Alberto Purpura

E-mail: alb.purpura@gmail.com | LinkedIn: albpurpura

EDUCATION

UNIVERSITY OF PADOVA
PH.D. IN INFORMATION ENGINEERING
Oct. 2018 - Oct. 2021 | Padova, IT
Thesis: Deep Neural Models for Documents Retrieval and Ranking.
Topics: Studied deep learning models in the context of Neural Information Retrieval and Learning to Rank.

MENG IN COMPUTER ENGINEERING
Oct. 2015 - Feb. 2018 | Padova, IT
Final Grade: 110/110 cum laude,
Thesis (officially deemed as *distinguished* by the thesis committee): Weakly and Semi-Supervised Approaches for Aspect-Based Sentiment Analysis on Online Reviews

BENG IN INFORMATION ENGINEERING
Oct. 2012 - Sep. 2015 | Padova, IT
Thesis: Design and Implementation of a Coding Interface for Textual Analysis Software

SKILLS

PROGRAMMING
Production-Level:
Python • Swift (iOS) • React Native
Academic Knowledge:
Go • Java • Matlab • VBA
Basic Knowledge:
C • C# • R

FRAMEWORKS
Production-Level:
Tensorflow • Pytorch • Docker • Kubernetes • PySpark • Python NLTK • Spacy • ElasticSearch • SQLite • Postgres • Huggingface transformers • IBM Cloud Codengine
Academic Knowledge:
Stanford CoreNLP • Apache Lucene • Terrier

OTHER

REVIEWER/PC MEMBER
Information Processing and Management (Elsevier), Quality and Quantity (Springer), Applied Soft Computing (Elsevier), WWW 2021, ECIR 2020, ECAI 2020, CIKM 2019, EMNLP 2021, EMNLP 2022, EMNLP 2023, ACL 2023, CIKM 2023.

MENTORSHIP
SwEng Industry Project 2023 mentor for IBM at Trinity College Dublin

EXPERIENCE

IBM RESEARCH EUROPE - DUBLIN RESEARCH LAB | RESEARCH SCIENTIST
Apr. 2022 - Present | Dublin, Ireland
Member of the NLP Reasoning in Healthcare team. I developed novel solutions for information extraction, management and retrieval from scientific literature e.g., tools for entity disambiguation, relation extraction, fine-tuned LLMs for different tasks and created pipelines for information management and retrieval using Postgres and ElasticSearch. I also built UIs to allow a small pool of users to access the services. I presented many of these developed solutions at different academic conferences.

DATA SCIENTIST/FULL STACK DEVELOPER | FREELANCE
Oct. 2018 – Apr. 2022 | Remote
Developed different solutions related to text and data analytics for business customers. I communicated directly with the customers translating their needs into technical solutions. These projects involved: the creation of backend APIs to search a company's internal data; the development of text classification APIs, websites creation, and other data analysis tasks.

HUAWEI IRELAND RESEARCH CENTER | PH.D. INTERN
Oct. 2021 - Apr. 2022 | Dublin, Ireland
Developed new solutions for search on relational databases in the context of AIOps and models for the conversion of natural language questions to SQL.

APPLE INC. | SIRI SEARCH INTERN
Mar. 2020 – Sep. 2020 | Cupertino, USA
Studied and developed deep learning models for learning to rank (LETOR) and other statistical approaches to trigger their retraining. I also proposed an approach for feature selection tailored for neural LETOR models and other strategies to improve their efficiency. **This work** was later presented at ECIR 2021.

SIGNAL AI | VISITING RESEARCHER
Sep. 2019 – Dec. 2019 | London, UK
Developed a Aspect-Based Sentiment Analysis model for the company's reputation monitoring system. Planned all the steps from the task definition to data collection, annotation and model development. The final model decreased by 50% the number of manual corrections of sentiment scores made by users of the platform.

SELECTED PUBLICATIONS

Complete list on **Google Scholar**.

- **A. Purpura**, J. Bettencourt-Silva, N. Mulligan, et al., 2023. Automatic Mapping of Terminology Items with Transformers, AMIA 2023.
- **A. Purpura**, J. Bettencourt-Silva, F. Bonin, et al., 2023. Automated Feature Selection from Medical Literature, MEDINFO 2023.
- **A. Purpura**, J. Bettencourt-Silva, F. Bonin, 2022. Accelerating the discovery of semantic associations from medical literature: Mining relations between diseases and symptoms. EMNLP 2022.
- **A. Purpura**, G. Silvello, G.A. Susto, 2022. Learning to Rank from Relevance Judgments Distributions. JASIST, Wiley.
- **A. Purpura**, G.Sartori, D. Giorgianni et al. Identifying Faked Responses in Questionnaires with Self-Attention-Based Autoencoders, 2022, Informatics.
- **A. Purpura**, K. Buchner, G. Silvello, G.A. Susto, 2021. Neural Feature Selection for Learning to Rank. Proc. of ECIR 2021.
- A. Fabris, **A. Purpura**, G. Silvello, G.A. Susto. 2020. Gender Stereotype Reinforcement: Measuring the Gender Bias Conveyed by Ranking Algorithms. IP & M, Elsevier.

OPEN SOURCE CONTRIBUTIONS

- Contributed to the IBM zshot library for zero-shot entity linking and relation classification
- Contributed to the The Open-Source IR Replicability Challenge (OSSIRC) library to reproduce experimental results from neural information retrieval models

My personal page; GitHub; BitBucket; Apple Developer Page.