasons are many. First of all, is the fact that this sector is one of the last that started to activate itself to face the problem of cybersecurity. Many hospitals have installed on their devices outdated software like Windows XP that is not supported by their producers anymore; furthermore, lots of endpoint or medical devices have installed software for which a vulnerability has been discovered or exposed by the vendor itself, and because of that, an update is required. This fact makes it easier for an attacker to violate the IT of this sector. The second reason is that medical information is most valuable on the black market. This is because this information can be used to access drug delivery or expose the health state of relevant patients. Having access to this information means having the power to affect politics and markets. Medical information and credentials on the dark web are sold up to thousands of dollars in Bitcoin. A third reason is undoubtedly the fact that dealing with the health and lives of people makes the healthcare organizations, above all the others, more inclined to accept the paying of the ransom in order to thwart any fatal consequence.

In order to respond to the crisis of cyber attacks, each state defines laws, acts, regulations, and decrees that each actor that performs digital actions needs to follow. Moreover, in the last ten years, institutes were born in order to create standards, guidelines, and best practices to help organizations evaluate risk management in the cyber domain. An example is the National Institute of Standards and Technology (NIST), which is nowadays a well-known cybersecurity framework that, thanks to its elasticity, is used by the vast majority of US organizations. Therefore, the cyber-risk assessment has become a requirement for all the organizations that deal with the digital world, Healthcare first of all.

## 2.2 Related Works

Numerous studies have explored threats and vulnerability analysis employing a variety of methods. This section offers a summary of pertinent literature concerning our study. Specifically, we refer to threat modeling, cyber-attacks within the healthcare industry, and threat analysis based on machine learning techniques.

### Threat Modeling and Cyber Attacks in the Healthcare Sector

Threat modeling is crucial in grasping the potential risks tailored to a particular system. PASTA and Attack Tree are among the established methods in this domain [53]. PASTA focuses on risk and uncovers security vulnerabilities, aiming to determine the right countermeasures based on the potential impact. This method emphasizes the collaboration between analysts and businesses to evaluate, document, and suggest measures depending on the attack probability. Attack Tree employs a hierarchical tree format to outline a system's security. The top node represents the ultimate objective, whereas the subordinate nodes elucidate potential system attacks. This structure offers insight into specific attack strategies, detailing threats and potential defensive tactics against them. Research from the Centre for Internet Security (CIS) indicates that cyber threats like ransomware, DDoS, insider threats, and data breaches frequently target the healthcare industry [10]. A recent investigation found that roughly 20% of medical device manufacturers faced ransomware or malware assaults in the preceding 20 months [21]. Cyber adversaries may aim at specific medical equipment like infusion pumps [23, 37] or broader healthcare services like medication distribution systems [4]. Literature underscores the significance of patch and incident management measures to bolster hospital security.

### Threat and Vulnerability Analysis Using Machine Learning Models

Several recent studies concentrate on threat and vulnerability identification and analysis through Machine Learning (ML) models. Ghaffarian et al. [20] provide a review of ML and Data Mining techniques designed to minimize software vulnerability damage, identifying four principal vulnerability prediction categories: (i) Supervised ML approaches utilizing Software Metrics-based Prediction Models; (ii) Anomaly Detection Approaches employing unsupervised ML techniques to model normality or mine rules from source code, subsequently identifying vulnerabilities deviating from the norm; (iii) Vulnerable Code Pattern Recognition, leveraging supervised ML to identify vulnerable code segment

patterns; and (iv) Miscellaneous Approaches, encompassing various uncategorized AI and ML techniques.

In another study [65], authors introduce a cyber supply chain threat analysis method combining Random Forest and XGBoost algorithms for precise threat forecasting, focusing on the Tactics, Techniques, and Procedures (TTP) of cyber attacks with demonstrated high accuracy.

SHChecker is a unique threat analysis framework presented in [24], merging ML with formal Smart Healthcare Systems (SHSs) analysis. This research zeroes in on the Internet of Medical Things (IoMT), incorporating various ML algorithms such as Decision Trees (DT), Artificial Neural Networks (ANN), K-means, among others, finding in experiments that DT-based algorithms outperform NN-based ones in accuracy.

A paper by different authors [68] outlines a method analyzing the severity of Computer Security (CS) threats by examining the language in CS-related tweets through Deep Learning (DL), with experiments involving 6000 tweets describing software vulnerabilities. This method also links the vulnerabilities reported in tweets to CVEs and NVD KBs, demonstrating high precision in forecasting severe vulnerabilities.

Satyapanich et al. [51] introduced a semantic schema to narrate CS events, applying a Deep Learning-based Information Extraction (IE) pipeline for automatic data breach, ransomware, and phishing attack information extraction, as well as vulnerability discovery and patching. This supports the extraction of custom information for the CS area, exploiting taxonomies, sharing standards, and ontologies [35] as part of cybersecurity's threat intelligence knowledge base.

A Named Entity Recognition (NER) method is also proposed, employing a BiLSTM-CRF architecture with a multi-head self-attention neural network and word embeddings trained on CS-specific texts [55]. This approach, paired with Knowledge Bases (KBs), aids in identifying asset details involved in CS issues.

Furthermore, [33] details an NLP DL-based architecture to identify pertinent CS information, including vulnerability exploitations and advanced persistent threats. This architecture comprises a word-embedding layer, a BiLSTM layer, and a CRF layer, with another BiLSTM serving as the output layer, showing improvement over baseline results in experiments.

Lastly, Nikoloudakis et al. [44] showcase an ML-based framework enhancing situational awareness by detecting entities in an IoT environment using real-time awareness features from the Software-Defined Networking (SDN) paradigm. This system, continuously mon-



itoring assessed entities with an ML-based Intrusion Detection System (IDS) trained on an enriched dataset, demonstrated superior prediction accuracy compared to traditional systems.

In another noteworthy study [57], authors approached software vulnerability detection as an NLP problem, treating source code as text. They employed recent DL NLP models for automated software vulnerability detection and compared various models based on accuracy. The top-performing model achieved a 95% accuracy rate, and the proposed method could also classify the vulnerabilities of the source codes.

Recently, architectures rooted in the Transformer framework [60], including BERT [14], have been adopted within the cybersecurity (CS) sector, specifically for devising Named Entity Recognition (NER) methodologies. These methodologies excel at pinpointing threats, vulnerabilities, and attacks in unstructured textual content. One such model is CyBERT, introduced by [3], designed for semi-automated vetting in Industrial Control Systems (ICS). This model benefited from training on a purpose-built corpus of annotated sequences from diverse ICS device documentation, yielding superior outcomes compared to models trained on broader domains. Additionally, [11] showcased a fine-tuned BERT-based model for the CS NER task, enhancing results using domain-specific dictionaries.

CyNER [2], another Transformer-based model highlighted in the research, employs the XLM RoBERTa-large language model [31], pre-trained on threat analyses and refined for the CS NER function. It also integrates various strategies, such as priority-based merging for entity extraction. Specifically, it combines regular expressions, knowledge bases (KBs), a generic domain ML model, and a Flair-based [1] NER model to enhance results. Furthermore, a method described in [67] for CS NER amalgamates the BERT and BiLSTM-CRF Deep Learning architectures, surpassing the benchmark performance.

Our research distinguishes itself from these contributions by primarily targeting cyber attacks in the healthcare domain and medical devices. Furthermore, we employ NLP to identify and evaluate threats and vulnerabilities from textual data, thereby determining their severity.



## Chapter 3

# The Healthcare Domain

This chapter describes all the relevant topics related to the healthcare sector.

### 3.1 Clinical Data

The healthcare domain, like any other sector that started to incorporate technology in its field, deals with generating, storing, and managing vast data that need to be kept secure. Nonetheless, Healthcare data differs from other information precisely because of their source domain. Every department in a hospital handles Personally Identifiable Information (PII) and Protected Health Information (PHI). The first one refers to any information that could be used to identify an individual (e.g., address, telephone number, email, social security number), and the second refers to particular health data of a patient (e.g., patient's name, date of birth, health conditions, medical identification number, exam results). All healthcare providers (e.g., nurses, pharmacists, technicians, dietitians, physical therapists, assistants, and doctors) use Electronic Health Records (EHR), e-prescribing software, remote patient monitoring, laboratory information systems; scheduling and administration departments work with clinical data on scheduling software; the billing office works with insurance and financial information through medical billing software and so on. Moreover, all the patient's clinical data needs to be transmitted, stored, and visualized when needed, and for this purpose, integrated systems were created to manage all those tasks. An example is the Radiology Information System (RIS), which, with the Picture Archiving and Communication System (PACS), manages all the transmission flow, storage, visualization, and delivery of medical reports and diagnostic images [64][63]. These information systems generate, in turn, other data that are attached to medical reports; another example is the ECG monitor that records patient's vital signals.

After this overview of the different actors who generate and manage clinical data, we can

now see how cybersecurity is adopted in this sector.

Differently from other fields (e.g., financial, banking), where this issue has been confronted for decades, in which policies have been redacted, and resources have been invested in security, the healthcare sector is relatively new to this problem and, moreover, it is very cost constrained and limited resources are allocated to IT security. The type of data exchanged makes public health a particular sector because patient safety is constantly at risk. In addition, we have to consider the hyper-connectivity of hundreds of thousands of medical devices that need to cooperate to keep information available and up to date. This network comprehends a multitude of DMs, mobile apps, endpoints, and IoMTs, and this introduces healthcare ICT to many vulnerabilities and cyber threats.

If we take the example of a stolen credit card, we know that once the victim is aware of the theft, he can immediately block the card, cancel his old data, and issue a new one. This is impossible if the patient's PHI is stolen; the patient cannot change the birth date, health information, or diagnostic data. So, the immutability of healthcare data makes them more valuable than any other sector's information because they never expire; moreover, once stolen, health information is widely applicable and valuable for crimes from identity theft to medical fraud. To have an idea of their value compared to the one of any other sector is enough to know that on the dark web, an individual's health information can be sold from 10 up to 20 times than a credit card number [9] [32].

Clinical data and, in general, all information that belongs to the healthcare domain are regulated by laws and guidelines that will be better specified in Section 3.4 and 3.4.1 .

## 3.2 Medical device

We keep mentioning Medical Devices, but what are they? Which are the DMs that we kept in count for this project?

The Global Harmonization Task Force (GHTF) [58] defined a DM in the 'international guidance documents. According to this document, a medical device is:

Any instrument, apparatus, implement, machine, appliance, implant, reagent for in vitro use, software, material, or other similar or related article intended by the manufacturer to be used, alone or in combination, for human beings, for one or more of the specific medical purpose(s) of:

- diagnosis, prevention, monitoring, treatment, or alleviation of disease;
- diagnosis, monitoring, treatment, alleviation of, or compensation for an injury;
- investigation, replacement, modification, or support of the anatomy or a physiological process;
- supporting or sustaining life;
- control of conception;
- disinfection of medical devices;

and which does not achieve its primary intended action by pharmacological, immunological, or metabolic means in or on the human body but may be assisted in its intended function by such means [48].

The WHO Global Model Regulatory Framework for medical devices [46] supports Member States to develop and implement regulatory controls and regional guidelines for good manufacturing to ensure the quality, safety, and efficacy of medical devices available in their countries. The Organization also works with Member States and collaborating centers to develop guidelines and tools, including norms and standards on medical devices [62].

Since in a hospital DMs are very numerous and of different categories, from autopsy tables to computed tomography scanning systems, there is the need to individuate a subclass that considers only the needed DMs. This subclass collects all the DMs that can connect to each other, interface themselves with the hospital ICT network infrastructure, and all the devices and software that generate, store, send, and visualize clinical information. Since these DMs are in the order of hundreds, we cannot list all of them; instead, we can show the categories to which they belong. Table 3.1 list all of them:

DM's categories
ANALYZERS
ANALYZER
MAMMOGRAPH
ANGIOGRAPH
PHONOANGIOGRAPH
FLUROANGIOGRAPH
ECHOTOMOGRAPH
TOMOGRAPH
MAGNETIC RESONANCE IMAGING
POLYGRAPH
ELECTROMYOGRAPH
HOLTER
ECHENCEPHALOGRAPH
ELECTROENCEPHALOGRAPH
MONITORING CENTER
MONITORING SYSTEM
X-RAY
RADIOLOGY
FLUOROSCOPY
NEURONAVIGATION SYSTEM
NAVIGATION SYSTEM
TELEMETRY
PRINTER
BIOIMAGES
SOFTWARE
DIGITAL RADIOLOGY
POLYSOMNOGRAPH

Table 3.1: Categories of DM taken into consideration

### 3.3 Cyber attacks, threats and vulnerabilities in HT domain

Up to now, we have just said that healthcare is one of the most targeted sectors by cyber-attacks; this section will explore stats, trends, and future forecasts about cyber threats in the HT industry. According to the World Economic Forum, cybercrime and cyber insecurity are new entrants into the top-10 rankings of the most severe global risks over the next decade (now taking the eighth spot and standing side-by-side with threats including climate change and involuntary migration)[61].

Cybersecurity Ventures expects global cybercrime costs to grow by 15 percent per year over the next three years, reaching \$8 trillion globally this year and \$10.5 trillion annually by 2025, up from \$3 trillion in 2015.

In 2018, the U.S. Department of Justice reported less than one in seven cybercrimes. In some countries, the reported rate was even lower. Cybersecurity Ventures believes that reporting practices concerning illegal cyber activity are improving. However, in 2023, we still face a situation where less than 25 percent of cybercrimes committed globally are reported to law enforcement.

According to Cloudwards research (2021), every 14 seconds, a new organization gets hit by ransomware. Schools, healthcare providers, and even government institutions have all become victims of ransomware attacks by cybercriminals. With even crucial public services shut down, ransomware is now a global threat to organizations and individuals. With the COVID-19 pandemic, the rate of cybercrime increased by 600%, and the total cost of all cybercrime damages in 2021 is expected to amount to about \$6 trillion worldwide[40]. Hospitals and other organizations in the healthcare industry were already suffering from a widespread lack of staff and budget to deal with cyber security risks, and the abrupt changes caused by the pandemic only worsened existing IT weaknesses. In this context, a staggering 90% of healthcare staff in 2020 did not receive any updated training on cyber security best practices after the COVID-19 pandemic forced them to work from home [7].

Among all types of attacks, ransomware has been the one that has increased a lot in recent years. The global cost of ransomware was predicted to reach \$20 billion in 2021, up from \$325 million in 2015. Cybersecurity Ventures expects ransomware damage costs to exceed \$265 billion annually by 2031. CNA Financial made the biggest ransomware

payout on record to know the business behind this illegal activity: the Chicago-based company paid \$40 million to the Phoenix cybercriminal group (Cybersecurity Ventures profiled 42 ransomware gangs in the latest edition of its “Who’s Who In Ransomware” quarterly report). Ransomware attacks do not just cause monetary damage; in the United States alone, 764 organizations in the healthcare sector temporarily stopped operations because of ransomware in 2019.

From January 2021 through May 2021, the Health Sector Cybersecurity Coordination Center (HC3) documented a total of 82 ransomware attacks around the world, with 48 of these attacks taking place within the United States healthcare sector[26]. The 2023 Data Breach Investigation Report published insights into the HT sector.

Figure 3.1 shows how ransomware remains a top action type in breaches. It is important to notice that even if it did not grow much from 2022, it still holds statistically steady at 24%. Other essential insights are shown in Figure 3.3, which highlights essential HT trends.

Along with ransomware, data breaches represents a huge problem for all sectors, but as we said previously, HT is one of the most hit industry due to the value of clinical data on the black market.

IBM claims the average cost of a data breach in healthcare — comprising of hospitals and clinics — increased by nearly 1 million USD to \$10.10 million in 2022 [28][41].

The cost for a compromised healthcare provider to recover a lost or stolen record can reach up to \$408. The advertising costs of alleviating reputational damages can go up to \$1.75 million [29]. Medical records are one of the most profitable items cybercriminals can steal due to the large amount of personal information contained therein and the large ransoms they can extort from struggling hospitals desperate to get their patients’ confidential data back.

Finally, in Figure 3.2, some interesting HT insights are shown that will help understand the context of this sector. Those stats refer to data breaches and, are up to date and are provided by the DBIR (2023).

As can be seen from the figure, 74% of security breaches have a human component, where individuals play a role either through mistakes, misuse of privileges, use of pilfered credentials, or susceptibility to social engineering. External actors are implicated in 83% of these breaches, and the main incentive behind the attacks is predominantly financial,



representing 95% of these incidents. The top three methods attackers use to penetrate an organization include pilfered login details, phishing schemes, and taking advantage of vulnerabilities.

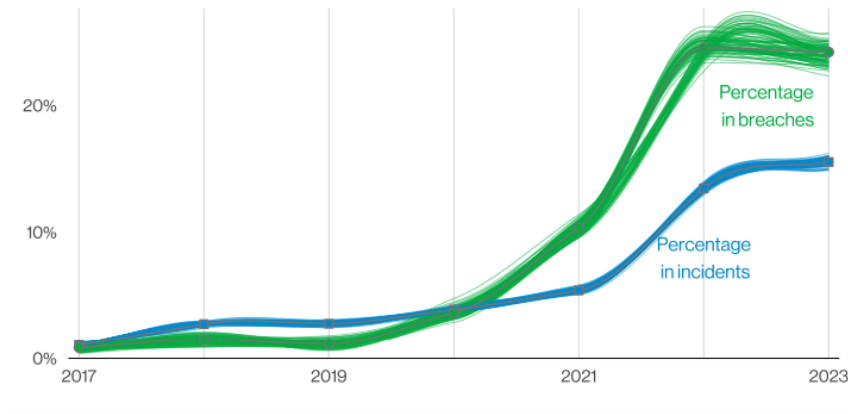


Figure 3.1: Ransomware action variety over time

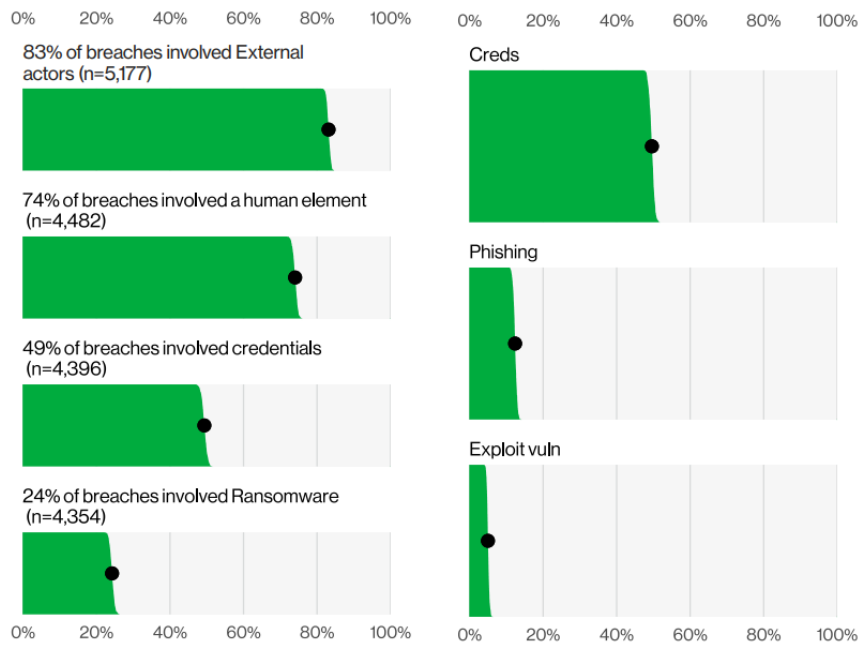


Figure 3.2: Data Breach trends in healthcare - DBIR 2023

<b>Frequency</b>	525 incidents, 436 with confirmed data disclosure
<b>Top patterns</b>	System Intrusion, Basic Web Application Attacks and Miscellaneous Errors represent 68% of breaches
<b>Threat actors</b>	External (66%), Internal (35%), Multiple (2%) (breaches)
<b>Actor motives</b>	Financial (98%), Espionage (2%), Fun (1%), Ideology (1%) (breaches)
<b>Data compromised</b>	Personal (67%), Medical (54%), Credentials (36%), Other (17%) (breaches)
<b>What is the same?</b>	The top three patterns remain the same, although the order has changed. Internal actors making mistakes continue to trouble this sector.

Figure 3.3: CyberSecurity trends in healthcare

### 3.4 CyberSecurity's laws, regulations and frameworks in healthcare domain

This section analyzes the evolution of laws, guidelines, and frameworks relating to cybersecurity in the healthcare domain.

Over the past twenty years, PHI and PII guidelines have continuously evolved. In the US, the Health Insurance Portability & Accountability Act (HIPAA) of 1996 was enacted to safeguard health information and improve health insurance portability, aiming to reduce fraud and simplify insurance administration [5]. In 2009, the Health Information Technology for Economic and Clinical Health (HITECH) Act supplemented HIPAA, intensifying penalties for HIPAA violations, mandating breach notifications, and encouraging the use of electronic health records (EHRs) while also providing rights for patients to access their EHRs[54] .

The cybersecurity landscape in the European Union's healthcare sector is primarily shaped by the Network and Information Systems (NIS) Directive [17], a pivotal piece of legislation aimed at enhancing the security of network and information systems across the EU. This directive, along with its updated version, NIS 2 [18], imposes stringent requirements on member states to adopt national network and information system security strategies. It explicitly identifies operators in critical sectors, including healthcare, mandating them to implement robust measures to manage cybersecurity risks. These directives are complemented by the General Data Protection Regulation (GDPR ).

The GDPR, or Privacy Regulation EU 2016/679, substituted previous rules by introducing measures and stipulations related to every EU citizen's personally identifiable information (PII), encompassing clauses for breach alerts and the imposition of penalties [16]. It is the most well-known European law, and it also defines the methods of processing personal data and requires organizations to adopt appropriate technical and organizational measures to ensure their security. The risks associated with processing information include significant risks to the fundamental rights and freedoms of the individual (profiling), risk of accidental or unlawful destruction, risk of breaches or loss of data (data breach), risk of unintended modification (data falsification), and risk of unauthorized communication and dissemination. The purpose of this classification is to assist in understanding how to classify information and how to assess the risk and progressively reduce it.

When discussing medical devices and healthcare software, it is crucial to emphasize the importance of adhering to specific regulations and safety standards. These regulations play a central role in ensuring the safety and efficacy of such products:

- **European Regulation on Medical Devices (MDR):** It governs medical devices and provides marketing, use, and surveillance requirements after they enter the market. The regulation sets safety and performance standards for medical devices, such as implants, diagnostic devices, and medical equipment. The MDR entered into force in May 2017, replacing the Medical Device Directive (MDD) and the Active Implantable Medical Device Directive. The MDR applies from 26 May 2021, except for some provisions (Article 123). The new regulation is a response to technological progress in the development of medical devices. It also addresses problems with the previous regulatory system revealed in scandals involving unsafe medical devices, and it harmonizes the rules for medical devices in the EU [43].
- **European Regulation on In Vitro Diagnostic Medical Devices (IVDR):** This concerns in vitro diagnostic medical devices, that is, devices used for the diagnosis of diseases or for assessing a person's health status, such as laboratory tests. The regulation establishes specific requirements for the marketing, use, and surveillance of in vitro diagnostic medical devices.

In order to make a correct and reliable assessment of IT risk, one must rely on methods recognized by the scientific community and adopt a model that allows for a precise, reliable, and certifiable risk assessment. Let us quickly look at the laws and regulations on which to base the assessment of IT risk; such laws should be viewed from two different aspects: authoritative (which determines the mandatory nature of the rule based on the authority that issued it) and operational (which defines the necessary steps for the application of the rule) [6].

- **Legal provisions:** They define terms and general principles. They can be consulted for a general overview of the set objective.
  - Authoritative: They have the highest importance since all other rules must be consistent with these provisions and never in conflict.
  - Operational: Low impact as they define the general lines to follow.
- **Government regulations:** They are more specific and go into context. For instance, for all public entities in Italy, there are the "minimum ICT security measures" (March/April 2017) issued by AgID (discussed in the following). They are a

practical reference for evaluating and improving public administrations' cybersecurity level, aiming to counter the most frequent cyber threats. These rules explain which evaluations to make.

- **Guidelines:** They are collections of recommendations, mainly of a behavioral nature, and define the best practices to follow since they have proven to be the most efficient in certain operational contexts and to achieve specific objectives.
  - Authoritative: Low impact as they are not mandatory procedures.
  - Operational: High impact because they find widespread practical application.
- **Company regulation:** Related to IT tools, the company regulation is that document in which all provisions, behaviors, and organizational measures required of employees and company collaborators are contained to counter IT risks.
  - Authoritative: Relative impact as it is only valid within the company scope.
  - Operational: High impact as employees are obliged to sign and respect it.
- **National or international standards:** An example is the GDPR, which many companies have not yet adopted.
  - Authoritative: High impact as they are written and mandatory.
  - Operational: Medium as they are not always implemented in detail.

Frameworks are not exhaustive at the authoritative level with all the features of a standard and cannot, in any way, be considered tools for compliance with current regulations. However, they are widely applied as they have both an evaluation component and a component related to action and correction of areas most exposed to risk. Cybersecurity frameworks provide an organized approach to managing security risks, mitigating potential vulnerabilities, and improving overall digital defense. As enterprises continue integrating digital technologies into their operations, staying up-to-date with the most current cybersecurity frameworks is increasingly important.

A cybersecurity framework is a set of policies, practices, and procedures implemented to create an effective security posture. These frameworks provide organizations with the guidance to protect their assets from cyber threats by identifying, assessing, and managing risks that could lead to data breaches, system outages, or other disruptions.

Cybersecurity frameworks help organizations develop and maintain an effective security strategy that meets the specific needs of their environment. Through evaluating current security practices and identifying gaps in protection, these frameworks help cybersecurity teams implement appropriate safeguards to protect critical assets.

The following list mentions the nine most known cybersecurity frameworks [49]:

- **NIST:** The National Institute of Standards and Technology (NIST) is a U.S. governmental body dedicated to setting technological and security benchmarks within the country. NIST's Cybersecurity Framework offers a roadmap for organizations to navigate, safeguard, recognize, counteract, and rebound from cyber threats. Initially rolled out in 2014 for federal agency guidance, its recommendations are versatile enough for virtually any entity aiming to fortify its digital safeguards. Currently in its updated iteration, NIST's framework serves as an exhaustive list of ideal practices for entities striving to enhance their cybersecurity measures. It provides in-depth insights into areas like risk assessment, resource management, identity governance, incident reaction strategies, and supply chain management, among others.
- **ISO 27001 and ISO 27002:**

ISO 27001 and ISO 27002 are among the foremost standards in information security management in today's landscape. These standards present an all-encompassing framework for entities wishing to safeguard their data through stringent policies and recommended practices.

Crafted by the International Organization for Standardization (ISO), these standards articulate practices and tenets, ensuring that organizations adopt suitable safeguards for their data. They cover aspects from resource oversight and access governance to incident handling and sustaining business operations, offering organizations an intricate blueprint for network security.

ISO 27001 introduces an international benchmark that encapsulates a methodical strategy for evaluating risks, choosing controls, and their execution. It entails criteria for instituting an Information Security Management System (ISMS).

Conversely, ISO 27002 is a practical guideline that enumerates more granular and nuanced security measures. Employing both standards simultaneously equips organizations with a holistic methodology for managing information security.

ISO 27001 outlines the requirements to establish and manage a company's information security management system. This system operates in stages: initially, it identifies assets related to information management and then ensures that the necessary controls, based on selected objectives, have been implemented. This process culminates in a risk analysis, followed by mitigation suggestions. Periodic checks of the framework are advised to gauge and potentially reduce risk exposure.

Both these systems can be applied by state and private organizations alike. By adopting ISO 27001, entities can assure their stakeholders that their data remains

secure. The national cybersecurity framework and ISO 27001 provide clear guidelines on mandatory documentation and essential measures. These tools specify the outcomes without dictating the methods of implementation.

- **CIS Controls:** The Center for Internet Security (CIS) Control Framework encompasses 20 controls that span various security domains, including access management, resource oversight, and incident management. The CIS Controls are categorized into three tiers: Basic, Foundational, and Organizational.

Basic Controls emphasize core cybersecurity actions every organization should adopt, like consistent updates and antivirus safeguards. Foundational Controls represent enhanced steps beyond basic security procedures, encompassing measures like dual-factor authentication and continuous surveillance of log files for unusual activities. Organizational Controls are tailored to offer protections that cater to an organization's unique requirements, such as fostering user awareness and providing training.

- **SOC2:** The Service Organization Control (SOC) framework serves as an audit standard employed by external auditors to evaluate a company's systems and services in terms of security, availability, processing integrity, confidentiality, and privacy. Among these, SOC2 stands out, specifically crafted for cloud service vendors.

Under the SOC standard, companies must present comprehensive documentation concerning their internal procedures associated with security, availability, processing reliability, confidentiality, and privacy. Documents adhering to SOC requirements should encompass strategies like access management, cryptographic data measures, and emergency response blueprints.

Furthermore, companies must showcase proof of the efficacy of their protective measures, such as inspection records or results from security breach tests. This ensures the company's security mechanisms operate as intended and offer a robust defense against cyber adversities.

- **PCI-DSS:** A consortium of leading payment processors formulated the Payment Card Industry Data Security Standard (PCI-DSS) to shield consumers' payment card details. This standard lays out a detailed set of stipulations to assist entities in securing their systems and thwarting unsanctioned access to client details.

The PCI-DSS structure stipulates 12 essential criteria that organizations must fulfill to ensure customer information safety. These criteria address access management, network defense, and data preservation, particularly tailored for payment processing. The framework also introduces protective measures for customer payment card details, encompassing techniques such as encryption and tokenization.

- **COBIT** Formulated by the Information Systems Audit and Control Association (ISACA), Control Objectives for Information and related Technology (COBIT) presents a holistic framework crafted to aid organizations in optimizing their IT assets. This framework prescribes best practices in governance, risk oversight, and cybersecurity.

The COBIT structure is segmented into five domains: Plan & Organize, Acquire & Implement, Deliver & Support, Monitor & Evaluate, and Manage & Assess. Each domain encompasses distinct processes and tasks tailored to bolster efficient IT resource administration.

Moreover, COBIT integrates exhaustive guidelines related to data security and safeguarding. This spans access management, user verification, cryptographic methods, audit trails, and strategies for handling security incidents. Through these guidelines, organizations receive a thorough toolkit to defend their infrastructure against cyber adversities.

- **HITRUST Common Security Framework** The Health Information Trust Alliance (HITRUST) Common Security Framework (CSF) is a robust security blueprint tailored specifically for the healthcare sector. It prescribes best practices to ensure the confidentiality of patient information, touching upon aspects like access management, identity verification, cryptographic measures, audit documentation, and strategies for tackling security incidents.

The HITRUST CSF provides an in-depth guide on cybersecurity governance, risk oversight, and compliance mandates. This assists organizations in adhering to pertinent regulatory standards while safeguarding their infrastructures from prospective cyber challenges.

- **Cloud Control Matrix** The Cloud Security Alliance's (CSA) Cloud Control Matrix (CCM) is a comprehensive security framework for cloud-based systems and applications that covers access control, user authentication, encryption, audit logging, and incident response.

Similar to HITRUST, the CCM also includes detailed guidelines for security governance and risk management and is aimed at helping organizations meet relevant regulatory standards.

- **CMMC 2.0** The Cybersecurity Maturity Model Certification (CMMC) is the latest version of the US Department of Defense's (DOD) framework, announced in 2021. This was designed to protect national security information by creating consistent cybersecurity standards for any organization working with the DOD.



In Italy, we have the National Framework for Cybersecurity second edition [34], which is starting to be recognized and applied by organizations. The current version introduces contributions aimed at addressing key aspects related to data protection as stipulated by the GDPR.

The EU sector regulations are four and serve as models and norms for accurate and consistent assessments in line with current laws.

- **Cybersecurity ACT:** It has validity at the European level and defines a regulatory framework for protecting industrial and civil use devices. This regulation concerns the cybersecurity of European ICT (information and communication technology) products and services. Thanks to this, companies can self-certify their products/services where they can submit a declaration of conformity and be recognized as one of the three levels of compliance (basic, substantial, high), which corresponds to the ability to resist attacks, effectively acquiring a guarantee criterion in the security field.
- **EIDAS Regulation:** A 2014 regulation has enabled the creation of European standards for everything financial that moves online (e.g., money, electronic authentication), with the aim of giving the digital document the equivalent legal value reserved for the paper one.
- **NIS Directive:** It aims to make national computer systems more secure. Compared to previous regulations, in addition to standardizing such rules at the European level, it makes specific monitoring, preparation, response, and recovery measures mandatory for essential service operators (of which healthcare is a part).
- **GDPR or EU Privacy Regulation 2016/679:** Already explained above.

Let us now talk about AGID. The Agency for Digital Italy is an Italian public agency that carries out functions and tasks to achieve the highest level of technological innovation in the organization and development of public administration for the benefit of citizens and businesses. A significant responsibility of AgID is to accredit or authorize entities that perform certain activities in the digital domain.

In the healthcare sector, AGID plays a coordinating role and provides technical support for digitization activities and the development of health information systems, aiming to enhance the efficiency and effectiveness of the national health system. Additionally, AGID promotes cybersecurity and the protection of sensible data in the healthcare sector by adopting regulations and guidelines for managing IT risk.

Specifically, AGID collaborates with the Ministry of Health and other relevant authorities to promote the adoption of technological and security standards for health information systems, such as the implementation of the SPID (Public System of Digital Identity) for accessing online health services and defining regulations for the management of patients' sensible data.

### 3.4.1 CyberSecurity regulations in Italy for IT risk management

In Italy, European Union regulations and national laws represent the primary reference legislation concerning IT risk management in the healthcare sector. The former establishes provisions for protecting personal data, including sensible data, in the healthcare sector. Based on these provisions, healthcare facilities must adopt technical, organizational, and procedural security measures to protect personal data collected, processed, and transmitted. These measures must be suitable to prevent unauthorized access, distribution, loss, alteration, and dissemination of personal data.

The latter deals with public administrations' methods of managing electronic services, including those in the healthcare sector. In particular, it sets out the procedures and responsibilities related to managing information systems and communication networks and the methods of accessing data by authorized users.

Furthermore, the Ministry of Health has issued numerous guidelines and circulars related to IT risk management in the healthcare sector, including the "Technical Guide to Minimum Security Measures" and the "Guide to Clinical Risk Management in a Healthcare Environment."

To conduct an IT risk analysis in the healthcare sector, one must refer to the current regulations on personal data protection and cybersecurity specific to the healthcare sector:

- **GDPR:** as an EU member, Italy adheres to the GDPR, which sets stringent rules for the processing and protecting of personal data. In healthcare, this involves safeguarding patient data, requiring consent for its use, and ensuring its confidentiality and integrity.
- **Italian Privacy Code:** this national law aligns with the GDPR and includes specific provisions for data protection in Italy. It covers aspects such as data processing, rights of the data subjects, and obligations of data controllers and processors.
- **Italian Legislative Decree 101/2018:** this decree aligns the Italian Privacy Code with the GDPR, ensuring that national legislation is consistent with EU regulations.
- **Cybersecurity Laws and Directives:** this includes the EU's NIS Directive, which sets security requirements for network and information systems in critical sectors, including healthcare. Italy implements these directives through national legislation and specific cybersecurity policies.
- **ISO 27001, ISO 27002:** those technical standards have already been explained in depth in Section 3.4.



## Chapter 4

# Natural Language Processing and Neural Networks

This chapter explains the context of Natural Language Processing and details its tasks. Then, Deep Learning and neural networks are explained, in particular, considering the architecture of transformer BERT models.

### 4.1 NLP definition

Natural language processing (NLP) refers to the branch of computer science, and more specifically, the branch of artificial intelligence or AI, concerned with giving computers the ability to understand text and spoken words as human beings can.

NLP combines computational linguistics and rule-based modeling of human language with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in text or voice data and ‘understand’ its whole meaning, complete with the speaker or writer’s intent and sentiment [27].

Classic use cases of NLP are:

- Customer service chatbot
- Speech-to-text dictation software
- Language Translation
- Text and document analysis

and so on.

## 4.2 NLP tasks

Human language is filled with ambiguities that make it incredibly difficult to write software that accurately determines the intended meaning of text or voice data. Homonyms, homophones, sarcasm, idioms, metaphors, grammar and usage exceptions, and variations in sentence structure are just a few of the irregularities of human language that take people years to learn, but that programmer must teach to natural language-driven applications to recognize and understand it accurately.

Several NLP tasks break down human text and voice data in ways that help the computer make sense of what it is ingesting. Some of these tasks include the following:

- *Classification*: it is one of the most common tasks in NLP. It consists of assigning a label or category to a given text. For example, we can classify emails as spam or not spam (Text Classification) and tweets as positive or negative (Sentiment Analysis) [36]
- *Text Preprocessing*: it involves transforming text into a clean and consistent format that can then be fed into a model for further analysis and learning. Text preprocessing techniques may include Feature Extraction, Part of Speech Tagging, Grammatical Error Correction, and Word Sense Disambiguation [50]
- *Chatbots*: it is a software program that can understand and respond to human speech. Bots powered by NLP allows people to communicate with computers naturally and human-likely, mimicking person-to-person conversations. Examples of chatbot uses are Slot Filling, Intent Detection, and Keyword Extraction [59]
- *Text-to-Text Generation*: It creates text by using a neural network to generate new text from a given input. With this power, a model can perform Machine Translation, Text Generation, Text Summarization, Text Simplification, and Lexical Normalization [13]
- *Text-to-Data and vice versa*: the base principle is the same as Text2Text Generation, but what changes are that input and output have different domains. Examples of Text2Data uses are Text2Speech, Text2Image, Text2Data and vice versa
- *Information Retrieval and Document Ranking*: it consists of retrieving information inside a sentence or a document through a Question Answering method or a Sentence Similarity Analysis
- *Knowledge Bases, Entities, and relations*: this task gives us humans the perception of intelligence behind a model. The machine gives the illusion of really understand-

ing the semantics and the meaning of words through Relation Extraction, Relation Prediction, Named Entity Recognition, and Entity Linking

- *Text Reasoning*: in NLP, it refers to a category of challenges where a system must apply logical reasoning or understanding to process and generate text
- *Topics & KeyWords*: it refers to tasks like keywords extraction, which consists of identifying words or phrases from the text that can be deemed as highly representative of the content, and topic modeling that involves identifying the underlying thematic structure in a collection of texts.

Figure 4.1 shows a map of all the NLP tasks and their belonging families.



Figure 4.1: Map of Natural Language Process tasks

### 4.3 Named Entity Recognition

Among all those tasks, we focus our attention on Named Entity Recognition (NER), which is the core of this project; in fact, the outcomes of our work depend on the performance of a Neural Network trained to perform NER on cybersecurity and the healthcare domain.

Named Entity Recognition (NER) aims to recognize mentions of rigid designators from text belonging to predefined semantic types such as person, location, organization [42]. NER not only acts as a standalone tool for information extraction (IE) but also plays an essential role in a variety of natural language processing (NLP) applications such as text understanding [66][12], information retrieval [22][47], automatic text summarization [45], question answering [39], machine translation [38], and knowledge base construction [15].

A *named entity* is a word or phrase identifying one item from a set of other items with similar attributes [52]. Namely named entities include organization, person, and location names in the general domain or gene, protein, drug, and disease names in the biomedical domain. NER is locating and classifying named entities in text into predefined entity categories.

Formally, given a sequence of tokens  $s = \langle w_1, w_2, \dots, w_N \rangle$ , NER gives as output a list of tuples  $s = \langle I_s, I_e, t \rangle$ , each of which is a named entity mentioned in  $s$ . Here,  $I_s \in [1, N]$  and  $I_e \in [1, N]$  are the start and the end indexes of a named entity mentioned;  $t$  is the entity type from a predefined category set. Figure 4.2 shows an example where a NER system recognizes three named entities from the given sentence [30]. A practical example of the output produced by a pre-trained NER model is shown in Figure 4.3.



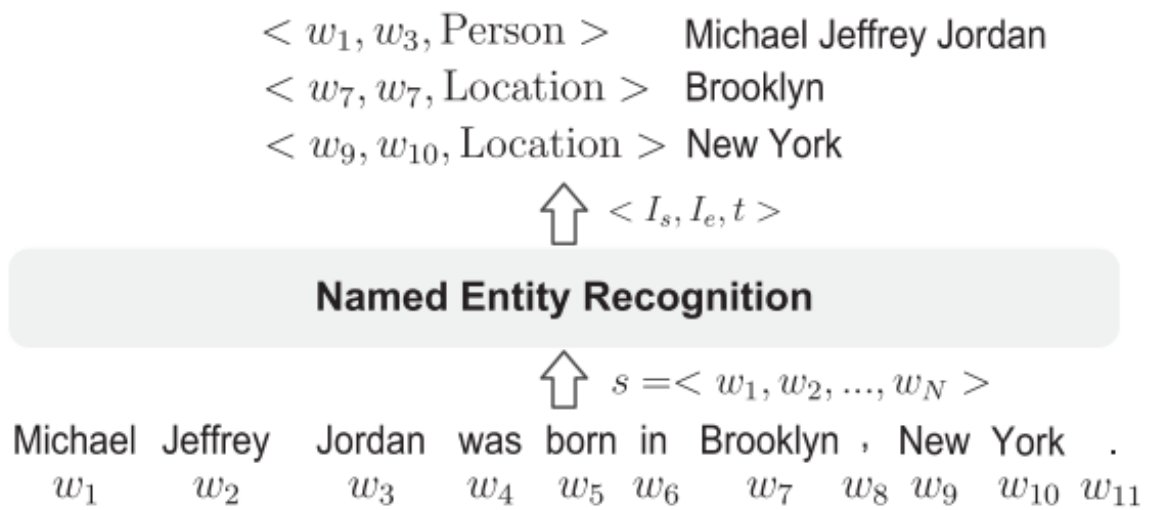


Figure 4.2: Visualization of named entity recognition task

Michael Jeffrey Jordan **PER** was born in Brooklyn **LOC** , New York **LOC** .

Figure 4.3: Output sample of a NER model trained on a generic domain

## 4.4 Evaluation Metrics

NER systems are usually evaluated by comparing their outputs against human annotations. The comparison can be quantified by either an exact or relaxed match. NER requires pinpointing both the limits of the entity and its classification. In the "exact-match evaluation" method, an entity is deemed correctly identified only when its boundaries and category align perfectly with the reference data.

Indeed, the "exact-match evaluation" criteria are less stringent. A relaxed match might allow for partial credit in cases where the entity type is correct, but the entity boundaries are not perfectly matched.

Classic metrics used in the literature and adopted in our model to evaluate a NER model's performance are:

- **PRECISION:** measures the ability of a NER system to present only correct entities

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \quad (4.1)$$

- **RECALL:** measures the ability of a NER system to recognize all entities in a corpus

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \quad (4.2)$$

- **F1-SCORE:** is the harmonic mean of precision and recall, and the balanced F-score is most commonly used

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.3)$$

In particular, True Positive is the number of entities that are recognized by NER and match ground truth, False Positive is the number of entities that are recognized by NER but do not match ground truth, and finally, False Negative is the number of entities annotated in the ground truth that NER does not recognize.

An efficient way to visualize the accuracy of a NER model is by building the confusion matrix that shows how many times each entity is matched correctly or mismatched with other entities. Figure 4.4 shows a confusion matrix example:

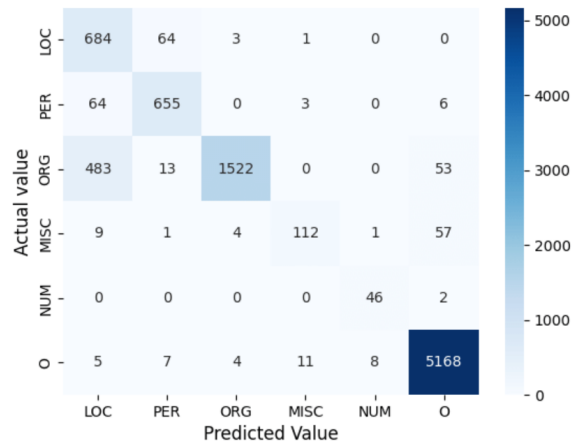


Figure 4.4: Example of a confusion matrix in a generic domain

## 4.5 NER challenges

NER models deal with some logistic problems. First of all, there is a lack of domain-specific annotated corpora. This is because the process of annotating takes a very long time because it is done manually. Over time, new semi-supervised frameworks and tools have been proposed to automate this as much as possible, but human supervision is always needed. Robert Bridges et Al. (2013) proposed an automatic labeling method for entity extraction from several data sources by leveraging article-specific structured data [8]. Stefano Silvestri et Al. (2022) proposed an iterative annotation method based on active learning (human effort plus machine learning techniques) and distant supervision (automatic procedure exploiting dictionaries and knowledge bases) activities [56].

Another challenge is that understanding natural languages is a hard goal due to all cultures' semantic and syntactic variations. Nevertheless, why do we have to take this road full of complications? Why can't we perform a perfect match analysis on a document or sentence instead of taking a neural network off the shelf? The response is simple. Performing a perfect match analysis implies having a vast dictionary with all the variations, abbreviations, synonyms, and genres of the possible words we want to detect. Moreover, this kind of analysis is purely syntactical and static and does not consider the context. A deep learning model, indeed, can convert the input text into a particular input that can capture the semantics of a sentence, considering also the words that appear before and after and, therefore, the context. This way allows us to reach a higher level of accuracy, asking to change a robust computation obtained through the state-of-the-art transformer neural networks.

## 4.6 NER tools: SpaCy

There are many NER tools available online with pre-trained models. Figure 4.5 summarizes popular ones for English NER by academia (top) and industry (bottom).

NER System	URL
StanfordCoreNLP	<a href="https://stanfordnlp.github.io/CoreNLP/">https://stanfordnlp.github.io/CoreNLP/</a>
OSU Twitter NLP	<a href="https://github.com/aritter/twitter_nlp">https://github.com/aritter/twitter_nlp</a>
Illinois NLP	<a href="http://cogcomp.org/page/software/">http://cogcomp.org/page/software/</a>
NeuroNER	<a href="http://neuroner.com/">http://neuroner.com/</a>
NERsuite	<a href="http://nersuite.nlplab.org/">http://nersuite.nlplab.org/</a>
Polyglot	<a href="https://polyglot.readthedocs.io">https://polyglot.readthedocs.io</a>
Gimli	<a href="http://bioinformatics.ua.pt/gimli">http://bioinformatics.ua.pt/gimli</a>
spaCy	<a href="https://spacy.io/api/entityrecognizer">https://spacy.io/api/entityrecognizer</a>
NLTK	<a href="https://www.nltk.org">https://www.nltk.org</a>
OpenNLP	<a href="https://opennlp.apache.org/">https://opennlp.apache.org/</a>
LingPipe	<a href="http://alias-i.com/lingpipe-3.9.3/">http://alias-i.com/lingpipe-3.9.3/</a>
AllenNLP	<a href="https://demo.allennlp.org/">https://demo.allennlp.org/</a>
IBM Watson	<a href="https://natural-language-understanding-demo.ng.bluemix.net/">https://natural-language-understanding-demo.ng.bluemix.net/</a>
FG-NER	<a href="https://fgner.alt.ai/extractor/">https://fgner.alt.ai/extractor/</a>
Intellexer	<a href="http://demo.intellexer.com/">http://demo.intellexer.com/</a>
Repustate	<a href="https://repustate.com/named-entity-recognition-api-demo/">https://repustate.com/named-entity-recognition-api-demo/</a>
AYLIEN	<a href="https://developer.aylien.com/text-api-demo">https://developer.aylien.com/text-api-demo</a>
Dandelion API	<a href="https://dandelion.eu/semantic-text/entity-extraction-demo/">https://dandelion.eu/semantic-text/entity-extraction-demo/</a>
displaCy	<a href="https://explosion.ai/demos/displacy-ent">https://explosion.ai/demos/displacy-ent</a>
ParallelDots	<a href="https://www.paralleldots.com/named-entity-recognition">https://www.paralleldots.com/named-entity-recognition</a>
TextRazor	<a href="https://www.textrazor.com/named_entity_recognition">https://www.textrazor.com/named_entity_recognition</a>

Figure 4.5: Tools list for Named Entity Recognition task

In our project, we adopted SpaCy, a free, open-source library for advanced Natural Language Processing (NLP) in Python. SpaCy is explicitly designed for production use and helps build applications that process and “understand” large volumes of text. It can be used to build information extraction or natural language understanding systems or to pre-process text for deep learning. SpaCy provides many components that let us execute different activities. Those components are put one after another in a so-called pipeline, as shown in Figure 4.6.

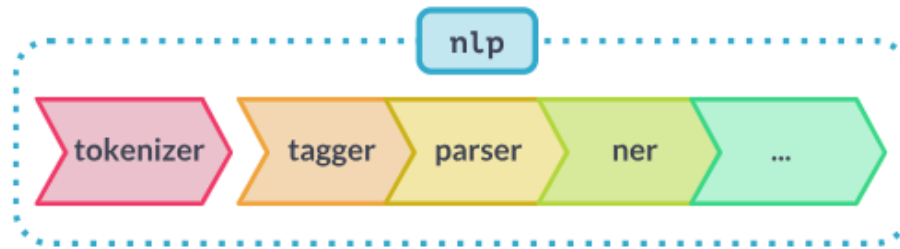


Figure 4.6: SpaCy's pipeline

SpaCy's tagger, parser, text categorizer, and many other components are powered by statistical models. Every “decision” these components make – for example, which part-of-speech tag to assign or whether a word is a named entity – is a prediction based on the model's current weight values. The weight values are estimated based on examples the model has seen during training. Training data on examples of text and the labels we want the model to predict is needed to train a model. This could be a part-of-speech tag, a named entity, or any other information.

Training is an iterative process in which the model's predictions are compared against the reference annotations to estimate the gradient of the loss. The gradient of the loss is then used to calculate the gradient of the weights through backpropagation. The gradients indicate how the weight values should be changed so that the model's predictions become more similar to the reference labels over time. Those concepts will be better explained in Section 4.7.

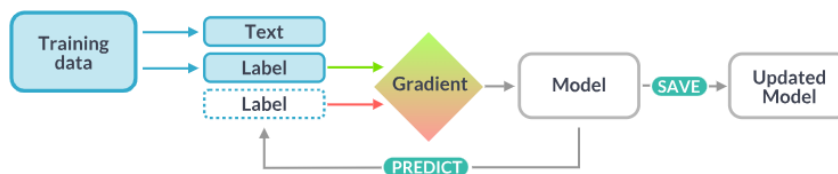


Figure 4.7: Training workflow in SpaCy

When training a model, we do not just want it to memorize our examples – we want it to develop a theory that can be generalized across unseen data. After all, we do not just want the model to learn that this one instance of “Amazon” right here is a company – we want it to learn that “Amazon”, in contexts *like this*, is most likely a company. That is why the training data should always be representative of the data we want to process. A model trained on Wikipedia, where sentences in the first person are sporadic, will likely perform

poorly on Twitter. Similarly, a model trained in romantic novels will likely perform poorly on legal text.

This also means that in order to know how well the model is performing and whether it is learning the right things, we do not only need training data – we will also need evaluation data. If we only test the model with the data it was trained on, we will have no idea how well it generalizes. If we want to train a model from scratch, we usually need at least a few hundred examples for both training and evaluation [19].

SpaCy supports several transfers and multi-task learning workflows that can often help improve the pipeline’s efficiency or accuracy. Transfer learning refers to techniques such as word vector tables and language model pretraining. These techniques can be used to import knowledge from raw text into the pipeline so that models can generalize better from annotated examples. Specifically, we used BERT transformers to train our model. In the following, we will analyze this complex and robust network in detail.

## 4.7 Artificial Neural Network

Artificial neural networks (ANNs) [25] are computational models inspired by the neuronal structure of animal brains. These networks consist of artificial neurons or nodes connected by edges, much like the synaptic connections in biological brains. Each neuron processes incoming signals and can pass the output to other neurons. The strength of these signals is modulated by weights that are adjusted during learning, with the possibility of a threshold for activation. To perform complex computations, ANNs are typically organized in layers that transform inputs and pass signals through the network from the input layer to the output layer, often with multiple passes through the layers.

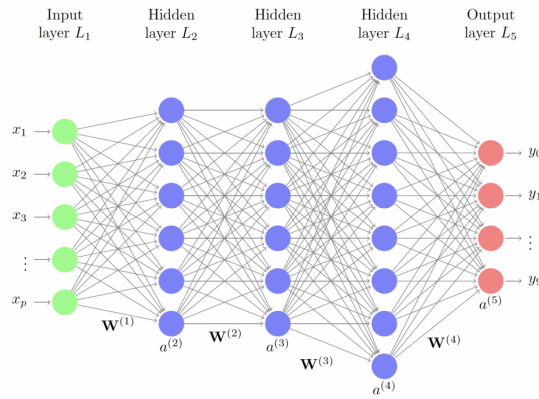


Figure 4.8: Typical structure of an ANN

Figure 4.8 shows a typical structure of an artificial neural network.

The transition from layer  $k-1$  to layer  $k$  is:

$$\mathbf{Z}^{(k)} = \mathbf{W}^{(k-1)} \mathbf{a}^{(k-1)} \quad (4.4)$$

$$\mathbf{a}^{(k)} = g^{(k)}(\mathbf{Z})^{(k)} \quad (4.5)$$

where:

- $\mathbf{W}^{(k-1)}$  represents the matrix of weights that go from layer  $L_{k-1}$  to layer  $L_k$
- $\mathbf{a}^{(k)}$  is the vector of activations at layer  $L_k$  where  $a_l^{(1)} = x_l$



### 4.7.1 Working of ANN

The input layer receives the raw data. Each neuron in this layer represents a feature of the input data. After the input layer, there can be one or more hidden layers where the actual processing is done through a system of weighted connections. The layers are "hidden" because they are not directly exposed to the input or output. The last layer is the output layer, which presents the final output. The structure of the output layer depends on the specific task (e.g., classification, regression). Neurons use activation functions to transform the input into an output. Typical functions include the `sigmoid`, hyperbolic tangent (`tanh`), and rectified linear unit (ReLU). These functions introduce non-linear properties to the network, allowing it to learn complex data patterns.

### 4.7.2 Learning process

Learning is described as the process of the network adjusting its weights and, if applicable, thresholds to better perform a specific task based on sample data. This adjustment minimizes errors between the network's predictions and actual outcomes. Learning is complete when additional data does not significantly reduce the error rate. Typically, even after learning, the error rate will not be zero. This learning process is often guided by a cost function, which measures how well the network performs. The learning continues as long as the value of the cost function is decreasing. The cost function is usually a statistical measure that can only be approximated, not calculated directly. Since the network outputs are numerical, a low error rate means the network's numerical output is close to the correct value. The goal of learning is to minimize the sum of differences between the network's predictions and the correct answers across all observations.

In the following, the main methods used to improve the learning process are explained:

- **Backpropagation:** it is the central algorithm for learning in ANNs. The theoretical foundation of backpropagation is deeply rooted in calculus. The process involves using derivatives to compute the gradient of a loss function, which measures the difference between the network's prediction and the actual target values. The network gradually improves performance by adjusting the weights in a direction that minimally decreases this loss. The backpropagation process unfolds through several stages. Initially, input is fed through the network in the forward pass. This input travels through various layers, each consisting of neurons that apply a weighted sum to their inputs and pass the result through a non-linear activation function. The final output is then compared to the target value, and a loss is computed. Follow-

ing this, the backward pass begins. Here, the algorithm calculates the gradient of the loss function with respect to each weight in the network. This calculation is conducted backward, starting from the output layer and moving towards the input layer. The chain rule of calculus is the mathematical principle that facilitates this gradient computation. Once the gradients are determined, the network's weights are updated. This is usually achieved through optimization algorithms like gradient descent. The updates are made in the opposite direction of the computed gradient, and the magnitude of these updates is governed by a key parameter known as the learning rate. The process of forward pass, loss calculation, backward pass, and weight update is iteratively repeated across numerous epochs, each involving multiple inputs and their corresponding targets. Through this repetitive adjustment, the network incrementally reduces loss and improves its ability to predict or classify data accurately.

- **Forward Propagation:** data is passed through the network from the input layer to the output layer. As the data moves from the input layer to the next, every neuron in the subsequent layer computes a weighted sum of its inputs. These weights are parameters reflecting the strength of connections between neurons, which the network aims to learn during its training phase. A bias term is added along with this weighted sum, serving as another adjustable parameter in the learning process. After computing the weighted sum and adding the bias, the neuron applies an activation function to this result. The activation function, often non-linear, introduces complexity and non-linearity into the network, enabling it to capture and model more complex patterns in the data. Standard activation functions include `sigmoid`, `hyperbolic tangent (tanh)`, and `Rectified Linear Unit (ReLU)`. The output from the activation function of each neuron then becomes the input for the next layer. With each layer performing its calculations, this forward movement of data continues until the data reaches the output layer. In the output layer, the process remains essentially the same. However, the activation function here may differ based on the specific task, such as using a `softmax` function for classification tasks or a linear function for regression tasks.
- **Gradient Descent:** This optimization algorithm adjusts weights to minimize the loss function. The process begins with the initialization of the model parameters, often randomly. Then, the algorithm computes the gradient of the loss function at the current point in the parameter space. This gradient is a vector consisting of the partial derivatives of the loss function with respect to each parameter and points in the direction of the steepest ascent. In the next step, the algorithm

updates the parameters by moving them a small step in the opposite direction of the gradient. The size of this step is determined by the learning rate, a hyperparameter that controls how big a step the algorithm takes. A high learning rate allows the algorithm to converge quickly but risks overshooting the minimum. Conversely, a low learning rate ensures a more precise convergence but at the cost of increased computational time. The updated parameters are then used in the next iteration, where the loss and gradient are recalculated. This process is repeated until the algorithm converges to a minimum, indicated by a negligible change in the loss or the gradient over iterations. Variants of gradient descent, such as Stochastic Gradient Descent (SGD), are commonly used.

- **Loss Function:** it measures the difference between the network's prediction and the actual target values. This measure of error is used to adjust the weights during training.
- **Hyperparameters:** a hyperparameter is a constant parameter whose value is set before the learning process begins. The values of parameters are derived via learning. Examples of hyperparameters include learning rate, the number of hidden layers, and batch size. The values of some hyperparameters can be dependent on those of other hyperparameters.

## 4.8 BERT: Deep Bidirectional Transformers for Language Processing

Transformers are a family of neural network architectures that compute dense, context-sensitive representations for document tokens. Downstream models in the pipeline can then use these representations as input features to improve their predictions.

Jacob Devlin et Al. [14] proposed in 2019 the BERT model, which still is the state-of-art deep learning model for a range of tasks of NLP and, among them, Named Entity Recognition. Its architecture is a multi-layer bidirectional Transformer encoder based on the original implementation described in Vaswani et al. (2017) [60] and released in the tensor2tensor library. Figure 4.9 shows in detail its components:

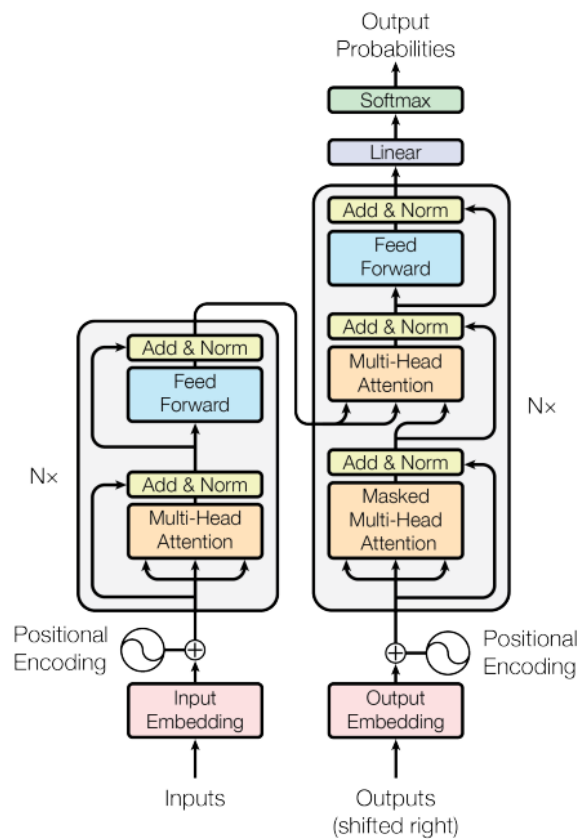


Figure 4.9: BERT's architecture

### 4.8.1 Input Embeddings & Positional Encoding

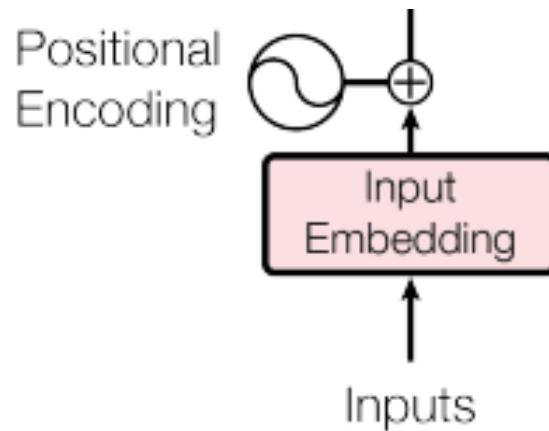


Figure 4.10: Embedding & encoding components

This phase transforms a text input, a string, into something the machine understands: an embedding. An embedding consists of encoding a word's features, which could be the position in the text, the following words, and the semantics. Since a neural network is a complex maths algorithm, embeddings are numeric vectors from which it is easy for the machine to discover features under the footsteps of patterns.

There exist different types of embeddings that take into consideration different aspects of a word inside a text:

- Bag Of Words: each value of the vector represents the number of words in a sentence or a document
- One Hot Encoding: each word is identified with a vector in which every element is a word of the dictionary. A word has a corresponding vector with all zeros and only one corresponding to its position in the dictionary
- Word2Vec: it captures the associations between words based on the hypothesis that close words have semantic similarities. CBOV and SKIPGRAM are two different variants of this embedding
- Deep Autoencoder: it reduces the input size with a complex representation that keeps the most important features of the data.

This last embedding is the one selected for BERT and can capture a text's semantics. It considers features of the single word, the sentence in which a word is present, and the word's position inside a text. The positional encoding is significant since it substitutes

the temporal dimension with a positional one that gives us information about the order of the words inside a text. Figure 4.11 helps to visualize the complexity of the embedding:

Token Embeddings and Segment Embedding consist of a simple association of a number to a word, which is the number of the position of the word inside the dictionary and the number of the segment inside a text. Positional Encoding for position  $p$  and dimension  $i$  is defined as:

$$PE(p, 2i) = \sin\left(\frac{p}{10000^{\frac{2i}{d}}}\right) \quad (4.6)$$

Where:

- $p$  is the position of the word/token in the sequence.
- $i$  ranges from 0 to  $\frac{d}{2} - 1$ , where  $d$  is the dimension of the model.

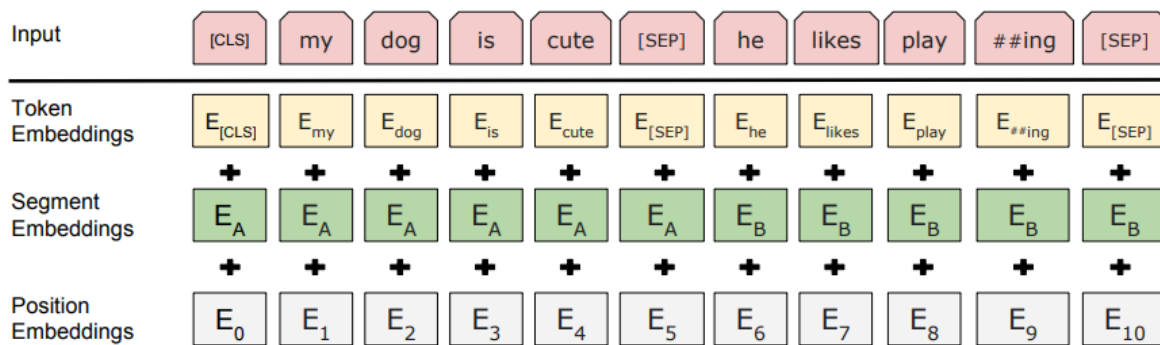


Figure 4.11: Embeddings of a natural language input

## 4.8.2 Multi-Head Attention and Add&Norm

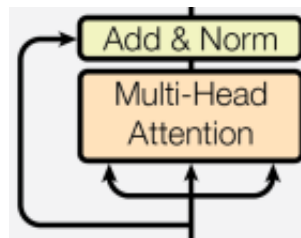


Figure 4.12: Multi-Head Attention & Add&Norm components

The attention mechanism is the true core of BERT. In this phase, a query vector  $Q$  is compared with a set of keys to determine their compatibility. Each vector key  $K$  is coupled with a value vector  $V$ . The higher the compatibility between a query and a key, the higher the influence of the corresponding value on the output of the attention mechanism.  $Q$ ,  $K$ , and  $V$  are learned during the training of the model. Considering, for example, the first word of a text, a dot product is performed between  $Q$  and each element of  $K$ ; the higher the product, the higher the compatibility between the query and the key. In this way, we obtain a weights vector  $A = \langle \alpha_0, \alpha_1, \dots, \alpha_n \rangle$  normalized through a softmax function.

$$\text{softmax}\left(\frac{\langle \alpha_0, \alpha_1, \dots, \alpha_n \rangle}{\sqrt{d_n}}\right) \quad (4.7)$$

where:

- $d_n$  is the dimensionality of the  $Q$  and  $K$  vectors.

In this way, the weights become nonnegative, and their dimensionality is scaled up to one of the query and key vectors to avoid problems with the gradient descent.

Once normalized, a linear combination between  $A$  and  $V$  is created, representing the output of the attention mechanism. The output is a matrix in which each row is an updated representation of the word in the corresponding position in the sequence. We can summarize the attention mechanism with the following function:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_K}}\right) \cdot V \quad (4.8)$$

Up to now, we have talked about referring only to a single attention block. In reality, this operation is applied to simultaneous representations of the input. Finally, normalization is applied to add the input to the output to get mean  $\mu = 0$  and variance  $\delta = 1$ .

### 4.8.3 FeedForward and Add&Norm

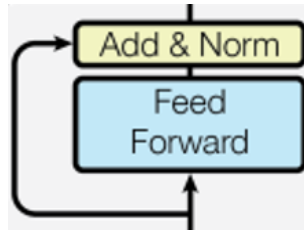


Figure 4.13: Feed Forward & Add&Norm components

This block consists of two layers of fully connected hidden nodes with a RELU activation function specified below.

$$RELU(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.9)$$

Another input addition and normalization, like in the previous step, is applied to the output of the feed-forward network.

A stack of  $N_X$  of these components could be allocated to have different encoders for each input representation. In this way, the model can combine different learning in different settings and embedding variations. For the purposes of our research, the decoder block will not be discussed in detail since, for the NER task, only the decoder block is needed. Nevertheless, it is important to highlight that a Masked Language Model (MLM), which randomly masks some of the tokens from the input and the target label, is used to predict the original vocabulary ID of the masked word based only on its context. Unlike left-to-right language model pre-training, the MLM objective enables the representation to fuse the left and the proper context, which allows us to pre-train a deep bidirectional Transformer.



## Chapter 5

# Proposed method

This chapter explains the materials and methods fundamental to our research. It also highlights the techniques used to obtain the data to train the model. In Sections 5.5 and 5.6, our model is described with all the hyperparameters used.

Here, Figure 5.1 and Figure 5.2 show the workflow of the proposed method.

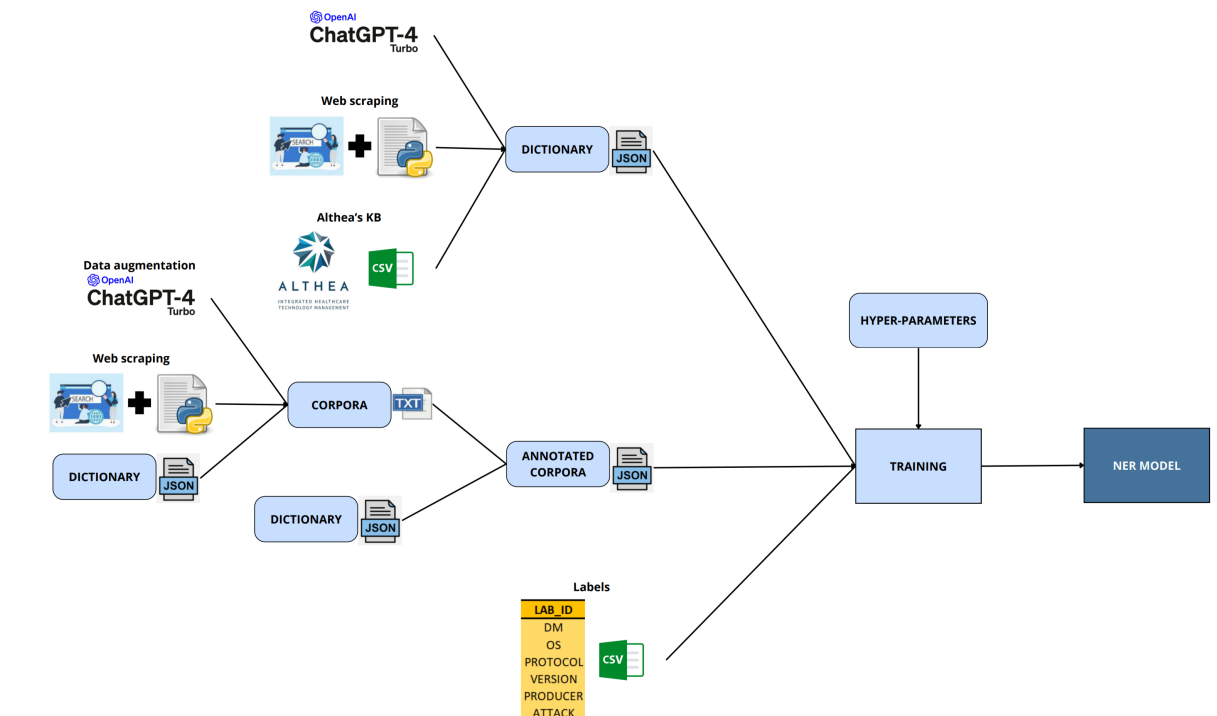


Figure 5.1: Work Flow - NER Model

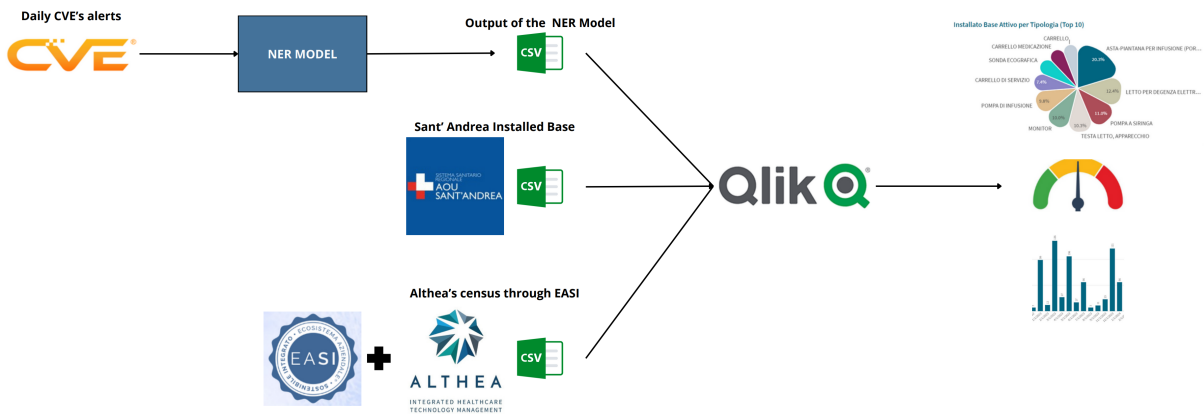


Figure 5.2: Work Flow - Business Intelligence

## 5.1 Data Requirements Identification

In order to train a Neural Network, we need to define and prepare the following data:

1. **Labels:** they are the model's output and represent the classes of terms that the model aims to identify.
2. **Dictionaries:** for each label, a dictionary of words is needed to provide some examples to learn from. The more exhaustive a dictionary is, the higher the number of words covered and the higher the precision of the model.
3. **Training & Test sets:** they consist of an annotated corpora, a text in which words are marked with their class label. The dataset will be split into Training, Validation, and Test sets following the 70%-15%-15% proportion.

## 5.2 Labels

The labels that we want the model to predict are six:

1. Medical Device
2. Operating System
3. Protocol
4. Version
5. Producer
6. Cyber-Attack

The selection of the labels is important since those attributes will be used as input parameters for evaluating the risk index in the business intelligence phase. Therefore, we choose to consider those features:

- Producer, Version, Protocols, and Name of operating systems and Software Medical Devices (SDM)
- Producer, Version and Name of Hardware Medical Devices (HDM)
- Attack type, Vulnerability, and threat of DMs

Table 5.1 wraps up all the information the model extracts split by category.

Entity List		
Hardware DM	Software DM	CyberSecurity
Producer	Producer	Attack
Version	Version	
HDM Name	SDM Name	
	Protocol Name	

Table 5.1: Information extracted by category

The response variable and their explanation are showed in Table 5.2:

LAB_ID	Explanation
DM	Medical device
OS	Operating system
PROTOCOL	Protocol
VERSION	Version
PRODUCER	Software or hardware producer
ATTACK	Cyber attack, vulnerability or threat

Table 5.2: Label Id and their explanation

### 5.3 Dictionaries

A Dictionary is a set of words from each class collected in a JSON list. For every label, we have identified words of its specific domain. To obtain the list of words, we followed the steps outlined below:

1. Website scraping through a specific Python script based on the type of the queried site (e.g., Wikipedia) and words looked for (e.g., versions)
2. Enriching the words' list with specific and supervised prompts to Chat-GPT4
3. Merging all the previous results and cleaning the final words' list with a supervised Python script

Those steps were followed for all the categories, obtaining six consistent and numerous dictionaries. We could enrich the dictionaries with information belonging to Althea's official catalogs and censuses from Sant'Andrea Hospital for the DM, OS, and PROD labels. In this way, we could add perfect matching names in the dictionaries.

Table 5.3 shows the size of each dictionary.

LAB_ID	Dictionary's Name	Size
DM	medical_devices.json	212
OS	operating_systems.json	90
PRTCL	protocols.json	394
VER	versions.json	146
PROD	producers.json	429
ATT	cyber_attacks.json	141

Table 5.3: Dictionary's size by Label

### 5.3.1 Scraping instruments

#### Chat-GPT4

Chat-GPT4 has been used in the data augmentation phase, and the interactions have been crafted using the "Prompt Perfect" plugin, which allowed us to query the AI model always using the best prompt possible to obtain an exhaustive and comprehensive list of words. Figure 5.3 shows a sample query crafted with the Prompt Perfect plugin for words belonging to the DM producers' family.

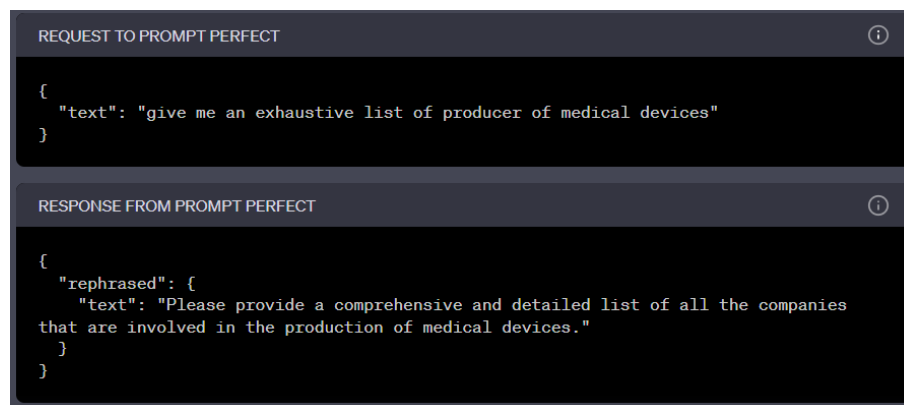


Figure 5.3: Sample of Chat-GPT4 prompt

## Beautiful-Soup Library

Words have been collected with a scraping Python code that exploits the BeautifulSoup library, which allows us to obtain the XML structure of an internet page, knowing its URL. Once we obtain the page's XML, we can browse its elements and content, exploiting the `.text()` attribute. In order to scrape the list of entities that our model has to recognize, we have accessed Wikipedia, which usually groups them in 2 main standard ways: by standard list and by table. Knowing that information, we could write a Python code that browses and extracts the content of `<li></li>` tags and `<td></td>` tags, which denote, respectively, a list element and a cell element. After this operation, we have a list of elements that need to be cleaned. For this purpose, a supervised method is used with an automated method that exploits regular expression, which aims to clean and standardize as much as possible all the elements scraped. Since we are creating six different dictionaries, an ad-hoc Python code must be executed for each label category.

In the end, some common manipulation has been performed:

- **Duplicate elimination:** duplicated items in the dictionary list are deleted
- **Data enrichment:** some keywords are added before or after the word element, like 'OS' ('Windows' becomes 'Windows OS') or 'protocol' ('TCP' becomes ' TCP protocol')

In particular, the data enrichment operation allows us to add some context to words, trying to forecast a future occurrence in a brand-new input sample.



## 5.4 Training and Test Set

This section explains how a complete dataset was obtained starting from a blank slate. A NLP dataset consists of a text, called corpora, whose words are annotated with a specific annotating template that considers the label and the word's position inside a text.

The annotations need to follow a specific SpaCy format consisting of a list of couples that, for a test, report the label of the entities identified and the initial and final index of the initial and final corresponding character. Here is an example:

*The endpoint has installed Windows 98 and communicates with a Pulse Oximeter from Philips. {[ 'OS', (27, 36)], [ 'DM', (62, 68)], [ 'PRODUCER', (82, 88)]}*

Unfortunately, the state-of-the-art annotated datasets we can find online on different data science platforms are relative to specific domains that consist of the annotation of 4 entities: PERSON, GPE, DATE, and MISC. Finding an already annotated corpora with a consistent size for our specific domain of Healthcare technologies is difficult, so we created our annotated corpora.

The first step was to obtain a domain-specific text, which should have had an important size with sufficient occurrences of the dictionary's words to train the model. In order to reach this goal, we followed two ways:

1. Scraping from cybersecurity websites:  
this first method allowed us to cover a consistent proportion of OS, PRODUCER, VERSION, PROTOCOL, and ATTACK dictionaries
2. Querying Chat-GPT4 to generate custom text:  
this method, indeed, lets us craft a particular domain text with the assurance that every word present in the dictionary would have been cited a requested number of times

The forum selected to perform web scraping is packetstormsecurity, an information security website offering current and historical computer security tools, exploits, and security advisories. It is operated by a group of security enthusiasts that publish new security information and offer tools for educational and testing purposes.

Table 5.4 reports some insight into the scraping task.

Source	Total pages	Article per page	Total article scraped	Total words
PacketStorm forum	5125	25	128100	4133655

Table 5.4: Size of scraped corpora

Meanwhile, Chat-GPT4 has been queried with five prompts, reported in Table 5.5, per label, and the requests were used to generate a domain-specific text with the requirement of citing at least two times each word of a provided list.

Prompts submitted to generate a customize text	
<b>P1:</b>	<i>Write a long text using more than once the words contained in the following list: [...]</i>
<b>P2:</b>	<i>Do it again using all the elements in the list</i>
<b>P3:</b>	<i>Do it again</i>
<b>P4:</b>	<i>Do it again, be exhaustive</i>
<b>P5:</b>	<i>Do it again, be repetitive</i>

Table 5.5: Chat-GPT4 prompts used for text augmentation

With P1, we assign the task to the model providing the batched list; P2 guarantees we do not forget any word. P3, P4, and P5 make the model generate new texts with more possible repetitions.

Table 5.6 shows the size of the data augmented corpora.

Source	Total sentences	Words per sentence
GPT4/versions_corpora	49	9958
GPT4/protocols_corpora	50	50281
GPT4/producers_corpora	50	30267
GPT4/operating_systems_corpora	58	7894
GPT4/medical_device_corpora	56	16417
GPT4/cyber_attacks_corpora	54	19227

Table 5.6: Corpora's size by label

ChatGPT4 is the most efficient language model up to now able to generate a long sequence of text with a bit of context, so we decided to use it.

Nevertheless, the problem of all NLP models is attention, which means that an input prompt cannot be as long as we want because, otherwise, the machine will not "understand" all of it and will lose some information. Because of that, we could not give as input the complete list of each dictionary, but we had to split them up into batches and perform five queries on a single batch at a time. Table 5.7 reports the batch sizes for each dictionary list:

Dictionary Name	Dictionary size	Batch size
medical_devices.json	212	35
operating_systems.json	90	30
protocols.json	394	35
versions.json	146	30
producers.json	429	35
cyber_attacks.json	141	20

Table 5.7: Size of bathced items used for each Chat-GPT4 prompt

Once a corpus is obtained, dictionaries have been used to annotate it by marking all the words with the corresponding label among the initial six plus the 'O' labels that stand for non-entity. This way of tagging is necessary and required by the SpaCy pipeline, which needs to receive input patterns. A pattern is an association between a word and its label; if a word does not have a class, then it is marked with the O label.

The dataset has finally been split in training, validation, and test sets following the proportion 70%-15%-15%.

## 5.5 The model

As we have seen before, a traditional SpaCy pipeline is composed by:

- **Tokenizer:** to break the input text into individual tokens. It takes a string of text and produces a sequence of token objects. Each token object contains information about the token's text, position in the original string, and more
- **Tagger:** to assign part-of-speech tags to each token. It assigns part-of-speech tags based on the definitions from the Universal Dependencies scheme.
- **Parser:** to identify syntactic dependencies between tokens. It assigns syntactic dependency labels to token pairs, describing their grammatical relationship, and it can also determine the hierarchical structure of phrases or sentences.
- **Named Entity Recognizer:** to identify and classify named entities in the text into predefined categories. Based on context, it uses statistical models to predict the boundaries and types of named entities.
- **Others:** other blocks that can be integrated to perform different tasks.

These blocks allow SpaCy's users to perform a wide variety of NLP tasks; however, we only need the tokenizer and the named entity recognizer to perform the NER task. The other components generate information that we do not want to process.

The input is a raw text, simply a string of text we want to process. It can be a sentence, a paragraph, or even longer documents. The raw text does not inherently have linguistic annotations; it is just a sequence of characters. Once the raw text is processed through the pipeline, it is converted into a `Doc` object containing many linguistic annotations. The `Doc` object is a central data structure in SpaCy. Here are the primary features:

- The `Doc` object is a sequence of token objects. Each token corresponds to a word or punctuation from the original text.
- Each token has an associated part-of-speech tag, accessed with the `token.pos_` attribute.
- Syntactic relationships between tokens can be accessed using `token.dep_` and `token.head`.
- Named entities are stored as spans within the `Doc` object, accessible with `doc.ents`. Each entity has a label accessed with the `ent.label_` attribute.
- The `Doc` object can be segmented into sentences using `doc.sents`.

- If word vectors are loaded, each token and the Doc can have an associated vector.
- The Doc object and tokens have other attributes like lemma (`token.lemma_`), morphological analysis (`token.morph`), etc.
- Custom metadata can be added to the Doc, Token, and Span objects through extensions.

## 5.6 The parameters

Hyperparameters are variables whose values control the learning process and determine the values of model parameters that a learning algorithm learns. Changing one or more of them can directly impact the model's performance. To find the best parameters for our project, we performed fine-tuning starting from the values suggested by SpaCy's website; in fact, it provides users with a free tool to automatically generate a configuration file based on their needs. Parameters like `batch size`, `hidden width`, and `max steps` have been selected by the needs of the computational resources that we had available, indeed, the model has been trained in Google colab environment, taking advantage of GPU. Despite the high and fast performances of the processor, we had to keep low the previous mentioned parameters to avoid an excessive usage of memory.

The final model implements the following hyperparameters.

- Dropout = 0.2
- Patience = 1600
- Evaluation frequency = 200
- Max steps = 20000
- Optimizer = Adam.v1
- Learning rate = 0.001
- Hidden width = 64
- Activation function = Maxout()
- Batch size: 1000
- Use upper = false

**Dropout** is a regularization technique used in neural networks to prevent overfitting. The value 0.2 means that during training, a random 20% of the nodes in the network are "dropped out" or temporarily deactivated at each iteration. This helps the model generalize better to unseen data.

**Patience** is an early stopping criterion. It refers to the number of evaluations without improvement on the development set before halting training. Training will be stopped if the model's performance does not improve for 1600 evaluations. This prevents overtraining and saves computational resources.

**Evaluation frequency** specifies how often the model's performance should be evaluated on the development set during training. For example, if it is set to 200, the model's performance will be evaluated every 200 update steps.

**Max steps** sets the maximum number of update steps the training process will take. Training will stop after 20,000 update steps, regardless of other stopping criteria.

The optimizer determines how the model updates its weights based on the error it produces. **Adam** is a popular optimization algorithm for NER that adjusts the learning rate for each parameter individually.

The **learning rate** controls the size of the updates to the model's weights. A value of 0.001 means the model will make relatively small weight adjustments with each update, ensuring more stable and gradual learning.

**Hidden width** sets the width (or size) of the hidden layers in the neural network. A hidden width of 64 means that 64 nodes (or units) are in the hidden layer.

The activation function introduces non-linearity into the model, allowing it to learn more complex patterns. **Maxout** is a piecewise linear activation function that takes the maximum of its inputs. It works well in deep learning models and reduces the risk of vanishing gradients.

**Batch size** indicates the number of examples (documents) processed together in one batch during training. A batch size of 1000 means 1000 examples are processed at once.

The **use\_upper** parameter refers to whether the model should consider the tokens' case (uppercase/lowercase). We decided to set it false since we do not need the model to differentiate between, for example, "apple" and "Apple".





## Chapter 6

# Experimental evaluations

This chapter presents the experiment results conducted as part of this research study. Here, the performance of the trained models on the test data is shown and commented on the improvements made to the network.

In the final sections, we explain the method used to evaluate the cyber risk index with all the procedures used to compute the static and dynamic indexes of risk.

### 6.1 Network output and evaluation

The first models trained outlined good performance. Despite that, when giving input to some samples of the CVE, the network did not work well, missing some entities. This is because of overfitting during the training. Table 6.1 shows the network's precision, recall, and F1-score performances.

Table 6.1: First NER model biased by overfitting

Entity Type	Precision	Recall	F1-Score
Overall	0.9867	0.9884	0.9875
DM	0.9968	0.9963	0.9965
ATTACK	0.9856	0.9973	0.9914
PROTOCOL	0.9758	0.9262	0.9504
VERSION	0.9748	0.9990	0.9867
PRODUCER	0.9450	0.9476	0.9463
OS	0.9631	0.8757	0.9173

Another indicator of overfitting is the almost perfect confusion matrix shown in Figure 6.1. Only protocols seemed to be mispredicted as cyber attacks.

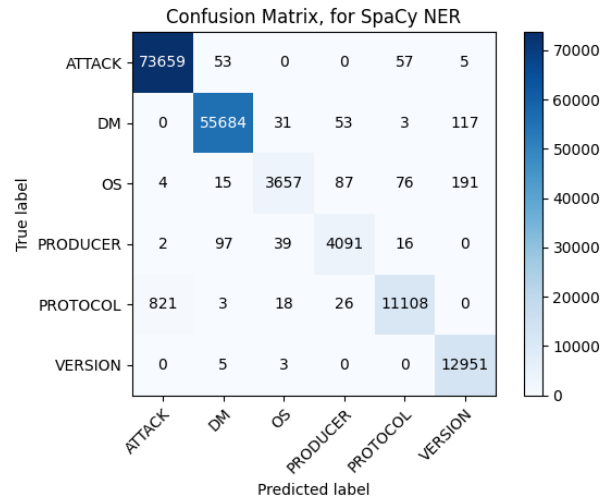


Figure 6.1: Confusion matrix of the initial model

The following graph highlights the trend of the loss function associated with the `tok2vec` and `NER`. The loss function represents the difference between predicted and actual output. A lower loss indicates that the model's predictions are closer to the actual values.

The `tok2vec` layer is responsible for converting tokens (words, punctuation) into vectors (numerical representations). This loss value indicates how well this layer is performing its task.

The loss for the `NER` (Named Entity Recognition) component of the pipeline indicates how well the `NER` model is performing its task of recognizing named entities.

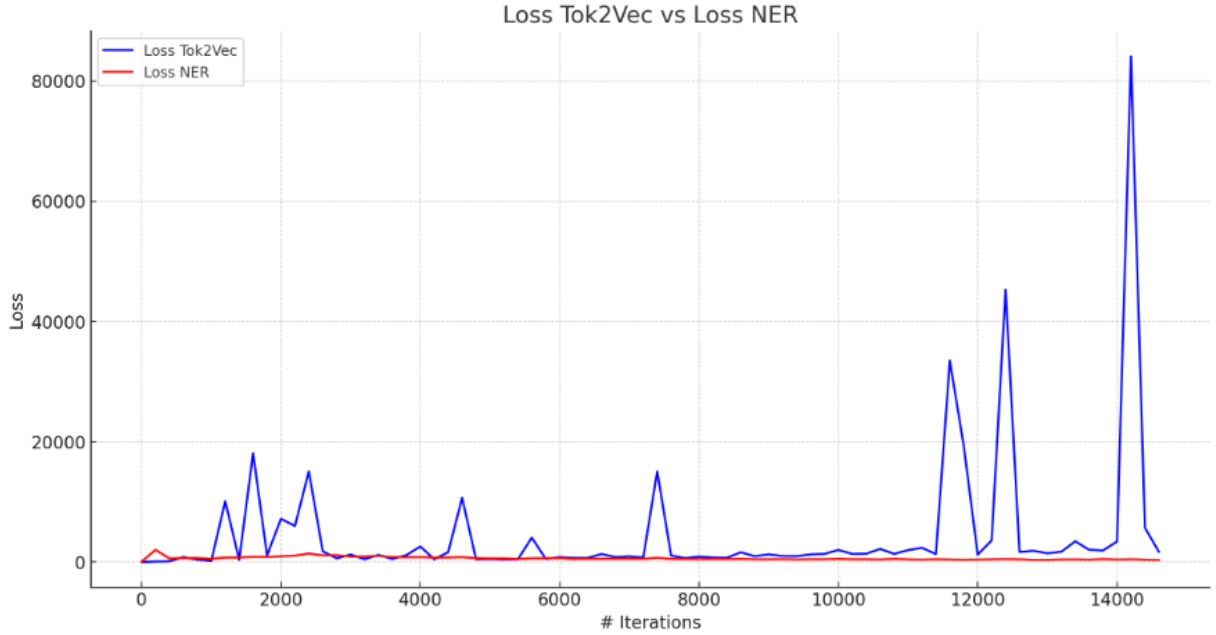


Figure 6.2: Loss functions trends of the initial model

Due to overfitting, we decided to clear the dictionaries by erasing the stop words like articles, prepositions, symbols, and so on. We also decided to remove some words that were too specific domain and, instead, we left more generic words. We then trained the network on smaller but more consistent dictionaries. Moreover, we converted all the words from uppercase to lowercase but capitalized. This improved the training, limiting the overfitting and giving more realistic performances, as shown in Table 6.2.

Table 6.2: NER Model Performance Metrics after dictionary cleaning

Entity Type	Precision	Recall	F1-Score
Overall	0.9774	0.9778	0.9776
ATTACK	0.9892	0.9947	0.9920
OS	0.9531	0.8872	0.9190
PROTOCOL	0.9935	0.9984	0.9959
PRODUCER	0.9722	0.8756	0.9214
VERSION	0.9927	0.9919	0.9923
DM	0.8697	0.9462	0.9064

The related confusion matrix is reported in Figure 6.3.

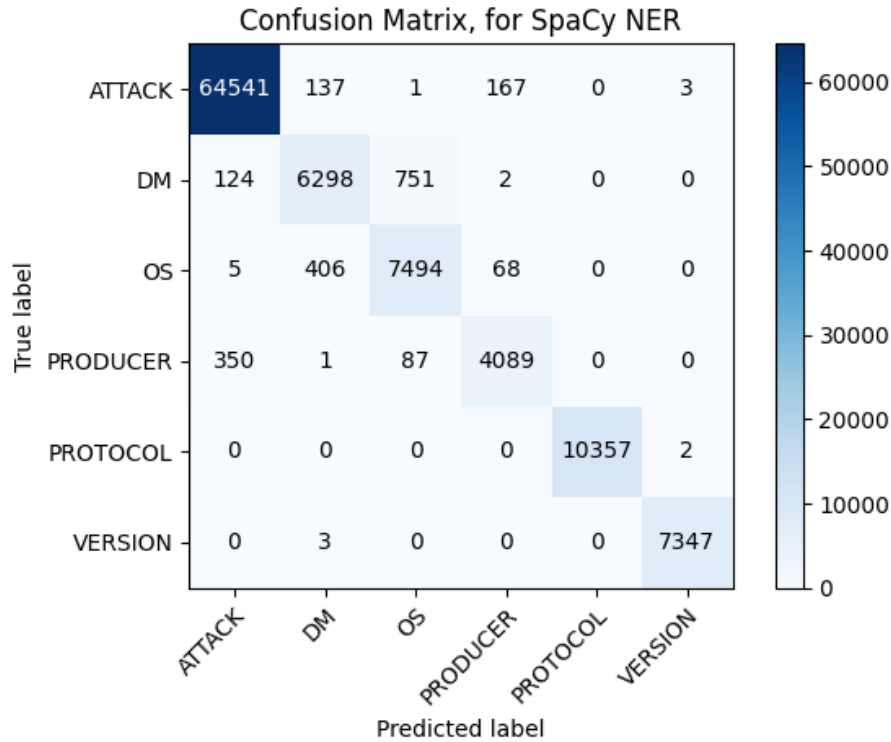


Figure 6.3: Confusion matrix of the model after dictionary cleaning

The graph of the loss functions shown in Figure 6.4 has a better behavior with respect to the first models trained.

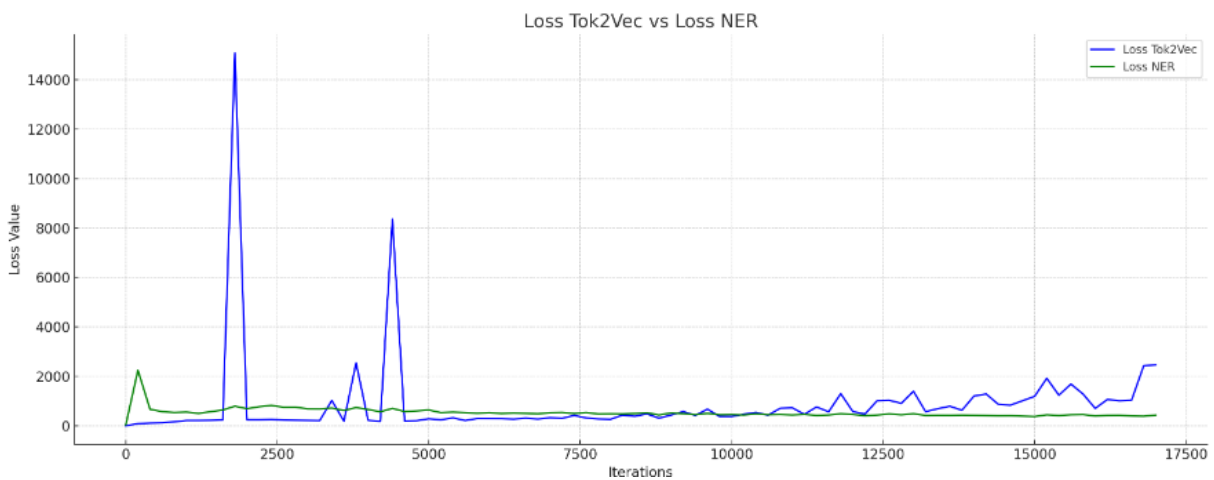


Figure 6.4: Loss functions of the model after dictionary cleaning

After those improvements, the model responded pretty well on the CVE samples but

continued to miss recognize words with upper case characters.

So we decided to set as false the `use_upper` parameter, which is responsible for the learning of different cased words.

Moreover, we raised the `dropout` value to 2. This assured a reduction in the overfitting and gives the following results.

Table 6.3: Final NER Model Performance Metrics

Entity Type	Precision	Recall	F1-Score
Overall	0.9378	0.9410	0.9394
ATTACK	0.9604	0.9767	0.9685
PROTOCOL	0.9435	0.9275	0.9354
OS	0.9106	0.8872	0.8987
VERSION	0.9310	0.8690	0.8990
PRODUCER	0.9810	0.8429	0.9067
DM	0.6907	0.7987	0.7408

The confusion matrix of the final model is shown in Figure 6.5.

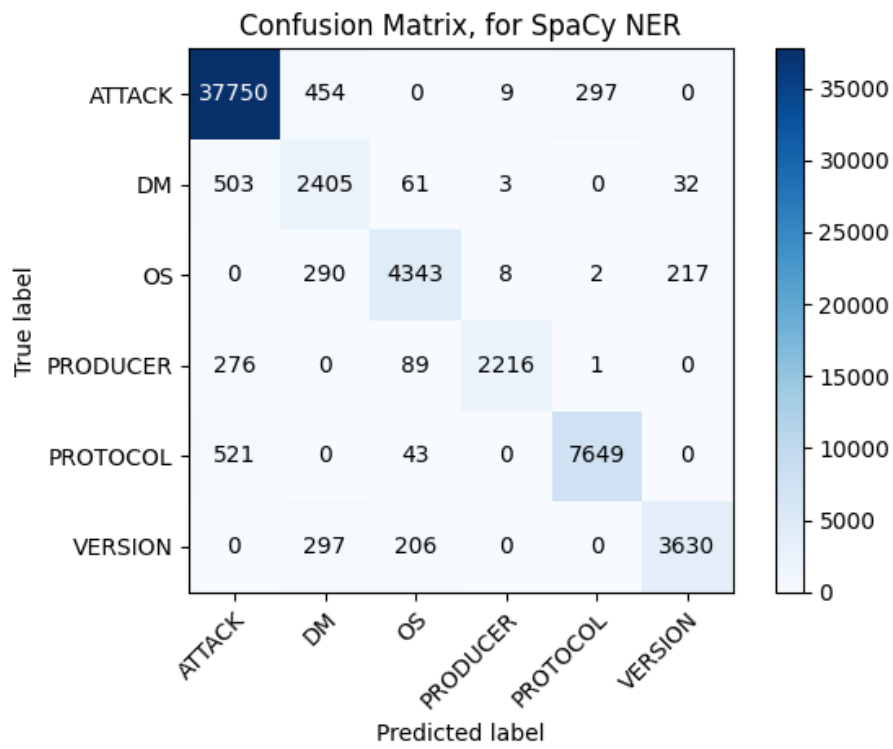


Figure 6.5: Confusion matrix of the final model

The remaining misprediction is because many acronyms are used in the IT domain. It happens lots of times that some acronyms of different entities are identical. Also, medical devices use acronyms in their names, which explains why 500 out of  $\sim 40000$  have been mispredicted as cyber attacks.

In Figure 6.6, the plot of the loss functions of the final network.

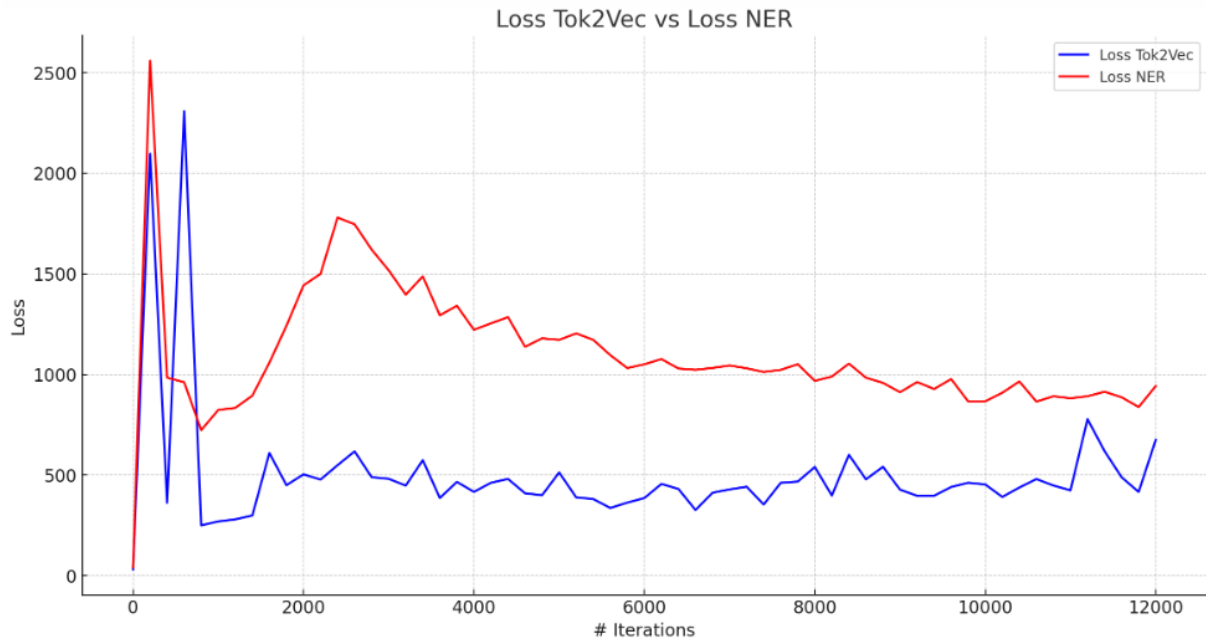


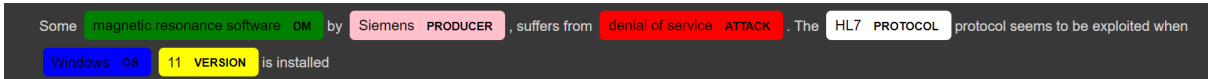
Figure 6.6: Loss functions of the final model

Finally, a recap of the models trained is shown in Table 6.4. It is important to note that fine-tuning has been performed for each model to find the best hyper-parameters.

Table 6.4: Comparison of NER Model Performance Metrics

Model	Precision	Recall	F1-Score
First NER model biased by overfitting	0.9867	0.9884	0.9875
NER Model after dictionary cleaning	0.9774	0.9778	0.9776
Final NER Model Performance Metrics	0.9378	0.9410	0.9394

In order to have a better visual representation of the model's outcome, we have reported in Figure 6.7 what the output looks like. We used the `displacy` library to simplify everything.



Some magnetic resonance software DM by Siemens PRODUCER , suffers from denial of service ATTACK . The HL7 PROTOCOL protocol seems to be exploited when Windows OS 11 VERSION is installed

Figure 6.7: Outcome of the NER model



## 6.2 Business intelligence

The Business Intelligence process can be defined as "the process and technology required to transform data and information into knowledge." The Althea BI module aims to enable healthcare managers to make strategic decisions by providing accurate, up-to-date, and relevant information in the given context. This module allows data to be explored freely, following one's logical paths to uncover new strategic insights. The Business Intelligence module is built on the QlikSense engine, one of the most advanced BI technologies available on the market. The module is structured through the provision of Dashboards.

We decided to include the samedashboard-based data visualization in our project and use Qlik to show our results. They are visible in Section 6.4.

The following sections explain the metrics used and implemented in the BI module to extract information from the outcome of the network and compute the risk index of the Sant'Andrea IB.

### 6.3 Index of risk

Our goal is to compute an index of risk for each asset of the Sant' Andrea hospital. From now on, we call the index of risk as  $IR_{tot}$  that depends on  $IRs$  and  $IRd$ :

- $IRs$  is the static index of risk and denotes the starting index for each asset depending on its features
- $IRd$  is the dynamic index of risk and denotes the increment of risk for each asset depending on the outcome of our network

Both  $IRs$  and  $IRd$  will be explained in the next sections.

For the moment, it is enough for us to know that  $IRs$  and  $IRd$  can assume three values: *LOW*, *MEDIUM*, *HIGH*. As it is easily understandable, *LOW* means that the asset is subjected to low risk, *MEDIUM* means that the asset is subjected to medium risk, and *HIGH* means that the asset is subjected to high risk.

In order to compute  $IR_{tot}$  we applied the logic specified in Table 6.5:

$IRs$	$IRd$	$IR_{tot}$
Low	Low	Low
Low	Medium	Medium
Low	High	High
Medium	Low	Medium
Medium	Medium	Medium
Medium	High	High
High	Low	High
High	Medium	High
High	High	High

Table 6.5: Logic behind the computation of  $IR_{tot}$

Following this table, the dynamic risk index can only worsen the starting static risk index.

This is how  $IR_{tot}$  is computed. The following sections describe the computation of  $IRs$  and  $IRd$ .

### 6.3.1 Static Index of Risk - IRs

The IRs represent the base index of risk of the currently active DMs in Sant' Andrea hospital. Its computation depends on a Risk Portfolio and information about the DMs.

The risk portfolio consists of a set of  $M = 26$  cyber risks concerning hospitals; they have been selected with the help of Althea's data and security manger. Table 6.6 lists all of those risks.

Risk ID	Risk
<b>R1</b>	Immissione di ransomware
<b>R2</b>	Immissione di malware informatici
<b>R3</b>	Perdita grave di dati clinici
<b>R4</b>	Attacchi Dos or DDoS ai sistemi interni
<b>R5</b>	Manomissione segni vitali medici di un paziente
<b>R6</b>	Violazione dei Dati Personali
<b>R7</b>	Manomissione dati cartelle cliniche
<b>R8</b>	Manomissione immagini dignostiche paziente
<b>R9</b>	Manomissione dati test di laboratorio
<b>R10</b>	Manomissione parametri erogatori di terapie
<b>R11</b>	Violazione dei dati ingiustificata o eccessiva
<b>R12</b>	Raccolta di dati ingiustificata o eccessiva
<b>R13</b>	Uso inappropriato o improprio di dati personali
<b>R14</b>	Furto o smarrimento PC oWS
<b>R15</b>	Accesso non autorizzato via Internet
<b>R16</b>	Accesso non autorizzato tramite WiFi
<b>R17</b>	Indisponibilità di fornitore o servizio critico
<b>R18</b>	Furto o smarrimento Dispositivo Medico
<b>R19</b>	Malfunzionamento di applicazione software
<b>R20</b>	Malfunzionamento di apparati hardware
<b>R21</b>	Effetti collaterali da upgrade di sw o firmware
<b>R22</b>	Effetti collaterali da change su apparati
<b>R23</b>	Errori del personale IT
<b>R24</b>	Errori degli utenti

Table 6.6: Table of risks

It is important to note that each risk is associated with an id  $R_i$  with  $1 \leq i \leq 26$ .

This notation will be useful for us in the following.

The information about DMs has been collected through a census conducted by hospital technicians, who, machine by machine, filled out an online form with the requested information. All the assets with their information have been stored in EASI, an online platform created by Consulnet Italia, which helps companies like Althea in systems management. We end up with  $N = 28$  fields listed in table 6.7.

Fields ID	Field
C1	Dispositivo medico
C2	Dispositivo embedded
C3	Licenza software
C4	Dispositivo standalone
C5	Software medicale
C6	Dispositivo collegato in rete
C7	Software di supporto alla diagnosi
C8	Ambito di applicazione del software
C9	Tipologia connessione
C10	Dichiarazione di end of life del produttore
C11	Possibilità di navigazione internet
C12	Autenticazione utente software
C13	Dispositivo sotto dominio
C14	Connessione LAN
C15	Tipologia utente
C16	Porte USB attive
C17	Presenza di autologin
C18	Autenticazione utente OS
C19	Esportabilità del dato
C20	Presenza di un database dedicato
C21	Tipologia Utente OS
C22	Presenza di autologin OS
C23	Tipologia di dato archiviato
C24	Presenza nel software di cache locale per l'archiviazione
C25	Sistema Operativo Installato
C26	Possibilità di accesso condiviso al database
C27	Protocollo di comunicazione utilizzati
C28	Possibilità di collegamento in VPN per accesso remoto

Table 6.7: Table of fields

Even in this case, it is important to note that each field is associated with a label  $C_j$  with  $1 \leq j \leq 28$ .

Risks  $R_i$  and fields  $C_j$  have then been used to create a matrix  $M \times N$  in which for each  $R_i$ , the corresponding  $C_j$  is marked with 1 if the field is relevant for the risk, otherwise it is

marked with 0. The decision to mark a field as relevant is based on the experience gained during the internship and on the advice received from Althea's experts.

This matrix helped us to find a subset of relevant fields for each risk. The next step was to compute, for each asset, a score for each risk  $R_i$ . Then, we computed the score of each risk  $R_i$  applying the following algorithm has been used for the computation:

---

**Algorithm 6.1** Risk's Score Computation Algorithm

---

```

1: for All the risks  $R_i$  do
2:   for All the fields  $C_j$  do
3:     Set the score of the first risk:  $score_i = 0$ .
4:     if the values of a field  $C_j$  is positive (which means that the value with which
       the field is filled makes the field relevant for the risk itself). then
5:       update the score:  $score_i = score_i + 1$ .
6:     end if
7:     else  $score_i = score_i + 0$ .
8:   end for
9: end for

```

---

We now want to do a weighted sum for each asset of each risk. In order to compute the weights  $WR_i$ , we adopted the *Threshold Algorithm* (TA) with  $Sum()$  as a scoring function and  $k = 24$ . TA has been executed on the risks, ordered with two different criteria. The first ranking is based on the cardinalities of relevant fields; the second consists of a ranking based on the importance of the risks.

Again, the decision to mark a risk more or less important than another is based on the experience gained during the internship and on the advice from Althea's experts.

---

**Algorithm 6.2** Threshold Algorithm - TA

---

```

1: Input: Integer  $k$ , a monotone function  $S$  combining ranked lists  $R_1, R_2, \dots, R_m$ .
2: Output: The top- $k$   $\langle$ object, score $\rangle$  pairs.
3:
4: Do a sorted access in parallel in each list  $R_i$ .
5: for each object  $o$  do
6:   Do random accesses in the other lists  $R_j$ , extracting score  $s_j$ .
7:   Compute overall score  $S(s_1, \dots, s_m)$ .
8:   if value is among the  $k$  highest seen so far then
9:     Remember  $o$ .
10:  end if
11: end for
12: Let  $s_{L_i}$  be the last score seen under sorted access for  $R_i$ .
13: Define threshold  $T = S(s_{L_1}, \dots, s_{L_m})$ .
14: if score of the  $k$ -th object is worse than  $T$  then
15:   Go to step 1.
16: end if
17: return the current top- $k$  objects.

```

---

Once ordered, we assigned a raw weight  $WR_i$  to each risk starting with 24 for the most important, down to 1 for the less important.

Then, each weight has been normalized such that:

- $0 < WR_i < 1$
- $\sum WR_i = 1$

The formula used to perform the normalization is:

$$WR_{i\_normalized} = \frac{WR}{SMAX1} \quad (6.1)$$

where:

- $SMAX1$  is the sum of all the weights,  $SMAX = 300$ .

Table 6.8 shows all the weights  $WR_i$  associated to each risk  $R_i$ :

Risk ID	Weight
R1	10
R2	9
R3	8
R4	7
R5	6
R6	5
R7	4
R8	17
R9	13
R10	11
R11	14
R12	2
R13	20
R14	1
R15	24
R16	23
R17	22
R18	21
R19	15
R20	12
R21	3
R22	18
R23	19
R24	16

Table 6.8: List of risks  $R_i$  with their weights  $WR_i$

Finally,  $IRs$  is computed with the weighted sum:

$$IRs = \sum_{m=1}^{m=24} R_m * WR_m \quad (6.2)$$



As the last step, we normalized the result on a scale from 0 to 3. The normalization formula is the following:

$$IRs\_normalized = \frac{IRs * 3}{SMAX2} \quad (6.3)$$

where:

- $SMAX2$  is the value of  $IRs$  in the best case, that is when all the risks  $R_i$  have all the fields relevant, which means that all the fields  $C_j$  have relevant values;  $SMAX2 = 9.26$ .

In the end, a fuzzy logic has been applied to extract a label from the values so that we can express  $IRs$  as *LOW*, *MEDIUM*, *HIGH*. Table 6.9 shows the logic adopted.

VALUE	LABEL
$2 \leq IRs < 3$	HIGH
$1 \leq IRs < 2$	MEDIUM
$0 \leq IRs < 1$	LOW

Table 6.9: Fuzzy logic for  $IRs$

We have finally obtained the *static index of risk*.

### 6.3.2 Dynamic Index of Risk - IRd

The IRd represents the increment of the risk index of the DMs currently active in Sant' Andrea Hospital. Its computation depends on the output of our NER model. This output results from giving the model the daily CVEs reports from 2015 as input. Table 6.10 shows this output's structure.

Fields of the output of the NER model					
DM	PRODUCER	OS	VERSION	PROTOCOL	ATTACK

Table 6.10: Structure of the output of the NER model

The computation of IRd is pretty simple. We have calculated the IDF metric for the first five fields. This metric takes into account how many times a terms  $t$  is found inside a collection of document  $d$ .

IDF metric is specified as:

$$IDF(t, D) = \log \left( \frac{\text{number of document } d}{1 + \text{number of document } d \text{ that contain } t} \right) \quad (6.4)$$

where:

- number of document  $d$  corresponds to the *numberofrows* of the output of our model
- $t$  a the value of a field of the output table

We then calculated the IRd as the sum of all the IDFs, and in the end, we normalized such value in a scale from 0 to 3. The normalization formula is the following:

$$IRs\_normalized = \frac{IRs * 3}{SMAX3} \quad (6.5)$$

where:

- SMAX3 is the sum of all the IDF values in the worst case when we have zero occurrences for each field in the model's output. Since the output has  $\approx 20000$  rows and  $\log_{10} 20000 \approx 4,5$ , so  $SMAX3 \approx 4,5 * 5 = 22,5$ .

In the end, a fuzzy logic has been applied to extract a label from the values so that we can express IRd as *LOW*, *MEDIUM*, *HIGH*. Table 6.11 shows the logic adopted.

VALUE	LABEL
$0 < IRd \leq 1$	HIGH
$1 < IRd \leq 2$	MEDIUM
$2 < IRd \leq 3$	LOW

Table 6.11: Fuzzy logic for IRd

Note that since we are using IDF metric, a high value means that we have few occurrences, so the risk is lower for the asset.

We have finally obtained the *dynamic index of risk*.

Usually, in a hospital, some medical devices need special attention since their use is vital for patient safety. Therefore, even if few occurrences of attacks or vulnerabilities are found inside the collection of CVEs, their risk index must be HIGH. As a demonstration, we selected three classes of medical devices in which the dynamic risk index is incremented by 50%. The classes of device selected are ECT, SR4, and TAC, which respectively represents: Ultrasound Scanning System, Software for Magnetic Resonance Imaging , and Computed Tomography Scanning System.

## 6.4 Outcome of the use case

This section presents the dashboard realized in Qlik and created for Sant' Andrea. It consists of four dynamic sheets, each containing graphs, measures, and information to have a clear vision of all the assets in the structure. All the views presented are interactable and are dynamically updated by selecting any filter or any field of the graphs. The selection of a field performs a drill down in real-time, modifying all the other graphs and the related measures.

The first sheet shows an overview of the assets under management in the hospital. Information such as when the warranty will expire in the coming months is shown in the bar graph, while in the treemap, we can easily see the number of assets by class, manufacturer, and model. Information such as location, asset type, and ownership type are shown in pie charts. Finally, we have inserted a filter box to filter the information based on the desired search criterion.

The second sheet shows a general view of all the assets from the point of view of the risk indices calculated according to the criteria explained in the previous chapter. In this section, we have included the average scores of IRS, IRD, IRT, and their graphic representation using pie charts. Tables show the details of each asset with the relative score. Finally, bar graphs show the trend of the various risk indices on all the assets surveyed. For illustrative purposes, the total risk index section shows the details of some of the classes under consideration, i.e. ECT, SR4, and TAC, whose dynamic risk index has been set to HIGH. This is because these categories of devices represent an asset of particular importance for the customer, and therefore, even a minimal risk must be marked as high.

The third sheet shows the details of the critical software and operating systems. A pie chart shows three dimensions: model, manufacturer, and software version. The table then shows the details of each software with information from the census.

The last sheet shows the details of the medical devices, and the pie charts show how they are divided based on the dimensions shown above, such as Communication protocols used, type of connection, and operating system installed. As usual, a table is shown to show the details for each asset.

For demonstration purposes, Figure 6.12 shows an example of a dynamic view, on the sheet of the cyber security of medical devices. Applied filters are shown in the top left corner; in this case, Windows 7 was selected as the operating system and HL7 as the communication protocol.

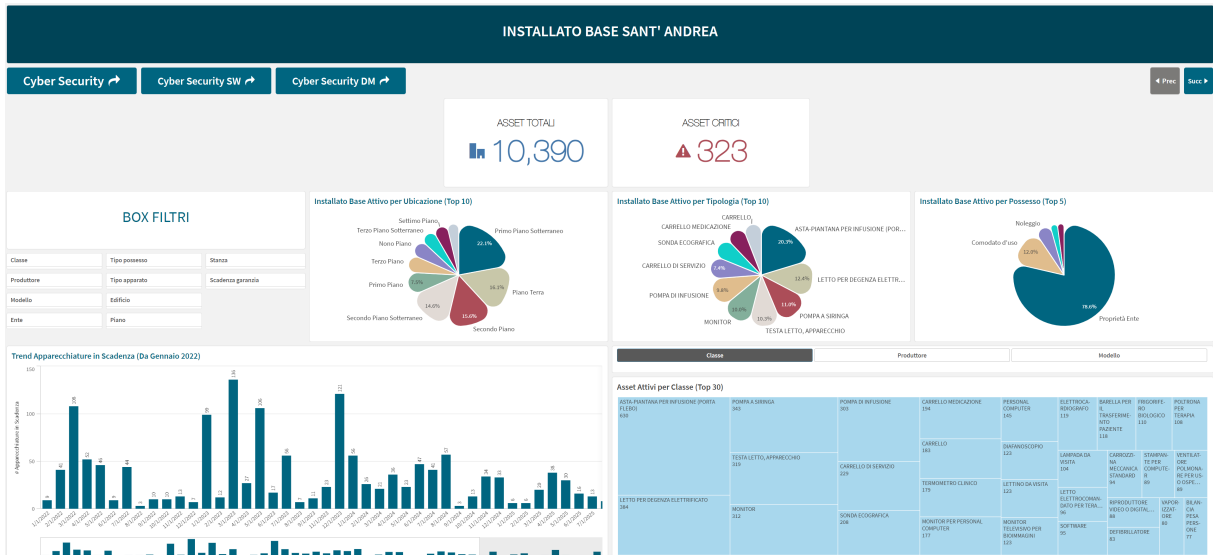


Figure 6.8: Sant'Andrea's Dashboard - Installato Base Sant'Andrea

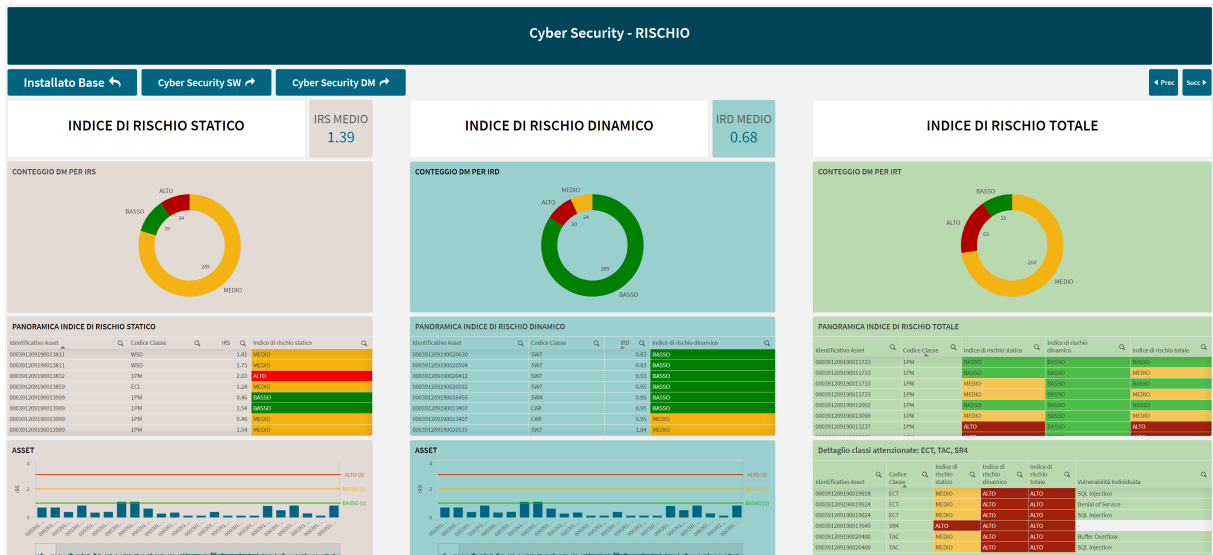


Figure 6.9: Sant'Andrea's Dashboard - Cyber Security, Risk

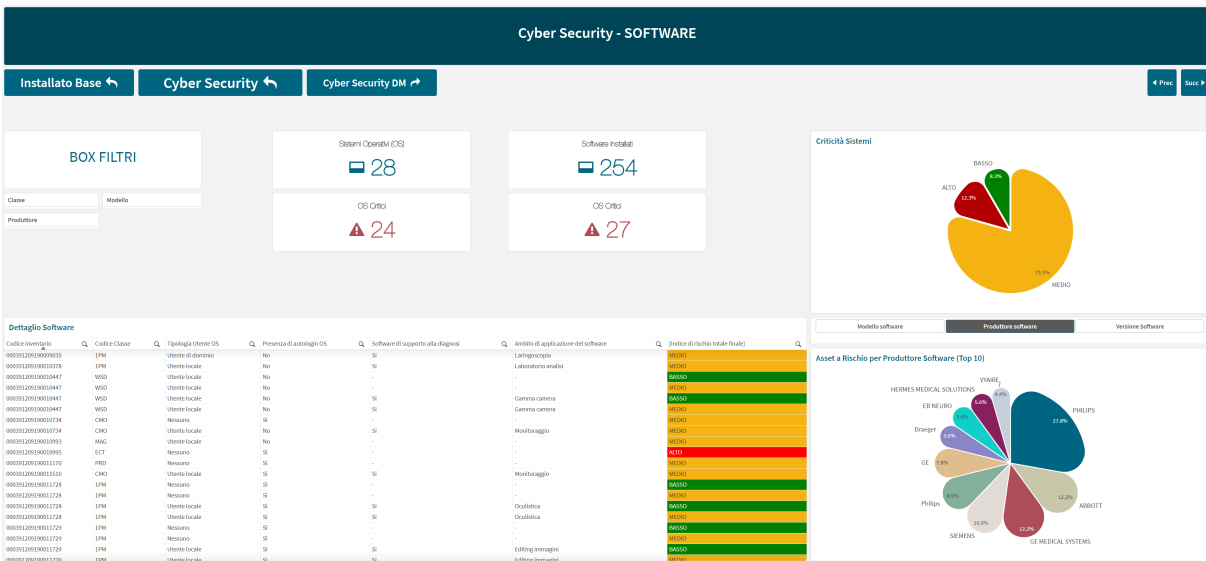


Figure 6.10: Sant'Andrea's Dashboard - Cyber Security, Software

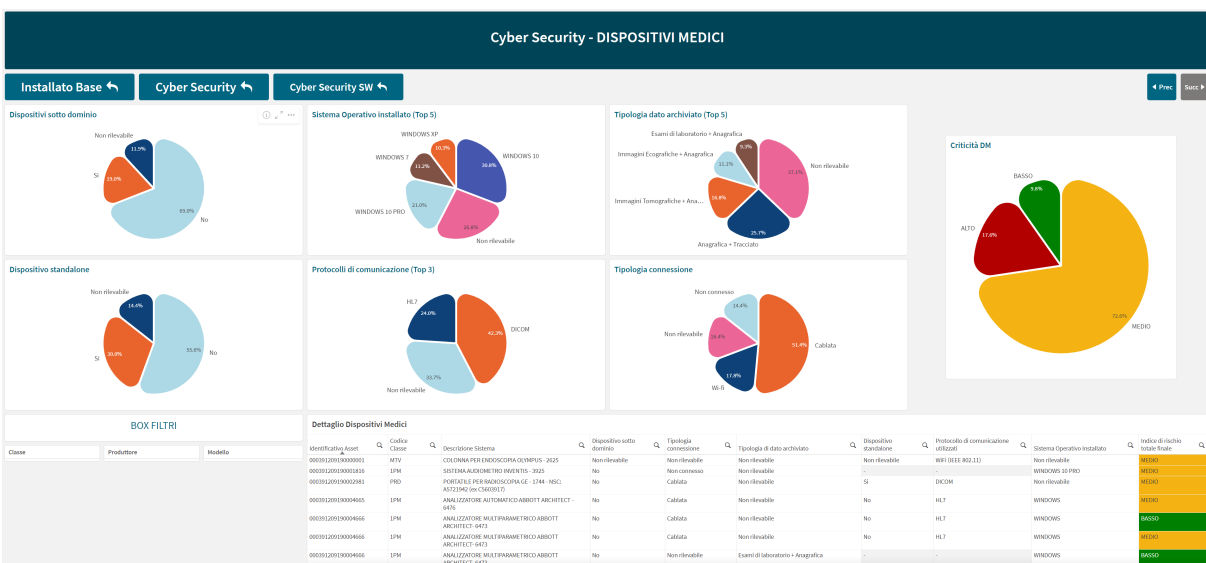


Figure 6.11: Sant'Andrea's Dashboard - Cyber Security, Dispositivi Medici

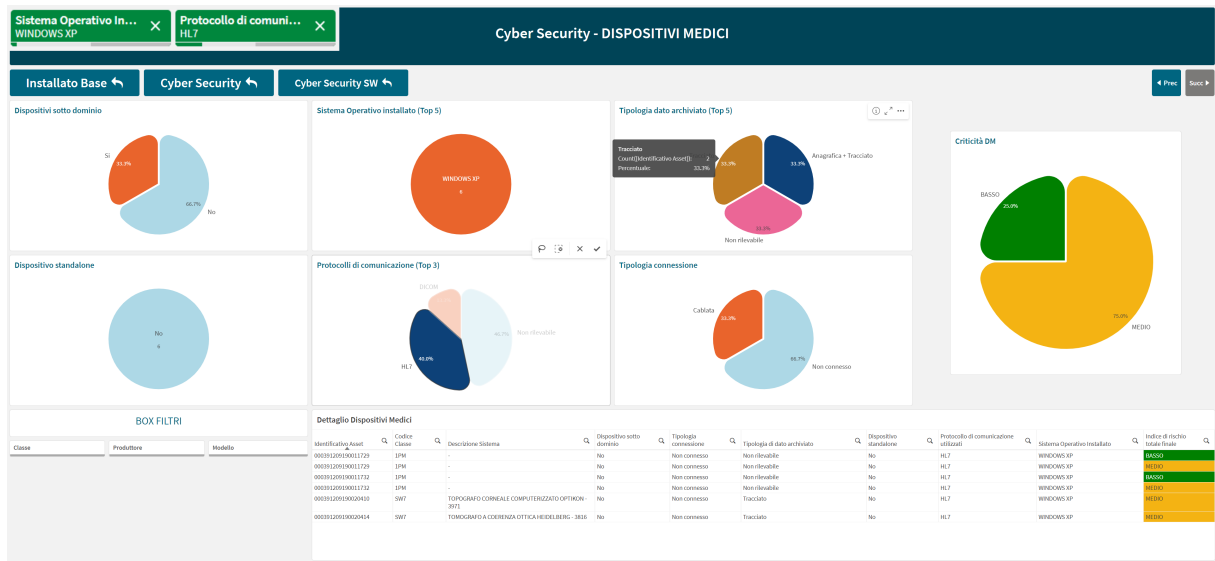


Figure 6.12: Sant'Andrea's Dashboard - Drill-down example





## Chapter 7

# Conclusions and future works

Addressing the cyber risk in the healthcare domain is challenging due to the variety of assets and the never-ending IT innovation research. Personally Identifiable Information, Protected Health Information, and Electronic Health Records must be kept safe to protect patient safety and well-being. Our project aimed to achieve these goals by inspecting the vulnerabilities and threats to which medical devices are exposed daily to cybercriminals. Cyber criminality is nearly impossible to prevent, but reacting quickly to a disaster is important. The model we provided succeeded in automating the risk assessment process to reduce the time of information extraction in healthcare and cyber domain through neural networks. With an overall F1-score of 0.93 and a precision and recall score of 0.93 and 0.94, the model can recognize information like operating systems, protocols, versions, medical devices, producers, and cyber-attacks and vulnerabilities. Finally, through a process of business intelligence performed in Qlik Sense, the information extracted from the daily CVE is compared with the installed base of our use case, the AOU Sant'Andrea, and with mathematical logic, a risk index is computed. Our work increased the security inside healthcare organizations and companies, but the cyber landscape remains varied and continuously exposed to threats. Reducing the risk to zero is impossible since new vulnerabilities arise every day, and the medical devices we can find inside a hospital are a lot and can change continuously in a short time. Instead, we can reduce the impact of risk by studying the medical devices a hospital has in charge and acting to keep them as safe as possible. We leave as future work the improvements of annotated corpora that lack in volume; using ChatGPT4 to perform data augmentation is not enough, and new annotated corpora need to be built both manually and with the help of supervised machine learning methods. This will help to increase the performance of the network. Moreover, new cyber security blogs related to biotechnology could be considered to monitor the current vulnerabilities of medical devices.



## Bibliography

- [1] A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter, and R. Vollgraf. Flair: An easy-to-use framework for state-of-the-art nlp. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics (demonstrations)*, pages 54–59, 2019.
- [2] M. T. Alam, D. Bhusal, Y. Park, and N. Rastogi. Cyner: A python library for cybersecurity named entity recognition. *ArXiv*, abs/2204.05754, 2022. URL <https://api.semanticscholar.org/CorpusID:248118916>.
- [3] K. Ameri, M. Hempel, H. Sharif, J. Lopez Jr, and K. Perumalla. Cybert: Cybersecurity claim classification by fine-tuning the bert language model. *Journal of Cybersecurity and Privacy*, 1(4):615–637, 2021.
- [4] S. T. Argaw, J. R. Troncoso-Pastoriza, D. Lacey, M.-V. Florin, F. Calcavecchia, D. Anderson, W. Burleson, J.-M. Vogel, C. O’Leary, B. Eshaya-Chauvin, et al. Cybersecurity of hospitals: discussing the challenges and working towards mitigating the risks. *BMC medical informatics and decision making*, 20:1–10, 2020.
- [5] ASPE Office of the Assistant Secretary for Planning and Evaluation. Health insurance portability and accountability act of 1996, 1996. URL <https://aspe.hhs.gov/reports/health-insurance-portability-accountability-act-1996>.
- [6] Associazione Data Protection Officer. Norme e regolamenti per gestire correttamente le informazioni, 2020. URL <https://www.assodpo.it/2020/09/21/norme-e-regolamenti-per-gestire-correttamente-le-informazioni/>.
- [7] Black Book Market Research LLC. State of the healthcare cybersecurity industry, 2020.
- [8] R. A. Bridges, C. L. Jones, M. D. Iannacone, and J. R. Goodall. Automatic labeling for entity extraction in cyber security. *ArXiv*, abs/1308.4941, 2013. URL <https://api.semanticscholar.org/CorpusID:15487201>.
- [9] Caroline Humer, Jim Finkle. Your medical record is worth more to hack-

- ers than your credit card, 2014. URL <https://www.reuters.com/article/us-cybersecurity-hospitals-idUSKCN0HJ21I20140924>. [Online; accessed 3-October-2023].
- [10] Center for Internet Security (CIS). Cyber attacks: In the healthcare sector. Technical report, Center for Internet Security (CIS), East Greenbush, NY, USA, 2017.
- [11] Y. Chen, J. Ding, D. Li, and Z. Chen. Joint bert model based cybersecurity named entity recognition. In *2021 The 4th International Conference on Software Engineering and Information Management*, pages 236–242, 2021.
- [12] P. Cheng and K. Erk. Attending to entities for better text understanding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 7554–7561, 2020.
- [13] W. contributors. Use cases that writer covers with its generative ai capabilities, mar 2023. URL <https://dev.writer.com/docs/use-cases#:~:text=Text2text%20generation%20is%20a%20method,text%20from%20a%20given%20input>.
- [14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *North American Chapter of the Association for Computational Linguistics*, 2019. URL <https://api.semanticscholar.org/CorpusID:52967399>.
- [15] O. Etzioni, M. Cafarella, D. Downey, A.-M. Popescu, T. Shaked, S. Soderland, D. S. Weld, and A. Yates. Unsupervised named-entity extraction from the web: An experimental study. *Artificial intelligence*, 165(1):91–134, 2005.
- [16] European Parliament. Regulation (eu) 2016/679 of the european parliament and of the council, 2016.
- [17] European Parliament and Council of the European Union. Directive (eu) 2016/1148 of the european parliament and of the council of 6 july 2016. Official Journal of the European Union, 2016.
- [18] European Parliament and Council of the European Union. Directive (eu) 2020/823 of the european parliament and of the council. Official Journal of the European Union, 2020.
- [19] Explosion AI. spacy 101: Everything you need to know. <https://spacy.io/usage/spacy-101>, 2023. Accessed: 2023-11-08.
- [20] S. M. Ghaffarian and H. R. Shahriari. Software vulnerability analysis and discovery

- using machine-learning and data-mining techniques: A survey. *ACM Computing Surveys (CSUR)*, 50(4):1–36, 2017.
- [21] N. Goud. Malware and ransomware attack on medical devices. Technical report, Cybersecurity Insiders, Baltimore, MD, USA, 2017.
- [22] J. Guo, G. Xu, X. Cheng, and H. Li. Named entity recognition in query. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 267–274, 2009.
- [23] D. Halperin, T. S. Heydt-Benjamin, B. Ransford, S. S. Clark, B. Defend, W. Morgan, K. Fu, T. Kohno, and W. H. Maisel. Pacemakers and implantable cardiac defibrillators: Software radio attacks and zero-power defenses. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, pages 129–142. IEEE, 2008.
- [24] N. I. Haque, M. A. Rahman, M. H. Shahriar, A. A. Khalil, and S. Uluagac. A novel framework for threat analysis of machine learning-based smart healthcare systems. *ArXiv*, abs/2103.03472, 2021. URL <https://api.semanticscholar.org/CorpusID:232135033>.
- [25] T. Hastie, R. Tibshirani, and J. Friedman. *Neural Networks*, pages 389–416. Springer New York, New York, NY, 2009.
- [26] HHS contributors. Health sector cybersecurity coordination center (hc3), 2023. URL <https://www.hhs.gov/about/agencies/asa/ocio/hc3/index.html>.
- [27] IBM. What is natural language processing? URL <https://www.ibm.com/topics/natural-language-processing>. [Online; accessed 3-October-2023].
- [28] IBM contributors. Cost of a data breach report 2023, 2023. URL <https://www.ibm.com/reports/data-breach>.
- [29] K. Lab. Kaspersky cyber pulse report 2019. Pdf file, 2019. URL [https://media.kasperskydaily.com/wp-content/uploads/sites/85/2019/08/16121510/Kaspersky-Cyber-Pulse-Report-2019\\_FINAL.pdf](https://media.kasperskydaily.com/wp-content/uploads/sites/85/2019/08/16121510/Kaspersky-Cyber-Pulse-Report-2019_FINAL.pdf). Accessed: 2023-11-08.
- [30] J. Li, A. Sun, J. Han, and C. Li. A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*, 34(1):50–70, 2020.
- [31] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. Roberta: A robustly optimized bert pretraining ap-

- proach. *ArXiv*, abs/1907.11692, 2019. URL <https://api.semanticscholar.org/CorpusID:198953378>.
- [32] R. Luna, E. Rhine, M. Myhra, R. Sullivan, and C. S. Kruse. Cyber threats to health information systems: A systematic review. *Technology and Health Care*, 24(1):1–9, 2016. doi: 10.3233/THC-151102.
- [33] P. Ma, B. Jiang, Z. Lu, N. Li, and Z. Jiang. Cybersecurity named entity recognition using bidirectional long short-term memory with conditional random fields. *Tsinghua Science and Technology*, 26(3):259–265, 2020.
- [34] C. C. Marco Angelini. Framework nazionale cybersecurity data protection, CIS Sapienza, 2019. URL [https://www.cybersecurityframework.it/sites/default/files/framework2/Framework\\_nazionale\\_cybersecurity\\_data\\_protection.pdf](https://www.cybersecurityframework.it/sites/default/files/framework2/Framework_nazionale_cybersecurity_data_protection.pdf). Accessed: 2023-11-10.
- [35] V. Mavroeidis and S. Bromander. Cyber threat intelligence model: an evaluation of taxonomies, sharing standards, and ontologies within cyber threat intelligence. In *2017 European Intelligence and Security Informatics Conference (EISIC)*, pages 91–98. IEEE, 2017.
- [36] D. A. MBCS. Understanding text classification in natural language processing—a beginners’ guide. feb 2023. URL <https://www.linkedin.com/pulse/understanding-text-classification-natural-language-david-adamson-mbcs#:~:text=Text%20classification%20is%20one%20of,relevant%20to%20a%20given%20topic>.
- [37] D. McKee and P. Laulheret. McAfee enterprise atr uncovers vulnerabilities in globally used b. braun infusion pump. Technical report, Milpitas, CA, USA, 2021.
- [38] M. Modrzejewski. *Improvement of the Translation of Named Entities in Neural Machine Translation*. PhD thesis, Karlsruhe Institute of Technology, 2020.
- [39] D. Mollá, M. Van Zaanen, and D. Smith. Named entity recognition for question answering. In *Proceedings of the Australasian language technology workshop 2006*, pages 51–58, 2006.
- [40] S. Morgan. Global cybercrime damages predicted to reach \$6 trillion annually by 2021. *Cybersecurity Ventures Official Annual Cybercrime Report*, oct 2020. URL <https://cybersecurityventures.com/annual-cybercrime-report-2020/>. Northport, N.Y.
- [41] S. Morgan. 2023 cybersecurity almanac: 100 facts, figures, predictions, and statis-

- tics. *Cybercrime Magazine*, may 2023. URL <https://cybersecurityventures.com/cybersecurity-almanac-2023/>. Sausalito, Calif.
- [42] D. Nadeau and S. Sekine. A survey of named entity recognition and classification. *Linguisticae Investigationes*, 30(1):3–26, 2007.
- [43] E. Niemiec. Will the eu medical device regulation help to improve the safety and performance of medical ai devices? *Digital Health*, 8:20552076221089079, 2022.
- [44] Y. Nikoloudakis, I. Kefaloukos, S. Klados, S. Panagiotakis, E. Pallis, C. Skianis, and E. K. Markakis. Towards a machine learning based situational awareness framework for cybersecurity: an sdn implementation. *Sensors*, 21(14):4939, 2021.
- [45] M. E. Okurowski, H. Wilson, J. Urbina, T. Taylor, R. C. Clark, and F. Krapcho. Text summarizer in use: Lessons learned from real world deployment and evaluation. In *NAACL-ANLP 2000 Workshop: Automatic Summarization*, 2000.
- [46] W. H. Organization et al. Who global model regulatory framework for medical devices including in vitro diagnostic medical devices. World Health Organization, 2017.
- [47] D. Petkova and W. B. Croft. Proximity-based document representation for named entity retrieval. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 731–740, 2007.
- [48] PresentationEze. Medical device. <https://www.presentationeze.com/blog/medical-device/>, 2023. Accessed: 2023-11-12.
- [49] J. Ryerse. Top 11 cybersecurity frameworks in 2023. mar 2023. URL <https://www.connectwise.com/blog/cybersecurity/11-best-cybersecurity-frameworks>.
- [50] M. Saeed. A guide to text preprocessing techniques for nlp, feb 2022. URL <https://exchange.scale.com/public/blogs/preprocessing-techniques-in-nlp-a-guide>.
- [51] T. Satyapanich, F. Ferraro, and T. Finin. Casie: Extracting cybersecurity event information from text. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8749–8757, 2020.
- [52] R. Sharnagat. Named entity recognition: A literature survey. *Center For Indian Language Technology*, pages 1–27, 2014.
- [53] N. Shevchenko, T. A. Chick, P. O’Riordan, T. P. Scanlon, and C. Woody. Threat modeling: a summary of available methods. Technical report, Carnegie Mellon University Software Engineering Institute Pittsburgh United, 2018.

- [54] Shon Tyler. The impact of hipaa and hitech. Pdf file, 2016. URL <https://silotips/download/the-impact-of-hipaa-and-hitech#>. Accessed: 2023-11-08.
- [55] S. Silvestri, F. Gargiulo, and M. Ciampi. Improving biomedical information extraction with word embeddings trained on closed-domain corpora. In *2019 IEEE symposium on computers and communications (ISCC)*, pages 1129–1134. IEEE, 2019.
- [56] S. Silvestri, F. Gargiulo, and M. Ciampi. Iterative annotation of biomedical ner corpora with deep neural networks and knowledge bases. *Applied Sciences*, 12(12): 5775, 2022.
- [57] K. Singh, S. S. Grover, and R. K. Kumar. Cyber security vulnerability detection using natural language processing. In *2022 IEEE World AI IoT Congress (AIIoT)*, pages 174–178. IEEE, 2022.
- [58] Study Group 1 of the Global Harmonization Task. Principles of in vitro diagnostic (ivd) medical devices classification. *GHTF public archives*, 2008. URL <https://www.imdrf.org/sites/default/files/docs/ghtf/final/sg1/procedural-docs/ghtf-sg1-n045-2008-principles-ivd-medical-devices-classification-080219.pdf>.
- [59] ultimate AI contributors. What is an nlp chatbot — and how do nlp-powered bots work?, aug 2023. URL <https://www.ultimate.ai/blog/ai-automation/how-nlp-text-based-chatbots-work>.
- [60] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [61] WEF contributors. Global gender gap report 2023, 2023. URL [https://www.weforum.org/reports/global-gender-gap-report-2023/?gclid=CjwKCAjw9-6oBhBaEiwAHv1QvAm9G4AHUmcEcogrheH9YJ1nwN5oMG91VDDQ6c\\_2LeF5PhzL9wXa7BoC-1cQAvD\\_BwE](https://www.weforum.org/reports/global-gender-gap-report-2023/?gclid=CjwKCAjw9-6oBhBaEiwAHv1QvAm9G4AHUmcEcogrheH9YJ1nwN5oMG91VDDQ6c_2LeF5PhzL9wXa7BoC-1cQAvD_BwE).
- [62] WHO contributors. Medical devices. URL [https://www.who.int/health-topics/medical-devices/#tab=tab\\_2](https://www.who.int/health-topics/medical-devices/#tab=tab_2). [Online; accessed 3-October-2023].
- [63] Wikipedia contributors. Picture archiving and communication system — Wikipedia, the free encyclopedia, 2023. URL [https://en.wikipedia.org/w/index.php?title=Picture\\_archiving\\_and\\_communication\\_system&oldid=1158697462](https://en.wikipedia.org/w/index.php?title=Picture_archiving_and_communication_system&oldid=1158697462). [Online; accessed 3-October-2023].
- [64] Wikipedia contributors. Radiological information system — Wikipedia, the



- free encyclopedia, 2023. URL [https://en.wikipedia.org/w/index.php?title=Radiological\\_information\\_system&oldid=1170866192](https://en.wikipedia.org/w/index.php?title=Radiological_information_system&oldid=1170866192). [Online; accessed 3-October-2023].
- [65] A. Yeboah-Ofori, H. Mouratidis, U. Ismai, S. Islam, and S. Papastergiou. Cyber supply chain threat analysis and prediction using machine learning and ontology. In *Artificial Intelligence Applications and Innovations: 17th IFIP WG 12.5 International Conference, AIAI 2021, Hersonissos, Crete, Greece, June 25–27, 2021, Proceedings 17*, pages 518–530. Springer, 2021.
- [66] Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, and Q. Liu. Ernie: Enhanced language representation with informative entities. *ArXiv*, abs/1905.07129, 2019. URL <https://api.semanticscholar.org/CorpusID:158046772>.
- [67] S. Zhou, J. Liu, X. Zhong, and W. Zhao. Named entity recognition using bert with whole world masking in cybersecurity domain. In *2021 IEEE 6th International Conference on Big Data Analytics (ICBDA)*, pages 316–320. IEEE, 2021.
- [68] S. Zong, A. Ritter, G. Mueller, and E. Wright. Analyzing the perceived severity of cybersecurity threats reported on social media. *ArXiv*, abs/1902.10680, 2019. URL <https://api.semanticscholar.org/CorpusID:67855269>.



## List of Figures

3.1	Ransomware action variety over time . . . . .	17
3.2	Data Breach trends in healthcare - DBIR 2023 . . . . .	17
3.3	CyberSecurity trends in healthcare . . . . .	18
4.1	Map of Natural Language Process tasks . . . . .	31
4.2	Visualization of named entity recognition task . . . . .	33
4.3	Output sample of a NER model trained on a generic domain . . . . .	33
4.4	Example of a confusion matrix in a generic domain . . . . .	35
4.5	Tools list for Named Entity Recognition task . . . . .	37
4.6	SpaCy's pipeline . . . . .	38
4.7	Training workflow in SpaCy . . . . .	38
4.8	Typical structure of an ANN . . . . .	40
4.9	BERT's architecture . . . . .	44
4.10	Embedding & encoding components . . . . .	45
4.11	Embeddings of a natural language input . . . . .	46
4.12	Multi-Head Attention & Add&Norm components . . . . .	47
4.13	Feed Forward & Add&Norm components . . . . .	48
5.1	Work Flow - NER Model . . . . .	49
5.2	Work Flow - Business Intelligence . . . . .	50
5.3	Sample of Chat-GPT4 prompt . . . . .	55
6.1	Confusion matrix of the initial model . . . . .	66
6.2	Loss functions trends of the initial model . . . . .	67
6.3	Confusion matrix of the model after dictionary cleaning . . . . .	68
6.4	Loss functions of the model after dictionary cleaning . . . . .	68
6.5	Confusion matrix of the final model . . . . .	70
6.6	Loss functions of the final model . . . . .	71
6.7	Outcome of the NER model . . . . .	72
6.8	Sant'Andrea's Dashboard - Installato Base Sant'Andrea . . . . .	85
6.9	Sant'Andrea's Dashboard - Cyber Security, Risk . . . . .	85

6.10 Sant'Andrea's Dashboard - Cyber Security, Software . . . . .	86
6.11 Sant'Andrea's Dashboard - Cyber Security, Dispositivi Medici . . . . .	86
6.12 Sant'Andrea's Dashboard - Drill-down example . . . . .	87

# List of Tables

3.1	Categories of DM taken into consideration . . . . .	14
5.1	Information extracted by category . . . . .	52
5.2	Label Id and their explanation . . . . .	53
5.3	Dictionary's size by Label . . . . .	54
5.4	Size of scraped corpora . . . . .	58
5.5	Chat-GPT4 prompts used for text augmentation . . . . .	58
5.6	Corpora's size by label . . . . .	58
5.7	Size of bathced items used for each Chat-GPT4 prompt . . . . .	59
6.1	First NER model biased by overfitting . . . . .	65
6.2	NER Model Performance Metrics after dictionary cleaning . . . . .	67
6.3	Final NER Model Performance Metrics . . . . .	69
6.4	Comparison of NER Model Performance Metrics . . . . .	71
6.5	Logic behind the computation of IR_tot . . . . .	74
6.6	Table of risks . . . . .	75
6.7	Table of fields . . . . .	77
6.8	List of risks Ri with their weights WRi . . . . .	80
6.9	Fuzzy logic for IRs . . . . .	81
6.10	Structure of the output of the NER model . . . . .	82
6.11	Fuzzy logic for IRd . . . . .	83



## Acknowledgements

First, I would like to thank Professor Zanero for being my supervisor and guiding me during my internship.

I would also like to thank Antonio and Ramon for following me and helping me during my internship experience. They always provided me with all the necessary information to continue my research as quickly as they could. They brightened my days with their jokes and nuggets of wisdom.

Additionally, I would like to thank my classmates Maria Chiara, Alessia, Pietro, Francesco, Tommaso, Stefano, Nicolò, and Alessio. They lit up every single day with their friendliness, happiness, and energy.

A heartfelt thanks to Andrea, Nicolò, Daniele, and Ivan, who are my second family; they have supported me in every choice I made, helping me always to choose what was right for me and to follow my desires.

A special thanks to Marta and Davide, who always found the right words in moments of despair when nothing seemed to be going right; you were fundamental and precious in finding a solution to everything.

The most important thanks goes to my family, who have always followed and supported me in my long journey of studies, which has finally come to an end. They have always known how to advise me and help me overcome any difficulty. A special thanks to my mother, my grandmother, and my sister, who have listened to me repeat any subject since I was young.

