

# Luca Ragazzi

“Things don't happen by closing your eyes. Help yourself and you will be helped.”



## Personal

17/01/1996   Forlì   luca.ragazzi9696@gmail.com

## Research & Academy

Scholar   Website   LinkedIn   GitHub  
 DISI-UniBO-NLP   ORCID   l.ragazzi@unibo.it

## Short Bio

Luca Ragazzi, a third-year PhD student, joined the Department of Computer Science and Engineering at the University of Bologna in 2020. Having earned bachelor's and master's (with honors) degrees from the same faculty, he specializes in the field of natural language processing, with a focus on text summarization and generation in low-resource regimes. Luca has presented numerous original papers at esteemed international conferences such as AAAI, ACL, and ECAI. He also contributed as a session chair for AAAI 2023. Currently, his research is centered on utilizing up-to-date cutting-edge large language models to advance artificial intelligence applications, particularly in high social-impact domains such as law and biomedicine.

## Education

- 2020 – now   **Ph.D.** in Computer Science and Engineering, University of Bologna, Cesena, Italy.  
Topics: *Natural Language Processing – Text Generation – Text Summarization – Question Answering – Low Resource Regimes – Large Language Models.*  
Thesis title: *Generative Text Summarization in Low-Resource Regimes.*
- 2018 – 2020   **M.Sc.** in Computer Science and Engineering, University of Bologna, Cesena, Italy.  
110L/110 - Magna cum Laude  
Thesis title: *Abstractive Summarization on Legal Case Reports: New State-of-the-art Solutions with Transformer-based Language Models.*
- 2015 – 2018   **B.Sc.** in Computer Science and Engineering, University of Bologna, Cesena, Italy.  
Thesis title: *Design and Development of an Offline Web Application with Angular Service Worker and IndexedDB.*
- 2010 – 2015   **Secondary High School Diploma** in Liceo Scientifico Statale Fulcieri Paulucci di Calboli, Forlì, Italy.

## Work Experience

- 2023   **Visiting Researcher** at EURECOM, Biot, France.  
Topic: *Automatic Data Generation for Computational Fact-Checking.*
- 2022   **Scientific and Technological Consultant** for L&G Solution, Foggia, Italy.  
Topic: *Retrieval-based Italian Chatbot.*
- 2018   **Web Designer & Developer.** Curricular Internship at Librasoft, Forlì, Italy.  
Topic: *Web application design and development with React and Angular frameworks.*

## Languages


- Italian   Mother language.
- English   B2 level.


## Skills


Coding	Python, Bash, Java, Scala, $\text{\LaTeX}$ , JSON
ML Libraries	PyTorch, HuggingFace, TensorFlow, NumPy, Pandas.
Web Dev	JavaScript, TypeScript, HTML, CSS, Angular, Vue.
Sw. & Tools	Docker, Git, Slurm.
Operating Systems	Mac OS X, Microsoft Windows, Linux.
Misc.	Leadership, Teamwork, Motivation, Problem Solving, Perseverance, Calmness.

## Research Publications

### Journal Articles

- 1 G. Moro, N. Piscaglia, L. Ragazzi, and P. Italiani, “Multi-language transfer learning for low-resource legal case summarization,” *Artificial Intelligence and Law*, pp. 1–29, 2023.  DOI: 10.1007/s10506-023-09373-8.

Analyzing and evaluating legal case reports are labor-intensive tasks for judges and lawyers, who usually base their decisions on report abstracts, legal principles, and commonsense reasoning. Thus, summarizing legal documents is time-consuming and requires excellent human expertise. Moreover, public legal corpora of specific languages are almost unavailable. This paper proposes a transfer learning approach with extractive and abstractive techniques to cope with the lack of labeled legal summarization datasets, namely a low-resource scenario. In particular, we conducted extensive multi- and cross-language experiments. The proposed work outperforms the state-of-the-art results of extractive summarization on the Australian Legal Case Reports dataset and sets a new baseline for abstractive summarization. Finally, syntactic and semantic metrics assessments have been carried out to evaluate the accuracy and the factual consistency of the machine-generated legal summaries.
- 2 G. Moro and L. Ragazzi, “Align-then-abstract representation learning for low-resource summarization,” *Neurocomputing*, vol. 548, p. 126 356, 2023.  DOI: 10.1016/J.NEUCOM.2023.126356.

Generative transformer-based models have achieved state-of-the-art performance in text summarization. Nevertheless, they still struggle in real-world scenarios with long documents when trained in low-resource settings of a few dozen labeled training instances, namely in low-resource summarization (LRS). This paper bridges the gap by addressing two key research challenges when summarizing long documents, i.e., long-input processing and document representation, in one coherent model trained for LRS. Specifically, our novel align-then-abstract representation learning model (ATHENA) jointly trains a segmenter and a summarizer by maximizing the alignment between the chunk-target pairs in output from the text segmentation. Extensive experiments reveal that ATHENA outperforms the current state-of-the-art approaches in LRS on multiple long document summarization datasets from different domains.
- 3 G. Moro, L. Ragazzi, L. Valgimigli, G. Frisoni, C. Sartori, and G. Marfia, “Efficient memory-enhanced transformer for long-document summarization in low-resource regimes,” *Sensors*, vol. 23, no. 7, p. 3542, 2023.  DOI: 10.3390/S23073542.

Long document summarization poses obstacles to current generative transformer-based models because of the broad context to process and understand. Indeed, detecting long-range dependencies is still challenging for today’s state-of-the-art solutions, usually requiring model expansion at the cost of an unsustainable demand for computing and memory capacities. This paper introduces EMMA, a novel efficient memory-enhanced transformer-based architecture. By segmenting a lengthy input into multiple text fragments, our model stores and compares the current chunk with previous ones, gaining the capability to read and comprehend the entire context over the whole document with a fixed amount of GPU memory. This method enables the model to deal with theoretically infinitely long documents, using less than 18 and 13 GB of memory for training and inference, respectively. We conducted extensive performance analyses and demonstrate that EMMA achieved competitive results on two datasets of

different domains while consuming significantly less GPU memory than competitors do, even in low-resource settings.

## Conference Proceedings

- 1 G. Moro, L. Ragazzi, and L. Valgimigli, “Carburacy: Summarization models tuning and comparison in eco-sustainable regimes with a novel carbon-aware accuracy,” in *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, B. Williams, Y. Chen, and J. Neville, Eds., AAAI Press, 2023, pp. 14 417–14 425. [DOI: 10.1609/AAAI.V37I12.26686](https://doi.org/10.1609/AAAI.V37I12.26686).

Generative transformer-based models have reached cutting-edge performance in long document summarization. Nevertheless, this task is witnessing a paradigm shift in developing ever-increasingly computationally-hungry solutions, focusing on effectiveness while ignoring the economic, environmental, and social costs of yielding such results. Accordingly, such extensive resources impact climate change and raise barriers to small and medium organizations distinguished by low-resource regimes of hardware and data. As a result, this unsustainable trend has lifted many concerns in the community, which directs the primary efforts on the proposal of tools to monitor models’ energy costs. Despite their importance, no evaluation measure considering models’ eco-sustainability exists yet. In this work, we propose Carburacy, the first carbon-aware accuracy measure that captures both model effectiveness and eco-sustainability. We perform a comprehensive benchmark for long document summarization, comparing multiple state-of-the-art quadratic and linear transformers on several datasets under eco-sustainable regimes. Finally, thanks to Carburacy, we found optimal combinations of hyperparameters that let models be competitive in effectiveness with significantly lower costs.

- 2 G. Moro, L. Ragazzi, and L. Valgimigli, “Graph-based abstractive summarization of extracted essential knowledge for low-resource scenarios,” in *ECAI 2023 - 26th European Conference on Artificial Intelligence, September 30 - October 4, 2023, Kraków, Poland - Including 12th Conference on Prestigious Applications of Intelligent Systems (PAIS 2023)*, K. Gal, A. Nowé, G. J. Nalepa, R. Fairstein, and R. Radulescu, Eds., ser. *Frontiers in Artificial Intelligence and Applications*, vol. 372, IOS Press, 2023, pp. 1747–1754. [DOI: 10.3233/FAIA230460](https://doi.org/10.3233/FAIA230460).

Although current summarization models can process increasingly long text sequences, they still struggle to capture salient related information spread across the lengthy size of inputs with few labeled training instances. Today’s research still relies on standard input truncation without considering graph-based modeling of multiple semantic units to summarize only crucial facets. This paper proposes G-SEEK, a graph-based summarization of extracted essential knowledge. By representing the long source with a heterogeneous graph, our method extracts and provides salient sentences to an abstractive summarization model to generate the summary. Experimental results in low-resource scenarios, distinguished by data scarcity, reveal that G-SEEK consistently improves both the long- and multi-document summarization performance and accuracy across several datasets.

- 3 G. Moro, L. Ragazzi, L. Valgimigli, and L. Molfetta, “Retrieve-and-rank end-to-end summarization of biomedical studies,” in *Similarity Search and Applications - 16th International Conference, SISAP 2023, A Coruña, Spain, October 9-11, 2023, Proceedings*, O. Pedreira and V. Estivill-Castro, Eds., ser. *Lecture Notes in Computer Science*, vol. 14289, Springer, 2023, pp. 64–78. [DOI: 10.1007/978-3-031-46994-7\\_6](https://doi.org/10.1007/978-3-031-46994-7_6).

An arduous biomedical task involves condensing evidence derived from multiple interrelated studies, given a context as input, to generate reviews or provide answers autonomously. We named this task context-aware multi-document summarization (CA-MDS). Existing state-of-the-art (SOTA) solutions require truncation of the input due to the high memory demands, resulting in the loss of meaningful content. To address this issue effectively, we propose a novel approach called RAMSES, which employs a retrieve-and-rank technique for end-to-end summarization. The model acquires the ability to (i) index each document by modeling its semantic features, (ii) retrieve the most relevant ones, and (iii) generate a summary via token probability marginalization. To facilitate the evaluation, we introduce a new dataset, FAQSUMC19, which includes the synthesizing of multiple supporting papers to answer questions related to Covid-19. Our experimental findings demonstrate that RAMSES achieves notably superior ROUGE

scores compared to state-of-the-art methodologies, including the establishment of a new SOTA for the generation of systematic literature reviews using MS2. Quality observation through human evaluation indicates that our model produces more informative responses than previous leading approaches.

- 4 G. Moro and L. Ragazzi, "Semantic self-segmentation for abstractive summarization of long documents in low-resource regimes," in *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, AAAI Press, 2022, pp. 11 085–11 093. [DOI: 10.1609/AAAI.V36I10.21357](https://doi.org/10.1609/AAAI.V36I10.21357).

The quadratic memory complexity of transformers prevents long document summarization in low computational resource scenarios. State-of-the-art models need to apply input truncation, thus discarding and ignoring potential summary-relevant contents, leading to a performance drop.

Furthermore, this loss is generally destructive for semantic text analytics in high-impact domains such as the legal one. In this paper, we propose a novel semantic self-segmentation (Se<sub>3</sub>) approach for long document summarization to address the critical problems of low-resource regimes, namely to process inputs longer than the GPU memory capacity and produce accurate summaries despite the availability of only a few dozens of training instances. Se<sub>3</sub> segments a long input into semantically coherent chunks, allowing transformers to summarize very long documents without truncation by summarizing each chunk and concatenating the results. Experimental outcomes show the approach significantly improves the performance of abstractive summarization transformers, even with just a dozen of labeled data, achieving new state-of-the-art results on two legal datasets of different domains and contents. Finally, we report ablation studies to evaluate each contribution of the components of our method to the performance gain.

- 5 G. Moro, L. Ragazzi, L. Valgimigli, and D. Freddi, "Discriminative marginalized probabilistic neural method for multi-document summarization of medical literature," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, S. Muresan, P. Nakov, and A. Villavicencio, Eds., Association for Computational Linguistics, 2022, pp. 180–189. [DOI: 10.18653/V1/2022.ACL-LONG.15](https://doi.org/10.18653/V1/2022.ACL-LONG.15).

Although current state-of-the-art Transformer-based solutions succeeded in a wide range for single-document NLP tasks, they still struggle to address multi-input tasks such as multi-document summarization. Many solutions truncate the inputs, thus ignoring potential summary-relevant contents, which is unacceptable in the medical domain where each information can be vital. Others leverage linear model approximations to apply multi-input concatenation, worsening the results because all information is considered, even if it is conflicting or noisy with respect to a shared background. Despite the importance and social impact of medicine, there are no ad-hoc solutions for multi-document summarization. For this reason, we propose a novel discriminative marginalized probabilistic method (DAMEN) trained to discriminate critical information from a cluster of topic-related medical documents and generate a multi-document summary via token probability marginalization. Results prove we outperform the previous state-of-the-art on a biomedical dataset for multi-document summarization of systematic literature reviews. Moreover, we perform extensive ablation studies to motivate the design choices and prove the importance of each module of our method.

## Under-Review: Journal Articles

- 1 J. F. Bussotti, L. Ragazzi, G. Frisoni, G. Moro, and P. Papotti, *Unknown claims: Generation of fact-checking training examples from unstructured and structured data*, Submitted to *Trans. Assoc. Comput. Linguistics*, 2024.

Computational fact-checking relies on supervised models to verify whether a claim is supported or refuted based on evidence provided. A substantial volume of meticulously annotated data is required for training, implying a resource-intensive and time-consuming annotation endeavor. We introduce UNOWN, a novel framework designed to automatically generate training instances for fact-checking systems from heterogeneous knowledge sources. Specifically, UNOWN ingests tabular and/or textual content, constructs evidence documents, and builds artificial claims that support or refute evidence. UNOWN is conceived to be efficient and module-centric, accommodating various evidence selection and

claim generation strategies. Quantitative and qualitative experiments on FEVEROUS and SCIFACT, two benchmarks from different domains, reveal that our synthetic examples exhibit a quality similar to that of expert-labeled data.

2 L. Ragazzi, G. Frisoni, G. Moro, P. Italiani, L. Molfetta, and V. Folin, *COMMA: A multi-task and multi-lingual dataset of constitutional verdicts*, Submitted to *Comput. Linguistics*, 2024.  
Large-scale public datasets are vital for driving the progress of abstractive summarization, especially in law, where documents have highly specialized jargon. However, the available resources are English-centered, limiting research advancements in other languages. This paper introduces LAWSU-IT, a collection of 14K Italian legal rulings with expert-authored abstractive maxims drawn from the Constitutional Court of the Italian Republic. LAWSU-IT presents an arduous task with lengthy source texts and evenly distributed salient content. We offer extensive experiments and propose SEGSUMM, a simple yet effective language-agnostic segmentation-based approach that outperforms previous work in full and low-resource settings. We openly release LAWSU-IT to foster the development and automation of real-world legal applications.

3 L. Ragazzi, G. Moro, S. Guidi, and G. Frisoni, *LAWSU-IT: Abstractive summarization of constitutional legal rulings*, Submitted to *Artificial Intelligence and Law*, 2024.  
Large-scale public datasets are vital for driving the progress of abstractive summarization, especially in law, where documents have highly specialized jargon. However, the available resources are English-centered, limiting research advancements in other languages. This paper introduces LAWSU-IT, a collection of 14K Italian legal rulings with expert-authored abstractive maxims drawn from the Constitutional Court of the Italian Republic. LAWSU-IT presents an arduous task with lengthy source texts and evenly distributed salient content. We offer extensive experiments and propose SEGSUMM, a simple yet effective language-agnostic segmentation-based approach that outperforms previous work in full and low-resource settings. We openly release LAWSU-IT to foster the development and automation of real-world legal applications.

## Under-Review: Conference Proceedings

1 G. Frisoni, L. Ragazzi, D. Cohen, G. Moro, A. Carbonaro, and C. Sartori, *Abstractive summarization through the PRISM of decoding strategies*, Submitted to *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*.

In the realm of natural language generation, abstractive summarization (AS) is at the center of an unparalleled evolution driven by transformer-based language models (LMs). However, the significance of decoding strategies is often neglected, despite their influence on the generated summaries. Given the abundance of token selection heuristics and their accompanying hyperparameters, the community needs directions to steer well-founded decisions based on the task and the target metrics at hand. To fill this gap, we comparatively assess the effectiveness and efficiency of decoding-time techniques for short, long, and multi-document AS. We explore more than 2500 combinations of 3 widely used million-scale autoregressive encoder-decoder models, 6 datasets, and 9 decoding settings. Our findings shed light on the field, demonstrating that optimized decoding choices can yield substantial performance enhancements. In addition to human evaluation, we quantitatively measure effects using 10 automatic metrics, including dimensions such as semantic similarity, factuality, compression, redundancy, and carbon footprint. We introduce PRISM, a pioneering dataset that pairs AS gold input-output examples with LM predictions under a wide array of decoding options.

2 P. Italiani, L. Ragazzi, and G. Moro, *ACE-ATTORNEY: Large language model distillation for legal question answering*, Submitted to *The 2024 Conference of the North American Chapter of the Association for Computational Linguistics, NAACL 2024, Mexico City, Mexico, June 16-21, 2024*.

Legal question answering (LQA) relies on supervised methods to automatically handle law-related queries. These solutions require a significant amount of carefully annotated data for training, which makes the process very costly. Although large language models (LLMs) show promise in zero-shot QA, their computational demands limit their practical use, making specialized small language models (SLMs) more favorable. Furthermore, the growing interest in synthetic data generation has recently

surged, spurred by the impressive generation capabilities of LLMs. This paper presents *ACE-ATTORNEY*, an LLM distillation framework devised to develop LQA data and supervised models without human annotation. Given a textual prompt, a frozen instruction-tuned LLM generates artificial examples that are used as knowledge to train a student SLM with an order of magnitude fewer parameters. Extensive experiments prove the effectiveness and efficiency of distilled models on *SYN-LEQA* and *PRIVACYQA*, our synthetic dataset and a public human-annotated corpus, respectively. Notably, our best SLM achieves LLM-comparable performance on *PRIVACYQA* using just 30 training samples and  $\approx 1200\%$  less CO<sub>2</sub> emissions.

- 3 G. Moro, L. Ragazzi, L. Valgimigli, and F. Vincenzi, *REVELIO: Interpretable long-form question answering*, Submitted to The Second Tiny Papers Track at ICLR 2024, Tiny Papers @ ICLR 2024, Vienna, Austria, May 7, 2024.

The black-box architecture of pretrained language models (PLMs) hinders the interpretability of lengthy responses in long-form question answering (LFQA). Prior studies use knowledge graphs (KGs) to enhance output transparency, but mostly focus on non-generative or short-form QA. We present *REVELIO*, a new layer that maps PLM's inner working onto a KG walk. Tests on two LFQA datasets show that *REVELIO* supports PLM-generated answers with reasoning paths presented as rationales while retaining performance and time akin to their vanilla counterparts.

## ☰ Scientific Activities



### Participation in Research Groups

2020 - now  **DISI-UniBo-NLP**.



### Certifications

2022  **NLP Specialization**. Awarded by DeepLearning.AI.  Certificate.






### Speaker (Paper Presentation)

- 2023  **Association for the Advancement of Artificial Intelligence, AAAI 2023, Washington DC, USA, In-person**; *Carburacy: Summarization Models Tuning and Comparison in Eco-Sustainable Regimes with a Novel Carbon-Aware Accuracy*.
- 2022  **Association for the Advancement of Artificial Intelligence, AAAI 2022, Virtual**; *Semantic Self-Segmentation for Abstractive Summarization of Long Documents in Low-Resource Regimes*.

### Reviewing for International Journals

- 2023  **IEEE Transactions on Systems, Man and Cybernetics: Systems**; SJR Class: Q1.
- 2020  **Artificial Intelligence and Law**; SJR Class: Q1.


### Reviewing for Conferences

- 2024  **European Semantic Web Conference, ESWC 2024**; GGS Class: 2, GGS Rating: A-.
- 2023  **Association for the Advancement of Artificial Intelligence, AAAI 2024**; GGS Class: 1, GGS Rating: A++.
-  **ACM Symposium on Applied Computing, ACM SAC 2024**; GGS Class: 2, GGS Rating: A-.
-  **Empirical Methods on Natural Language Processing, EMNLP 2023**; GGS Class: 1, GGS Rating: A+.
-  **Neural Information Processing Systems, NeurIPS 2023**; GGS Class: 1, GGS Rating: A++.

## ☰ Scientific Activities (continued)

- 2022  **Association for the Advancement of Artificial Intelligence, AAI 2023**; GGS Class: 1, GGS Rating: A++.

### Session Chair






- 2023  **Association for the Advancement of Artificial Intelligence, AAI 2023**, Washington DC, USA; GGS Class: 1, GGS Rating: A++.

## Teaching





### Seminars

- 2023  *Automatic Text Summarization: from Theory to Practice*. “Data Mining, Text Mining and Big Data Analytics” M.Sc. course, Artificial Intelligence, University of Bologna, December 15, 2023.
-  *Long Document Summarization in Low-Resource Regimes*. “Data Mining, Text Mining and Big Data Analytics” M.Sc. course, Artificial Intelligence, University of Bologna, October 18, 2023.
-  *Retrieval-based Italian Chatbot*. “Data Intensive Applications” B.Sc. course, Computer Science and Engineering, University of Bologna, June 8, 2023.
-  *Long and Multi-Document Abstractive Summarization in Low-Resource Regimes*. “EURECOM”, March 21, 2023, Biot, France.
- 2022  *Long and Multi-Document Abstractive Summarization in Low-Resource Regimes*. “Data Mining” M.Sc. course, Computer Science and Engineering, University of Bologna, December 1, 2022.
- 2021  *Long Document Summarization in Low-resource Regimes with Applications in the Legal Domain*. “Data Mining” M.Sc. course, Computer Science and Engineering, University of Bologna, December 16, 2021.
- 2020  *Natural Language Processing for Automatic Text Summarization: an Overview*. “Data Mining” M.Sc. course, Computer Science and Engineering, University of Bologna, December 12, 2020.

### Thesis with Co-Supervisor Role







- 2023  *Explaining Generative Model for Long-form Question Answering with Reasoning Graph*. Candidate: Fabian Vincenzi, Supervisor: Gianluca Moro, Co-supervisor: Luca Ragazzi, Lorenzo Valgimigli. October, 2023. M.Sc. in Artificial Intelligence, University of Bologna.
-  *Summarization Astrattiva di Lunghi Articoli Scientifici mediante Estrazione di Frammenti Rilevanti*. Candidate: Filippo Di Pietro, Supervisor: Gianluca Moro, Co-supervisor: Luca Ragazzi, Paolo Italiani, 2023. B.Sc. in Computer Science and Engineering, University of Bologna.
-  *Survey on Few-Shot Summarization*. Candidate: Emanuele Artegiani, Supervisor: Gianluca Moro, Co-supervisor: Luca Ragazzi, Giacomo Frisoni. November, 2023. B.Sc. in Computer Science and Engineering, University of Bologna.
-  *Graph Neural Network Benchmark per la Selezione di Contenuto Rilevante nella Low-Resource Summarization*. Candidate: Riccardo Fiorani, Supervisor: Gianluca Moro, Co-supervisor: Luca Ragazzi, Lorenzo Valgimigli. November, 2023. B.Sc. in Computer Science and Engineering, University of Bologna.
-  *SCI-LAY: Un Nuovo Dataset per Long Document Summarization Scientifica e Divulgativa di Studi Biomedici*. Candidate: Mattia Panni, Supervisor: Gianluca Moro, Co-supervisor: Luca Ragazzi, Paolo Italiani, Giacomo Frisoni. November, 2023. B.Sc. in Computer Science and Engineering, University of Bologna.

## Teaching (continued)





- 2022  *Generazione di Riassunti Fattuali mediante Parsing Semantico*. Candidate: Luca Grandi, Supervisor: Gianluca Moro, Co-supervisor: *Luca Ragazzi*, Giacomo Frisoni. March, 2023. B.Sc. in Computer Science and Engineering, University of Bologna.
- 2022  *Sviluppo di Metodi di Soft Labeling per la Multi-Document Summarization in Ambito Legale*. Candidate: Luca Rubboli, Supervisor: Gianluca Moro, Co-supervisor: *Luca Ragazzi*. December, 2022. B.Sc. in Computer Science and Engineering, University of Bologna.
-  *Sviluppo di Retrieval-based Chatbot per l'Italiano con Transformer*. Candidate: Luca Morlino, Supervisor: Gianluca Moro, Co-supervisor: *Luca Ragazzi*. December, 2022. B.Sc. in Computer Science and Engineering, University of Bologna.
-  *LAWSU-IT: Un Nuovo Dataset Giudiziario Italiano per Long Document Summarization con Baseline Estrattive e Astrattive*. Candidate: Stefano Guidi, Supervisor: Gianluca Moro, Co-supervisor: *Luca Ragazzi*. October, 2022. B.Sc. in Computer Science and Engineering, University of Bologna.
- 2021  *Sintesi Generativa Multi-documento con Discriminazione della Rilevanza Mediante Probabilità Marginale: Una soluzione Neurale End-to-End per la Letteratura Medica*. Candidate: Davide Freddi, Supervisor: Gianluca Moro, Co-supervisor: *Luca Ragazzi*. December, 2021. B.Sc. in Computer Science and Engineering, University of Bologna.
-  *Abstractive Long Document Summarization: Studio e Sperimentazione di Modelli Generativi Retrieval-Augmented*. Candidate: Veronika Folin, Supervisor: Gianluca Moro, Co-supervisor: *Luca Ragazzi*. October, 2021. B.Sc. in Computer Science and Engineering, University of Bologna.

## Projects





### Machine Learning

- 2020  Design and implementation of a deep learning project on time series forecasting. 
-  Design and implementation of a text mining project on article knowledge discovery. 
- 2018  Design and implementation of a machine learning project on gold market trend. 















### Data Science

- 2020  Design and implementation of a big data project on accident severity analysis. 
-  Design and implementation of a semantic web project on the expansion and modeling of an ontology for road management by an Ego vehicle. 

### Web Development

- 2019  Design and implementation of the AlmaNotes web application. 
- 2018  Design and implementation of the Jestergest web app for the bachelor thesis. 

### Software Engineering





- 2020  Design of a project management project on the development of a basketball application. 
- 2019  Design and implementation of concurrent and distributed programming projects. 
-  Design and implementation of a new compiler. 
- 2018  Design and implementation of a high-performance computing project. 
-  Design and implementation of embedded systems and internet of things projects. 
- 2017  Design and implementation of the GeoQuiz geography quiz game. 
-  Design and implementation of a database project. 




## Extracurricular

---

### Sports Activities

- 2024 – now  **Bocce**, D series.
- 2018 – now  **Scuba Diving**. Advanced level (3 patents).
- 2004 – now  **Basketball**. Shooting guard, regional division 2.  Website

### Sports Achievements

- 2023  **National Academic Bocce Championship First Place**. First position out of 42 participants for the bocce's national academic championship, September 2023, Padua, Italy.