



Tommaso Soru

Data Scientist • AI Researcher

London, United Kingdom

tom@tommaso-soru.it

+44 7894 456 626

<http://tommaso-soru.it>

I'm an AI enthusiast with the dream of giving my small contribution to the quest for intelligence, bridging the gap between language models and symbolic systems. My research topics focus on Large Language Models and Knowledge Graphs.

EXPERIENCE

Serendipity AI Ltd., London, UK

ARTIFICIAL INTELLIGENCE LEAD > SEPTEMBER 2021 – PRESENT

- Research & development of an event forecasting system. **Python, OpenAI GPT-4, Llama**
- Signal & Trend Extraction for AI-driven strategic foresight via Natural Language Understanding and Large Language Models. **Python, Torch, Transformers, Neo4j, Docker, AWS, OpenAI GPT-4, Llama**
- Design & implementation of time-series forecasting models. **Python, TensorFlow, Keras**

Scalpel Ltd., London, UK

SPEECH & NLP ENGINEER > AUGUST 2020 – SEPTEMBER 2021

- Deep learning model for distant Automatic Speech Recognition in a surgery room. **Python, Torch, fairseq**
- Design & implementation of real-time Medical Information Extraction from text. **Python, AWS, Docker**

Data Lens Ltd. and Semantic Integration Ltd., London, UK

DATA SCIENTIST & ARCHITECT > NOVEMBER 2018 – AUGUST 2020

- NLP Pipeline for scalable Information Extraction from unstructured data. **Python, Java, AWS, Docker**
- Development of a Named Entity Recognition algorithm for the Fintech domain with 90% accuracy.
- Design & implementation of a semantic reasoner for an international retail company. **Node.js, SPARQL**
- Design of a data architecture for knowledge discovery for a global firm in the FTSE 250 Index. **RDF(S)**

AKSW, University of Leipzig, Leipzig, Germany

DOCTORAL RESEARCHER > JULY 2012 – OCTOBER 2018

- PhD thesis titled "Metric Learning for Link Discovery" under the supervision of [Prof. Dr. Axel Ngonga](#).
- R&D of two scalable methods for Link Prediction in directed graphs. **Python, Java, SPARQL, RDF(S)**
- Design and development of [ROCKER](#), algorithm for RDF Key Discovery, and its web-based GUI. **Java, MySQL, RDF(S), AWS, JS, Docker**
- Machine Learning techniques (KG embeddings, SVMs, MLP) for Entity Resolution. **Java, Keras, Weka**
- Contributions in research projects [LinkLion](#), [LIMES 1.0](#), [ML-Schema](#). **Java, JSP, OWL, SPARQL, RDF(S)**
- 30+ publications in Web conferences (WWW, ISWC, ESWC) and AI workshops (WSDM, ICML, AAAI).

BuzzTweet, Berlin, Germany

LEAD DEVELOPER > APRIL 2015 – JUNE 2017

- R&D of a real-time Sentiment Analysis framework to predict KPIs. **Java, Weka ML, AWS, Docker**

GroopLab, Berlin, Germany

DATA SCIENTIST, SHAREHOLDER > JUNE 2014 – JUNE 2015

- R&D of Machine-Learning methods for event recommendation and NER. **Python, scikit-learn, Java, AWS**

ADDITIONAL PROJECTS

Liber AI, London, UK

CO-FOUNDER, RESEARCHER > OCTOBER 2017 – PRESENT

- Open-source research project [Neural SPARQL Machines](#): Question Answering over Knowledge Graphs using Recurrent Neural Networks (200+ stars and 80+ forks on Github). **Python, TensorFlow, SPARQL**

DBpedia Association, Leipzig, Germany

MENTOR, ORGANISATION ADMIN (REMOTE) > FEBRUARY 2017 – PRESENT

- Org Admin for [Google Summer of Code](#) (GSoC) at [DBpedia](#), a Multilingual Semantic Knowledge Graph.
- Mentor of 13 students in open-source projects on Deep Learning for and over Knowledge Graphs. 85% of them joined the programme as co-mentors the next year. **Python, TensorFlow, Keras, SPARQL**

EDUCATION

University of Leipzig, Leipzig, Germany

PHD IN COMPUTER SCIENCE > JULY 2012 – OCTOBER 2018

University of Milan-Bicocca, Milan, Italy

MASTER OF SCIENCE IN ARTIFICIAL INTELLIGENCE > NOVEMBER 2008 – MARCH 2012

Polytechnic of Milan, Milan, Italy

BACHELOR IN COMPUTER ENGINEERING > SEPTEMBER 2003 – SEPTEMBER 2008

ABOUT ME

- Team player, self-starter, lateral thinker, good at presentations, excellent at brainstorming.
- [Drummer](#) since 1996. 200+ live shows, 4 albums and 3 EPs published.
- Interests: progressive music, genealogy, quantified self, association football, science fiction.
- EU citizen, settled status in the UK, no VISAs/sponsorships needed.
- Languages: Italian (native), English (full proficiency), French & German (basic).

SELECTED PUBLICATIONS & AWARDS

Complete list on [Google Scholar](#).

Tommaso Soru, Jim Marshall, [Anticipating the Future with Large Language Models](#). To appear, IEEE ICSC, 2025. **Industry paper; AI-driven strategic foresight**

Tommaso Soru, Jim Marshall, [Trend Extraction and Analysis via Large Language Models](#). In proceedings, IEEE ICSC, 2024. **First industry paper; AI-driven strategic foresight**

Anand Panchbhai, **Tommaso Soru**, Edgard Marx, [Exploring Sequence-to-Sequence Models for SPARQL Pattern Composition](#). In proceedings, KGSWC, 2020. **First publication as a supervisor**

Diego Moussallem, **Tommaso Soru**, Axel Ngonga, [Neural Translation and Enrichment of Knowledge Graphs](#). In proceedings, ISWC, 2019. **First-of-its-kind research application**

Muhammad Saleem, Alexander Potocki, **Tommaso Soru**, Olaf Hartig, Axel Ngonga, [CostFed: Cost-Based Query Optimization for SPARQL Endpoint Federation](#). In proceedings, SEMANTICS, 2018. **Best paper award**

Tommaso Soru, Edgard Marx, Diego Moussallem, Gustavo Publio, André Valdestilhas, Diego Esteves, Ciro Baron Neto, [SPARQL as a Foreign Language](#). SEMANTICS, 2017. **Best poster nominee; most cited poster ever of the SEMANTICS conference; first-ever use of sequence-to-sequence models on knowledge graphs**

Tommaso Soru, Edgard Marx, Axel Ngonga, [ROCKER: A Refinement Operator for Key Discovery](#). In proceedings, WWW, 2015. **14% acceptance rate; 40+ citations**

Tommaso Soru, Axel Ngonga, [A comparison of supervised learning classifiers for link discovery](#). In proceedings, SEMANTICS, 2014. **40+ citations**