Job Adverts Analyzer for Cybersecurity Skills Needs Evaluation

Sara Ricci*
ricci@vutbr.cz
Brno University of Technology
Brno, Czech Republic

Marek Sikora
marek.sikora@vut.cz
Brno University of Technology
Brno, Czech Republic

Simon Parker
simon.parker@dkfz-heidelberg.de
Deutsches Krebsforschungszentrum
Heidelberg, Germany

Imre Lendak
lendak@uns.ac.rs
Faculty of Technical Sciences,
University of Novi Sad
Novi Sad, Serbia

Yianna Danidou
y.danidou@euc.ac.cy
European University Cyprus
Nicosia, Cyprus

Argyro Chatzopoulou
ac@apiroplus.solutions
APIROPLUS Solutions Ltd.
Limassol, Cyprus

Remi Badonnel
badonnel@loria.fr
University of Lorraine, Loria-Inria
Villers-lès-Nancy, France

Donatas Alksnys
doalksnys@stud.mruni.eu
Mykolas Romeris University
Vilnius, Lithuania

ABSTRACT

This article presents a new free web-based application, the Cybersecurity Job Ads Analyzer, which has been created to collect and analyse job adverts using a machine learning algorithm. This algorithm enables the detection of the skills required in advertised cybersecurity work positions. The application is both interactive and dynamic allowing for automated analyses and for the underlying database of job adverts to be easily updated. Through the Cybersecurity Job Ads Analyzer, it is possible to explore the skills required over time, and thereby enable academia and other training providers to better understand and address the needs of the industry. We will describe in detail the user interface and technical background of the application, as well as highlight the preliminary statistical results we have obtained from analysing the current database of job adverts.

CCS CONCEPTS

• Social and professional topics → Computing education; **job ads analysis**; **• Applied computing → Education**; **• Human-centered computing →** *Human computer interaction*; **• Computing methodologies →** *Machine learning*.

KEYWORDS

Cybersecurity Education, Skills, Work Roles, Machine Learning, Job Ads Analyzer

ACM Reference Format:
Sara Ricci, Marek Sikora, Simon Parker, Imre Lendak, Yianna Danidou, Argyro Chatzopoulou, Remi Badonnel, and Donatas Alksnys. 2022. Job Adverts Analyzer for Cybersecurity Skills Needs Evaluation. In *The 17th International Conference on Availability, Reliability and Security (ARES 2022), August 23–26, 2022, Vienna, Austria.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3538969.3543821

1 INTRODUCTION

Alongside the rapid development and growth of Information Technology (IT), cybersecurity has also become a key area of interest worldwide leading to high demand across the labour market for roles with specific technical skills. In 2021, the (ISC)^2 [16] estimated that there was a global shortfall of approximately 2.7 million cybersecurity experts. Moreover, this issue has also been highlighted by job market researchers [2] and governments [8, 11].

However, despite this pressing need, recognition of the cybersecurity skills needed, and consequently appropriate coordination of training pathways, remains a challenge due to the lack of shared understanding and taxonomy of cybersecurity roles and skills. In particular, ESCO [8] identified a lack of consensus between academia and industry and an inability to produce enough graduates with the skills that reflect the needs of the market.

In order to help education and training providers as well as industries with the identification of skills needed in current cybersecurity work roles, we have designed and implemented a free web-based application called the Cybersecurity Job Ads Analyzer. This application is built upon a well-known NIST NICE standard [30] and enables users to identify the skills required in cybersecurity job advertisements through a Machine Learning (ML) algorithm.

Furthermore, because cybersecurity is by its nature an area in development, the classification and the analysis of skills are strongly dependent on when the data are collected. The web application is a dynamic tool that allows the performing of analyses in an automated way and its database of collected job adverts can be easily updated. In doing so, a new skill needs evaluation, and a comparison with previous years, can be performed.

The rest of this article is organised as follows. Section 2 reviews the state-of-the-art of cybersecurity skills needs analyses. Section 3 introduces the methodology used for the skills needs analysis with the Cybersecurity Job Ads Analyzer. Section 4 describes the

ARES '22, August 23–26, 2022, Vienna, Austria
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9670-7/22/08…$15.00
https://doi.org/10.1145/3538969.3543821

implemented of the web application, focusing on its usability and technical specifications. Section 5 presents some of our preliminary analyses of job adverts. The final section contains our conclusions about the work of this project so far.

2 RELATED WORK

Within the cybersecurity literature, there have been several studies exploring skills frameworks but we were unable to locate of a large body of literature that has specifically focused on evaluating skill needs in cybersecurity. A study by the UK Ipsos MORI Social Research Institute [20] focused on understanding the UK cybersecurity skills labour market. This report includes an extensive literature review, interviews with industry experts, a quantitative survey, and qualitative interviews from private, public, and charitable sectors. One of the foremost challenges highlighted by this research was that a cybersecurity skills definition that could serve as a starting point to help organisations to better understand their skills needs and individuals to better understand their job roles was lacking. Moreover, considering the rapidly changing IT environment, the authors argued that the focus has to be put not only on current skills needs but also on potential future skills needs. Recommendations provided in this document regarding skills needs primarily consist of actions to establish a framework and extend cybersecurity awareness. For example, they propose the adoption of a definition of cybersecurity skills, the creation of cybersecurity skills career pathways including what skills and qualifications are required for each role at different levels, a cyber-certified professional scheme, raising the overall cyber security awareness in organisations, and fostering collaboration and partnerships. Such measures would establish common ground and help to better elucidate cybersecurity skills needs in different organisations.

The survey [20] (conducted in 2018) was conducted again in 2021 by Ipsos MORI on behalf of the UK Department for Digital, Culture, Media and Sport (DCMS) [17]. This updated research builds upon previous surveys and explores the nature and extent of cybersecurity skills gaps (people lacking the appropriate skills) and skills shortages (a lack of people available to work in cybersecurity job roles) using a mixture of: representative surveys with cyber-sector businesses as well as the wider population of UK organisations, qualitative research with recruitment agents, cyber-firms, and large organisations in various sectors, and a secondary analysis of cybersecurity job postings on the Burning Glass Technologies database. The results of the survey were broadly comparable to the previous survey, though there were small improvements seen in the representative surveys and qualitative research. With regards to the analysis of cybersecurity job postings, the analysis covered vacancies from September 2019 to the end of December 2020, supplementing the work performed in the 2020 study. This analysis provided, amongst others, insights on the job vacancies within the cybersecurity, the most common skills requirements mentioned in job descriptions (information security skills, network security skills, and knowledge of ISO 27001) as well as the top skills requested for core cyber-related job roles.

The ECHO project [28] aimed to address the needs and skills gaps of cybersecurity professionals by aligning with know-how from existing frameworks. Their goal was to provide a flexible mechanism to design and develop of learning outcome-based training program. Within this process, they identified the necessity of a stakeholders' skills analysis. In particular, the ECHO project proposed that the identification of competence needs could start from the comparative analysis of multi-sector and inter-sector challenges. Their study remains on high-level analysis. Focusing on hospitals, energy companies, ship crews, and outsourced (third-party) Security Operation Centres (SOC), the project identified several cybersecurity professional roles.

The SPARTA project provides a blueprint for a European Skills Framework [26] and its applicability to education [27]. In particular, the SPARTA Framework is based upon the Joint Research Centre's (JRC) cybersecurity domains taxonomy and the US-based National Initiative for Cybersecurity Education (NICE). As part of deliverable D9.2 'Curricula descriptions' the project identified skills that should be taught in cybersecurity curricula. Organised as a high-level taxonomy, 11 fundamental topics, 16 cybersecurity topics, and 2 new trends were identified for comparing existing cybersecurity study programs. Using this taxonomy, several recommendations on the needed skills were given. SPARTA topics can therefore be viewed as a skills needs analysis from an educative perspective. The relevance of the recommendations were analysed by comparing a selection of existing cybersecurity study programs.

The CyberSec4Europe project dedicated a work package to the identification and prioritisation of the cyber skills needed for education at the university-level [13]. To do so, they conducted a survey to identify the skills needed within the workforce. The ACM Cybersecurity Curricula framework and NICE Cybersecurity Workforce Framework were considered as inputs for the creation of a skills taxonomy. Afterwards a review of European university graduate programs in cybersecurity was conducted to support the identification and prioritisation of the cybersecurity skills required.

As part of Task T3.4 (Establishing an European Education Ecosystem for Cybersecurity), the CONCORDIA project proposed a methodology [6] for designing content for adult learners while considering the specificity of the cybersecurity area. Based on this methodology, new courses could be further developed targeting mainly industry mid-level management and/or executives. The methodology defines five stages that need to be followed sequentially (i.e. ENGAGE, DEFINE, PRODUCE, VALIDATE and DELIVER. Within the first stage (ENGAGE), there is a special focus on the competencies intended to be covered by the course. These competencies are carried across the rest of the stages. In DEFINE, the needed competencies of the selected profile are finalised, the PRODUCE stage builds the curriculum and associated materials, and in VALIDATE new needs for skills development for the given profile are identified.

For the identification and definition of competencies, the methodology proposes several approaches e.g. 1). perform market research by checking for existing courses addressing the different needs, 2). check for trends either by market analysis or by considering new research areas mentioned in EU calls, 3). cluster the needs per industry, 4). check for needs per role profile, 5). seniority matters, and 6). utilising feedback loops. However, these methods were not compared and only one strategy was implemented as an example. Specifically, for the pilot implementation of the methodology within the CONCORDIA project, a course for cybersecurity was implemented. For the analysis of competencies, an initial survey of knowledge, skills, abilities and tasks was carried out based on the NIST NICE framework [30]. These competencies were refined through a workshop supported by a specialised tool, which allowed the visualisation and aggregation of a high number of entries and facilitated their ranking in terms of importance [7]. The latter was extremely important since the totality of the competencies, based on the NICE framework, may be difficult to conceptualise.

The European Cyber Security Organisation (ECSO), the European Commission's contractual public-private partnership includes various cybersecurity stakeholders, large companies, small and medium enterprises, research centres, universities and others. The ECSO prepared two studies which are highly relevant in the context of this research. The ECSO document titled 'European Cybersecurity Education and Professional Training: Minimum Reference Curriculum' [10] tackles the ambitious task of describing four high-level (cybersecurity) curriculum clusters and defines courses titles and corresponding outlines with required topic lists necessary to actually transfer the necessary knowledge to prospective students. The ECSO report entitled 'Information and Cyber Security Professional Certification' [9] reviews a rich set of certifications offered in Europe and worldwide and emphasises the need for a common skills framework on the European-level.

ENISA [12] has recently published a report providing an analysis of the existing capabilities and best practices followed by EU Member States to raise citizen's awareness and consequently assist Member States by providing recommendations on how to improve their cybersecurity capacities. Due to the increasing dependence on ICT in all aspects of society, it is clear that the need for cybersecurity awareness and enhanced skills/competences is critical, and yet the report does not provide any specific analysis on this. As part of this study, 20 structured interviews were conducted with the relevant authorities, in which evidence was collected and additional analytical data were gathered from desktop research. With regards to the improvement of cybersecurity competences, only a limited number of Member States seem to have provisioned actions to strengthen cybersecurity skills through education and to promote digital education and literacy.

ENISA has also produced a report [1] regarding the challenges and recommendations of cybersecurity for Small Medium Enterprises (SMEs), that are all too often insufficiently prepared to address security intrusions and attacks. The recommendations of the report have been developed based on extended state-of-the-art research, together with a two-month survey, where 249 European SMEs provided their feedback on their state of digital security and preparedness for crises such as the COVID-19 pandemic, and with targeted interviews with selected participants. These are structured in three main categories, including people recommendations (including responsibility, employee buy-in, employee awareness, cybersecurity training, cybersecurity policies, and third-party management), process recommendations (including audits, incident planning and response, passwords, software patches, and data protection), and technical recommendations (including network security, antivirus, encryption, security monitoring, physical security, and secure backups). These are mapped onto major security threats and complemented by recommendations directed towards regional, national and European authorities with the objective of assisting SMEs in applying cybersecurity solutions. The report highlights that several recommendations can be implemented by SMEs without having to invest significantly financially or in terms of resources. A majority of these lower cost measures focus on ensuring roles and responsibilities are assigned to appropriate users, and making sure that these users are aware of cybersecurity risks, and how to identify and protect against them. The benefits of such cybersecurity upskilling are critically important for these enterprises, who may be unable to survive a significant a cyber attack.

2.1 Our Contribution

In this paper, we describe our REWIRE methodology for analysing cybersecurity skill needs. Next, we describe the novel, free, web application that applies our methodology and can be used to collect and evaluate the skills needs found in job advertisements. Furthermore, we provide details about the technical implementation of the application and we will demonstrate how it can be used. Finally, we present our plans for further development of the web application.

3 ANALYZING SKILLS NEEDS

The Cybersecurity Skills Alliance - REWIRE project [23] highlighted the need for the development of a European sectorial skills strategy for cybersecurity. This need, and the shortfall in the number of trained cybersecurity professionals available to the labour market, motivated us to develop a methodology for the analysis of skills needed in cybersecurity work roles. This methodology incorporates a series of steps implemented with contribution from all REWIRE partners. The methodology consists of 1). the identification of cybersecurity skills to be used in the analysis, 2). the collection of job advertisements, 3). the automated analysis of the gathered data, and 4). the development of a dynamic web application. Indeed, because cybersecurity is an area in continuous development the classification and the analysis of skills are strongly dependent on the year when the underlying data are collected. The web application allows performing the analysis in an automated way and its database of collected job adverts can be updated allowing new skill needs evaluation and comparison with previous years. Sections 3.1 and 3.2 introduce the used classification of skills and automated machine learning algorithm deployed in the analysis, respectively. The web application is described more fully in Section 4.

3.1 Classification of Skills

Any analysis of cybersecurity skills needs requires a detailed cybersecurity skills taxonomy, i.e. a skills framework, to facilitate the detection of required skills. A skills framework relies on an exhaustive classification of roles, functions, and tasks covering the scope of work performed in daily activities. These role definitions should include the complete scope of 'what are specialists doing in the organisation, unit or role'. The National Initiative for Cybersecurity Education (NICE) Framework [19] provides detailed descriptions of tasks that have to be performed in the fields of cybersecurity in a variety of organisations. It connects theoretical concepts with real-world practice. The NICE Framework identifies a list of 585 Knowledge (K) items, 368 Skills (S) items and 177 Abilities (A) items. In 2020, competencies were introduced to the framework. Specifically, NISTIR 8355 [30] mentions that 'competencies offer flexibility

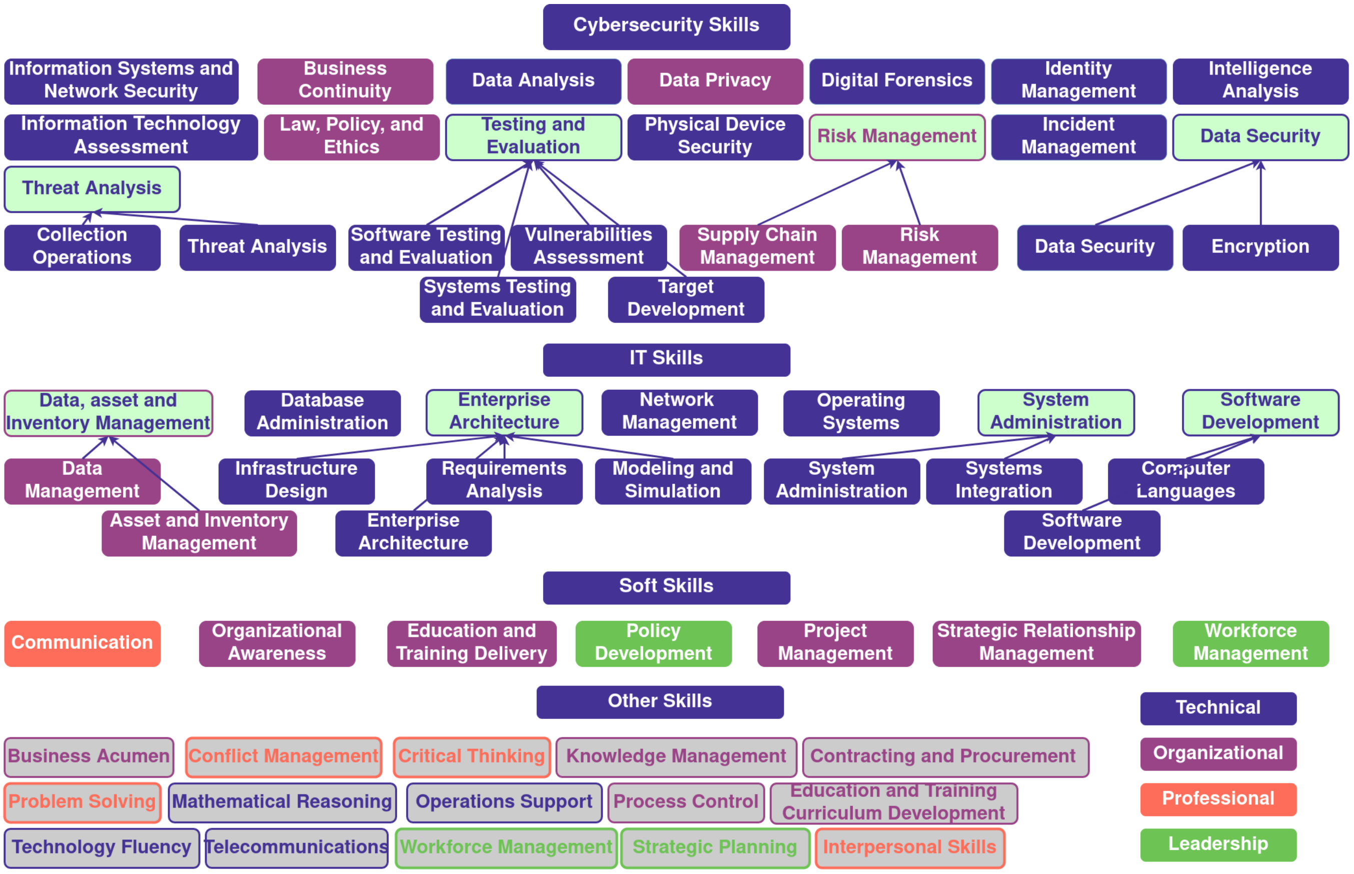


Figure 1: Mapping from NICE to REWIRE competencies.

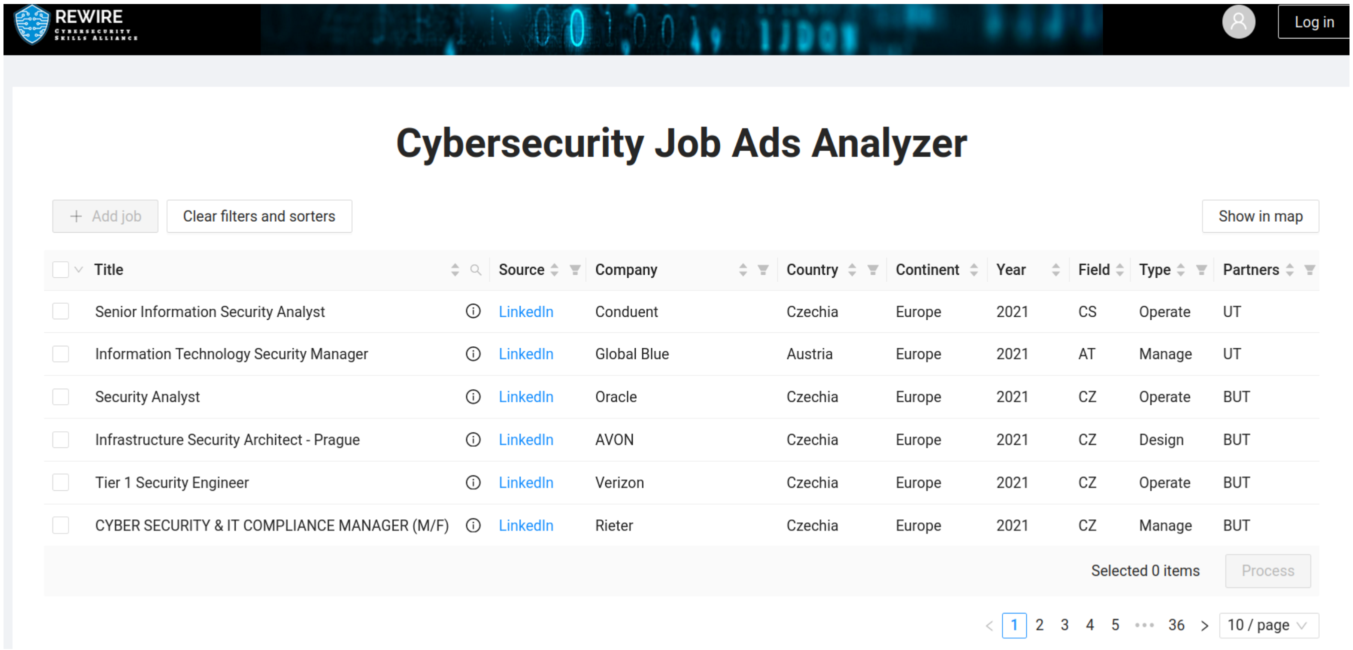
by allowing organizations to group together various Tasks, Knowledge, and Skills (TKS) statements into overarching areas that define a broad need'. The NICE competencies document presents a detailed description of the competencies needed in work roles and, therefore, a direct connection between skills and jobs. 55 competencies divided into Technical, Organisational, Professional, and Leadership categories were identified, with Technical accounting for the majority of the competencies (over 55%).These competencies are of particular interest for our analyses since they allow 'education and training providers to be responsive to employer or sector needs' as mentioned in [19].

The European Union Agency for Cybersecurity (ENISA) started the development of a European Cybersecurity Skills Framework that is expected to report in 2022. The framework will 'take into account the needs of the EU and each one of its Member States' and 'is considered an essential step towards Europe's digital future'. This framework was not available at the time of writing this paper.

Since, currently, there is no universally recognised and accepted European skills framework, we selected the NICE Competencies classification as a starting point. Then an analysis was conducted, taking as an input the groupings (in some cases also mentioned as competencies) introduced from other similar skills frameworks (e.g., e-CF [5], Singapore's [18], Australia's [14], and ISACA's [15]). The objective of the analysis was to identify commonalities and to refine the competencies to a more complete and suitable set. We refer the interested reader to [21] for further information.

As a result of this process, 29 competencies were identified and further utilised in the development of the application. It is the intention of the project team to extend this analysis further and create a final set of competencies. It is important to note that the proposed analysis is not skills-dependent and as such it could be applied to any skills framework. The identified skills were divided into three categories: 1). Cybersecurity Skills, 2). IT Skills, and 3). Soft Skills. Cybersecurity Skills are those Knowledge, Skills, and Abilities (KSAs) that should be known if working in a cybersecurity role while IT Skills represent more general and fundamental Information Technology knowledge. Finally, Soft Skills refer to non-technological KSAs. IT and Soft Skills should also be considered necessary since they are required for the good performance of cybersecurity work roles. It should be remembered that cybersecurity is a multidisciplinary field and therefore requires knowledge beyond simply computer science.

Figure 1 depicts the map from NICE competencies to skills defined by REWIRE. The colour of skill rectangles reflect the NICE groupings, i.e., blue for Technical, purple for Organisational, orange for Professional, and green for Leadership. Merged competencies



Cybersecurity Job Ads Analyzer

Title	Source	Company	Country	Continent	Year	Field	Type	Partners
Senior Information Security Analyst	LinkedIn	Conduent	Czechia	Europe	2021	CS	Operate	UT
Information Technology Security Manager	LinkedIn	Global Blue	Austria	Europe	2021	AT	Manage	UT
Security Analyst	LinkedIn	Oracle	Czechia	Europe	2021	CZ	Operate	BUT
Infrastructure Security Architect - Prague	LinkedIn	AVON	Czechia	Europe	2021	CZ	Design	BUT
Tier 1 Security Engineer	LinkedIn	Verizon	Czechia	Europe	2021	CZ	Operate	BUT
CYBER SECURITY & IT COMPLIANCE MANAGER (M/F)	LinkedIn	Rieter	Czechia	Europe	2021	CZ	Manage	BUT

Figure 2: Cybersecurity Job Ads database environment.

have a light-green rectangle with a border of another colour. Furthermore, they are connected by arrows to the NICE competencies that were merged. "Other Skills" group are those NICE competencies that were considered to be non-essential for this analysis.

3.2 Machine Learning Algorithm

Machine learning (ML) is a technology increasingly used across a wide range of different analytical tasks; including image analysis and time-series forecasting [3, 4]. The ML approach chosen is highly dependant on the nature of analytical task, for instance, Recurrent Neural Networks (RNN) are often used for time series analysis.

RNN sequentially take input data and compress the input data into a context vector [24]. The compressed information can be used to perform tasks such as time-series forecasting. Moreover, RNNs can be applied to Natural language processing (NLP) [31]. To do so, a sentence is interpreted as a time series by dividing it into words and mapping the words into integers. Then the series of integers can be given as input to the RNN model. Importantly, the RNN approach tends to outperform other algorithms for NLP.

Furthermore, the RNN approach has also been used for sentiment analysis, i.e., a task classifying the negative or positive sentiments of a sentence by using ML algorithms, and has shown great performance. The classification of required skills for job advertisements can be made analogously to sentiment analysis methodology. For this project, the ambition was to analyse advertisements with an RNN model and thereby automate the analysis of job advertisements.

Our model uses a pre-built word-piece tokeniser for Bidirectional Transformer (BERT) [3]. With the tokens from the word-piece tokeniser, a model is used to concatenate the mean of all intermediate context vectors and the last context vector. The concatenated vector is fed to an Artificial Neural Network (ANN) to classify security skills. The hidden unit of Long Short-Term Memory (LSTM) is 128, and the hidden layer of ANN is 256.

The training consisted of testing the model's ability to predict the required security skills using a test dataset. The ground truths were manually labeled and the elements of the labeled dataset were randomly assigned 90% and 10% as training dataset and test dataset respectively. Furthermore, the LSTM model was only trained on the training dataset to prevent the model from learning the test dataset and thereby biasing the training results.

It is important to notice that two ML models and a dictionary-based analysis were compared before choosing the proposed model. In our preliminary comparisons and analyses, we considered two training datasets: 1) A small dataset (i.e., 31 ads), and 2) A medium dataset (i.e., 87 ads). The collected sets were sufficient to evaluate the models. In our training, the learning rates were set to 0.001, batch size set to 10, and epochs set to 50. We refer to [21] for more details. Note that the batch size is small due to the limited number of elements in main job adverts dataset (currently, we have collected 266 job ads). Only English-written job ads were considered.

In order to evaluate the performance of the selected model, we computed the accuracy by comparing the model prediction and the ground truths of the test dataset. In particular, the accuracy was defined as the fraction of correct predictions in the entire prediction,

$$\text{Accuracy} = \frac{\text{correct}_{pred}}{\text{correct}_{pred} + \text{wrong}_{pred}},$$

where correct_{pred} and wrong_{pred} are the number of correct and wrong predictions, respectively. Note that the accuracy is evaluated after each epoch. The mean and standard deviation (namely, StDev) of the accuracy was computed over 10 repetitions. The model tested on the medium dataset showed higher performance (i.e., Train data: Mean 83.31 and StDev 0.63, and Test Data: Mean 77.03 and StDev 1.21) when compared to the model trained on the small dataset (i.e., Train data: Mean 80.15 and StDev 0.98, and Test Data: Mean 68.82.0 and StDev 2.94). The analysis indicated that the size of the data used and the model's accuracy should be prioritised when retraining. We refer to [21] for more details.

Accordingly, we suggest the following two approaches:

Accuracy-based approach. This can be implemented in three ways. Firstly, after performing an analysis, a user may choose to manually correct the suggested skills. This can be regarded as accuracy-based training for the model. Secondly, the model can be run periodically on up-to-date advertisements and calculate the difference between the corrected labels and the predictions from our ML model. However, the period may need to vary depending on the seasonality of job markets. The model should not be retrained to account for a limited number of new job advertisements. Finally, if the accuracy is lower than a certain threshold, we can trigger retraining.

Data-based approach. A buffer of job advertisements to train the model can be defined. Therefore, when a certain amount (i.e., the size of the buffer) of job advertisements are recorded, we can automatically trigger the retraining of the model.

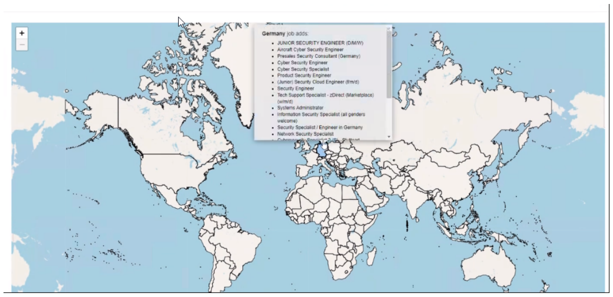
3.3 Advantages

The ML learning approach we have taken for our analytical strategy offers a number of advantages. Textual analysis, particularly when performed manually on a large number of documents, can be labour- and time-intensive. Whilst there is a need to perform some manual analysis and train a model to accurately assess texts, once trained, it becomes far easier to use the ML model on additional data, even when accounting for additional training and optimisation. As such, we are able to analyse a greater number of job advertisements without significant extra cost. An alternative approach to manual assessment may have been to utilise a regex-matching to simply count the occurrences of certain phrases or words across the job advertisements dataset. However, this approach would fail to take account for the nuance in the way that certain sentiments were phrased, for example, a simple search for 'digital forensics' would flag both 'experience of digital forensics is required' and 'no experience in digital forensics is expected' when clearly the needs expressed are different. Machine Learning, because it is learning from manual assessments is better able to detect these differences and categorise them appropriately. A third advantage of our approach is that it allows a user to easily add their own data to the dataset and include them in analyses, without significant work on their behalf. This increases the flexibility of the tool, as well as its usability.

An alternative approach may have been to utilise a survey of businesses to understand cybersecurity skills needs. However, conducting such survey presents a number of problems. It would be necessary for example, to identify businesses that are likely to be recruiting cybersecurity positions across a number of European nations. This would be very time-consuming and resource-intensive to do effectively. Furthermore, having identified potential businesses, there may be issues due to low response rates when requesting that they complete a survey. As such, we were able to obtain more data with less difficulty by accessing publicly available job advertisements.

4 JOB ADS ANALYZER: PURPOSE, TARGETED USERS, USABILITY

The web application allows users to add new job advertisements, select any advert present in the database, and run a ML algorithm to identify the most frequently mentioned cybersecurity skills in the selected sample. Using the tool and its deployed ML algorithm, it is easy to research the cybersecurity skills currently being requested by the labour market and how these requirements have evolved over time. We refer to Section 3 for more details on the ML algorithm.



Germany job adds:

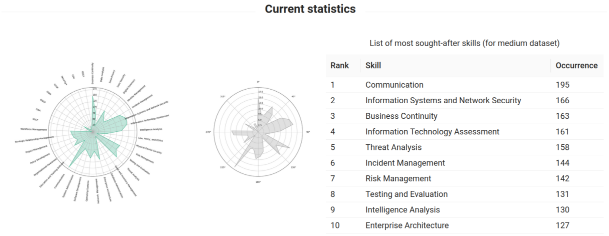
- JUNIOR SECURITY ENGINEER (D/M/W)
- Aircraft Cyber Security Engineer
- Presales Security Consultant (Germany)
- Cyber Security Engineer
- Cyber Security Specialist
- Product Security Engineer
- (Junior) Security Cloud Engineer (f/m/d)
- Security Engineer
- Tech Support Specialist - zDirect (Marketplace) (w/m/d)
- Systems Administrator
- Information Security Specialist (all genders welcome)
- Security Specialist / Engineer in Germany
- Network Security Specialist

Figure 3: Cybersecurity Job Ads map environment.

4.1 Using the Tool

The tool has three main environments: 1) the database, 2) the map, and 3) the ML results that are depicted in Figures 2, 3 and 4, respectively. The database allows users to add job adverts and filter the adverts using the fields. The map is a visual representation of the database where job adverts are split per country and several filter options are also available. Finally, the ML results show the identified cybersecurity skills required by employers within the selected job adverts.

4.1.1 The Database Environment. Currently, the database comprises 266 job ads from countries located in Europe. Figure 2 shows the job adverts database. Advertisements are divided into columns according to the title, source, company name, country, continent, year, field, type, and partner who added the advert to the database. This allows any user to filter the job adverts on demand, for example, if one is interested in identifying which cybersecurity skills are required for work roles in Austria. Each advertisement also contains an information icon. If the user clicks on this icon, the job description will expand and the original text can be seen. Moreover, if the user logs in, they are able to add advertisements using the 'Add job' button in the upper left corner. The database field 'partner' enables a user to easily select the adverts that they have uploaded to the database in case they wish to analyse only those adverts.



Current statistics

List of most sought-after skills (for medium dataset)

Rank	Skill	Occurrence
1	Communication	195
2	Information Systems and Network Security	166
3	Business Continuity	163
4	Information Technology Assessment	161
5	Threat Analysis	158
6	Incident Management	144
7	Risk Management	142
8	Testing and Evaluation	131
9	Intelligence Analysis	130
10	Enterprise Architecture	127

Figure 4: Cybersecurity Job Ads ML results environment.

4.1.2 The Map Environment. The web page also presents a second environment as depicted in Figure 3. The map opens when the user clicks on 'Show in map' button in the upper right corner of the database environment. The map shows a preview of the jobs available in each country that have been uploaded to the database. Here, the job adverts can again be filtered by country.

4.1.3 The ML Results Environment. Once the jobs adverts to be analysed have been selected, the user can run the ML algorithm and the web page displays the result in the form of a table showing the distribution of the required skills found in the sample. An Example of experimental results is shown in Figure 4.

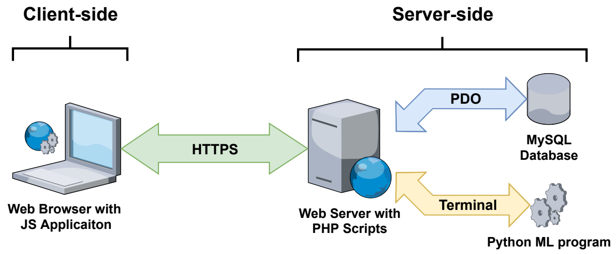


Figure 5: Application structure.

4.2 Technical Implementation

The web application is divided into two parts: 1). the client and 2). the server with the database. The structure is shown in Figure 5. In order to increase security, client and server parts are only accessible through an encrypted HTTPS protocol.

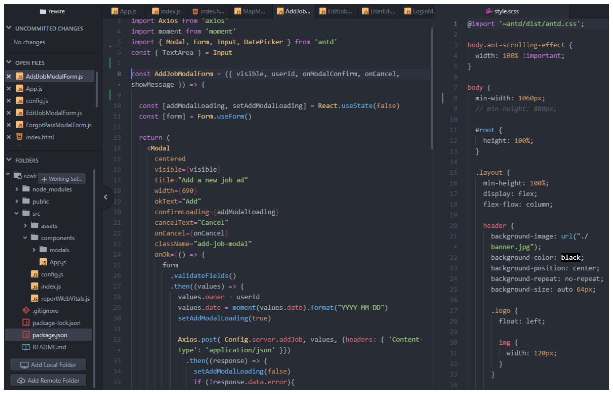
4.2.1 Client-side. The client part is implemented as a front-end application written in JavaScript language (ECMAScript 6), which runs in the client's (visitor's) web browser. The core of the application is based on the React framework. The following programs and libraries were also used:

- Antd – Graphical design, components, and structure,
- Crypto – Cryptography features,
- Moment – Date formatting,
- Node-sass – Style preprocessor,
- Leaflet – Map component,
- React-highlight-words – Search feature,
- Axios – Connection to the data server,
- NPM - Node Packet Manager.

The entire React application is divided into sub-components that inherit basic React features and behaviours such as *states* and *props*. The <App> component, during initialisation, downloads all job data from the server part of the application using an HTTP GET request. The data are stored in the *states* of the <App> component and are updated only in the case of data manipulation (addition, removal, or modification of the job advert). The data are distributed to other components through the *props* feature.

The <App> component contains a table from the Antd library, which displays all available job adverts. The table allows filtering and sorting based on several criteria and the selection of an advert for processing by the ML algorithm. The table contains only a few of the most important parameters: name, source, company, country, continent, date, field, type, and partners. The user interface also has buttons for logging in and adding new job adverts.

The application is designed as a one-page dynamic application, so all other sections and functions are created as pop-up modals. Modal components handle user actions and data editing. These events are followed by HTTP requests to the server. The server then processes the request and returns a processing status response or data. These events include, for example, user authentication, registration, modification, deletion, and adding job adverts. A code sample of one of the modals in the Atom development environment is shown in Figure 6.



```
import Axios from 'axios'
import moment from 'moment'
import { Modal, Form, Input, DatePicker } from 'antd'
const { TextArea } = Input

const AddJobModalForm = ({ visible, userId, onModalConfirm, onCancel,
showMessage }) => {

  const [addModalLoading, setAddModalLoading] = React.useState(false)
  const [form] = Form.useForm()

  return (
    <Modal
      centered
      visible={visible}
      title="Add a new job ad"
      width={690}
      okText="Add"
      confirmLoading={addModalLoading}
      cancelText="Cancel"
      onCancel={onCancel}
      className="add-job-modal"
      onOk={() => {
        form
          .validateFields()
          .then((values) => {
            values.owner = userId
            values.date = moment(values.date).format("YYYY-MM-DD")
            setAddModalLoading(true)

            Axios.post( Config.server.addJob, values, {headers: { 'Content-
            Type': 'application/json' }})
              .then((response) => {
                setAddModalLoading(false)
                if (!response.data.error){
```

```
@import '~antd/dist/antd.css';

body.ant-scrolling-effect {
  width: 100% !important;
}

body {
  min-width: 1060px;
  // min-height: 880px;

  #root {
    height: 100%;
  }

  .layout {
    min-height: 100%;
    display: flex;
    flex-flow: column;

    header {
      background-image: url("./
      banner.jpg");
      background-color: black;
      background-position: center;
      background-repeat: no-repeat;
      background-size: auto 64px;

      .logo {
        float: left;

        img {
          width: 120px;
        }
      }
```

Figure 6: Atom Development Environment.

4.2.2 Server-side. The server part of the application consists of several PHP (i.e., PHP: Hypertext Preprocessor) scripts, located on the web server next to the source JavaScript codes of the client part. These server scripts handle requests coming from clients' Internet browsers and serve as an intermediary for data manipulation. Each script is for a specific action. They are triggered by a query for a specific URL such as *https://SERVER_URL/addJob.php*. The scripts require specific data in the query body according to the required operation, as well as an authorisation token. They will read or write data to the application database based on a valid request. The connection to the database is provided by PHP Data Objects (PDO), which helps to prevent SQL injection and character set inconsistencies by design.

4.2.3 Database. The application database is implemented on the same machine via the MySQL server MariaDB. For security reasons, access to the database is restricted to localhost only, i.e., only to users and programs running on the server. The database currently contains only a table with users and a table with job adverts. The job table contains, in addition to the values displayed on the web, dozens of other parameters that act as raw truths for the ML algorithm.

4.2.4 Machine Learning Implementation. The ML algorithm for skill prediction and keyword processing is implemented as a Python program. The program consists of a tokeniser and an LSTM model. The machine learning model can also be re-created and re-learned. The ML program is stored on the webserver in a sub-directory of web content.

When requesting data processing by the ML algorithm, an HTTP request is sent from the client application to a PHP script which, in the first step, reads all the necessary data from the database and creates a dataset in a temporary directory on the server. In the second step, the PHP script executes the terminal command using `exec` function to run the Python ML program. After performing the ML analysis, the program outputs the return code back to the PHP script and creates the analysis output files in a temporary folder. The PHP script then loads these files, processes them, and sends the results back to the client application for viewing by the user.

5 EVALUATION OF JOB ADS

5.1 Methodology

For our preliminary analyses we adopted two approaches:

1. Generating radar plots to display skills occurrences.
2. A Rank Biased Overlap (RBO) comparison of skills occurrences.

The database environment currently stores 266 adverts from 26 countries across Europe, and from this collection of job adverts we used the ML algorithm to generate ranked skills occurrence results for three countries. The results are ranked by the frequency of occurrence and as such a ranking of 1 should be interpreted as the most commonly occurring skill. Results were generated for Czechia, Germany, and Hungary as these countries had the greatest number of available job adverts. We also utilised the ranked skills occurrence result for the overall database as a baseline for our analyses. From these results, we generated radar plots so that we could visually compare the skills occurrence rankings across countries and against the baseline.

The rankings we obtained for each country were also compared using the Rank Biased Overlap (RBO) measure [29]. The RBO measure was developed to calculate the similarity between ranked lists that are both indefinite and require weighting. This was selected as an appropriate measure for our analyses as not all skills were identified in the job adverts from each country, making the list of rankings indefinite with missings, and because of the skewness of the underlying frequencies. Analysis of the full job adverts database identified 26 skills with a total frequency of 1,926 occurrences. However, the uppermost 13 skills account for 1,743 of these, approximated 90.5% of the total number. This means that the difference in the frequencies between, for example, the skills ranked 19th and 20th is likely to be very different to the difference in the frequencies between the skills ranked 3rd and 4th. As such, we cannot think of the difference in ranks between consistent across the lists. We used the implementation of RBO in the *gespeR* [25] package in the *R* statistical environment [22] to calculate the RBO for each country in comparison to all others. We weighted the upper rankings at 0.905 to account for them representing 90.5% of the total occurrence frequencies.

5.2 Results

The rankings for Czechia, Germany, Hungary and the overall database can be seen in Table 1. The skills are ordered in Table 1 by their ranking in the overall database.

Table 1: Summary of Rankings

Skill	Rankings		
	Cz	De	Hu
1. Communication	1	1	1
2. Threat Analysis	2	3	2
3. Data Security	3	2	3
4. Information Systems and Network Security	4	4	4
5. Risk Management	5	5	6
6. Testing and Evaluation	6	6	5
7. Operating Systems	7	7	7
8. Incident Management	8	8	8
9. Enterprise Architecture	9	9	9
10. Information Technology Assessment	11	11	10
11. Business Continuity	12	10	13
12. Project Management	13	14	14
13. Intelligence Analysis	10	12	11
14. Organisational Awareness	14	13	12
15. Identity Management	15	15	15
16. Software Development	16	20	16
17. Law, Policy, and Ethics	19	25	19
18. System Administration	17	16	17
19. Policy Development	21	19	20
20. Strategic Relationship Management	20	17	18
21. Data Analysis	18	24	19
22. Digital Forensics	23	18	-
23. Workforce Management	26	23	-
24. Database Administration	24	21	-
25. Network Management	25	22	-
26. Education and Training Delivery	22	-	-
Number of adverts	84	24	19

These values were then plotted as radar plots. In the radar plots, the closer a point is to the outer edge, the higher the underlying factor was in the skills occurrence rating. By contrast, skills which are positioned more closely to the centre of the radar plot, appeared lower in the occurrence rankings. The plot for Czechia compared to the overall rankings can be seen in Figure 7.

As we can see, the plot of rankings for Czechia very closely follows the overall rankings for the most commonly occurring skills. Indeed the 9 most commonly occurring skills in the jobs adverts from Czechia exactly match the 9 most commonly occurring skills across all job adverts. There is greater variation for the least commonly skills though with the greatest difference being for *Education and Training Delivery* which was the lowest occurring skill overall but ranked 4 place higher at 22nd in Czechia. A similar pattern was

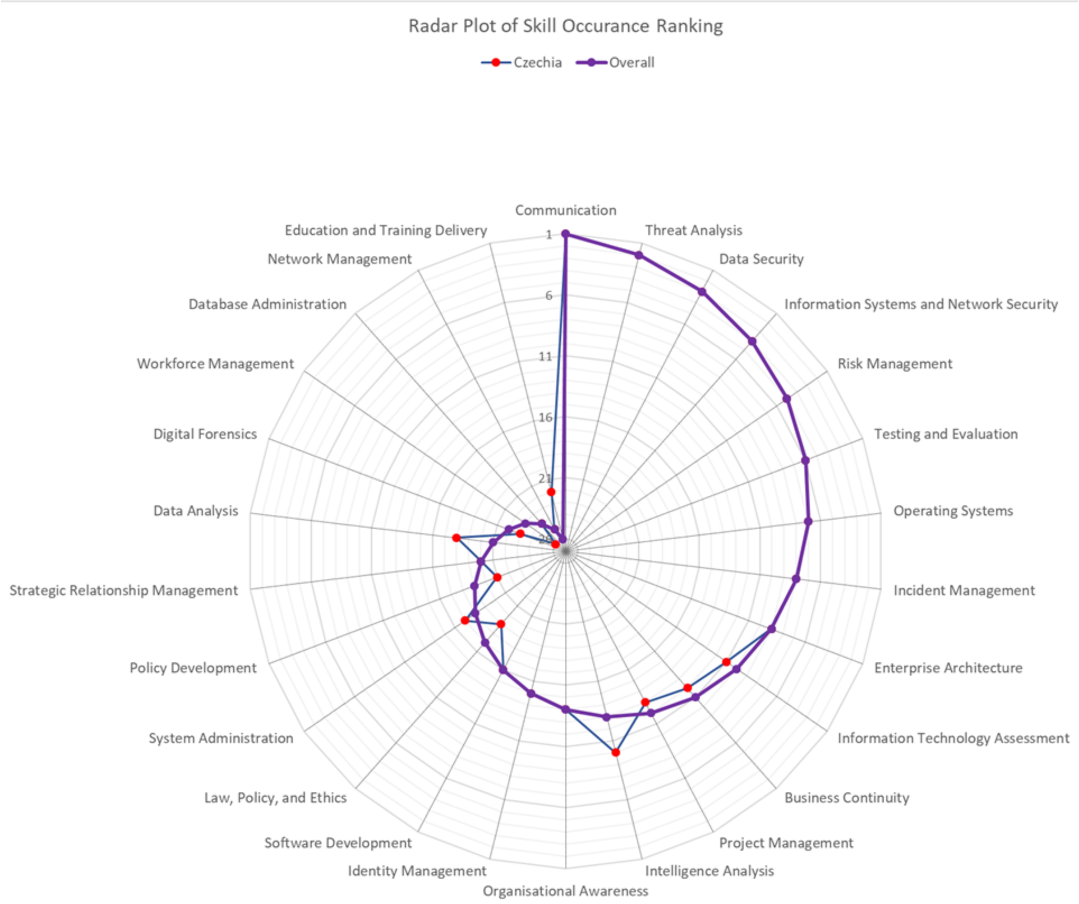


Figure 7: Radar Plot of Skills Occurrence Ranking for Czechia

observed for Germany and Hungary, with the greatest variation seen for the least commonly occurring skills.

To ascertain how similar the ranked lists are, we used the RBO measure to compare the rankings obtained for Czechia, Germany, Hungary, and the overall database. The RBO measure is scored between 0 and 1, with a score of 0 indicating that there is no similarity between two lists, and a score of 1 indicating that the lists are perfectly aligned. The results from this analysis can be seen in Table 2.

Table 2: Rank Biased Overlap Scores

	Czechia	Germany	Hungary	Overall
Czechia	1	-	-	-
Germany	0.6075	1	-	-
Hungary	0.3051	0.5385	1	-
Overall	0.7926	0.5864	0.3051	1

The results in Table 2 do support the view that the cybersecurity jobs adverts from Czechia are similar to the overall rankings, with an RBO score of 0.7926. This corresponds with the radar plot of the rankings seen in 7. For Germany and Hungary, the RBO scores are lower suggesting that they are less alike the overall rankings than Czechia. Hungary in particular appears to vary substantially to the overall ranking, which may be a result of a number of skills that were not detected by the ML algorithm. This missingness is not necessarily the result of there being fewer Hungary-based job adverts, but may suggest a tendency to Hungary for adverts to focus on a narrower set of skills.

We should be cautious however with these preliminary results as they may be inaccurate due to the relatively small number of job advertisements currently available. It is quite possible that differences between countries are the result of an unrepresentative sample. We must also consider that the decision to focus on English job adverts may bias the sample towards larger, multinational companies, that may be less reflective of the attitudes of businesses in the country from which the advert collected. It may be expedient in the future to compare the English and non-English job advertisements in certain countries to ascertain how significant this potential issue is for our analyses.

Additional jobs adverts for each country are required for us to explore in more depth the differences in cybersecurity job adverts across Europe and to better identify any trends that might exist.

Increasing the number of job adverts available for each country would also reduce the potential inaccuracy caused by using limited samples.

6 CONCLUSION

In this paper, we described a practical interactive tool for the identification of skills needed for current cybersecurity work roles. The tool aims to help education and training providers as well as industry in mapping job profiles and required skills. To do so, we developed a methodology for skills needs analysis based on a machine learning algorithm able to recognise the skills referred to in job advertisements. The dynamism of our analytical approach through an online application is essential due to the evolving nature of the cybersecurity labour market. Currently, our web application is based on the modified NIST NICE cybersecurity competencies framework. As a next step, we plan to extend the tool to also support the European Cybersecurity Skills Framework once published by ENISA. In this way, a user will be able to choose the most relevant framework to use when analysing adverts, either a US or a European framework. Furthermore, we plan to make the ingest of job advertisements automatic via link addresses with a duplicate detection feature, and to extend the database to include approximately 1,000 job advertisements to improve the consistency of the analysis and to better identify new and emerging skill trends.

ACKNOWLEDGMENTS

The following funding sources are gratefully acknowledged: the ERASMUS+ programme of the European Union (grant 621701-EPP-1-2020-1-LT-EPPKA2-SSA-B 'REWIRE'), the Ministry of the Interior of the Czech Republic (grant VJ01030001), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – NFDI 1/1 "GHGA – German Human Genome-Phenome Archive", and the Ministry of Education, Science and Technological Development of the Republic of Serbia through project no. 451-03-68/2022-14/200156 "Innovative scientific and artistic research from the FTS (activity) domain".

REFERENCES

[1] Georgia Bafoutsou Anna Sarri, Viktor Paggio. 2021. Cybersecurity for SMEs: Challenges and Recommendations. https://www.enisa.europa.eu/publications/enisa-report-cybersecurity-for-smes

[2] ANSSI. 2021. SecNumedu, labeling of higher education courses in cybersecurity. https://www.ssi.gouv.fr/en/cybersecurity-in-france/formations/secnumedu-labeling-of-higher-education-courses-in-cybersecurity/

[3] Stuart Berg, Dominik Kutra, Thorben Kroeger, Christoph N Straehle, Bernhard X Kausler, Carsten Haubold, Martin Schiegg, Janez Ales, Thorsten Beier, Markus Rudy, et al. 2019. Ilastik: interactive machine learning for (bio) image analysis. *Nature Methods* 16, 12 (2019), 1226–1232.

[4] Gianluca Bontempi, Souhaib Ben Taieb, and Yann-Aël Le Borgne. 2012. Machine learning strategies for time series forecasting. In *European business intelligence summer school*. Springer, 62–77.

[5] CEN. 2020. European e-Competence Framework 3.0. https://itprofessionalism.org/app/uploads/2019/11/User-guide-for-the-application-of-the-e-CF-3.0_CEN_CWA_16234-2_2014.pdf

[6] Felicia Cutas. 2020. CONCORDIA: Methodology for the creation and deployment of new courses and/or teaching materials for cybersecurity professionals. https://www.concordia-h2020.eu/wpcontent/uploads/2020/06/CONCORDIA-methodology-courses-professionals-for-publication.pdf

[7] Felicia Cutas. 2020. CONCORDIA Workshop on Education for cybersecurity professionals - post workshop report -. https://www.concordia-h2020.eu/wp-content/uploads/2020/07/CONCORDIAWorkshoponEducation2020-forpublication.pdf

[8] European Cyber Security Organisation (ECSO). 2018. Gaps in European Cyber Education and Professional Training. https://ecs-org.eu/documents/publications/5fdb282a4dcbd.pdf

[9] European Cyber Security Organisation (ECSO). 2020. Infromation and Cyber Security Professional Certification. https://ecs-org.eu/documents/publications/60101ad752a50.pdf

[10] European Cyber Security Organisation (ECSO). 2021. European Cybersecurity Education and Professional Training: Minimum Reference Curriculum. https://ecs-org.eu/documents/publications/62164c38c8139.pdf

[11] ENISA. 2020. Cybersecurity Skills Development in the EU. https://www.enisa.europa.eu/publications/the-status-of-cyber-security-education-in-the-european-union

[12] e-Governance Academy (EGA) European Union Agency for Cybersecurity (ENISA). 2021. *Raising awareness of cybersecurity*. Number November. 53 pages. www.enisa.europa.eu.

[13] Cyber Security for Europe. 2020. D6.2 Education and Training Review. https://cybersec4europe.eu/wp-content/uploads/2020/02/D6.2-Education-and-Training-Review-V1.2-Submtted.pdf

[14] Australian Governament. 2020. ASD Cyber Skills Framework. https://www.cyber.gov.au/sites/default/files/2020-09/ASD-Cyber-Skills-Framework-v2.pdf

[15] ISACA. 2022. State of Cybersecurity 2022 Report. https://www.isaca.org/go/state-of-cybersecurity-2022

[16] (ISC)^2. 2021. (ISC)^2 Cybersecurity Workforce Study. https://www.isc2.org//-/media/ISC2/Research/2021/ISC2-Cybersecurity-Workforce-Study-2021.ashx

[17] Darragh McHenry, Tania Borges, Alex Bollen, Jayesh Navin, Sam Donaldson, David Crozier, and Steven Furnell. 2021. Cyber security skills in the UK labour market 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1042429/Cyber_skills_in_the_labour_market_report_v6_.pdf

[18] INFOCOMM media development Authority. 2021. Skills Framework for ICT. https://www.imda.gov.sg/cwp/assets/imtalent/skills-framework-for-ict/index.html#

[19] William Newhouse, Stephanie Keith, Benjamin Scribner, and Greg Witte. 2017. National initiative for cybersecurity education (NICE) cybersecurity workforce framework. *NIST special publication* 800, 2017 (2017), 181.

[20] Daniel Pedley, Darragh McHenry, Helen Motha, and J Shah. 2018. Understanding the UK cyber security skills labour market. *United States Sentencing Commission, Sentencing Guidelines for United States Courts, http://www. ussc.gov/FEDREG/05_04_notice. pdf* (2018).

[21] REWIRE project. 2021. WP2 Methodology to anticipate future needs. https://rewireproject.eu/wp-content/uploads/2021/12/R2.2.3-MethodologyToAnticipateFutureNeeds-FINAL.pdf

[22] R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

[23] REWIRE. 2020. REWIRE: Cybersecurity Skills Alliance - A new Vision for Europe. https://rewireproject.eu/

[24] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. 1986. Learning representations by back-propagating errors. *nature* 323, 6088 (1986), 533–536.

[25] Fabian Schmich. 2021. *gespeR: Gene-Specific Phenotype EstimatoR*. http://www.cbg.ethz.ch/software/gespeR R package version 1.26.0.

[26] SPARTA. 2019. D9.1 - Cybersecurity skills framework. https://www.sparta.eu/assets/deliverables/SPARTA-D9.1-Cybersecurity-skills-framework-PU-M12.pdf

[27] SPARTA. 2020. D9.2 - Curricula descriptions. https://www.sparta.eu/assets/deliverables/SPARTA-D9.2-Curricula-descriptions-PU-M18.pdf

[28] Pavel Varbanov. 2021. D2.6 ECHO CYBERSKILLS FRAMEWORK. https://echonetwork.eu/wpcontent/uploads/2021/03/ECHO_D2.6_Cyberskills-Framework.pdf

[29] William Webber, Alistair Moffat, and Justin Zobel. 2010. A Similarity Measure for Indefinite Rankings. *ACM Transactions on Information Systems* 28, 4 (2010).

[30] Karen A. Wetzel. 2021. NICE Framework Competencies: 19 Assessing Learners for Cybersecurity Work. https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8355-draft.pdf

[31] Wenpeng Yin, Katharina Kann, Mo Yu, and Hinrich Schütze. 2017. Comparative study of CNN and RNN for natural language processing. *arXiv preprint arXiv:1702.01923* (2017).