

PERSONAL INFORMATION

Livio Robaldo

✉ livio.robaldo@swansea.ac.uk

🌐 www.liviorobaldo.com

Full CV available at <http://www.liviorobaldo.com/LivioRobaldoCV.pdf>

Gender M | Date of birth 23/06/1979 | Nationality Italian

PERSONAL STATEMENT

Researcher and software developer with fifteen years of experience in **LegalTech, Natural Language Processing (NLP), Knowledge Representation (KR), Artificial Intelligence (AI)**.

WORK EXPERIENCE

1.5.2024 – 28.2.2026

Senior Lecturer in LegalTech

School of Law, Swansea University.

Research activities in NLP, KR, Semantic Web, Big Data, AI, etc. applied to the legal domain:

- Innovate UK project "Odyssey" (www.liviorobaldo.com/odyssey.html)

Work package leader of the project. Co-developer of a metadata vocabulary to enrich UK National Archives (<https://www.nationalarchives.gov.uk>) datasets to improve Large Language Models accuracy and interoperability

1.6.2015 – 31.5.2020

Research Associate

University of Luxembourg, Interdisciplinary Centre for Security, Reliability and Trust (SnT)

Research, development, and management activities in the context of funded projects in NLP, KR, Semantic Web, Big Data, AI, etc. applied to the legal domain:

- H2020 Marie Curie IF project "ProLeMAS" (www.liviorobaldo.com/prolemas.html)

Principal investigator of the project. Inventor of a novel machine-readable format for representing norms in natural language, coming from existing legislation. Developer of NLP procedures for tagging documents in LegalDocML (<https://www.oasis-open.org/committees/legaldocml>). 6-months secondments at APIS EOOD.

- H2020 Marie Curie RISE project "MIREL" (www.liviorobaldo.com/mirel.html, www.mirelproject.eu)

Principal author of the project (see statement at <http://www.liviorobaldo.com/MIRELstatement.pdf>). Management and coordination of the 16 partners: I organized the secondments and the research activities among the partners of the MIREL consortium, I joined some of these activities by co-authoring 12 publications in legal reasoning, Semantic Web, multi-agent systems, etc., I authored, contributed to, or monitored the writing of all reports and deliverables while ensuring they were delivered on due time and with high quality, I chaired project meetings and I organized dissemination initiatives such as conferences, workshops, tutorials, seminars, journal special issues, and ultimately a book devoted to the project. 3-months secondment at Stanford University, partner in MIREL.

- H2020 Marie Curie ITN project "LAST-JD RoE" (www.liviorobaldo.com/lastjd.html, www.last-jd-rioe.eu)

Responsible of the LAST-JD PhD students in cotutelle with the University of Luxembourg and contributor of the project writing. I followed and advised the students in their research activities, I gave classes of NLP techniques applied to legal documents, I assisted them with administrative procedures, I allowed six of them to spend visiting periods at universities in the MIREL consortium and I co-authored eleven publications with five of them in information retrieval, concept recognition, legal ontologies, and textual inference.

- FNR/CORE national project "DAPRECO" (www.liviorobaldo.com/dapreco.html, www.fnr.lu/projects/data-protection-regulation-compliance)

Leader of the Work Package devoted to KR and NLP. The main result of the project is the DAPRECO knowledge base (D-KB, <https://link.springer.com/article/10.1007/s10849-019-09309-z>), which formalizes the GDPR in the XML standard LegalRuleML (<https://www.oasis-open.org/committees/legalruleml>) and which I mainly authored and developed. Contributor of the OWL Privacy Ontology (PrOnto), the biggest and most exhaustive OWL ontology for the privacy and data protection domain at the time of the project. Developer of NLP procedures to convert legal documents in the data protection domain into LegalDocML XML files, while performing entity linking and concept recognition with respect to the D-KB and PrOnto.

1.3.2014 – 20.1.2021 Stakeholder and developer in Nomotika SRL

Nomotika SRL was a spin-off of the University of Turin and Augeos SPA (www.augeos.it), leader in the Italian ICT market for the financial domain.

Researcher and developer of novel NLP procedures for entity linking, concept recognition, semantic role labelling and text classification within the main company's products (Eunomos, described in the article https://link.springer.com/chapter/10.1007/978-3-642-35731-2_9). Researcher and developer of NLP tools to extract data from prospectuses and other financial documents from banks in the client portfolio of Augeos SPA. Researcher and developer of machine learning procedures for textual entailment between ISO and NIST security standards controls.

1.10.2017 – 31.12.2017 Visiting Scholar

Stanford University (Palo Alto, USA).

Visiting scholar in the context of the project MIREL. Scientific referent: prof. Cleo Condoravdi. The visiting period was devoted to disseminate my research results while collecting feedbacks and suggestions for improvement from researchers working at Stanford University, in particular about the DAPRECO knowledge base which I mainly developed during this secondment.

1.10.2015 – 31.12.2015 Seconded Researcher and Developer

1.03.2017 – 31.05.2017

Temporary researcher and developer in the private company APIS EOOD (<https://www.apis.bg/en>), specialized in ICT solutions for the LegalTech domain, in the context of the project ProLeMAS.

Researcher and developer of NLP solutions for the products marketed by the company. Specifically, I contributed to design, implement, and disseminate a novel NLP textual similarity technique called Unifying Similarity Measure, which combines cosine similarity, n-gram similarity, and approximate string matching, in order to recognize transposition of EU directives in national legislation.

1.09.2014 – 31.12.2014 Start-up accelerator (WCAP)

Winner of the WCAP start-up grant, project SentiTagger (see <http://bit.ly/2P2DEh2>).

WCAP is a Telecom Italia grant to support new start-ups. In 2014, the selection was highly competitive: only 40 projects out of about 1,300 submitted ones were selected. Each selected project was granted 25,000 euros, VAT excluded, from Telecom Italia.

Inventor and developer of a hybrid (rule-based and statistical) NLP framework based on the TULE parser. Start-up project proposal to use it for Sentiment Analysis, together with the OpinionMining-ML XML format (see <https://www.sciencedirect.com/science/article/abs/pii/S0920548912001225>).

Eventually, the start-up was not created in that the notification of the Marie Curie fellowship for the project ProLeMAS arrived when I was working in the WCAP incubator. I then decided to abandon the start-up project in order to get the fellowship. As the NLP hybrid framework used in SentiTagger was invented in the context of my PhD and other projects at the University of Turin, I asked and obtained the permission from the Patent Office of that University to release it under GNU/LGPL. It is available, together with a demo and a user manual, at <http://www.liviorobaldo.com/sdftagger.html>.

1.9.2004 – 31.05.2015

PhD and Post-doctoral Researcher

University of Turin (Italy), Department of Computer Science.

Researcher and developer in NLP, KR, and AI. Basic research on several topics (see publications at www.liviorobaldo.com/publications.html). Applied research in the following financed industrial projects:

- European FP7 project "EUCases" (www.liviorobaldo.com/eucases.html, www.eucases.eu)
Researcher and developer of the NLP toolkit for the EUCases platform, for the Italian language (see statement at www.liviorobaldo.com/EUCasesStatement.pdf). The toolkit was able to crawl legal documents from the Web and transform them in LegalDocML while recognizing cross-references and other named entities. Contributor of the European Legal Taxonomy Syllabus ontology, used within the Eunomos legal document management system. Contributor of several project deliverables (see <http://www.eucases.eu/deliverables.html>).
- e-Justice project "BO-ECLI" (<https://bo-ecli.eu>)
Contributor of the research and dissemination activities of the European Case Law Identifier (ECLI).
- Regione Piemonte project ICT4LAW (www.liviorobaldo.com/ict4law.html)
Researcher and developer of rule-based NLP procedures for recognizing modificatory provisions and statistical NLP procedures for text classification. In the context of this project, I also had a short-term contract with the private company SSB progetti, partner in the project. The ICT4LAW project led to the design and the implementation of Eunomos, an advanced legal document management system, and to the creation of the spin-off Nomotika SRL (see <https://link.springer.com/article/10.1007/s10506-016-9184-3>).
- Smart City Social Innovation project "FirstLife" (www.firstlife.org), Regione Piemonte projects "SEeS@W" (www.firstlife.org/projects/seesaw), and "Librare" (www.firstlife.org/librare-libri-in-movimento)
Contributor of the writing and implementation of research tasks in Sentiment Analysis, in relation to the WCAP project SentiTagger (see below).
- Regione Piemonte project LIS4ALL (www.bepseng.it/ricerca/lis4all-2011-2014) and Regione Piemonte project ATLAS (www.crit.rai.it/eletel/2011-3/113-4.pdf)
Researcher and developer of machine-translation tools and corpora from/to Italian language and Italian Sign language, based on the TULE parser (see <https://github.com/alexmazzei/TULE>). In the context of this project, I also had a short-term contract with the private company Virtual Reality & Multi Media Park, partner in the project.
- FIRB project TOCAI.AI (<http://bit.ly/38pwZq5>)
Researcher and developer of NLP procedures based on the TULE parser. Annotator of the Wikipedia section of the Turin University Treebank. Contributor of the development of a bidirectional dependency parsing trained on the Turin University Treebank.
- FP6-IST project HOPS (<https://cordis.europa.eu/project/id/507967>)
Researcher and developer of NLP procedures based on the TULE parser for ontology-driven natural language interfaces to the HOPS databases.

5.02.2007 – 31.07.2007

1.06.2009 – 31.07.2009

Visiting Scholar

University of Pennsylvania (Philadelphia, USA).

Visiting scholar in the context of activities related to research on Generalized Quantifiers, which I carried out in the context of my PhD thesis, and new research activities started at the University of Pennsylvania on the Penn Discourse Treebank (PDTB). Contributor to the PDTB 2.0 annotation manual (see <https://www.seas.upenn.edu/~pdtb/PDTBAPI/pdtb-annotation-manual.pdf>). Scientific referents: prof. Maribel Romero, prof. Aravind Joshi, and prof. Eleni Miltsakaki.

EDUCATION AND TRAINING

1.9.2004 – 1.2.2007

Ph.D. in Computer Science / Artificial Intelligence

PhD in computer science, with a research thesis titled "Dependency Tree Semantics (DTS)". DTS was an AI underspecified format built on the top of the dependency trees built by the TULE parser.

1.9.1998 – 1.7.2003

MSc in Computer Science

MSc in computer science, with a master thesis titled “Use of lexical resources for the semantic interpretation of the analysis of natural language”. Final grade: 110/110 cum laude and mention for the exceptional curriculum. In addition, I was awarded of “Premio Optime”, Confindustria award given to the best graduates of the University and Polytechnic of Turin, for the Academic year 2003-2004.

1.9.2014 – 31.10.2014

WCAP 2014 Program

Working Capital 2014 program: 2-months intensive seminar series devoted to start-up foundation, pitch preparation, marketing analysis, etc., organized by Telecom Italia for WCAP 2014 winners.

PERSONAL SKILLS

Mother tongue(s) Italian

Other language(s)

	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken interaction	Spoken production	
English	C1	C2	C1	C1	C2
French	A1	A1	A1	A1	A1

Communication skills

- Excellent communication skills gained through experience in academia: dissemination in conferences, meetings with industry, etc.
- Excellent writing skills gained through submission of research project proposals that were retained for funding (twelve projects so far, three as main author) as well as publications in top scientific journals and conferences, workshops, book chapters, and technical reports.
- Excellent teaching skills gained through teaching experience in academia (990 hours in total)

Organisational / managerial skills

- Leadership skills gained through supervision of Ph.D., bachelor, and master students.
- Responsible for the University of Luxembourg of the LAST-JD program (<https://www.last-jd-rioe.eu>)

Computer skills and competences

Talented programmer since the high school (1995)

Specialized in Java since 15 years; programming languages currently known:

- Java (excellent)
- Python (good)
- C/C++/C# (good)
- LISP (good)
- Prolog (good)
- HTML/CSS (very good)
- JavaScript (basic)

Good knowledge of main DBMSs, both relational (e.g., MySQL) and NoSQL (e.g., MongoDB).

Good knowledge of RDF models, computational ontologies, and related formats (XML, OWL, JSON etc.), in particular those specific for the LegalTech domain (LegalDocML and LegalRuleML). Inventor of a novel XML format for Sentiment Analysis called OpinionMining-ML (mentioned above).

Ability to learn a new programming language or a new library, environment, editor, methodology, etc. for the languages listed above in very short time.

Driving licence B

ADDITIONAL INFORMATION

Publications
Citations

More than 120 publications, see <http://www.liviorobaldo.com/publications.html>
1835 citations on Scopus (<https://www.scopus.com/authid/detail.uri?authorId=15045771000>)