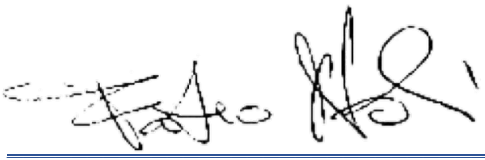
PhD in Information Technology and Electrical Engineering
Università degli Studi di Napoli Federico II

PhD Student: Fabio Noli

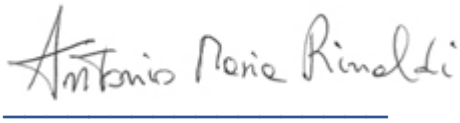
Cycle: XXXIX

Training and Research Activities Report

Year: First



Tutor: prof. Antonio Maria Rinaldi



Co-Tutor: prof. Luigi Verolino

Date: October 31, 2024

1. Information:

- **PhD student: Fabio Noli**
- **DR number: DR997220**
- **Date of birth: 21/05/1984**
- **Master Science degree: Software System Security University: University of Molise**
- **Doctoral Cycle: XXXIX**
- **Scholarship type: no scholarship**
- **Tutor: prof. Antonio Maria Rinaldi**
- **Co-tutor: prof. Luigi Verolino**

2. Study and training activities:

Activity	Type[1]	Hours	Credits	Dates	Organizer	Certificate[2]
La scorciatoia: come le macchine sono diventate intelligenti senza pensare in modo umano	Seminar	2	0.4	25/11/2023	PROF. MARIO DI BERNARDO	Y
Ensuring Electronic Reliability	Seminar	2	0.4	01/12/2023	Prof. Francesco Fienga	Y
Energy-Efficient Data Science	Seminar	2	0.4	13/12/2023	Prof. Elio Masciari	Y
Robotics meets AI & 5G: il Futuro è adesso!	Seminar	2	0.4	14/12/2023	prof.ssa Daniela Montesarchio	Y
Organization and design of surveys on applied artificial intelligence, big data technologies and digital forensics	Research		4	01/11/2023 - 31/12/2023		Y
Strategic Orientation For Stem Research & Writing.	Course	24	5	Dec 2023/ Jan 2024 / Feb 2024	Adriana D'Auria - DIETI-Unina	Y
How To Boost Your Phd.	Course	18	5	Jan 2024 / Feb 2024	Prof.ssa Antigone Marino	Y
HOMINIS	Seminar	5	1	21/02/2024	Prof. Carlo Sansone	Y
32nd Eacsl Annual Conference On Computer Science Logic 2024	Seminar	9	1	23/02/2024	PROF. ANIELLO MURANO	Y

Activity	Type[1]	Hours	Credits	Dates	Organizer	Certificate[2]
Organization and design of surveys on applied artificial intelligence, big data technologies and digital forensics	Research		2	01/01/2024 - 29/02/2024		Y
Using Deep Learning properly	Course	12	4	Jan 2024 / Feb 2024	Prof. Apicella Andrea	Y
Virtualization technologies and their applications	Course	20	5	Jan 2024 / Feb 2024	Prof. Luigi De Simone	Y
Analytic center selection of optimization-based controllers for robot ecology	Seminar	1	0.2	09/04/2024	prof. Bruno Siciliano	Y
Organization and design of surveys on applied artificial intelligence, big data technologies and digital forensics	Research		2	01/03/2024 - 31/05/2024		Y
I pilastri della trasformazione digitale	Course	12	3	04/2024	Dott. Francesco Tortorelli	Y
Big Data Architecture and Analytics	Course	21	5	06/2024	Prof. G. Sperlì	Y
Analytic center selection of optimization-based controllers for robot ecology	Seminar	1	0.2	09/05/2024	prof. Bruno Siciliano	Y
IEEE Authorship and Open Access Symposium: Tips and Best Practices to Get Published from IEEE Editors	Seminar	2	0.4	07/05/2024	CARE-CRUI and IEEE	Y
Rewire The Brain: The Potential Of Neuroplasticity	Seminar	2	0.4	09/05/2024	Prof.ssa Antonia Maria Taulino	Y
5G & DIGITAL TRASFORMATION	Seminar	4	0.8	10/05/2024	Prof.ssa Antonia Maria Taulino	Y

Activity	Type[1]	Hours	Credits	Dates	Organizer	Certificate[2]
Regolazione In Tema Di Ai Alla Luce Dell'AI Act	Seminar	5	1	13/05/2024	Prof.ssa Antonia Maria Taulino	Y
Intelligenza Artificiale E Regole Del Mercato	Seminar	2	0.4	14/05/2024	Prof.ssa Antonia Maria Taulino	Y
Verso una gestione smart della risorsa idrica con il supporto della digital innovation	Seminar	1	0.2	14/05/2024	Prof.ssa Antonia Maria Taulino	Y
Sustainable IT: Strategies and best practice for a green engineering future	Seminar	5	1	27/05/2024	Prof.ssa Antonia Maria Taulino	Y
GEN AI for SW engineering : strategies, impacts and practical applications	Seminar	5	1	29/05/2024	Prof.ssa Antonia Maria Taulino	Y
Social Network Analysis: Methods and Applications	Seminar	2	0.4	07/06/2024	Prof. G. Sperlì	Y
Introduction to Large Language Models: Evolution and the current state	Seminar	2	0.4	10/06/2024	Prof. G. Sperlì	Y
Innovation and Entrepreneurship	Course	12	4	07/2024	Prof. Pierluigi Rippa	Y
organization and design of surveys on applied artificial intelligence, big data technologies and digital forensics	Research		1	01/09/2024 - 31/10/2024		Y

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	-	1.6	4	-	5.6
Bimonth 2	10	2	2	-	14
Bimonth 3	9	0.2	2	-	11.2
Bimonth 4	8	6.2	-	-	14.2
Bimonth 5	4	-	1	-	5
Bimonth 6	-	-	1	-	1
Total	31	10	10	-	51
Expected	20 - 40	5 - 10	10 - 35	0 – 1.6	

3. Research activity:

My research activity this year, focused on "Artificial Intelligence and Big Data Technologies for Digital Forensics", led me to delve deeper into the topic of data lakes [5, 13] and their potential impact in the field of digital forensics.

It is undeniable that today's forensic investigations are faced with managing an unprecedented amount of data. This explosion of digital information requires a paradigm shift in the approach to forensic analysis. Data lakes [5, 13] emerge as a promising solution, offering the ability to efficiently manage and process huge amounts of heterogeneous data from different sources, such as mobile devices, computers, and the cloud.

I conducted a thorough analysis of the scientific literature, consulting academic databases such as IEEE Xplore, ACM Digital Library, ScienceDirect, and Google Scholar. The goal was to understand how data lakes can overcome the limitations of traditional data warehouses [6, 7], such as schema rigidity and dfficulty in managing unstructured data.

My research has shown that data lakes possess key features that make them particularly suitable for managing big data [1, 2]. Their flexibility and scalability allow for efficient data storage and processing, supporting a wide range of use cases, from real-time analysis [3] to machine learning [11].

A key aspect that differentiates data lakes from data warehouses is the "schema-on-read" approach [4]. Unlike data warehouses, which impose a predefined schema before data loading ("schema-on-write"), data lakes allow the schema to be applied only at query time. This feature translates into greater flexibility in integrating new data sources and adapting to changing analysis needs.

I also examined the different types of data lake architectures [9], from the most basic mono-zone to the more complex multi-zone ones, such as the multi-zone functional architecture [9, 12], which is based on technologies like Hadoop [8] and Spark. These advanced architectures allow data to be structured according to its processing level and purpose, improving data management, security, and quality.

Particular attention has been paid to the emergence of the data lakehouse model [10, 14], which combines the flexibility of data lakes with the reliability and integrity guarantees of data warehouses. This hybrid architecture represents an innovative solution for advanced data analysis and integration with artificial intelligence, opening new perspectives for research and application in various fields, including digital forensics.

In the next two years, i intend to further explore this topic. I will focus on applying data lakes to the domain of digital forensics, exploring how these technologies can support the acquisition, analysis, and management of large amounts of data from different digital sources. The goal will be to develop new AI-based techniques and tools for forensic data analysis within a data lake environment.

I am convinced that this line of research can significantly contribute to improving the effectiveness of digital investigations, reducing analysis time and increasing the ability to identify relevant information within massive amounts of data.

Ref.

[1] John, T.; Misra, P. Data Lake for Enterprises; Packt Publishing Ltd.: Birmingham, UK, 2017.

[2] LaPlante, A. Architecting Data Lakes; O'Reilly Media: Sebastopol, CA, USA, 2016.

[3] Liu, R.; Isah, H.; Zulkernine, F. A big data lake for multilevel streaming analytics. In Proceedings of the 2020 1st International Conference on Big Data Analytics and Practices (IBDAP), Bangkok, Thailand, 25–26 September 2020; pp. 1–6.

[4] Couto, J.; Borges, O.T.; Ruiz, D.D.; Marczak, S.; Prikladnicki, R. A Mapping Study about Data Lakes: An Improved Definition and Possible Architectures. In Proceedings of the SEKE, Lisbon, Portugal, 10–12 July 2019; pp. 453–578.

[5] Sawadogo, P.; Darmont, J. On data lake architectures and metadata management. J. Intell. Inf. Syst. 2021, 56, 97–120.

[6] Zagan, E.; Danubianu, M. From Data Warehouse to a New Trend in Data Architectures–Data Lake. IJCSNS Int. J. Comput. Sci. Netw. Secur. 2019, 19, 30–35.

[7] El Aissi, M.E.M.; Benjelloun, S.; Loukili, Y.; Lakhrissi, Y.; Boushaki, A.E.; Chougrad, H.; Elhaj Ben Ali, S. Data lake versus data warehouse architecture: A comparative study. In Proceedings of the WITS 2020: 6th International Conference on Wireless Technologies, Embedded, and Intelligent Systems; Springer: Berlin/Heidelberg, Germany, 2022; pp. 201–210.

[8] White, T. Hadoop: The Definitive Guide; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2012.

[9] Ravat, F.; Zhao, Y. Data lakes: Trends and perspectives. In Proceedings of the Database and Expert Systems Applications: 30th International Conference, DEXA 2019 (Part I 30), Linz, Austria, 26–29 August 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 304–313.

[10] Armbrust, M.; Ghodsi, A.; Xin, R.; Zaharia, M. Lakehouse: A new generation of open platforms that unify data warehousing and advanced analytics. In Proceedings of the CIDR, Online, 21 January 2021; Volume 8, p. 28.

[11] Giebler, C.; Gröger, C.; Hoos, E.; Eichler, R.; Schwarz, H.; Mitschang, B. The data lake architecture framework: A Foundation for Building a Comprehensive Data Lake Architecture. Ges. für Inform. Bonn 2021.

[12] Zhao, Y.; Megdiche, I.; Ravat, F.; Dang, V.n. A zone-based data lake architecture for IoT, small and big data. In Proceedings of the 25th International Database Engineering & Applications Symposium, New York, NY, USA, 14–16 July 2021; pp. 94–102.

[13] Giebler, C.; Gröger, C.; Hoos, E.; Schwarz, H.; Mitschang, B. Leveraging the data lake: Current state and challenges. In Proceedings of the Big Data Analytics and Knowledge Discovery: 21st International Conference, DaWaK 2019 (Proceedings 21), Linz, Austria, 26–29 August 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 179–188.

[14] Hlupić, T.; Oreščanin, D.; Ružak, D.; Baranović, M. An overview of current data lake architecture models. In Proceedings of the 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 23–27 May 2022; pp. 1082–1087.

4. Research products:

-

5. Conferences and seminars attended

-

6. Activity abroad:

-

7. Activity in partner companies:

-

8. Tutorship

-