Curriculum Vitae

Croce Danilo

Last Update: May 2026

Birth Date: July 20^{th}, 1982

Birth Place: Marino (RM)

Current Position: Associate Professor (SSD INFO-01/A, ex INF/01)
Department of Enterprise Engineering, University of Rome "Tor Vergata"

Main Research Topics

- Large Language Models (LLMs) and Generative AI: instruction tuning, multi-task and continual learning, sustainable adaptation and evaluation;
- Machine Learning for Natural Language Processing: kernel methods, syntactic/semantic parsing, information extraction;
- Semantic Search and Evidence-based AI: retrieval-augmented generation (RAG), fact verification, adaptive information retrieval;
- Multimodality and Grounded AI: vision-language models for captioning, Visual Question Answering, and situated reasoning;
- Human-Robot Interaction: situated language understanding, adaptive and context-aware interfaces;
- Ethics and Transparency in AI: interpretability, fairness, and "ethics-by-design" methodologies.

Education: Master's Degree in Computer Engineering at the University of Rome, "Tor Vergata" obtained in 2008 (110/110 *cum laude*).

Qualified as Professional Engineer after passing the Italian State Examination

Ph.D. in Computer Science and Control Engineering, University of Rome, "Tor Vergata", Rome (June 2012), Italy. Ph.D. Thesis title: "*Structured Learning for Natural Language Semantic Processing*"

Bibliometric Indicators (H-index)
Google Scholar[1]: 26
Scopus[2]: 20

[1] https://scholar.google.it/citations?user=dXewdYAAAAAJ
[2] https://www.scopus.com/authid/detail.uri?authorId=27567467600

Contents

1 Professional experience

Tenure-track Assistant Professor (RTD-B) at the Department of Enterprise Engineering, University of Rome "Tor Vergata", scientific-disciplinary sector INF/01 from September 2022 to September 2025.

Assistant Professor (RTD-A) at the Department of Enterprise Engineering of the University of Rome, "Tor Vergata", Academic discipline ING/INF-05, from August 2015 to July 2020.

Postdoc at the Department of Enterprise Engineering of the University of Rome, "Tor Vergata". Research Topic: "*Statistical Methods of Lexical Semantics for Advanced Search Engines*", Academic discipline ING-INF/05, between 2013 and 2014.

Postdoc at the Department of Enterprise Engineering of the University of Rome, "Tor Vergata". Research Topic: "*Algorithms and Machine Learning methods for Advanced Information Retrieval*", Academic discipline ING-INF/05, in 2012.

Visiting Student at the Computational Linguistics and Computer Science Department, University of Colorado at Boulder, CO, USA (February - May 2011). Research Topic "*Automated Verb Class Classification and Clustering*". Supervisors: Alessandro Moschitti and Martha Palmer.

Member of the *Semantic Analytics Group* (SAG) at the University of Rome, "Tor Vergata" since 2008.

Member of the *ART* Group, *Artificial Intelligence at Tor Vergata*, at the University of Rome, "Tor Vergata" since 2008.

2 Research Activity

My research activity focuses on the study and application of advanced *Machine Learning* techniques, with particular emphasis on models based on **Large Language Models (LLMs)**, deep neural networks, and innovative kernel methods for complex problems in *Natural Language Processing* (NLP), Information Retrieval, and Computer Vision. My interests range from theoretical modeling to the practical realization of intelligent applications, with a strong emphasis on epistemological transparency and the integration of ethical principles into algorithmic decision-making. In particular, I have worked on the development of **geometric models of lexical semantics**, **kernel-based learning** techniques, and **deep learning** methods, as well as the study of complex linguistic phenomena typical of Big Data and social networks. I have contributed to the original development of advanced AI applications in areas such as *Question Answering*, *Semantic Search*, *Human-Robot Interaction*, *Fact Verification*, *Sentiment Analysis*, and **multimodality** (Image and Video Captioning, Visual Question Answering). Throughout my career, I have promoted both theoretical and applied approaches, participating in the development of large-scale linguistic systems and resources for Italian and other languages, and addressing current topics such as sustainability, portability, and ethics in machine learning models.

My research activity focuses on the development of transparent, adaptive, and linguistically grounded AI systems, spanning from theoretical models of language understanding to large-scale multimodal and socially responsible applications.

The main research topics can be divided into the following main areas.

Machine Learning, Large Language Models, and Advanced Methods for NLP and Computer Vision. The main lines of my research have focused on the study, development, and application of Machine Learning techniques for the automatic processing of natural language and computer vision, with an approach that combines both theoretical and practical aspects.

- Study of innovative kernel functions for the syntactic and semantic representation and processing of texts, using Support Vector Machines, structured kernels, and neural networks, to enable complex tasks such as automatic classification, semantic role labeling, and large-scale lexical knowledge acquisition.
- Combination of kernel methods and deep neural networks to integrate explicit linguistic information within neural learning paradigms, addressing phenomena typical of social networks and Big Data.
- Development of methods for epistemological transparency in neural networks and the integration of ethical principles into machine learning algorithms.
- Investigation of unsupervised learning paradigms (semi-supervised, boosting), scalability and portability of models, and online learning techniques for continuous and cognitively plausible model updates.
- Definition of models and algorithms for classical NLP tasks, such as semantic analysis, named entity recognition, question answering, and semantic search, often reformulated as classification or ranking problems.
- Development of methods based on Transformer models, including few-shot, zero-shot, multi-task, and continual learning paradigms, for both NLP and multimodal applications.
- Integration of linguistic and visual information in multimodal models for advanced tasks such as Image Captioning, Video Captioning, and Visual Question Answering, including for the Italian language.
- Creation of multimodal benchmarks and model suites (MM-IGLU, MM-IGLU-IT), as well as development of models tailored for different linguistic and application contexts.
- Development of LLM-based models for syntactic and semantic parsing (e.g., U-DepPLLaMA), and study of the sustainability, modularity, and scalability of neural models through techniques such as LoRA and PEFT.

Main Results: These activities have produced internationally recognized original contributions, including: the development of semantic tree kernels that integrate lexical and grammatical information [106, 107, 139, 140, 94, 132, 110, 99, 133, 117, 118]; the combination of kernel methods and neural networks [64, 63, 82]; approaches for increasing epistemological transparency and injecting ethical constraints in neural processes [65, 66, 58]; application of semantic kernels to the acquisition of linguistic knowledge such as frames in the Frame Semantics paradigm [152, 153, 133, 150, 144]; models for online learning and incremental learning [91, 93, 126, 125, 116, 109]; advances in the use of Transformer-based models for few-shot, zero-shot, multi-task, and continual learning [55, 47, 42, 64]; and the development of resources and systems for multimodality in Italian (Image Captioning, Video Captioning, VQA) [85, 68, 46]. Among the most recent and relevant results (2023-2025) are the development of the ExtremITA and U-DepPLLaMA models for advanced linguistic analysis, and the multimodal suite MM-IGLU and MM-IGLU-IT for grounded language understanding, as described in the following recent publications [17, 19, 22, 25, 24, 26, 30]. These lines of research have recently been extended through multimodal and situated reasoning frameworks, including the G-SRL model for robot command understanding [9], the Sanskrit Voyager system for unified linguistic resource integration [8], and dialogue-planning approaches for Human-Robot Interaction [15]. Further recent developments concern the detection of machine-generated text in Italian [4], visual word sense disambiguation through zero-shot Visual

Question Answering [5], and the learning of molecular structures from infrared spectra through latent evidence prediction [6].

Semantic Search, Adaptive Information Extraction, Fact Verification, and Retrieval-Augmented Generation. Research activities in this area have focused on the design of advanced tools for semantic search, adaptive information extraction, and automatic fact verification, combining traditional techniques (vector lexicons, word space models, matrix factorization) with modern approaches based on Large Language Models.

- Study and development of vector lexicons and distributed semantic representations through *word space* paradigms, co-occurrence analysis on large corpora, and algebraic factorization techniques, enabling large-scale *enterprise search* and conceptual querying.
- Design of flexible semantic search platforms, capable of managing both highly specialized lexicons and querying via concepts, sentences, document fragments, or keywords, applicable to vertical domains (tourism, banking) and the open Web.
- Adaptation, specialization, and evaluation of LLMs for semantic search and information retrieval in specific domains, including instruction-tuned models and pipelines leveraging external evidence.
- Development of Retrieval-Augmented Generation (RAG) systems, which combine large generative models with retrieval modules to produce informed responses grounded in external sources, enabling applications such as question answering and automated report generation based on up-to-date knowledge.
- Implementation of complete pipelines for automated fact-checking, particularly based on Wikipedia (e.g., the FEVER-it system), integrating retrieval modules, NLI verification, and claim validation.
- Study and implementation of automatic topic discovery techniques, using both probabilistic models (LDA, traditional topic modeling) and neural/embedding-based approaches, for unsupervised topic discovery and classification of large text collections.

Main results: These activities have led both to the development of industrial platforms for semantic search in various application domains (tourism, banking) and to methodological and applied contributions recognized in the literature. The results are discussed in [17, 25, 67, 62, 94, 111, 134, 131, 122, 120]. Recent work extends this paradigm to biomedical domains, with retrieval-augmented pipelines for question answering such as UniTor@BioASQ [12], and to graph-based reasoning and navigation tasks [13]. This research line was further extended in 2026 toward accurate, sustainable and trustworthy access to large-scale biomedical information collections [3], financial question answering based on agentic prompting [2], and retrieval-augmented linguistic reasoning for solving Italian crossword clues [7].

Social Media Analytics, Sentiment Analysis, Epidemic Intelligence, and Tourism. Research activities in this area have focused on the study and application of advanced Machine Learning techniques and linguistic-semantic models for the automatic analysis of opinions and social phenomena in texts produced on social media and the web. The focus has been on modeling polarity (sentiment), extracting opinions, and detecting signals in various application domains, including public health (epidemic intelligence) and digital tourism. Multiple media (forums, reviews, tweets, tourism portals) have been analyzed, in both Italian and English, using models ranging from lexical and semantic kernels to deep neural networks.

- Development of linguistic models for automatic classification of polarity, detection of subjective opinions, and identification of emerging topics in multilingual and multi-domain contexts.
- Study and implementation of techniques for the automatic acquisition of opinion lexicons and lexical resources for sentiment analysis.

- Creation of annotated benchmark corpora for the systematic evaluation of sentiment analysis systems in Italian.
- Analysis of emerging social phenomena via social media, with a special focus on applying NLP methods to epidemic intelligence: early detection of trends, events, and epidemiological signals through the automatic monitoring of textual data from social sources.
- Design of systems and pipelines for the automatic analysis of tourists' opinions and experiences, integrating data from social networks, online reviews, and thematic portals to monitor the digital reputation of destinations, discover new travel trends, and identify the perceptions and needs of travelers—supporting public bodies, regions, and tourism operators.

Main Results: These activities have led to the publication of relevant works and the creation of reference resources for the research community, including [17], [21], [33], [89], [90], [92], [117], [118], [121], [122], [114], [109], [80], [48], [88], [86]. Recent studies have also explored LLM-based sustainability and social responsibility reporting [11, 16], and fine-grained sentiment modeling grounded in frame semantics [14].

Natural Language Learning and Human-Robot Interaction (HRI). Research in this area has focused on the definition and implementation of advanced language workflows and neural frameworks for semantic understanding in robotic platforms, aiming for robust Human-Robot Interaction (HRI) across different domains and operational scenarios. These activities, carried out in collaboration with the Cognitive Robotics Laboratories at Sapienza University, have involved both theoretical and applied investigations:

- Design of general and modular linguistic workflows for robotic platforms, independent of the underlying system and operational tasks, leveraging psycholinguistic theories such as *Frame Semantics* for semantic parsing and command interpretation.
- Development of neural models (e.g., GrUT) for advanced semantic understanding of natural language commands, generating structured representations that connect linguistic meaning to the robot's internal knowledge base.
- Realization of context-aware human-robot interfaces that integrate semantic interpretation with the perception of the surrounding environment, enabling robots to adapt their responses to dynamic and situated contexts.
- Investigation of dynamic and developmental aspects of language understanding, including the integration of incremental learning and semantic adaptation as the robot interacts with humans and its environment.
- Application of HRI methodologies to the medical field, such as the development of robotic physiotherapist platforms capable of integrating physical and cognitive interaction ([45, 35]).

Main Results: The main achievements of this line of research are documented in international conferences and journals on robotics and computational linguistics. These include best paper awards ([101, 81]), the successful participation in the SemEval 2013 international evaluation campaign (Task: Spatial Role Labeling, [119]), and the development of context-aware processing pipelines such as `LU4R`. Key results are discussed in [87, 59, 113, 97, 95, 98, 100, 102, 112, 124], as well as recent contributions describing the GrUT model and medical HRI applications ([35, 45]). Recent research expands this direction through multimodal grounding and dialogue planning for Human-Robot Interaction [15, 9].

Ethics and Transparency in Machine Learning. Research activities in this area focus on the development of techniques and methodologies to make machine learning models not only more powerful and adaptable, but also more transparent, interpretable, and guided by explicit ethical principles. The goal is to address key challenges related to the trustworthiness of artificial intelligence systems, both from

an epistemological standpoint and in terms of social responsibility and sustainability.

- Study of algorithmic techniques to increase the epistemological transparency of neural models and facilitate the interpretability of decisions made by deep networks, including in complex scenarios such as NLP, computer vision, and multimodal systems.
- Development and experimentation of the EthicalNN framework, a PyTorch-based architecture that allows the integration of explicit ethical constraints ("truth-makers") into the learning process, fostering the design of models that learn to respect moral principles and requirements of fairness and equity.
- Proposal and implementation of "Ethics by Design" methodologies for intelligent adaptive systems, with the aim of making ethical design a structural—not merely post-hoc—element in AI systems, thereby promoting transparent, sustainable, and socially responsible solutions in both industrial and scientific contexts.
- Application and validation of the developed techniques on real datasets and scenarios characterized by high ethical sensitivity (e.g., evaluation of bias, fairness, and accountability in judicial, social, and healthcare datasets).

***Main results*:** The main theoretical and applied innovations in this area are summarized in works on the EthicalNN architecture ([58]), on approaches to neural network transparency ([65]), and on the "Ethics by Design" paradigm for intelligent adaptive systems ([39]). These efforts are closely connected to the development of socially responsible NLP applications, such as sustainability and impact-oriented language modeling [11].

Overall, these research activities contribute to the development of sustainable, interpretable, and human-centered AI systems, integrating linguistic, visual, and ethical dimensions across diverse domains.

2.1 Scientific Awards

- Honorable Mention, "Federico Sangati" Best System Award, EVALITA 2026 [4].
- Best Paper Award, 14th Joint Conference on Lexical and Computational Semantics (*SEM 2025) [10].
- Outstanding Paper Award, Workshop on NLP for Positive Impact (NLP4PI @ ACL 2025) [16]
- Best System Award, EVALITA 2023 [28]
- Best Paper Award, International Conference on Deep Learning Theory and Applications (DeLTA 2023) [33]
- Best Paper Award, Workshop on Natural Language for Artificial Intelligence (NL4AI, 6a edizione), 2022 [38]
- Best paper award, EVALITA 2020 [56]
- Best paper award, International Conference of the Italian Association for Artificial Intelligence (AIxIA 2019) [63]
- Best paper award, Fifth Italian Conference on Computational Linguistics (CLiC-it 2018) [72]
- Distinguished Young Paper, Third Italian Conference on Computational Linguistics, Napoli 2016. [95]
- IBM Best paper award COLING Conference (Rank A Conference) Dublin, August 2014. [117]
- Distinguished Young Paper, First Italian Conference on Computational Linguistics, Pisa 2014. [118]
- Best System Paper, RoboCup Symposium, Joao Pessoa, Brazil, 2014. [101]
- Best paper award, CICLING Conference, Mexico City, Mexico, March 2009. [147]

2.2 Systems Evaluated in International Evaluation Campaigns

The machine learning methods involved in the study of *Computational Natural Language Learning* and *Opinion Mining* have also been used for the participation in International Evaluation Campaigns (e.g., *SemEval* organized by the American Association of Computational Linguistics) or National Evaluation Campaigns (e.g., *EVALITA*, organized by the Italian Association for Artificial Intelligence and the Italian Association of Computational Linguistics), often with excellent results:

- EVALITA 2026 Task "Crossword Solving" [7]. (**Best system**)
- EVALITA 2026 Task "Enhanced Visual Word Sense Disambiguation for Italian" [5] (**Best system**)
- EVALITA 2026 Task "DEtection and SEgmentation of MAchine generated texts in Italian" [4] (**Honorable Mention**)
- EVALITA 2026 Task "Benchmarking Language Models in the Financial Domain" [2]
- BioASQ 2025 Task 13b: Participation with the **UniTor** system for modular biomedical question answering. [12] (**1st place** in Factoid and Ideal Answer tasks)
- EVALITA 2023: Participation in *all 13 tasks* with the **ExtremITA** system, based on Large Language Models trained on multiple Italian linguistic tasks. [28]
- EVALITA 2020 Task: "Task on Stance Detection", [56] (**Best System**)
- EVALITA 2020 Task: "Task on Meme Recognition and Hate Classification (Multimodal Artefacts Recognition)", [53] (**Best System**)
- EVALITA 2018 Task: "Task on Irony Detection in Italian Tweets", [75] (**Best System** in the Sarcasm Detection subtask)
- EVALITA 2018 Task: "itaLIan Speech acT labEliNg", [71] (**Best system**)
- SemEval 2016 Task: "*Community based Question Answering*", [94] (**Best system**)
- EVALITA 2016 Task: "*Sentiment Polarity Classification in Twitter*" (**Best System** on the "Sentiment Detection sub-task"), [89]
- SemEval 2014 Task: "*Aspect Based Opinion Mining*", [115] (**Second Best System** on the Topic Recognition sub-task)
- EVALITA 2014 Task: "*Sentiment Polarity Classification in Twitter*" (**Best System** on the "Irony Detection sub-task"), [114]
- SemEval 2013 Task: "*Spatial Role Labeling*", [119] (**Best system**)
- SemEval 2013 Task: "*Sentiment Analysis in Twitter*", [121]
- StarSem 2013 Task: "*Semantic Text Similarity*", [123] (**Best System** on "Semi-structured Text Similarity" recognition sub-task)
- SemEval 2012 Task: "*Semantic Text Similarity*", [130]
- EVALITA 2011 Task: "*Frame Labeling over Italian Texts*", [141] (**Best system**)

2.3 Major Invited Speeches

- Invited Teacher for the course "Large Language Models", held during the "Bertinoro International Spring School 2024 (BISS 2024)", Bertinoro, 2024.
  `https://cs.unibo.it/projects/BISS/2024/courses/`
  `https://github.com/crux82/BISS-2024`
- Lecture at the summer school "Advances in Artificial Intelligence 2024", organized by the Lake Como School of Advanced Studies, Como, 2024. Talk title: "Large Language Models (LLM)".
  `https://sites.google.com/unimib.it/advancesinai-2024/program`
- Invited talk titled "The phenomenon of hallucinations in LLMs", held at the Laboratory of the History of Linguistic Ideas, Department of Philosophy, "Sapienza" University of Rome, as part

of the series “The Babel of Golems - Conversations on human and AI languages”, Villa Mirafiori, Rome, 13 November 2024.
https://web.uniroma1.it/storiaideelinguistiche/seminari

– Invited tutorial titled “Large Language Models and How to Instruction Tune Them (in a Sustainable Way)”, presented at the Ninth Italian Conference on Computational Linguistics (CLiC-it 2023), Pisa, 2023.
https://clic2023.ilc.cnr.it/tutorial/
– Invited Tutorial entitled “Training Neural Architectures for NLP”, organized as part of the “Lectures on Computational Linguistics 2021” sponsored by the Italian Association of Computational Linguistics, June 2021.
https://www.ai-lc.it/en/lectures-2021/
https://github.com/crux82/AILC-lectures2021-lab
– Invited Research communication titled “Deep Learning in Semantic Kernel Spaces” at the Fourth Italian Conference on Computational Linguistics, Roma 2017.
– Invited Tutorial: “LU4R: Adaptive Spoken Language Understanding Chain For Robots" at the European Robotics League - Service Robots, Lisbon
– Invited Talk for the “IBM Best paper award" at the COLING 2014 conference, presenting the work “A context-based model for Sentiment Analysis in Twitter" [117].
– Invited Talk for the Best System at SemEval 2013, organized by the “Special Interest Group on Semantics of the American Association on Computational Linguistics", for the Task “Spatial Role Labeling".
– Invited Talk for the Best System at StarSem 2013, organized by the “Special Interest Group on Semantics of the American Association on Computational Linguistics", for the Task: "Semantic Text Similarity".

3 Academic and Professional Activities

Head of the Laboratory of the "Semantic Analytics Group" (SAG) since 2008[3]. The SAG carries out research activities in the area of Machine Learning for Natural Language Processing and collaborates with Italian and foreign research institutions for the organization of conferences (e.g. SemEval, EMNLP, Clic-It, Italian Workshop on Information Retrieval, EVALITA), the coordination of funded research projects and the promotion of Natural Language Processing in Italy and abroad.

Member of the Steering Committee of the Italian Association of Computational Linguistics (AILC) from 2022 to present, the largest network of research, academic institutions, and industrial organizations related to Computational Linguistics and applications of language technologies. AILC gathers about 200 individual members and companies, and is responsible for congress initiatives (Clic-It), editorial (the Italian Journal of Computational Linguistics) and evaluation campaigns (EVALITA) that gather very important initiatives of all Italian research in the area of language (such as linguistics, computational linguistics, artificial intelligence, physics of voice, cognitive sciences, digital humanities).
https://www.ai-lc.it/en/association/

3.1 Organizer/Chair of Conferences and Workshops

– Program Chair of the “Twelfth Italian Conference on Computational Linguistics” (CLiC-it 2026), Palermo, 2026.

[3] http://sag.art.uniroma2.it

https://clic2026.unipa.it/

- Member of the Organizing Committee - Demo Track Chair at the "Conference of the European Chapter of the Association for Computational Linguistics" (EACL 2026), [1].
  https://2026.eacl.org/committees/organization/
- Member of the Organizing Committee - Publication Chair at the "14th Joint Conference on Lexical and Computational Semantics" (*SEM / StarSem 2025).
  https://starsem2025.github.io/
- Organizing Committee Member - Special Track Organizer for the "Calamita" track at the 10th Italian Conference on Computational Linguistics (CLiC-it 2024). The track is dedicated to the evaluation of Large Language Models (LLMs) for Italian and received over 20 submissions.
  https://clic2024.ilc.cnr.it/
- Member of the Organizing Committee - Publication Chair at the "Tenth Italian Conference on Computational Linguistics" (CLiC-it 2024).
  https://clic2024.ilc.cnr.it/organization/
- Member of the Organizing Committee - Publication Chair at the "18th Conference of the European Chapter of the Association for Computational Linguistics" (EACL 2024).
  https://2024.eacl.org/committees/organization/
- Member of the Organizing Committee - Publication Chair at the "2024 Conference on Empirical Methods in Natural Language Processing" (EMNLP 2024).
  https://2024.emnlp.org/organization/
- Member of the Organizing Committee - Workshop & Tutorial Chair at the "22nd International Conference of the Italian Association for Artificial Intelligence" (AIxIA 2023).
  https://www.aixia2023.cnr.it/call/workshops
- Member of the Organizing Committee - Demo Track Chair of the "The 17th Conference of the European Chapter of the Association for Computational Linguistics" (EACL2023).
  https://2023.eacl.org/committees/organization/
- Member of the Organizing Committee - Publication Chair of the "2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics" (NAACL2022).
  https://2022.naacl.org/committees/organization/
- Member of the Organizing Committee - Publication Chair of the "60th Annual Meeting of the Association for Computational Linguistics" (ACL2022).
  https://www.2022.aclweb.org/organisers
- Chair of the Fifth Workshop on Natural Language for Artificial Intelligence (NL4AI) co-located with the AIxIA 2021 conference.
  http://ceur-ws.org/Vol-3015
- Publication Chair and Program co-Chair (Area Vision, Robotics, Multimodal and Grounding with Raffaella Bernardi) in the "Eight Italian Conference on Computational Linguistics" (CLiC-it 2021).
  http://ceur-ws.org/Vol-3033/xpreface.pdf
- Chair of EVALITA 2020, the Italian campaign for the systematic evaluation of automatic processing systems for written and spoken texts in the Italian language.
  http://ceur-ws.org/Vol-2765
- Chair of the Fourth Workshop on Natural Language for Artificial Intelligence (NL4AI) co-located with the AIxIA 2020 conference.
  http://ceur-ws.org/Vol-2735
- Program co-Chair (Area Explainability of Deep Learning models for NLP with Aurelie Herbelot) in the "Sixth Italian Conference on Computational Linguistics".

http://ceur-ws.org/Vol-2481/preface.pdf
- Chair of the Second Workshop on Natural Language for Artificial Intelligence (NL4AI) co-located with the AIxIA 2018 conference.
  http://ceur-ws.org/Vol-2244
- Program co-Chair (Area Machine Learning for Natural Language Processing) in the "Fifth Italian Conference on Computational Linguistics" (2018).
  http://ceur-ws.org/Vol-2253/preface.pdf
- Task Organizer: ABSITA: "Aspect-based Sentiment Analysis at EVALITA", [70]. The task is part of EVALITA, a national campaign for the systematic evaluation of automatic Natural Language Processing systems in Italian (2018).
  http://sag.art.uniroma2.it/absita/
- Chair of the First Workshop on Natural Language for Artificial Intelligence (NL4AI) co-located with the AIxIA 2017 conference. Editor of the Workshop Proceedings published by CEUR (2017).
  http://ceur-ws.org/Vol-1983
- Head of the Local Organizing Committee of the "Fourth Italian Conference on Computational Linguistics" (2017).
  http://sag.art.uniroma2.it/clic2017
- Program co-Chair (Area NLP for Web and Social Media, with Felice dell'Orletta) of the "Third Italian Conference on Computational Linguistics" (2016).
  http://ceur-ws.org/Vol-1749/preface.pdf
- Task Organizer: SENTIPOLC: "Sentiment Polarity Classification in Twitter" [41]. The task is part of EVALITA, a national campaign for the systematic evaluation of automatic NLP systems in Italian (2016).
  http://www.di.unito.it/~tutreeb/sentipolc-evalita16/
- Head of the Local Committee of the Italian Workshop on Information Retrieval, Roma, 2014.
  http://iir2014.uniroma2.it/

3.2 Program committee member for Conferences and Workshops

Member of the Program Committee of the main International and National Conferences in the area of Artificial Intelligence, Natural Language Processing, Machine Learning and Information Retrieval, such as:

- Conference on Artificial Intelligence (since AAAI 2015)
- Annual Meeting of the Association for Computational Linguistics (since ACL 2013)
- International Joint Conference on Artificial Intelligence (since IJCAI 2013)
- European Conference on Artificial Intelligence (since ECAI 2015)
- The International Conference on Computational Linguistics (since COLING 2014)
- Conference on Empirical Methods for Natural Language Processing (since EMNLP 2012)
- Conference on Computational Natural Language Learning (since CONLL 2015)
- The European Conference on Computational Linguistics (since EACL 2014)
- North Chapter of the Association for Computational Linguistics (since NAACL 2012)
- International Joint Conference on Natural Language Processing (since IJCNLP 2013)
- Conference on Lexical and Computational Semantics (since *SEM 2015)
- IEEE International Conference on Robot and Human Interactive Communication (since RO-MAN 2019)
- European Semantic Web Conference (since ESWC 2015)
- International Conference on Natural Language & Information Systems (since NLDB 2022)

- SIGIR Workshop on Semantic Matching in Information Retrieval (since SMIR 2014)
- International Conference on Language Resources and Evaluation (since LREC 2014)
- Italian Conference on Computational Linguistics (since CLIC-it 2014)
- Italian Information Retrieval Workshop (since IIR 2014)

3.3 Editorial Boards and Journal Reviewer

Guest editor of the Journal "Sensors" (Impact Factor 3.576) in the Special Issue "Deep Learning for Healthcare: Review, Opportunities and Challenges" (2022).

`https://www.mdpi.com/journal/sensors/special_issues/Deep_Learning_Healthcare_Sensors`

Member of the Editorial Board of the journal *Transactions of the Association for Computational Linguistics (TACL)*, published by the Association for Computational Linguistics (ACL), since 2018.

Technical Head of the Editorial Office of the "Italian Journal of Computational Linguistics", Academia Press.

`http://www.aaccademia.it/elenco-libri?aaidriv=3`

Peer-reviewer for journals in the area of Artificial Intelligence, Natural Language Processing, Machine Learning ed Information Retrieval:

- IEEE Transactions on Neural Networks and Learning Systems
- IEEE Transactions on Affective Computing
- Transactions of the Association for Computational Linguistics
- Information Retrieval Journal
- PLOS ONE
- Computational Intelligence and Neuroscience
- Knowledge-Based Systems Journal
- Information Processing and Management
- Intelligent Service Robotics
- Language Resources and Evaluation
- Journal of Web Semantics
- Applied Ontology
- MDPI Algorithms and MDPI Information
- Engineering Science and Technology, an International Journal
- Computers and Electrical Engineering
- Pattern Recognition Letters
- Expert Systems with Applications
- Italian Journal of Computational Linguistics

3.4 Software and Resources

GroundedSRL4HRI: Framework for *Grounded Semantic Role Labeling* (G-SRL) in Human-Robot Interaction. Introduces multimodal models for grounded semantic understanding and a synthetic data generation pipeline for domestic environments, based on linguistic and environmental constraints. The system combines diffusion-based image generation, automatic annotation, and MiniCPM-V training for situated command interpretation in robotics [9].

- `https://github.com/crux82/GroundedSRL4HRI`

Sanskrit Voyager: Unified web platform for interactive reading and linguistic analysis of Sanskrit texts. Integrates dictionary lookup, real-time text analysis, corpus search, and interactive reading for over 900 Sanskrit texts. Presented at the EMNLP 2025 System Demonstrations track [8].

- `https://www.sanskritvoyager.com/`
- Demo video: `https://www.youtube.com/watch?v=FCKlW4NKJEc`

BioASQ2025-UNITOR: Modular Biomedical Question Answering pipeline developed for the BioASQ 13b Challenge (CLEF 2025). The system integrates retrieval-augmented generation, synthetic snippet generation, and multi-task answer modeling to improve factual reliability and evidence traceability in biomedical QA [12]. Ranked among the top systems for Factoid and Ideal Answer tasks at BioASQ 2025.

- `https://github.com/crux82/BioASQ2025-UNITOR`

BioASQ2025 Benchmark Dataset: Curated datasets for training LLMs in biomedical information retrieval, snippet extraction, and multi-type answer generation. Built from the BioASQ 13b collection, the benchmark supports fine-tuning and evaluation of domain-adapted LLMs for evidence-based QA [12].

- Included within `https://github.com/crux82/BioASQ2025-UNITOR`

WikiGame-LLM-Eval: Experimental pipeline for evaluating *Large Language Models* (LLMs) on Wikipedia graph navigation, accompanying the paper "Evaluating Large Language Models on Wikipedia Graph Navigation: Insights from the WikiGame" [13]. The project provides a reproducible benchmark assessing models' ability in *multi-hop reasoning*, structural planning, and adherence to the real hyperlink graph of Wikipedia.

- `https://github.com/crux82/wikigame-llm-eval`

MM-IGLU-Dialogues: Repository accompanying the ACL Findings 2025 paper "Training Multi-Modal LLMs through Dialogue Planning for HRI" [15]. It provides code, data, and models for multimodal dialogue planning in Minecraft-like environments, enabling grounded interaction between a human "Architect" and an LLM-based "Builder." The system demonstrates how structured dialogue planning improves task executability and interaction clarity.

- `https://github.com/crux82/MM-IGLU-Dialogues`

BacKGen: Background Knowledge Generator — toolkit for augmenting textual datasets with retrieved knowledge snippets. Designed for data enrichment in retrieval-augmented generation (RAG) pipelines [14].

- `https://github.com/crux82/BacKGen`

MM-IGLU: Development of a benchmark and model suite for multimodal grounded language understanding in 3D environments. Includes BART models and multimodal architectures based on CLIP and LLaMA2-Chat-13B via LLaVA. The system supports command execution and interactive clarification generation in virtual or robotic contexts [24].

- `https://github.com/crux82/MM-IGLU`

MM-IGLU-IT: Italian version of the MM-IGLU system, with annotated datasets and fine-tuned models for natural command following in 3D multimodal environments. It is the first large-scale benchmark for grounded instruction following in Italian [19].

- `https://github.com/crux82/MM-IGLU-IT`

ExtremITA: Instruction-tuned LLM for Italian based on LLaMA and adapted via LoRA. Ranked first in 9 out of 22 tasks at EVALITA 2023. Highly modular, it supports sustainable fine-tuning via PEFT libraries [29].

– https://github.com/crux82/ExtremITA

U-DepPLLaMA: Framework for universal dependency parsing using autoregressive LLMs (LLaMA2). It casts parsing as a sequential task, achieving SOTA results in 26 languages without relying on dedicated architectures [23].

– https://github.com/crux82/u-deppllama

GrUT: Neural model for language understanding in Human-Robot Interaction contexts. Generates structured semantic representations (Frame Semantics) linked to the robot's internal knowledge, improving execution of complex spoken commands [37, 38].

– https://github.com/crux82/grut

FEVER-it: Full fact-checking system for Italian, based on Wikipedia. Includes a pipeline for evidence retrieval and NLI models to validate or refute claims in natural language [25].

– https://github.com/crux82/FEVER-it

EthicalNN: PyTorch-based neural architecture for integrating ethical decision-making into machine learning models. Supports learning guided by explicit moral principles ("truth-makers"), applied to sensitive datasets on fairness and bias [39].

– https://github.com/crux82/nn-ebd

KeLP: Kernel-based Learning Platform. I am one of the founders of the Kernel-based Learning Platform (KeLP) project, whose aim is the development of an open-source machine learning framework [74]. In particular, the aim of KeLP is to support the research within Kernel-based methods. KeLP in fact decouples learning algorithms from kernel functions, so as to encourage the development of new algorithms and/or kernels by inheriting each other functionality. KeLP has been entirely developed in Java and is available under the Apache 2.0 License.

– http://www.kelp-ml.org

GAN-BERT. Definition and implementation of a neural model for few-shot learning capable of applying Adversarial Learning methods to Transformer-based architectures. This activity is a collaboration with Amazon Seattle (Alexa Retail group) [55].

– https://github.com/crux82/ganbert
– https://github.com/crux82/ganbert-pytorch
– https://github.com/crux82/mt-ganbert

dats: Data augmentation toolkit for NLP. Provides a collection of augmentation strategies for text classification and sequence labeling tasks.

– https://github.com/crux82/dats

ACLPUB2. Responsible for the ACLPUB2 project with Ryan Cotterell (ETH Zürich & University of Cambridge) to develop a tool for generating Proceedings of all events organized by the American Association of Computational Linguistics (ACL).

– https://github.com/rycolab/aclpub2
– https://github.com/acl-org/aclpubcheck

GQA-IT: A large-scale Italian language dataset for training Deep Learning methods for Visual Question Answering. GQA-it contains more than 1 million question/answer pairs in Italian on 80K images. This resource is the result of the collaboration with the University of Pisa. [46].

– https://github.com/crux82/gqa-it

mscoco-it: A large-scale dataset in Italian for training Deep Learning methods for Automatic Image Captioning. The dataset contains more than 600,000 image/caption pairs. [68].

– https://github.com/crux82/mscoco-it

msr-vtt-it: A large-scale dataset in Italian for training Deep Learning methods for automatic video caption generation (*Automatic Video Captioning*). The dataset contains more than 200,000 video/caption pairs. [68].

– https://github.com/crux82/msr-vtt-it

SQUAD-IT. A large-scale dataset in Italian for the training and the systematic evaluation of Deep Learning methods for document-based Question Answering. The dataset contains more than 60,000 question/text/answer triples. [67, 73].

– https://github.com/crux82/squad-it

ABSITA. A large-scale Italian dataset for the training and the systematic evaluation of classification methods for the recognition of Opinions in touristic reviews. It is composed of more than 10,000 reviews. This dataset was used in the ABSITA competition organized in the context of EVALITA 2018. [70].

– http://sag.art.uniroma2.it/absita/

SENTIPOLC. A dataset for training classification methods for Sentiment recognition in microblog texts composed of more than 10,000 manually annotated tweets in Italian. This dataset has been used in the SENTIPOLC competition organized in the framework of EVALITA 2016. [41, 86].

– http://www.di.unito.it/~tutreeb/sentipolc-evalita16/

LU4R. Technical and scientific head of the team devoted to the development of an "adaptive spoken Language Understanding chain For Robots tool" (LU4R): a chain of the automatic interpretation of spoken commands to robots. This system is the result of the collaboration between the SAG group of the University of Rome, “Tor Vergata", and the Laboratory Ro.Co.Co. (Robot Cognitivi Cooperanti) of La Sapienza, University of Rome. The system is free to download via the Web and can be installed on a vast plethora of robotic architectures. LU4R has recently been made available to groups belonging to Robocup.

– http://sag.art.uniroma2.it/lu4r.html

3.5 Research Networks

– International collaboration since 2020 with Amazon Seattle (Alexa Retail group) to develop semi-supervised methods for automatic text classification based on adversarial learning paradigms. The results of this collaboration have been published in [55].
– Collaboration since 2021 with the University of Pisa for the study of Visual Question Answering methods. This collaboration has led to the construction of the first dataset for the training and evaluation of neural methods for VQA in the Italian language. [46].
– Collaboration from 2021 with ABI Lab to study neural methods for the automatic semantic enrichment of financial regulatory documents and study of methods for automatic generation of discrete structures reflecting banks’ regulatory processes [47].
– International collaboration with the Qatar Computing Research Institute (QCRI). I am one of the founders of the "Kernel-based Learning Platform" (KeLP) project whose aim is the development of a framework to support the implementation of machine learning systems focusing on Kernel-based learning. KeLP is the result of the international collaboration between the Semantic Analytics Group of the University of Rome “Tor Vergata” and the Qatar Computing Research Institute.
www.kelp-ml.org

- Responsible for hosting the portal of the Italian Association of Computational Linguistics (AILC) http://www.ai-lc.it/
- Collaboration since 2013 with the Cognitive Robotics Laboratories of the Sapienza University, for the study of robotic interfaces based on natural language for Human Robot Interaction. Scientific collaboration with the University of Rome, La Sapienza. I am one of the creators of LU4R, "adaptive spoken Language Understanding chain For Robots tool": a processing chain for the automatic interpretation of robotic spoken commands. LU4R is the result of the national collaboration between the Semantic Analytics Group of the University of Rome "Tor Vergata", and the Laboratory Ro.Co.Co. (Robot Cognitivi Cooperanti) of La Sapienza, University of Rome.
  http://sag.art.uniroma2.it/lu4r.html
  https://github.com/crux82/huric
- Member of the iFrame group, which involves the University of Trento, the Institute of Computational Linguistics of the CNR, the University of Pisa and the University of Bologna, for the development of a semantic resource based on the Frame Semantics theory for the Italian language.
- Collaboration with the Computational Linguistics and Computer Science Department (University of Colorado) and the University of Trento for the study of Kernel-based methods for the automatic recognition of verbal classes in texts (2011).

4 Teaching and Academic Supervision

4.1 Current Teaching Responsibilities

- **Machine Learning** — Master's Degree in Computer Science, University of Rome "Tor Vergata" (2025-present).
- **Information Retrieval** — Master's Degree in Computer Science, University of Rome "Tor Vergata" (2019-present).
- **Operating Systems and Computer Networks** — Bachelor's Degree in Computer Science, University of Rome "Tor Vergata" (2023-present).
- **Lecturer**, "Data Governance and Management", Second Level Master in *Data Science for Public Administration*, University of Rome "Tor Vergata" (2024-present).
  https://datasciencepa.uniroma2.it/docenti-master/

4.1.1 Previous Teaching Activities

- **Elements of Data Analytics** — Master's Degree in Management Engineering (2019-2023).
- **Java Programming for Mobile Devices** — Bachelor's Degree in Computer Science and Internet Engineering (2018-2025).
- **Teaching Assistantships:**
  - *Deep Learning* — Master's Degree in Computer Engineering and Computer Science (2022-present).
  - *Web Mining & Retrieval* — Master's Degree in Computer Engineering and Computer Science (2009-present).
  - *Database and Knowledge Management* — Bachelor's Degree in Computer Science, Computer and Management Engineering, and Internet Engineering (2012-present).
- **Teaching Assistant**, Second Level Master in *Big Data in Business* (2016-2017):
  Courses: *Text Mining and Document Analysis*, *Social Media Analysis and Recommendation Systems*.

4.2 Doctoral Programs

Faculty Member of the Ph.D. Program in Data Science, University of Rome "Tor Vergata".
`https://datasciencephd.uniroma2.it/collegio-docenti`

4.2.1 Doctoral Supervision and Co-Supervision

I have actively contributed to the supervision of several Ph.D. students in the fields of Natural Language Processing, Machine Learning, and Artificial Intelligence, both as main supervisor and co-tutor.

Completed Ph.D. Students:

- Simone Filice — Ph.D. in Computer Science, University of Rome "Tor Vergata" (2016)
- Giuseppe Castellucci — Ph.D. in Computer Science, University of Rome "Tor Vergata" (2016)
- Emanuele Bastianelli — Ph.D. in Computer Science, Sapienza University of Rome (2016)
- Andrea Vanzo — Ph.D. in Computer Science, Sapienza University of Rome (2018)
- Claudiu Daniel Hromei — Ph.D. in Artificial Intelligence (National Ph.D. Program in AI), University of Rome "Tor Vergata" (2025)

Ongoing Supervisions:

- Federico Borazio — Ph.D. in Data Science, University of Rome "Tor Vergata" (2^{nd} year)
- Shahid Iqbal Rai — Ph.D. in Data Science, University of Rome "Tor Vergata" (2^{nd} year)
- Seyed Alireza Mousavian Anaraki — Ph.D. in Data Science, University of Rome "Tor Vergata" (2^{nd} year)
- Natalia Pichierri — Ph.D. in Data Science, University of Rome "Tor Vergata" (2^{nd} year)
- Sergio José Peresson — Ph.D. in Data Science, University of Rome "Tor Vergata" (2^{nd} year)
- Giacomo De Luca — Ph.D. in Data Science, University of Rome "Tor Vergata" (1^{st} year)

4.3 Academic Events and Outreach

- **Local Organizer**, *Lectures on Computational Linguistics 2021* — annual school organized by the Italian Association for Computational Linguistics (AILC) on advanced topics in NLP for PhD students and young researchers.
  `https://www.ai-lc.it/en/lectures-2021/`

5 Research Projects

Co-founder and Academic Member of Reveal s.r.l., spin-off of the University of Rome "Tor Vergata" on language technologies and their applications in the Big Data scenarios. Reveal's customers include public institutions, banks (UniCredit, MPS), Italian and foreign industries (e.g., Aker Solutions, Norway). The solutions offered by Reveal are:

- RevNLT: the natural language toolkit for the linguistic processing and interpretation of unstructured data;
- Revealer: a suite for services for knowledge acquisition from texts, semantic indexing, natural language querying, automatic retrieval and semantic ranking;
- SentiRe: a distributed Service Oriented Architecture for Sentiment Analysis on the web, user-generated data streams and social networks.

Reveal is a member of Unindustria (Confindustria delle Imprese di Roma e del Lazio) since 2014, after being selected as a finalist, together with 14 other companies, for the 2014 Startup Award[4].

I have been involved in the following projects:

Academic Projects

- *Principal Investigator* of the project "Resources for Multimodal SEmanticS" (R4MSES), within the cascade call FAIR - Future Artificial Intelligence Research (PE00000013, SPOKE 5, CUP: B53C22003980006), funded by PNRR - NextGenerationEU (November 2024 - November 2025). The project is dedicated to the development of resources and frameworks for multimodal understanding and advanced human-machine interaction.
- *Principal Investigator* of the project "SHIELD – Secure Healthcare Information with Enhanced Learning and Data Integrity," a type B project funded at the Department of Enterprise Engineering (Tor Vergata) for a duration of 24 months (2024–2026). The project addresses the development and validation of advanced privacy-preserving machine learning techniques.
- *Technical lead* for the Tor Vergata unit in the project "Knowledge Integration for Nuclear Decommissioning", in collaboration with IAEA, aimed at document management and semantic indexing for transparency in nuclear decommissioning processes. (2024-2025)
- *Technical lead* for Tor Vergata in the project "CETERA" (*Progetto CETERA* MISE - CETERA - F/310151/05/X56 - CUP: B89J23001730005), focused on the development of LLM-based agents for education and digital learning. (2024-2026)
- *Scientific coordinator* for the Department of Enterprise Engineering, Tor Vergata, in collaboration with Aenduo s.r.l. (HomeRehab, PNRR), for the design of an AI-based virtual assistant supporting home rehabilitation for doctors and healthcare professionals. (2024-2025)
- *Participant* in the project "AI-driven Event Discovery: the ISS Epidemic Intelligence case" (2023-2024), in collaboration with the Istituto Superiore di Sanità. Development of LLM-based solutions for the detection and analysis of emerging epidemiological phenomena online, supporting public health initiatives.
- *Technical Manager* for the project "DECODE" (2022-2023) at the University of Rome Tor Vergata, in collaboration with ABILAB. The project focused on the application of Large Language Models for regulatory document analysis, supporting legal interpretation and compliance.
- *Technical and Scientific Manager* for the Department of Enterprise Engineering of the University of Rome, "Tor Vergata" in the "Deep2Net" project carried out at VEAS (a Norwegian company that manages the disposal of wastewater from Oslo and other Norwegian locations), for the application of neural methods for data-driven control of wastewater plants.
- *Scientific member* of the team participating to "*Semantic Search Engine: Enterprise Search and Process Management*" for the Department of Enterprise Engineering at the University of Rome "Tor Vergata", funded by UniCredit, Milan, December 2013-March 2015.
- *Technical Manager* of the Industrial project between the University of Rome "Tor Vergata" and Elettronica spa - Project: "ELT - Specific Emitter Verification" (2013).
- *Technical and Scientific Manager* for the FOXBIT unit, in the project DIVINO, coordinated by the Mastroberardino company, in the context of market intelligence activities in the food and wine domain, through the automation of Web Mining and Opinion Analysis processes as studied and developed in the context of Industria 2015. December 2009 - March 2014.
- *Scientific Manager* for the Roma Tor Vergata unit, in the project "*Progress-It*" (funded by the agency FILAS, FILAS-CR-2011-1089) from July 2012 to February 2014, for the engineering of an Enter-

[4] http://unirete.un-industria.it/2014/startup.php

prise Semantic Search solution dedicated to SMEs.
- *Scientific member* in the project European project *INSEARCH* (FP7-SME-2010-1, Research for the benefit of specific groups, GA n. 262491, December 2010-December 2012) for the study and development of an Enterprise Semantic Search system dedicated to Small and Medium Enterprises.
- *Technical Support* for the University of Rome *Tor Vergata*, in the project PRIN 2008, PARLI "Portale per l'Accesso alle Risorse Linguistiche per l'Italiano' (2009-2012).
- *Technical Manager* in the Industrial collaboration between the Department of Computer Science, Systems and Production of the University of Rome "Tor Vergata" and Elettronica spa - "design and implementation of algorithms for Automatic Recognition of sequence kernel-based radar emitters" (2008-2009).

Industrial Projects
- *Technical and Scientific Manager* for the unit of Reveal s.r.l. for the project "Opinion and Reputation Management", in collaboration with Target Reply for the definition, design and development of a system for Sentiment Analysis and Brand Reputation at ENEL (2022).
- *Technical and Scientific Manager* for the unit of Reveal s.r.l. for the project "ACI - Intelligent Image Processing License Plate Processing" for large-scale *quality assessment* of digitized documents at ACI (2021 - 2022).
- *Technical and Scientific Manager* for the unit of Reveal s.r.l. for the project "Healthcare Agents and Learning robots - HeAL9000", funded by the Lazio Region, under the "Progetti Strategici 2019 - Area di Specializzazione Scienze della Vita" (prot. A0320-2019-28108) for the design and development of a robotic physiotherapist (2020-2022).
- *Technical and Scientific Manager* for the unit of Reveal s.r.l. for the project "BAnking Semantic Search system" (BASS) for the design, development and deployment of a Semantic Search system for the internal regulatory documentation of the bank Monte dei Paschi di Siena (2021).
- *Technical and Scientific Manager* for the Reveal s.r.l. unit for the Project "*Re4CT - Revealer for Crime Tracking*", for the design, development and deployment of a Machine Learning and AI solution for the automatic analysis of texts, within the Italian Ministero degli Interno, Dipartimento della Pubblica Sicurezza (2020).
- *Technical and Scientific Manager* for the Reveal s.r.l. unit for the Project "*SIMOO - Surfing on the Map of Offers and Opportunities*": this project is in collaboration with the National Institute of Tourism Research (ISNART) for the definition, design and development of an intelligent system for the collection of tourism data on a national scale (2018).
- *Technical and Scientific Manager*, for the Reveal s.r.l., unit for the "*SARAI - Sentiment Analysis for RAI*" project, in collaboration with Target Reply from February to May 2018 for the definition, design and development of Sentiment Analysis system at RAI.
- *Technical and Scientific Manager*, for the Reveal s.r.l. unit, for the "*Internet for All (I4ALL)*" project, Horizon H2020 project - PON 2014/2020, in collaboration with Mediavoice srl and the University of Rome "La Sapienza" since January 2017 for the definition, design and development of an intelligent system to simplify the access to the Web by the visually impaired and the blind people.
- *Technical and Scientific Manager*, for the Reveal s.r.l. unit, for the "*Insideout*" project, in collaboration with the University of Bern and the University of Rome "La Sapienza" from January to December 2016 for the definition, design and development of an intelligent system for the analysis of adolescent behaviour on Social Networks.
- *Technical and Scientific Manager*, for the Reveal s.r.l. unit, for the "*Automated Systemic Support*" project, POR FESR Molise 2007-2013, since January 2015 for the definition, design and development of an intelligent system for the automation of Help Desk services.

– *Technical and Scientific Manager*, for the Reveal s.r.l., in the "*Aker Semantic Search*' project funded by Aker Solutions, Oslo Norway (2013-2015), for the definition and development of an Enterprise Semantic Search system dedicated to systems engineering for oil extraction in an underwater environment.

Publications

[1] Danilo Croce, Jochen L. Leidner, and Nafise Sadat Moosavi, editors. *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2026 - Volume 3: System Demonstrations, Rabat, Marocco, March 24-29, 2026*. Association for Computational Linguistics, 2026.

[2] Federico Borazio, Seyed Alireza Mousavian Anaraki, Shahid Iqbal Rai, Danilo Croce, and Roberto Basili. Unitor at PFB: constrained agentic prompting and self-consistency for financial multi-choice QA. In Francesco Cutugno, Alessio Miaschi, Alessio Palmero Aprosio, Giulia Rambelli, Lucia Siciliani, and Marco Antonio Stranisci, editors, *Proceedings of the 9th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2026), Bari, Italy, February 26th-27th, 2026*, CEUR Workshop Proceedings. CEUR-WS.org, 2026.

[3] Federico Borazio, Francesco Labbate, Danilo Croce, and Roberto Basili. Integrating AI and IR paradigms for sustainable and trustworthy accurate access to large scale biomedical information. In Ricardo Campos, Adam Jatowt, Yanyan Lan, Mohammad Aliannejadi, Christine Bauer, Sean MacAvaney, Avishek Anand, Zhaochun Ren, Suzan Verberne, Nan Bai, and Masoud Mansoury, editors, *Advances in Information Retrieval - 48th European Conference on Information Retrieval, ECIR 2026, Delft, The Netherlands, March 29 - April 2, 2026, Proceedings, Part III*, Lecture Notes in Computer Science, pages 398–412. Springer, 2026.

[4] Federico Borazio, Giacomo De Luca, Daniele Pasquini, Danilo Croce, and Roberto Basili. Unitor at desegma-it: Analyzing supervision and encoder representations for italian machine-generated text detection. In Francesco Cutugno, Alessio Miaschi, Alessio Palmero Aprosio, Giulia Rambelli, Lucia Siciliani, and Marco Antonio Stranisci, editors, *Proceedings of the 9th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2026), Bari, Italy, February 26th-27th, 2026*, CEUR Workshop Proceedings. CEUR-WS.org, 2026.

[5] Claudiu D. Hromei, Antonio Scaiella, Danilo Croce, and Roberto Basili. Unitor at EVWSD-ITA: zero-shot visual word sense disambiguation via visual question-answering. In Francesco Cutugno, Alessio Miaschi, Alessio Palmero Aprosio, Giulia Rambelli, Lucia Siciliani, and Marco Antonio Stranisci, editors, *Proceedings of the 9th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2026), Bari, Italy, February 26th-27th, 2026*, CEUR Workshop Proceedings. CEUR-WS.org, 2026.

[6] Sergio José Peresson, Danilo Croce, and Roberto Basili. Learning molecular structures from infrared spectra through latent evidence prediction. In Mitra Baratchi, Siegfried Nijssen, and Jan N. van Rijn, editors, *Advances in Intelligent Data Analysis XXIV - 24th International Symposium on Intelligent Data Analysis, IDA 2026, Leiden, The Netherlands, April 22-24, 2026, Proceedings*, Lecture Notes in Computer Science, pages 225–238. Springer, 2026.

[7] Andriy Shcherbakov, Danilo Croce, and Roberto Basili. Unitor at cruciverb-it: Retrieval-augmented two-step reasoning for italian crossword clue answering. In Francesco Cutugno, Alessio Miaschi, Alessio Palmero Aprosio, Giulia Rambelli, Lucia Siciliani, and Marco Antonio Stranisci, editors, *Proceedings of the 9th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2026), Bari, Italy, February 26th-27th, 2026*, CEUR Workshop Proceedings. CEUR-WS.org, 2026.

[8] Giacomo De Luca, Danilo Croce, and Roberto Basili. Sanskrit voyager: Unified web platform for interactive reading and linguistic analysis of Sanskrit texts. In Ivan Habernal, Peter Schulam, and Jörg Tiedemann, editors, *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 244–253, Suzhou, China, November 2025. Association for Computational Linguistics.

[9] Claudiu Daniel Hromei, Antonio Scaiella, Danilo Croce, and Roberto Basili. Grounded semantic role labelling from synthetic multimodal data for situated robot commands. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng, editors, *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 23758–23781, Suzhou, China, November 2025. Association for Computational Linguistics.

[10] Shahid Iqbal Rai, Danilo Croce, and Roberto Basili. Injecting frame semantics into large language models via prompt-based fine-tuning. In Lea Frermann and Mark Stevenson, editors, *Proceedings of the 14th Joint Conference on Lexical and Computational Semantics (*SEM 2025)*, pages 31–47, Suzhou, China, November 2025. Association for Computational Linguistics.

[11] Seyed Alireza Mousavian Anaraki, Danilo Croce, and Roberto Basili. Automatic gri–sdg annotation and llm-based filtering for sustainability reports. In Cristina Bosco, Elisabetta Jezek, Marco Polignano, and Manuela Sanguinetti, editors, *Proceedings of the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025)*, volume TBD, Cagliari, Italy, September 2025. CEUR Workshop Proceedings.

[12] Federico Borazio, Andriy Shcherbakov, Danilo Croce, and Roberto Basili. Unitor at bioasq 2025: Modular biomedical qa with synthetic snippets and multiple task answer generation. In Guglielmo Faggioli, Nicola Ferro, Allan Hanbury, and Evangelos Kanoulas, editors, *Working Notes of CLEF 2025 - Conference and Labs of the Evaluation Forum*, volume 4038, pages 165–177, Madrid, Spain, September 2025. CLEF Association, CEUR Workshop Proceedings.

[13] Daniele Margiotta, Danilo Croce, and Roberto Basili. Evaluating large language models on wikipedia graph navigation: Insights from the wikigame. In Cristina Bosco, Elisabetta Jezek, Marco Polignano, and Manuela Sanguinetti, editors, *Proceedings of the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025)*, volume TBD, Cagliari, Italy, September 2025. CEUR Workshop Proceedings.

[14] Muhammad Okky Ibrohim, Valerio Basile, Danilo Croce, Cristina Bosco, and Roberto Basili. Modeling background knowledge with frame semantics for fine-grained sentiment classification. In Giulia Rambelli, Filip Ilievski, Marianna Bolognesi, and Pia Sommerauer, editors, *Proceedings of the 2nd Workshop on Analogical Abstraction in Cognition, Perception, and Language (Analogy-Angle II)*, pages 22–36, Vienna, Austria, August 2025. Association for Computational Linguistics.

[15] Claudiu Daniel Hromei, Federico Borazio, Andrea Sensi, Elisa Passone, Danilo Croce, and Roberto Basili. Training multi-modal LLMs through dialogue planning for HRI. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, *Findings of the Association for Computational Linguistics: ACL 2025*, pages 16266–16284, Vienna, Austria, July 2025. Association for Computational Linguistics.

[16] Seyed Alireza Mousavian Anaraki, Danilo Croce, and Roberto Basili. Unsupervised sustainability report labeling based on the integration of the GRI and SDG standards. In Katherine Atwell, Laura Biester, Angana Borah, Daryna Dementieva, Oana Ignat, Neema Kotonya, Ziyi

Liu, Ruyuan Wan, Steven Wilson, and Jieyu Zhao, editors, *Proceedings of the Fourth Workshop on NLP for Positive Impact (NLP4PI)*, pages 151–162, Vienna, Austria, July 2025. Association for Computational Linguistics.

[17] Federico Borazio, Danilo Croce, and Roberto Basili. Adapting llms for domain-specific retrieval: A case study in nuclear safety. In Claudia Hauff, Craig Macdonald, Dietmar Jannach, Gabriella Kazai, Franco Maria Nardini, Fabio Pinelli, Fabrizio Silvestri, and Nicola Tonellotto, editors, *Advances in Information Retrieval - 47th European Conference on Information Retrieval, ECIR 2025, Lucca, Italy, April 6-10, 2025, Proceedings, Part V*, volume 15576 of *Lecture Notes in Computer Science*, pages 116–122. Springer, 2025.

[18] Giuseppe Attanasio, Pierpaolo Basile, Federico Borazio, Danilo Croce, Maria Francis, Jacopo Gili, Elio Musacchio, Malvina Nissim, Viviana Patti, Matteo Rinaldi, and Daniel Scalena. CALAMITA: challenge the abilities of language models in italian. In Felice Dell'Orletta, Alessandro Lenci, Simonetta Montemagni, and Rachele Sprugnoli, editors, *Proceedings of the Tenth Italian Conference on Computational Linguistics (CLiC-it 2024), Pisa, Italy, December 4-6, 2024*, volume 3878 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2024.

[19] Federico Borazio, Claudiu Daniel Hromei, Elisa Passone, Danilo Croce, and Roberto Basili. MM-IGLU-IT: multi-modal interactive grounded language understanding in italian. In Alessandro Artale, Gabriella Cortellessa, and Marco Montali, editors, *AIxIA 2024 - Advances in Artificial Intelligence - XXIIIrd International Conference of the Italian Association for Artificial Intelligence, AIxIA 2024, Bolzano, Italy, November 25-28, 2024, Proceedings*, volume 15450 of *Lecture Notes in Computer Science*, pages 64–78. Springer, 2024.

[20] Andrea Brunello and Danilo Croce. Special issue: Selected papers from the aixia 2023 workshops. *Intelligenza Artificiale*, 18(1):5–8, 2024.

[21] Danilo Croce, Artem Smirnov, Luigi Tiburzi, Serena Travaglini, Roberta Costa, Armando Calabrese, Roberto Basili, Nathan Levialdi Ghiron, and Gerry Melino. Ai-driven transcriptomic encoders: From explainable models to accurate, sample-independent cancer diagnostics. *Expert Syst. Appl.*, 258:125126, 2024.

[22] Claudiu D. Hromei, Danilo Croce, Rodolfo Delmonte, and Roberto Basili. La non canonica l'hai studiata? exploring llms and sentence canonicity in italian. In Felice Dell'Orletta, Alessandro Lenci, Simonetta Montemagni, and Rachele Sprugnoli, editors, *Proceedings of the Tenth Italian Conference on Computational Linguistics (CLiC-it 2024), Pisa, Italy, December 4-6, 2024*, volume 3878 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2024.

[23] Claudiu Daniel Hromei, Danilo Croce, and Roberto Basili. U-deppllama: Universal dependency parsing via auto-regressive large language models. *IJCoL*, 10(1):21–38, 2024.

[24] Claudiu Daniel Hromei, Daniele Margiotta, Danilo Croce, and Roberto Basili. MM-IGLU: multi-modal interactive grounded language understanding. In Nicoletta Calzolari, Min-Yen Kan, Véronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC/COLING 2024, 20-25 May, 2024, Torino, Italy*, pages 11440–11451. ELRA and ICCL, 2024.

[25] Antonio Scaiella, Stefano Costanzo, Elisa Passone, Danilo Croce, and Giorgio Gambosi. Leveraging large language models for fact verification in italian. In Felice Dell'Orletta, Alessandro Lenci, Simonetta Montemagni, and Rachele Sprugnoli, editors, *Proceedings of the Tenth Italian Conference*

on Computational Linguistics (CLiC-it 2024), Pisa, Italy, December 4-6, 2024, volume 3878 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2024.

[26] Antonio Scaiella, Daniele Margiotta, Claudiu Daniel Hromei, Danilo Croce, and Roberto Basili. Evaluating multimodal large language models for visual question-answering in italian. In Giovanni Bonetta, Claudiu Daniel Hromei, Lucia Siciliani, and Marco Antonio Stranisci, editors, *Proceedings of the Eight Workshop on Natural Language for Artificial Intelligence (NL4AI 2024) co-located with 23th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2024), Bolzano, Italy, November 26th-27th, 2024*, volume 3877 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2024.

[27] Danilo Croce and Luca Soldaini, editors. *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. EACL 2023 - System Demonstrations, Dubrovnik, Croatia, May 2-4, 2023*. Association for Computational Linguistics, 2023.

[28] Claudiu D. Hromei, Danilo Croce, Valerio Basile, and Roberto Basili. Extremita at EVALITA 2023: Multi-task sustainable scaling to large language models at its extreme. In Mirko Lai, Stefano Menini, Marco Polignano, Valentina Russo, Rachele Sprugnoli, and Giulia Venturi, editors, *Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023), Parma, Italy, September 7th-8th, 2023*, volume 3473 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2023.

[29] Claudiu D. Hromei, Danilo Croce, Valerio Basile, and Roberto Basili. Scaling large language models to the extreme: Neural semantic processing of multiple tasks in italian. In Roberto Basili, Domenico Lembo, Carla Limongelli, and Andrea Orlandini, editors, *AIxIA 2023 - Advances in Artificial Intelligence - XXIInd International Conference of the Italian Association for Artificial Intelligence, AIxIA 2023, Rome, Italy, November 6-9, 2023, Proceedings*, volume 14318 of *Lecture Notes in Computer Science*, pages 172–186. Springer, 2023.

[30] Claudiu D. Hromei, Danilo Croce, and Roberto Basili. Grounding end-to-end pre-trained architectures for semantic role labeling in multiple languages. *Intelligenza Artificiale*, 17(2):173–191, 2023.

[31] Claudiu Daniel Hromei, Danilo Croce, and Roberto Basili. End-to-end dependency parsing via auto-regressive large language model. In Federico Boschetti, Gianluca E. Lebani, Bernardo Magnini, and Nicole Novielli, editors, *Proceedings of the 9th Italian Conference on Computational Linguistics, Venice, Italy, November 30 - December 2, 2023*, volume 3596 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2023.

[32] Claudiu Daniel Hromei, Daniele Margiotta, Danilo Croce, and Roberto Basili. An end-to-end transformer-based model for interactive grounded language understanding. In Elisa Bassignana, Dominique Brunato, Marco Polignano, and Alan Ramponi, editors, *Proceedings of the Seventh Workshop on Natural Language for Artificial Intelligence (NL4AI 2023) co-located with 22th International Conference of the Italian Association for Artificial Intelligence (AIxIA 2023), Rome, Italy, November 6th-7th, 2023*, volume 3551 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2023.

[33] Daniele Margiotta, Danilo Croce, and Roberto Basili. Taxosbert: Unsupervised taxonomy expansion through expressive semantic similarity. In Donatello Conte, Ana Fred, Oleg Gusikhin, and Carlo Sansone, editors, *Deep Learning Theory and Applications - 4th International Conference, DeLTA 2023, Rome, Italy, July 13-14, 2023, Proceedings*, volume 1875 of *Communications in Computer and Information Science*, pages 295–307. Springer, 2023.

[34] Daniele Margiotta, Danilo Croce, Marco Rotoloni, Barbara Cacciamani, and Roberto Basili. Business knowledge and neural learning: organisation-specific transformer via semantic pretraining. In Fabrizio Falchi, Fosca Giannotti, Anna Monreale, Chiara Boldrini, Salvatore Rinzivillo, and Sara Colantonio, editors, *Proceedings of the Italia Intelligenza Artificiale - Thematic Workshops co-located with the 3rd CINI National Lab AIIS Conference on Artificial Intelligence (Ital IA 2023), Pisa, Italy, May 29-30, 2023*, volume 3486 of *CEUR Workshop Proceedings*, pages 500–505. CEUR-WS.org, 2023.

[35] Christian Tamantini, Francesco Scotto di Luzio, Claudiu Daniel Hromei, Lorenzo Cristofori, Danilo Croce, Marco Cammisa, Arcangela Cristofaro, Maria Vittoria Marabello, Roberto Basili, and Loredana Zollo. Integrating physical and cognitive interaction capabilities in a robot-aided rehabilitation platform. *IEEE Syst. J.*, 17(4):6516–6527, 2023.

[36] Danilo Croce, Simone Filice, Giuseppe Castellucci, and Roberto Basili. Learning to generate examples for semantic processing tasks. In Marine Carpuat, Marie-Catherine de Marneffe, and Iván Vladimir Meza Ruíz, editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022, Seattle, WA, United States, July 10-15, 2022*, pages 4587–4601. Association for Computational Linguistics, 2022.

[37] Claudiu Daniel Hromei, Lorenzo Cristofori, Danilo Croce, and Roberto Basili. Embedding contextual information in seq2seq models for grounded semantic role labeling. In Agostino Dovier, Angelo Montanari, and Andrea Orlandini, editors, *AIxIA 2022 - Advances in Artificial Intelligence - XXIst International Conference of the Italian Association for Artificial Intelligence, AIxIA 2022, Udine, Italy, November 28 - December 2, 2022, Proceedings*, volume 13796 of *Lecture Notes in Computer Science*, pages 472–485. Springer, 2022.

[38] Claudiu Daniel Hromei, Danilo Croce, and Roberto Basili. Grounding end-to-end architectures for semantic role labeling in human robot interaction. In Debora Nozza, Lucia C. Passaro, and Marco Polignano, editors, *Proceedings of the Sixth Workshop on Natural Language for Artificial Intelligence (NL4AI 2022) co-located with 21th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2022), Udine, November 30th, 2022*, volume 3287 of *CEUR Workshop Proceedings*, pages 24–38. CEUR-WS.org, 2022.

[39] Luca Squadrone, Danilo Croce, and Roberto Basili. Ethics by design for intelligent and sustainable adaptive systems. In Agostino Dovier, Angelo Montanari, and Andrea Orlandini, editors, *AIxIA 2022 - Advances in Artificial Intelligence - XXIst International Conference of the Italian Association for Artificial Intelligence, AIxIA 2022, Udine, Italy, November 28 - December 2, 2022, Proceedings*, volume 13796 of *Lecture Notes in Computer Science*, pages 154–167. Springer, 2022.

[40] Elena Cabrio, Danilo Croce, Lucia C. Passaro, and Rachele Sprugnoli, editors. *Proceedings of the Fifth Workshop on Natural Language for Artificial Intelligence (NL4AI 2021) co-located with 20th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2021), Online event, November 29, 2021*, volume 3015 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2021.

[41] Valerio Basile, Nicole Novielli, Danilo Croce, Francesco Barbieri, Malvina Nissim, and Viviana Patti. Sentiment polarity classification at EVALITA: lessons learned and open challenges. *IEEE Trans. Affect. Comput.*, 12(2):466–478, 2021.

[42] Claudia Breazzano, Danilo Croce, and Roberto Basili. MT-GAN-BERT: multi-task and generative adversarial learning for sustainable language processing. In Elena Cabrio, Danilo Croce, Lucia C.

Passaro, and Rachele Sprugnoli, editors, *Proceedings of the Fifth Workshop on Natural Language for Artificial Intelligence (NL4AI 2021) co-located with 20th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2021), Online event, November 29, 2021*, volume 3015 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2021.

[43] Claudia Breazzano, Danilo Croce, and Roberto Basili. Multi-task and generative adversarial learning for robust and sustainable text classification. In Stefania Bandini, Francesca Gasparini, Viviana Mascardi, Matteo Palmonari, and Giuseppe Vizzari, editors, *AIxIA 2021 - Advances in Artificial Intelligence - 20th International Conference of the Italian Association for Artificial Intelligence, Virtual Event, December 1-3, 2021, Revised Selected Papers*, volume 13196 of *Lecture Notes in Computer Science*, pages 228–244. Springer, 2021.

[44] Giuseppe Castellucci, Simone Filice, Danilo Croce, and Roberto Basili. Learning to solve NLP tasks in an incremental number of languages. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 2: Short Papers), Virtual Event, August 1-6, 2021*, pages 837–847. Association for Computational Linguistics, 2021.

[45] Lorenzo Cristofori, Claudiu D. Hromei, Francesco Scotto di Luzio, Christian Tamantini, Francesca Cordella, Danilo Croce, Loredana Zollo, and Roberto Basili. Heal9000: an intelligent rehabilitation robot. In Peter Lucas and Fabio Stella, editors, *Proceedings of the Workshop on Towards Smarter Health Care: Can Artificial Intelligence Help? co-located with 20th International Conference of the Italian Association for Artificial Intelligence (AIxIA2021), Anywhere, November 29th, 2021*, volume 3060 of *CEUR Workshop Proceedings*, pages 29–41. CEUR-WS.org, 2021.

[46] Danilo Croce, Lucia C. Passaro, Alessandro Lenci, and Roberto Basili. Gqa-it: Italian question answering on image scene graphs. In Elisabetta Fersini, Marco Passarotti, and Viviana Patti, editors, *Proceedings of the Eighth Italian Conference on Computational Linguistics, CLiC-it 2021, Milan, Italy, January 26-28, 2022*, volume 3033 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2021.

[47] Daniele Margiotta, Danilo Croce, Marco Rotoloni, Barbara Cacciamani, and Roberto Basili. Knowledge-based neural pre-training for intelligent document management. In Stefania Bandini, Francesca Gasparini, Viviana Mascardi, Matteo Palmonari, and Giuseppe Vizzari, editors, *AIxIA 2021 - Advances in Artificial Intelligence - 20th International Conference of the Italian Association for Artificial Intelligence, Virtual Event, December 1-3, 2021, Revised Selected Papers*, volume 13196 of *Lecture Notes in Computer Science*, pages 564–579. Springer, 2021.

[48] F. Fiori Nastro, D. Croce, S. Schmidt, R. Basili, and F. Schultze-Lutter. Insideout project: Using big data and machine learning for prevention in psychiatry. *European Psychiatry*, 64(S1):S343–S343, 2021.

[49] Pierpaolo Basile, Valerio Basile, Danilo Croce, and Elena Cabrio, editors. *Proceedings of the 4th Workshop on Natural Language for Artificial Intelligence (NL4AI 2020) co-located with the 19th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2020), Anywhere, November 25th-27th, 2020*, volume 2735 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020.

[50] Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors. *Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020), Online event, December 17th, 2020*, volume 2765 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020.

[51] Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro. EVALITA 2020: Overview of the 7th evaluation campaign of natural language processing and speech tools for italian. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020), Online event, December 17th, 2020*, volume 2765 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020.

[52] Silvia Brambilla, Danilo Croce, Fabio Tamburini, and Roberto Basili. Automatic induction of framenet lexical units in italian. In Johanna Monti, Felice Dell'Orletta, and Fabio Tamburini, editors, *Proceedings of the Seventh Italian Conference on Computational Linguistics, CLiC-it 2020, Bologna, Italy, March 1-3, 2021*, volume 2769 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020.

[53] Claudia Breazzano, Edoardo Rubino, Danilo Croce, and Roberto Basili. UNITOR @ DANKMEME: combining convolutional models and transformer-based architectures for accurate MEME management. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020), Online event, December 17th, 2020*, volume 2765 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020.

[54] Danilo Croce, Giuseppe Castellucci, and Roberto Basili. Adversarial training for few-shot text classification. *Intelligenza Artificiale*, 14(2):201–214, 2020.

[55] Danilo Croce, Giuseppe Castellucci, and Roberto Basili. GAN-BERT: generative adversarial learning for robust text classification with a bunch of labeled examples. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 2114–2119. Association for Computational Linguistics, 2020.

[56] Simone Giorgioni, Marcello Politi, Samir Salman, Roberto Basili, and Danilo Croce. UNITOR @ sardistance2020: Combining transformer-based architectures and transfer learning for robust stance detection. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020), Online event, December 17th, 2020*, volume 2765 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020.

[57] Lucia C. Passaro, Maria Di Maro, Valerio Basile, and Danilo Croce. Lessons learned from evalita 2020 and thirteen years of evaluation of italian language technology. *Italian Journal of Computational Linguistics*, 2020.

[58] Daniele Rossini, Danilo Croce, Sara Mancini, Massimo Pellegrino, and Roberto Basili. Actionable ethics through neural learning. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 5537–5544. AAAI Press, 2020.

[59] Andrea Vanzo, Danilo Croce, Emanuele Bastianelli, Roberto Basili, and Daniele Nardi. Grounded language interpretation of robotic commands through structured learning. *Artificial Intelligence*, 278, 2020.

[60] Fabio Massimo Zanzotto, Viviana Bono, Paola Vocca, Andrea Santilli, Danilo Croce, Giorgio Gambosi, and Roberto Basili. Gasp! generating abstracts of scientific papers from abstracts of cited papers. *CoRR*, abs/2003.04996, 2020.

[61] Danilo Croce, Roberto Basili, Vincenzo Lombardo, and Eleonora Ceccaldi. Automatic recognition of narrative drama units: A structured learning approach. In *Proceedings of Text2Story - 2nd Workshop on Narrative Extraction From Texts, co-located with the 41st European Conference on Information Retrieval, Text2Story@ECIR 2019, Cologne, Germany, April 14th, 2019*, pages 81–88, 2019.

[62] Danilo Croce, Giorgio Brandi, and Roberto Basili. Deep bidirectional transformers for italian question answering. In *Proceedings of the Sixth Italian Conference on Computational Linguistics, Bari, Italy, November 13-15, 2019*, 2019.

[63] Danilo Croce, Giuseppe Castellucci, and Roberto Basili. Kernel-based generative adversarial networks for weakly supervised learning. In *AI*IA 2019 - Advances in Artificial Intelligence - XVIIIth International Conference of the Italian Association for Artificial Intelligence, Rende, Italy, November 19-22, 2019, Proceedings*, pages 336–347, 2019.

[64] Danilo Croce, Simone Filice, and Roberto Basili. Making sense of kernel spaces in neural learning. *Computer Speech & Language*, 58:51–75, 2019.

[65] Danilo Croce, Daniele Rossini, and Roberto Basili. Auditing deep learning processes through kernel-based explanatory models. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 4035–4044, 2019.

[66] Danilo Croce, Daniele Rossini, and Roberto Basili. Neural embeddings: accurate and readable inferences based on semantic kernels. *Natural Language Engineering*, 25(4):519–541, 2019.

[67] Danilo Croce, Alexandra Zelenanska, and Roberto Basili. Enabling deep learning for large scale question answering in italian. *Intelligenza Artificiale*, 13(1):49–61, 2019.

[68] Antonio Scaiella, Danilo Croce, and Roberto Basili. Large scale datasets for image and video captioning in italian. *Italian Journal of Computational Linguistics*, 2(5):49–60, 2019.

[69] Pierpaolo Basile, Valerio Basile, Danilo Croce, Felice Dell'Orletta, and Marco Guerini, editors. *Proceedings of the 2nd Workshop on Natural Language for Artificial Intelligence (NL4AI 2018) co-located with 17th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2018), Trento, Italy, November 22nd to 23rd, 2018*, volume 2244 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2018.

[70] Pierpaolo Basile, Valerio Basile, Danilo Croce, and Marco Polignano. Overview of the EVALITA 2018 aspect-based sentiment analysis task (ABSITA). In *Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018) co-located with the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), Turin, Italy, December 12-13, 2018*, 2018.

[71] Danilo Croce and Roberto Basili. A markovian kernel-based approach for italian speech act labeling. In *Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018) co-located with the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), Turin, Italy, December 12-13, 2018*, 2018.

[72] Danilo Croce, Daniele Rossini, and Roberto Basili. On the readability of deep learning models: the role of kernel-based deep architectures. In *Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), Torino, Italy, December 10-12, 2018*, 2018.

[73] Danilo Croce, Alexandra Zelenanska, and Roberto Basili. Neural learning for question answering in italian. In *AI*IA 2018 - Advances in Artificial Intelligence - XVIIth International Conference of the Italian Association for Artificial Intelligence, Trento, Italy, November 20-23, 2018, Proceedings*, pages 389–402, 2018.

[74] S. Filice, G. Castellucci, G.D.S. Martino, A. Moschitti, D. Croce, and R. Basili. Kelp: A kernel-based learning platform. *Journal of Machine Learning Research*, 18:1–5, 2018.

[75] Andrea Santilli, Danilo Croce, and Roberto Basili. A kernel-based approach for irony and sarcasm detection in italian. In *Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018) co-located with the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), Turin, Italy, December 12-13, 2018*, 2018.

[76] P. Basile, D. Croce, and M. Guerini. Introduction to the first workshop on natural language for artificial intelligence. In *Proceedings of the 1st Workshop on Natural Language for Artificial Intelligence*, volume 1983, pages 1–3, 2017.

[77] R. Basili, S. Brambilla, D. Croce, and F. Tamburini. Developing a large scale framenet for italian: The iframenet experience. In *Proceedings of the Forth Italian Conference on Computational Linguistics (CLiC-it 2017)*, volume 2006, 2017.

[78] R. Basili and D. Croce. Structured knowledge and kernel-based learning: The case of grounded spoken language learning in interactive robotics. In *Proceedings of the AI*IA Workshop on Deep Understanding and Reasoning: A Challenge for Next-generation Intelligent Agents (URANIA 2016)*, volume 1802, pages 63–68, 2017.

[79] R. Basili, D. Croce, and G. Castellucci. Dynamic polarity lexicon acquisition for advanced social media analytics. *International Journal of Engineering Business Management*, 9:1–18, 2017.

[80] Roberto Basili, Valentina Bellomaria, Niels Jonas Bugge, Danilo Croce, Francesco De Michele, Federico Fiori Nastro, Paolo Fiori Nastro, Chantal Michel, Stefanie Schmidt, and Frauke Schultze-Lutter. Monitoring adolescents' distress using social web data as a source: the insideout project. In *Proceedings of the Fourth Italian Conference on Computational Linguistics (CLiC-it 2017), Rome, Italy, December 11-13, 2017.*, 2017.

[81] E. Bastianelli, G. Castellucci, D. Croce, R. Basili, and D. Nardi. Structured learning for spoken language understanding in human-robot interaction. *International Journal of Robotics Research*, 36(5-7), 2017.

[82] D. Croce, S. Filice, G. Castellucci, and R. Basili. Deep learning in semantic kernel spaces. In *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, volume 1, pages 345–354, 2017.

[83] Danilo Croce, Simone Filice, and Roberto Basili. Effective and scalable kernel-based language learning via stratified nyström methods. *Intelligenza Artificiale*, 11(2):93–116, 2017.

[84] Danilo Croce, Simone Filice, and Roberto Basili. On the impact of linguistic information in kernel-based deep architectures. In *AI*IA 2017 Advances in Artificial Intelligence - XVIth International Conference of the Italian Association for Artificial Intelligence, Bari, Italy, November 14-17, 2017, Proceedings*, pages 359–371, 2017.

[85] C. Masotti, D. Croce, and R. Basili. Deep learning for automatic image captioning in poor training conditions. In *Proceedings of the Forth Italian Conference on Computational Linguistics (CLiC-it 2017)*, volume 2006, 2017.

[86] Francesco Barbieri, Valerio Basile, Danilo Croce, Malvina Nissim, Nicole Novielli, and Viviana Patti. Overview of the evalita 2016 sentiment polarity classification task. In *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) & Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016), Napoli, Italy, December 5-7, 2016*, 2016.

[87] E. Bastianelli, D. Croce, A. Vanzo, R. Basili, and D. Nardi. A discriminative approach to grounded spoken language understanding in interactive robotics. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2016)*, volume 2016-January, pages 2747–2753, 2016.

[88] G. Castellucci, D. Croce, and R. Basili. A language independent method for generating large scale polarity lexicons. In *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, pages 38–45, 2016.

[89] Giuseppe Castellucci, Danilo Croce, and Roberto Basili. Context-aware convolutional neural networks for twitter sentiment analysis in italian. In *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) & Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016), Napoli, Italy, December 5-7, 2016*, 2016.

[90] Giuseppe Castellucci, Danilo Croce, Diego De Cao, and Roberto Basili. User mood tracking for opinion analysis on twitter. In *AI*IA 2016: Advances in Artificial Intelligence - XVth International Conference of the Italian Association for Artificial Intelligence, Genova, Italy, November 29 - December 1, 2016, Proceedings*, pages 76–88, 2016.

[91] Danilo Croce and Roberto Basili. Large-scale kernel-based language learning through the ensemble nystrom methods. In *Advances in Information Retrieval - 38th European Conference on IR Research, ECIR 2016, Padua, Italy, March 20-23, 2016. Proceedings*, pages 100–112, 2016.

[92] Danilo Croce, Giuseppe Castellucci, and Roberto Basili. Injecting sentiment information in context-aware convolutional neural networks. In *Proceedings of the 7th Italian Information Retrieval Workshop, Venezia, Italy, May 30-31, 2016*, 2016.

[93] Danilo Croce, Simone Filice, and Roberto Basili. Nyström methods for efficient kernel-based methods for community question answering. In *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) & Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016), Napoli, Italy, December 5-7, 2016*, 2016.

[94] S. Filice, D. Croce, A. Moschitti, and R. Basili. Kelp at semeval-2016 task 3: Learning semantic relations between questions and answers. In *SemEval 2016 - 10th International Workshop on Semantic Evaluation, Proceedings*, pages 1116–1123, 2016.

[95] Andrea Vanzo, Danilo Croce, Roberto Basili, and Daniele Nardi. Context-aware spoken language understanding for human robot interaction. In *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) & Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016), Napoli, Italy, December 5-7, 2016*, 2016.

[96] Andrea Vanzo, Danilo Croce, Emanuele Bastianelli, Roberto Basili, and Daniele Nardi. Robust spoken language understanding for house service robots. *Polibits*, 54:11–16, 2016.

[97] Andrea Vanzo, Danilo Croce, Emanuele Bastianelli, Guglielmo Gemignani, Roberto Basili, and Daniele Nardi. Dialogue with robots to support symbiotic autonomy. In *Dialogues with Social Robots - Enablements, Analyses, and Evaluation, Seventh International Workshop on Spoken Dialogue Systems, IWSDS 2016, Saariselkä, Finland, January 13-16, 2016*, pages 331–342, 2016.

[98] Andrea Vanzo, Danilo Croce, Giuseppe Castellucci, Roberto Basili, and Daniele Nardi. Spoken language understanding for service robotics in italian. In *AI*IA 2016: Advances in Artificial Intelligence - XVth International Conference of the Italian Association for Artificial Intelligence, Genova, Italy, November 29 - December 1, 2016, Proceedings*, pages 477–489, 2016.

[99] Roberto Basili, Paolo Annesi, Giuseppe Castellucci, and Danilo Croce. A compositional perspective in convolution kernels. In *Proceedings of the 6th Italian Information Retrieval Workshop, Cagliari, Italy, May 25-26, 2015*, 2015.

[100] E. Bastianelli, D. Croce, R. Basili, and D. Nardi. Using semantic maps for robust natural language interaction with robots. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, volume 2015-January, pages 1393–1397, 2015.

[101] E. Bastianelli, L. Iocchi, D. Nardi, G. Castellucci, D. Croce, and R. Basili. Robocup@home spoken corpus: Using robotic competitions for gathering datasets. In *In Proceedings of the Robot Soccer World Cup (RoboCup 2015)*, volume 8992, pages 19–30, 2015.

[102] Emanuele Bastianelli, Danilo Croce, Roberto Basili, and Daniele Nardi. Using semantic models for robust natural language human robot interaction. In *AI*IA 2015, Advances in Artificial Intelligence - XIVth International Conference of the Italian Association for Artificial Intelligence, Ferrara, Italy, September 23-25, 2015, Proceedings*, pages 343–356, 2015.

[103] G. Castellucci, D. Croce, and R. Basili. Acquiring an italian polarity lexicon through distributional methods. In *Proceedings of the 6th Italian Information Retrieval Workshop (IIR 2015)*, volume 1404, 2015.

[104] G. Castellucci, D. Croce, and R. Basili. Bootstrapping large scale polarity lexicons through advanced distributional methods. *Proceedings of the 14th Conference of the Italian Association for Artificial Intelligence (AIxIA 2015)*, 9336 LNCS:329–342, 2015.

[105] Giuseppe Castellucci, Danilo Croce, and Roberto Basili. Acquiring a large scale polarity lexicon through unsupervised distributional methods. In *Natural Language Processing and Information Systems - 20th International Conference on Applications of Natural Language to Information Systems, NLDB 2015 Passau, Germany, June 17-19, 2015 Proceedings*, pages 73–86, 2015.

[106] Danilo Croce, Roberto Basili, and Alessandro Moschitti. Semantic tree kernels for statistical natural language learning. In *Harmonization and Development of Resources and Tools for Italian Natural Language Processing within the PARLI Project*, pages 93–113. Springer, 2015.

[107] Danilo Croce, Simone Filice, and Roberto Basili. Distributional models for lexical semantics: An investigation of different representations for natural language learning. In *Harmonization and Development of Resources and Tools for Italian Natural Language Processing within the PARLI Project*, pages 115–134. Springer, 2015.

[108] S. Filice, G. Castellucci, D. Croce, and R. Basili. Kelp: A kernel-based learning platform for natural language processing. In *ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for*

Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Proceedings of System Demonstrations, pages 19–24, 2015.

[109] S. Filice, D. Croce, and R. Basili. A stratified strategy for efficient kernel-based learning. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI 2015)*, volume 3, pages 2239–2245, 2015.

[110] P. Annesi, D. Croce, and R. Basili. Semantic compositionality in tree kernels. In *CIKM 2014 - Proceedings of the 2014 ACM International Conference on Information and Knowledge Management*, pages 1029–1038, 2014.

[111] Roberto Basili, Andrea Ciapetti, Danilo Croce, Valeria Marino, Paolo Salvatore, and Valerio Storch. Enabling enterprise semantic search through language technologies: the progressit experience. In *Proceedings of the 5th Italian Information Retrieval Workshop, Roma, Italy, January 20-21, 2014*, pages 51–62, 2014.

[112] E. Bastianelli, G. Castellucci, D. Croce, L. Iocchi, R. Basili, and D. Nardi. Huric: A human robot interaction corpus. In *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, pages 4519–4526, 2014.

[113] Emanuele Bastianelli, Giuseppe Castellucci, Danilo Croce, Roberto Basili, and Daniele Nardi. Effective and robust natural language understanding for human-robot interaction. In *ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014)*, pages 57–62, 2014.

[114] Giuseppe Castellucci, Danilo Croce, Diego De Cao, and Roberto Basili. A multiple kernel approach for twitter sentiment analysis in italian. In *Evalita 2014*, 2014.

[115] Giuseppe Castellucci, Simone Filice, Danilo Croce, and Roberto Basili. Unitor: Aspect based sentiment analysis with structured learning. *SemEval 2014*, page 761, 2014.

[116] Simone Filice, Giuseppe Castellucci, Danilo Croce, and Roberto Basili. Effective kernelized online learning in language processing tasks. In *Advances in Information Retrieval - 36th European Conference on IR Research, ECIR 2014, Amsterdam, The Netherlands, April 13-16, 2014. Proceedings*, pages 347–358, 2014.

[117] A. Vanzo, D. Croce, and R. Basili. A context-based model for sentiment analysis in twitter. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)*, pages 2345–2354, 2014.

[118] Andrea Vanzo, Giuseppe Castellucci, Danilo Croce, and Roberto Basili. A context based model for sentiment analysis in twitter for the italian language. In *First Italian Conference on Computational Linguistics*, 2014.

[119] E. Bastianelli, D. Croce, D. Nardi, and R. Basili. Unitor-hmm-tk: Structured kernel-based learning for spatial role labeling. In *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, volume 2, pages 573–579, 2013.

[120] Diego De Cao, Valerio Storch, Danilo Croce, and Roberto Basili. INSEARCH: A platform for enterprise semantic search. In *Proceedings of the 4th Italian Information Retrieval Workshop, Pisa, Italy, January 16-17, 2013*, pages 104–115, 2013.

[121] G. Castellucci, S. Filice, D. Croce, and R. Basili. Unitor: Combining syntactic and semantic kernels for twitter sentiment analysis. In *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, volume 2, pages 369–374, 2013.

[122] D. Croce, F. Garzoli, M. Montesi, D. De Cao, and R. Basili. Enabling advanced business intelligence in divino. In *Proceedings of the 7th International Workshop on Information Filtering and Retrieval co-located with the 13th Conference of the Italian Association for Artificial Intelligence (AI*IA 2013)*, volume 1109, pages 61–72, 2013.

[123] D. Croce, V. Storch, and R. Basili. Unitor-core typed: Combining text similarity and semantic filters through sv regression. In *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, volume 1, pages 59–65, 2013.

[124] Bastianelli Emanuele, Giuseppe Castellucci, Danilo Croce, and Roberto Basili. Textual inference and meaning representation in human robot interaction. In *Joint Symposium on Semantic Processing.*, page 65, 2013.

[125] S. Filice, G. Castellucci, D. Croce, and R. Basili. Robust language learning via efficient budgeted online algorithms. In *Proceedings - IEEE 13th International Conference on Data Mining Workshops, ICDMW 2013*, pages 913–920, 2013.

[126] S. Filice, D. Croce, R. Basili, and F.M. Zanzotto. Linear online learning over structured data with distributed tree kernels. In *Proceedings - 2013 12th International Conference on Machine Learning and Applications, ICMLA 2013*, volume 1, pages 123–128, 2013.

[127] Simone Filice, Danilo Croce, and Roberto Basili. A robust machine learning approach for signal separation and classification. In *Pattern Recognition and Image Analysis - 6th Iberian Conference, IbPRIA 2013, Funchal, Madeira, Portugal, June 5-7, 2013. Proceedings*, pages 749–757, 2013.

[128] F. Garzoli, D. Croce, M. Nardini, F. Ciambra, and R. Basili. Robust requirements analysis in complex systems through machine learning. *Communications in Computer and Information Science*, 379 CCIS:44–58, 2013.

[129] P. Annesi, V. Storch, D. Croce, and R. Basili. Algebraic compositional models for semantic similarity in ranking and clustering. In *Proceedings of the Italian Information Retrieval Workshop (IIR 2012)*, volume 835, pages 155–166, 2012.

[130] D. Croce, P. Annesi, V. Storch, and R. Basili. Unitor: Combining semantic text similarity functions through sv regression. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, volume 2, pages 597–602, 2012.

[131] D. Croce and R. Basili. Proceedings of the italian information retrieval workshop (iir 2012). In *CEUR Workshop Proceedings*, volume 835, pages 133–143, 2012.

[132] D. Croce, S. Filice, and R. Basili. Distributional models and lexical semantics in convolution kernels. *International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2012)*, 7181 LNCS(PART 1):336–348, 2012.

[133] D. Croce, A. Moschitti, R. Basili, and M. Palmer. Verb classification using distributional similarity in syntactic and semantic structures. In *50th Annual Meeting of the Association for Computational Linguistics, ACL 2012 - Proceedings of the Conference*, volume 1, pages 263–272, 2012.

[134] D. Croce, V. Storch, P. Annesi, and R. Basili. Distributional compositional semantics and text similarity. In *Proceedings - IEEE 6th International Conference on Semantic Computing, ICSC 2012*, pages 242–249, 2012.

[135] Danilo Croce, Giuseppe Castellucci, and Emanuele Bastianelli. Structured learning for semantic role labeling. *Intelligenza Artificiale*, 6(2):163–176, 2012.

[136] M. Nardini, F. Ciambra, F. Garzoli, D. Croce, D. De Cao, and R. Basili. Machine learning technologies for the requirements analysis in complex systems. In *22nd Annual International Symposium of the International Council on Systems Engineering, INCOSE 2012 and the 8th Biennial European Systems Engineering Conference 2012, EuSEC 2012*, volume 1, pages 372–386, 2012.

[137] R. Basili, C. Giannone, D. Croce, and C. Domeniconi. Latent topic models of surface syntactic information. *The Congress of the Italian Association for Artificial Intelligence (AIxIA 2011)*, 6934 LNAI:225–237, 2011.

[138] D. Croce and R. Basili. Structured learning for semantic role labeling. *The Congress of the Italian Association for Artificial Intelligence (AIxIA 2011)*, 6934 LNAI:238–249, 2011.

[139] D. Croce, A. Moschitti, and R. Basili. Semantic convolution kernels over dependency trees: Smoothed partial tree kernel. In *Proceedings of the 20th ACM international conference on Information and knowledge management (CIKM 2011)*, pages 2013–2016, 2011.

[140] D. Croce, A. Moschitti, and R. Basili. Structured lexical similarity via convolution kernels on dependency trees. In *EMNLP 2011 - Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pages 1034–1046, 2011.

[141] Danilo Croce, Emanuele Bastianelli, and Giuseppe Castellucci. Structured kernel-based learning for the frame labeling over italian texts. In *Evaluation of Natural Language and Speech Tools for Italian, International Workshop, EVALITA 2011, Rome, Italy, January 24-25, 2012, Revised Selected Papers*, pages 220–229, 2011.

[142] S. Prezioso, D. Croce, and F.M. Zanzotto. Reading what machines "think": A challenge for nanotechnology. *Journal of Computational and Theoretical Nanoscience*, 8(10):2066–2071, 2011.

[143] R. Basili, D. Croce, C. Giannone, and D. De Cao. Acquiring ie patterns through distributional lexical semantic models. *Proceedings of the 11th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2010)*, 6008 LNCS:512–524, 2010.

[144] D. Croce, C. Giannone, P. Annesi, and R. Basili. Towards open-domain semantic role labeling. In *ACL 2010 - 48th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 237–246, 2010.

[145] D. De Cao, D. Croce, and R. Basili. Extensive evaluation of a framenet-wordnet mapping resource. In *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, pages 2752–2757, 2010.

[146] F.M. Zanzotto and D. Croce. Comparing eeg/erp-like and fmri-like techniques for reading machine thoughts. *Proceedings of the International Conference on Brain Informatics (BI 2010)*, 6334 LNAI:133–144, 2010.

[147] R. Basili, D.D. Cao, D. Croce, B. Coppola, and A. Moschitti. Cross-language frame semantics transfer in bilingual corpora. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5449 LNCS:332–345, 2009.

[148] R. Basili, D. Croce, D. De Cao, and C. Giannone. Learning semantic roles for ontology patterns. In *Proceedings - 2009 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IAT Workshops 2009*, volume 3, pages 291–294, 2009.

[149] C. Giannone, D. Croce, and R. Basili. Semantic word spaces for robust role labeling. In *8th International Conference on Machine Learning and Applications, ICMLA 2009*, pages 261–266, 2009.

[150] C. Giannone, D. Croce, R. Basili, and D. De Cao. A robust geometric model for argument classification. *International Conference of the Italian Association for Artificial Intelligence (AIxIA 2009)*, 5883 LNAI:284–293, 2009.

[151] F.M. Zanzotto and D. Croce. Reading what machines "think". *Proceedings of the International Conference on Brain Informatics (BI 2009)*, 5819 LNAI:159–170, 2009.

[152] Diego De Cao, Danilo Croce, Marco Pennacchiotti, and Roberto Basili. Combining word sense and usage for modeling frame semantics. In *Proceedings of the 2008 Conference on Semantics in Text Processing*, pages 85–101. Association for Computational Linguistics, 2008.

[153] M. Pennacchiotti, D. De Cao, R. Basili, D. Croce, and M. Roth. Automatic induction of framenet lexical units. In *EMNLP 2008 - 2008 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference: A Meeting of SIGDAT, a Special Interest Group of the ACL*, pages 457–465, 2008.