



**PhD in Information Technology and Electrical Engineering**  
**Università degli Studi di Napoli Federico II**

**PhD Student: Andrea Vignali**

---

**Cycle: XXXVIII**

**Training and Research Activities Report**

**Year: First**

*Andrea Vignali*

**Tutor: prof. Giancarlo Sperli**

*Giancarlo Sperli*

**Co-Tutor: prof. Simon Pietro Romano**

*Simon Pietro Romano*

**Date: October 18, 2023**

# Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle: XXXVIII

Author: Andrea Vignali

## 1. Information:

- PhD student: Andrea Vignali
- DR number: DR996624
- Date of birth: 03/06/1996
- Master Science degree: Computer Engineering
- University: Università degli Studi di Napoli Federico II
- Doctoral Cycle: XXXVIII
- Scholarship type: *PNRR – DM 352*
- Tutor: Giancarlo Sperli
- Co-tutor: Simon Pietro Romano

## 2. Study and training activities:

Activity	Type <sup>1</sup>	Hours	Credits	Dates	Organizer	Certificate <sup>2</sup>
On the challenges and impact of Artificial Intelligence in the Insurance domain	Course	12	3	30/11/2022	Ing. Lorenzo Riccardi Celsi	Y
Stabilizer Renyi Entropy and Quantum Complexity	Seminar	1	0.2	02/11/2022	Prof. Alioscia Hama	Y
Connecting the dots investigating an APT campaign using Splunk	Seminar	2	0.4	11/11/2022	Dr. Antonio Forzieri	Y
Data Mining the output of quantum simulators - from critical behavior to algorithmic complexity	Seminar	1	0.2	11/11/2022	Dr. Marcello Dalmonte	Y
CRASH COURSE on DATA EXCELLENCE – PART I	Seminar	2	0.4	14/11/2022	Roberto Maranca	N

# Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle: XXXVIII

Author: Andrea Vignali

<b>Cybercrime and information warfare: national and international actors</b>	<b>Seminar</b>	<b>2</b>	<b>0.4</b>	<b>18/11/2022</b>	<b>Dr. Pierluigi Paganini</b>	<b>Y</b>
<b>Privacy and Data Protection</b>	<b>Seminar</b>	<b>2</b>	<b>0.4</b>	<b>22/11/2022</b>	<b>Dr. Stefano Mele</b>	<b>Y</b>
<b>Automated Offensive Security: Intelligence is all you need</b>	<b>Seminar</b>	<b>1</b>	<b>0.2</b>	<b>28/11/2022</b>	<b>Prof. Simon Pietro Romano</b>	<b>N</b>
<b>Progettazione di strategie di controllo in ambiente Simulink</b>	<b>Seminar</b>	<b>3</b>	<b>0.6</b>	<b>01/12/2022</b>	<b>Dr. Gianfranco Fiore</b>	<b>Y</b>
<b>Game Theory for Information Engineering</b>	<b>Seminar</b>	<b>3</b>	<b>0.6</b>	<b>13/12/2022</b>	<b>Prof. Leonardo Badia</b>	<b>Y</b>
<b>From Cyber Situational Awareness to Adaptive Cyber Defense Leveling the Cyber Playing Field</b>	<b>Seminar</b>	<b>2</b>	<b>0.4</b>	<b>13/12/2022</b>	<b>Prof. Massimiliano Albanese</b>	<b>Y</b>
<b>Threat Hunting &amp; Incident Response</b>	<b>Seminar</b>	<b>2</b>	<b>0.4</b>	<b>13/12/2022</b>	<b>Vladimir Kurdin</b>	<b>Y</b>
<b>Malware Analysis</b>	<b>Seminar</b>	<b>2</b>	<b>0.4</b>	<b>15/12/2022</b>	<b>Dr. Gaetano Pellegrino</b>	<b>Y</b>
<b>Principi Architetture TOGAF 1</b>	<b>Seminar</b>	<b>3</b>	<b>0.6</b>	<b>30/01/2023</b>	<b>Alberto Curcio, Pietro Boscolo</b>	<b>N</b>
<b>Data Strategy</b>	<b>Seminar</b>	<b>3</b>	<b>0.6</b>	<b>03/02/2023</b>	<b>Lorenza Catalano</b>	<b>N</b>
<b>IoT Data Analysis</b>	<b>Course</b>	<b>12</b>	<b>4</b>	<b>09/02/2023</b>	<b>Prof. Raffaele Della Corte</b>	<b>Y</b>

# Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle: XXXVIII

Author: Andrea Vignali

<b>Blockchain and 5G business</b>	<b>Seminar</b>	<b>3</b>	<b>0.6</b>	<b>13/02/2023</b>	<b>Luca Confronto</b>	<b>N</b>
<b>Algorithm Unrolling: Efficient, Interpretable Deep Learning for Signal and Image Processing</b>	<b>Seminar</b>	<b>1</b>	<b>0.2</b>	<b>14/02/2023</b>	<b>Prof. Vishal Monga</b>	<b>Y</b>
<b>Scientific programming and visualization with Python</b>	<b>Course</b>	<b>20</b>	<b>2</b>	<b>23/02/2023</b>	<b>Prof. Alessio Botta</b>	<b>Y</b>
<b>Il cloud e gli hyperscalers + high performance computing</b>	<b>Seminar</b>	<b>3</b>	<b>0.6</b>	<b>28/02/2023</b>	<b>Giovanni Vendramel</b>	<b>N</b>
<b>-Open Access and Transformative Agreements in Italy: the Current State of the Art</b>  <b>-How to Publish Open Access Articles with IEEE under the CARE CRUI Agreement</b>  <b>-Additional Insights on Open Access Publishing</b>	<b>Seminar</b>	<b>1.5</b>	<b>0.3</b>	<b>31/03/2023</b>	<b>Prof. Davide Risso</b>	<b>N</b>
<b>MLOps: Achieving Operational Velocity with Faster Delivery and Collaboration</b>	<b>Seminar</b>	<b>1</b>	<b>0.2</b>	<b>02/03/2023</b>	<b>Prof. Tarry Singh</b>	<b>Y</b>
<b>Statistical Data Analysis for Data and Engineering</b>	<b>Course</b>	<b>12</b>	<b>4</b>	<b>18/04/2023</b>	<b>Prof. Roberto Pietrantuono</b>	<b>Y</b>

# Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle: XXXVIII

Author: Andrea Vignali

Traffic Engineering with Segmented Routing: optimally addressing popular use cases	Seminar	1	0.2	23/06/2023	Prof. Pascal Merindol	Y
BGP & Hot-Potato Routing: graceful and optimal convergence in case of IGP events	Seminar	1	0.2	29/06/2023	Prof. Pascal Merindol	Y
- Machine Learning and Big data  -Machine Learning for Engineering	Tutorship	7	0.3	31/05/2023 18/05/2023 01/06/2023		
Scienza moderna e disciplina giuridica dell'Intelligenza Artificiale	Course	24	6	20/07/2023	Prof. Lucio Franzese	Y
CNTC (Complex networks and telecommunications 3rd edition: Towards 6G) – PhD school	Doctoral School		4	27/02/2023		Y
Semantic artifacts and multimedia knowledge graphs for biodata integration	Course	10	2	10/10/2023	Prof. Cristiano Russo	Y
Artificial Intelligence and Natural Language Processing	Course	13	3	16/10/2023	Prof. Francesco Cutugno	Y
Big Data Architecture and Analytics	Course	20	5	17/10/2023	Prof. Giancarlo Sperli	Y

# Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle: XXXVIII

Author: Andrea Vignali

“RICERCA FORMAZIONE NELLA SOCIETÀ DELLA TRANSIZIONE DIGITALE”	E	Seminar	5	1	22/09/2023		N
--	---	---------	---	---	------------	--	---

- 1) Courses, Seminar, Doctoral School, Research, Tutorship
- 2) Choose: Y or N

## 2.1. Study and training activities - credits earned

	Courses	Seminars	Research	Tutorship	Total
Bimonth 1	3	4.6	2.4	0	10
Bimonth 2	6	2.6	1.4	0	10
Bimonth 3	4	0.5	5.5	0	10
Bimonth 4	0	0.4	9.3	0.3	10
Bimonth 5	10	0	0	0	10
Bimonth 6	10	1	1	0	12
<b>Total</b>	33	9.1	19.6	0.3	62
<b>Expected</b>	30 - 70	10 - 30	80 - 140	0 - 4.8	

## 3. Research activity:

During my first year of PhD, I carried out different research activities within my research field. In particular I investigated natural language processing and anomaly detection in cybersecurity deepening my knowledge about these topics.

**Natural Language Processing (NLP):** I investigated the application of Artificial Intelligence on the natural language. In particular I studied in deep the Named Entity Recognition (NER) task in few-shot scenarios and how to handle the data scarcity in specific domains such as the biomedical one. In my research products three main augmentation techniques have been proposed:

- Similarity based:** a vocabulary of named entities (i.e., diseases, genes or chemicals for the biomedical domain) is built starting from the original training set. Then, I used the Mention Replacement technique with replacing the entities in the original training set with the most similar entities in the vocabulary to create a new augmented training set. All the new examples generated in this manner will respect the semantic and the grammatic needs of the original sentence, without distorting its meaning, therefore being plausible examples.
- Active Learning (AL):** in the same manner of the first method, an augmented training set is built, though it can contain not useful or noisy elements that don't help the model in labeling new data. By applying AL cycles, training a simple and light model at each cycle, we can discriminate, using an uncertainty function, the most informative examples to select (i.e., the ones that are predicted by the model with the highest uncertainty) in order to build a better augmented training set. This technique not only improves the performances of the baselines, but it works well also in generic domains.

# Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle: XXXVIII

Author: Andrea Vignali

---

- (iii) **Reinforcement Learning:** it moves its first steps from the second technique, by leveraging a policy-based active learning to learn a selection policy that identifies the most informative augmented samples to enhance the NER model's generalization ability. More in deep, an agent selects, in base of the state, the augmented examples and receive a reward (that can be positive or negative) depending on the performance difference between the last cycle.

I studied NLP also to extract and represent knowledge from texts like questionnaires, to build decision support systems and to guide small and medium enterprises through the assessment of their maturity of the digital technologies.

**Anomaly detection (AD):** anomaly detection is a wide field that can be applied to many domains (e.g., cybersecurity, finance, healthcare, manufacturing, etc.) and has different kind input data (e.g., image, video, audio, time series), therefore it can be approached in many ways (e.g., statistical, distance-based, clustering, graph-based, neural networks). During this first year I focused on multivariate time series in cyber-physical systems (CPS) using both network and physical data, also investigating how to combine them to obtain a better prediction of the anomalies through unsupervised methods like autoencoders, variational autoencoders and generative adversarial networks. The AD in time series is the key step to apply to NLP since natural language sentences can be interpreted as time series that can present anomalous behavior like threats, hate and offensive speech, frauds, and crimes.

## 4. Research products:

- **Learning how to augment data: an application to biomedical NER** – Vincenzo Moscato, Marco Postiglione, Guido Maria Secondulfo, Giancarlo Sperli, Andrea Vignali – conference: 6th International Workshop on Knowledge Discovery from Healthcare Data (KDH-2023@IJCAI) – Published – 2023
- **Data Augmentation via Context Similarity: an application to biomedical Named Entity Recognition** – Iliaria Bartolini, Vincenzo Moscato, Marco Postiglione, Giancarlo Sperli, Andrea Vignali – journal: Information Systems – Published – 2023
- **An NLP-Based Approach to Assessing a Company's Maturity Level in the Digital Era** – Simon Pietro Romano, Giancarlo Sperli, Andrea Vignali – journal: Expert Systems With Applications – Submitted – 2023
- **CPS Security Unleashed: Anomaly Detection for Cyber-Physical Threats in Critical Infrastructures** – Roberto Canonico, Giovanni Esposito, Annalisa Navarro, Simon Pietro Romano, Giancarlo Sperli, and Andrea Vignali – journal: IEEE Transaction on Dependable and Secure Computing – Submitted – 2023
- **Network and Physical Data Fusion for Cyber-Physical Systems Protection** – Roberto Canonico, Giovanni Esposito, Annalisa Navarro, Simon Pietro Romano, Giancarlo Sperli, and Andrea Vignali – journal: IEEE Transactions on Industrial Informatics – Submitted – 2023
- **Active Learning based Data Augmentation for Named Entity Recognition** - Vincenzo Moscato, Marco Postiglione, Giancarlo Sperli, and Andrea Vignali – journal: Transactions on Knowledge Discovery from Data – Submitted – 2023

# Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle: XXXVIII

Author: Andrea Vignali

---

## 5. Conferences and seminars attended

*21st Mediterranean Communication and Computer Networking Conference (MEDCOMNET2023), Ponza, 13-15/06/2023 – Network Systems Testing Meets AI: How to Cut off Development Costs (PhD advancements)*

## 6. Activity abroad:

||

## 7. Tutorship

*Machine Learning and Big data 31/05 (3 h)*

*Machine Learning for Engineering 18/05 (2 h), 1/06 (2 h)*