Antonio Lanza

Date of birth: 13/12/1996

CONTACT

antoniolanza1996@gmail.com

(+39) 3405429172

www.linkedin.com/in/lanzaantonio

https://github.com/antoniolanza1996 **(GitHub)**

WORK EXPERIENCE

01/2021 – CURRENT Rende, Italy

AI Research Scientist & ML Engineer ALTILIA s.r.l.

- Working on the R&D team with the aim of building Altilia Intelligent Automation™ - a no/low-code, IPAaaS (Intelligent Process Automation as a Service) platform - that democratizes the adoption of AI, hyper automation, and decision intelligence at scale in modern organizations of any size.
- Conducting cutting-edge research in AI, including exploring new algorithms, techniques, and approaches for Intelligent Document Processing (IDP).
- Designing, implementing, and managing production-ready machine learning pipelines to deliver high-performance and reliable AI solutions that have a direct impact on business process efficiency.

06/2020 – 12/2020 Rende, Italy

AI Research Scientist Intern ALTILIA s.r.l.

Internship period in which Natural Language Processing (NLP) Transformer-based models were studied and implemented. In particular, Information Retrieval (IR) and Question Answering (QA) models were investigated in order to implement an Open Domain Question Answering (ODQA) system.

EDUCATION AND TRAINING

09/2021 – 09/2021

Passed the professional exam and licensed as a Information Engineer University of Calabria

01/2020 – 05/2020 Dublin, Ireland

Erasmus+ period University College Dublin (UCD)

Semester GPA: 4.04/4.2

09/2018 – 12/2020 Rende, Italy

Master's Degree in Computer Engineering University of Calabria

Field of study: Big Data Analytics e Data Science
Thesis title: Deep Learning techniques for Open Domain Question Answering
Final grade: 110/110 with honors

09/2015 – 09/2018 Rende, Italy

Bachelor's Degree in Computer Engineering University of Calabria

Thesis title: Use of Hidden Markov Models for the interpretation of low-level business process logs
Final grade: 110/110 with honors

09/2010 – 07/2015 Castrovillari, Italy

Science High School Diploma Liceo Scientifico E. Mattei

PROFESSIONAL SKILLS

Professional skills

- **Generative AI and LLMs**:
  - Extensive experience with generative AI across multiple applications:

        - **Retrieval-Augmented Generation (RAG) and AI Agents**: Designed and implemented RAG pipelines along with AI Agents, enabling effective information retrieval and automation of complex tasks.
        - **Prompt Engineering**: Skilled in designing prompts for optimal performance in zero/few-shot settings.
        - **Fine-Tuning of open-source LLMs**: Applied PEFT techniques (e.g. LoRA, QLoRA) for efficient LLM adaptation to specialized tasks.
    - *Libraries and frameworks*: PyTorch, Hugging Face's Transformers, Datasets, PEFT and TRL, LangChain, vLLM, OpenAI APIs.
- **Vector Databases and Information Retrieval**:
    - Extensive experience with both full-text (i.e. syntactic/sparse) and neural (i.e. semantic/dense) search techniques and vector databases to support high-performance AI-driven information retrieval and real-time semantic search.
    - *Libraries and frameworks*: UKPLab's Sentence Transformers, deepset's FARM and Haystack, Elasticsearch, Milvus, FAISS.
- **Other Machine and Deep Learning tasks**:
    - Extensive experience in developing, fine-tuning, and deploying a wide range of ML/DL models across various domains:
        - **Natural Language Processing**: Named Entity Recognition, Question Answering, Multi-Class and Multi-Label Text Classification, Sentiment Analysis, Aspect-based Sentiment Analysis.
        - **Computer Vision**: Document Layout Analysis, Image Classification, Object Detection/ Segmentation, Optical Character Recognition.
    - *Libraries and frameworks*: PyTorch, Hugging Face's Transformers and Datasets, Spacy, Scikit-learn, Weights & Biases, ClearML, Detectron2, OpenMMLab's MMCV and MMOCR.
- **DevOps and MLOps**
    - Familiar with Dev/MLOps practices for automating and managing the deployment of ML/DL models, with a focus on maintaining scalable and resilient infrastructure.
    - Good knowledge of using Git, Shell scripting and Docker.
    - Basic knowledge of Kubernetes and Ray framework.
- **Programming languages**: Python and Java.

LANGUAGE SKILLS

MOTHER TONGUE(S): Italian

Other language(s):

English

Listening B2
Spoken production B2
Reading B2
Spoken interaction B2
Writing B2

Levels: A1 and A2: Basic user; B1 and B2: Independent user; C1 and C2: Proficient user

HONOURS AND AWARDS

06/2020 DIMES department, University of Calabria

Path of Excellence

Winner of a scholarship issued by the DIMES department.

During this period, the Apache Storm framework was studied by creating an application that analyzes tweets published on Twitter in real time and provides, through appropriate graphs, statistics on hashtags and on the origin of users. Furthermore, it is also possible to perform sentiment analysis and spam detection tasks.

24/05/2016 PRAXIS MMT Corp, on behalf of Business Talents, Milan (MI)

BT Simulated Enterprise Program Finalist - 2015 Edition

09/04/2015 PRAXIS MMT Corp, on behalf of Young Business Talents, Milan (MI)

Winner of the YBT Simulated Enterprise program - 2014 Edition

21/03/2014 PRAXIS MMT Corp, on behalf of Young Business Talents, Milan (MI)

Finalist of the YBT Simulated Enterprise program - 2013 Edition

CONFERENCES AND SEMINARS

29/05/2024 – 30/05/2024 Ital-IA 2024 - Naples, Italy

Evaluating Retrieval-Augmented Generation for Question Answering with Large Language Models

Abstract: We present a comprehensive framework for evaluating retrieval-augmented generation (RAG) systems designed for questionanswering tasks using large language models (LLMs). The proposed framework integrates document ingestion, information retrieval, answer generation, and evaluation phases. Both ground truth-based and reference-free evaluation metrics are implemented to provide a multi-faceted assessment approach. Through experiments across diverse datasets like NarrativeQA and a proprietary financial dataset (FinAM-it), the reliability of existing metrics is investigated by comparing them against rigorous human evaluations. The results demonstrate that ground truth-based metrics such as BEM and RAGAS Answer Correctness exhibit a moderately strong correlation with human judgments. However, reference-free metrics still struggle to capture nuances in answer quality without predefined correct responses accurately. An in-depth analysis of Spearman correlation coefficients sheds light on the interrelationships and relative effectiveness of various evaluation approaches across multiple domains. While highlighting the current limitations of reference-free methodologies, the study underscores the need for more sophisticated techniques to better approximate human perception of answer relevance and correctness. Overall, this research contributes to ongoing efforts in developing reliable evaluation frameworks for RAG systems, paving the way for advancements in natural language processing and the realization of highly accurate and human-like AI systems.

Link https://ceur-ws.org/Vol-3762/495.pdf

29/05/2023 – 31/05/2023 Ital-IA 2023 - Pisa, Italy

Building a Platform for Intelligent Document Processing: Opportunities and Challenges

Abstract: Companies of any size and industry still struggle in automatic business processes where human cognitive and contextualization capabilities are required to read and understand complex documents. Ongoing progress in the fields of Computer Vision and Natural Language Processing, where (large) language models are becoming increasingly and freely available, have made possible to create a new generation of Intelligent Document Processing technologies that allow automatically analyzing and understanding both documents layout and contents. In this paper we present an Intelligent Document Processing platform that makes use of hybrid AI techniques to allow document reading comprehension by means of a combination of Document Layout Analysis and recognition, table recognition and detection, context free grammars, and question answering techniques. Such a technology combines also no-code principles with high performance computing based on micro-services to streamline the execution of tasks such as document and text classification, document segmentation, entity extraction, sentiment analysis, question answering, and more.

Link https://ceur-ws.org/Vol-3486/91.pdf

24/04/2023 – 26/04/2023 ICEIS 2023 - Prague, Czech Republic

ESG Data Collection with Adaptive AI

Abstract: The European Commission defines the sustainable finance as the process of taking Environmental, Social and Governance (ESG) considerations into account when making investment decisions, leading to more long-term investments in sustainable economic activities and projects. Banks, and other financial institutions, are increasingly incorporating data about ESG performances, with particular reference to risks posed by climate change, into their credit and investment portfolios evaluation methods. However, collecting the data related to ESG performances of corporate and businesses is still a difficult task. There exist no single source from which we can extract all the data. Furthermore, most important ESG data is in unstructured format, hence collecting it poses many technological and methodological challenges. In this paper we propose a method that addresses

the ESG data collection problem based on AI-based approaches. We also present the implementation of the proposed method and discus s some experiments carried out on real world documents.

Link https://www.scitepress.org/PublicationsDetail.aspx?ID=sCYU36yksCI=&t=1

08/11/2021 – 10/11/2021 Online Conference

Codemotion Online Tech Conference 2021 - Italian Edition | Autumn