Preface to the fifth Workshop on Natural Language for Artificial Intelligence (NL4AI)

Elena Cabrio[1], Danilo Croce[2], Lucia C. Passaro[3], and Rachele Sprugnoli[4]

[1] Université Côte d'Azur, Inria, CNRS, I3S, France
elena.cabrio@unice.fr
[2] Deptartment of Enterprise Engineering, University of Rome "Tor Vergata"
croce@info.uniroma2.it
[3] Department of Computer Science, University of Pisa, Pisa, Italy
lucia.passaro@unipi.it
[4] Università Cattolica del Sacro Cuore, Milano, Italy
rachele.sprugnoli@unicatt.it

Natural Language Processing (NLP) is an important research topic in Artificial Intelligence (AI), as it is the target of different scientific and industrial interests. Natural Language is at the crossroad of Learning, Knowledge Representation, and Cognitive Modeling. Several recent AI achievements have repeatedly shown their beneficial impact on complex inference tasks, with huge application perspectives in linguistic modeling, processing, and inferences. However, Natural Language Understanding is still a rich research topic, whose cross-fertilization spans a number of independent areas such as Cognitive Computing, Robotics as well as Human-Computer Interaction. For AI, Natural Languages are the research focus of paradigms and applications but, at the same time, they act as cornerstones of automation, autonomy, and learnability for most intelligent phenomena ranging from Vision to Planning and Social Behaviors. A reflection about such diverse and promising interactions is an important target for current AI studies, fully in the core mission of AI*IA. This workshop, supported by the Special Interest Group on NLP of AI*IA[5] and by the Italian Association of Computational Linguistics (AILC)[6], aims at providing a broad overview of recent activities in the eld of Human Language Technologies (HLT) in Italy.

In this context, the organization of NL4AI 2021 provided researchers with the opportunity to share experiences and insights about AI applications focused on NLP in several domains. The 2021 edition of NL4AI is co-located with the 20th International Conference of the Italian Association for Artificial Intelligence (AIxIA 2021), held online due to the COVID-19 pandemic. The program of the meeting is available on the official workshop website[7].

The call for papers attracted 12 submissions by 34 different authors from Italy (23), Germany (5), and France (4). After the review process, 10 of 12 papers were accepted for publication (acceptance rate 83%). Papers deal with

[5] https://sites.google.com/a/aixia.it/nlp/
[6] https://www.ai-lc.it/
[7] http://sag.art.uniroma2.it/NL4AI/

various languages, namely Italian, English, French and LIS (*Lingua Italiana dei Segni*, 'Italian Sign Language'). Moreover, different modalities have been taken into account, i.e. text, vision and speech.

Going into details, accepted papers address several topics from different perspectives. In the following, we provide a short overview of such works, grouping them by topics.

Many papers propose specific tools and applications related to AI and NLP. In particular, Benamar et al. introduce BERT-POS, a method that thanks to the concatenation of morpho-syntactic information into contextual embeddings, improves the semantic understanding of out-of-vocabulary words, like domain-specific data or misspellings. Barbara et al. describe an approach for relation linking leveraging on deep learning sequence-to-sequence models for performing natural language question answering over knowledge bases. Palmero Aprosio presents the latest version of Tint, an open-source NLP suite for Italian, based on the popular Stanford CoreNLP and including several modules for text processing. Coffrini et al. describe a system developed to play a word association game, namely "La Ghigliottina", proposed at EVALITA 2020.

Other contributions propose natural language applications as well, but focus on popular or emerging trends in NLP, that are AI for Society, Affective Computing and Bias in NLP.

Specifically, concerning AI for Society, Fazzinga et al. illustrate the integration of advanced NLP techniques in a dialogue system based on argumentation. The system aims to ensure privacy preserving interaction with conversational agents by automatically anonymizing and sanitizing user sentences. Breazzano et al. propose a new classifier architecture to perform multi-task generative adversarial learning with a methodology aimed at keeping the whole process sustainable in terms of the amount of training annotated data and the computational cost at classification time. In their paper, Fontana and Caligiore constitute an overview of the state-of-the-art in the area of Sign Language Recognition and Synthesis, addressing a topic strongly related to the societal challenge of improving inclusion of deaf citizens. This last paper introduces the topic of multimodality which has been addressed also by other authors.

In this context, Miaschi et al. propose a Transformer-based punctuation restoration model for Italian speech transcriptions, using a BERT-base model with several fine-tuning steps. The authors test the methodology with different training data and sizes for in both an in- and cross-domain scenario. Bondielli and Passaro propose a paper bringing together multimodality and affect. In particular, they challenged the Open-AI CLIP model to solve an emotion classification task on images. Evaluation is performed by comparing the results of a standard image classification task aimed at recognizing objects with a task aimed at recognizing subjective and emotive classes under zero-shot and fine-tuning settings. Another paper related to the topic of affective computing is the one by Fell et al., that focuses on the task of hate speech annotation. The paper describes the well-known topic of the bias of NLP systems. In this case, they

show that the subjective perception of hatred and abusive language may highly influence annotated data.

In addition to the oral presentation of the aforementioned 10 papers, we are delighted to have Professor Dirk Hovy (Università Bocconi, Milan, Italy) as keynote speaker with a talk titled "More than words – Integrating social factors into language modeling". The talk outlines several social dimensions that influence language use and how they affect NLP models and sheds light on the efforts that are already underway to incorporate them.

As a final remark, the program co-chairs would like to thank all the members of the Program Committee (listed below), as well as the organizers of the AI*IA 2021 Conference.