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

Debora Nozza[1], Lucia Passaro[2] and Marco Polignano[3]

[1]*Department of Computing Sciences, Bocconi University, Milan, Italy*
[2]*Department of Computer Science, University of Pisa, Pisa, Italy*
[3]*Department of Computer Science, University of Bari Aldo Moro, Italy*

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[1] and by the Italian Association of Computational Linguistics (AILC)[2], 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 [1] provided researchers with the opportunity to share experiences and insights about AI applications focused on NLP in several domains. The 2022 edition of NL4AI is co-located with the 21th International Conference of the Italian Association for Artificial Intelligence (AIxIA 2022), taking place on November 30th in Udine, Italy. The program of the meeting is available on the official workshop website[3]. We received 17 submissions, 13 of which were accepted after peer-review. In terms of topics, the contributions to the workshop range from pure NLP works to broader proposals bridging NLP with other AI applications.

The call for papers attracted 17 submissions by 52 different authors from Italy (44), UK (2), Algeria (4), and Germany (2). After the review process, 13 of 17 papers were accepted for publication (acceptance rate 76.47%). Going into details, accepted papers address several topics

[1]https://sites.google.com/a/aixia.it/nlp/
[2]https://www.ai-lc.it/
[3]http://sag.art.uniroma2.it/NL4AI/

from different perspectives. In the following, we provide a short overview of such works, grouping them by topics.

The majority of papers proposed specific AI approaches for NLP applications. Some authors focused on the task of fake news detection: La Barbera et al. [2] proposed a hybrid human-in-the-loop framework for fact-checking that relied on a combination of AI, crowdsourcing, and experts. Emotion and sentiment analysis was studied by Bellodi et al. [3] in the specific context of a novel collection of Italian anti-vaccination COVID-19 posts. Lucassen et al.[4] introduced a method for automatically extracting output probability distribution and correlating them with human uncertainty about the grounded interpretation of the question-answers pair. Basile [5] assessed the status of prompt-based learning applied to several text classification tasks in the Italian language. Amin et al.[6] employed natural language generation for data augmentation on logical inputs i.e., Discourse Representation Structures. Sucameli et al.[7] realized a framework for interfacing humans and databases, thus facilitating access in natural language to the data stored in databases. Bellan et al. [8] focused on the problem of extracting a Business Process Model from the textual content. Labruna and Magnini [9] presented a position paper arguing that automatic adaptation of training and test dialogues in conversational domains is key to simulating domain changes. Borghesi et al. [10] analyzed which linguistic and acoustic aspects of spoken language distinguish engagement potential in speech. De Ponte [11] proposes a model composed of three neural networks, able to directly map, without labels, visual elements collected from video sources into spoken utterances. Hromei et al. [12] proposed the application of a Transformer-based architecture that combines inputs with a linguistic description of the environment to improve natural language interactions between humans and robots. Papucci et al. [13] provided an extensive evaluation of the first text-to-text Italian Neural Language Model (IT5), also testing its performance in a few-shot learning scenario. Dusi et al. [14] visualized the presence of gender bias in the English base model of BERT through a novel weakly supervised approach, which only requires a list of gendered words that can be easily found in online lexical resources.

In addition to the oral presentation of the aforementioned 13 papers, we are delighted to have **Fabio Petroni** (Co-Founder & CTO at Samaya AI) as keynote speaker with a talk titled: *"Improving Wikipedia Verifiability with AI"*. Verifiability is a core content policy of Wikipedia: claims that are likely to be challenged need to be backed by citations. There are millions of articles available online and thousands of new articles are released each month. For this reason, finding relevant sources is a difficult task: many claims do not have any references that support them. Furthermore, even existing citations might not support a given claim or become obsolete once the original source is updated or deleted. Hence, maintaining and improving the quality of Wikipedia references is an important challenge and there is a pressing need for better tools to assist humans in this effort. In the talk, Fabio is going to present its recent research showing that the process of improving references can be tackled with the help of AI. The results indicate that an AI-based system could be used, in tandem with humans, to improve the verifiability of Wikipedia. More generally, he hopes that our work can be used to assist fact-checking efforts and increase the general trustworthiness of information online.

Moreover, **Giuseppe Attardi** will provide us with a concluding speech about the general trend and future challenges in the correlation between NLP and AI with a talk titled: *"Large Language Models are All You Need"*. LLMs are among the three scientific breakthroughs of Deep Learning applied to NLP in just ten years: word embeddings, attention, and prompt engineering. They have shown surprising effectiveness in all tasks and have also been adopted in other areas, including cross-cutting tasks such as generating images from text descriptions. However, LLMs have also raised perplexities, starting with doubts about the release of GPT-2 to Galactica's recent retirement. Their capabilities, which seem to grow with their size, have become a subject of study. We will discuss whether discrimination will continue to grow among groups of researchers capable of building them and about possible alternatives. Giuseppe Attardi is a full professor of Computer Science at the University of Pisa. He previously worked at MIT's AI Lab, Sony Paris Research Laboratory, ICSI Berkeley, and Yahoo Research Barcelona. He developed Omega, a precursor to ontology languages for the Web; CMM, the garbage collector used in Java; and DeSR, a grammar analyzer for several languages. He participated in the development of Arianna, the first Italian search engine, and iStella. He is the founder or partner of several startups, in Italy and Spain. He was involved in the implementation of the fiber optic networks of the JRC in Ispra, the University of Pisa, and the GARR national research network. He has promoted Internet deployment in Italy through the No TUT campaign to reduce network access costs. The main goal of his current research is to make computers capable of understanding human language, using Deep Learning techniques.

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 2022 Conference.

- Giuseppe Attanasio, Politecnico di Torino (Italy)
- Agnese Augello, ICAR - Istituto di Calcolo e Reti ad alte prestazioni Consiglio Nazionale delle Ricerche (Italy)
- Valerio Basile, University of Turin (Italy)
- Pierpaolo Basile, Dipartimento di Informatica - University of Bari (Italy)
- Alessandro Bondielli, Università di Pisa (Italy)
- Dominique Brunato, Institute of Computational Linguistics "A. Zampolli" (National Council of Research), Pisa (Italy)
- Pierluigi Cassotti, Università degli studi di Bari 'Aldo Moro' (Italy)
- Danilo Croce, Dept. of Enterprise Engineering - Univ. of Roma Tor Vergata (Italy)
- Marco de Gemmis, University of Bari Aldo Moro, Dept. of Computer Science (Italy)
- Pietro Dell'Oglio, University of Florence (Italy)
- Felice Dell'Orletta, Istituto di Linguistica Computazionale "Antonio Zampolli" - ILC CNR (Italy)
- Mauro Dragoni, Fondazione Bruno Kessler - FBK-IRST (Italy)
- Moreno La Quatra, Politecnico di Torino (Italy)
- Alessandro Lenci, University of Pisa (Italy)
- Alessio Miaschi, University of Pisa (Italy)
- Martina Miliani, Università di Pisa (Italy)

- Cataldo Musto, Dipartimento di Informatica - University of Bari (Italy)
- Fedelucio Narducci, Politecnico di Bari (Italy)
- Nicolò Penzo, Fondazione Bruno Kessler & University of Trento (Italy)
- Giulia Rambelli, University of Pisa (Italy)
- Andrea Amelio Ravelli, Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC) - CNR (Italy)
- Giovanni Semeraro, University of Bari (Italy)
- Lucia Siciliani, University of Bari (Italy)
- Rachele Sprugnoli, Università degli Studi di Parma (Italy)
- Irene Sucameli, Università di Pisa (Italy)
- Vincenzo Taccardi, University of Bari (Italy)

References

[1] C. Elena, C. Danilo, L. C. Passaro, S. Rachele, Preface to the fifth workshop on natural language for artificial intelligence (nl4ai), in: 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. URL: http://ceur-ws.org/Vol-3015/paper172.pdf.

[2] D. La Barbera, K. Roitero, S. Mizzaro, A hybrid human-in-the-loop framework for fact checking, in: D. Nozza, L. C. Passaro, M. Polignano (Eds.), 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), November 30, 2022, CEUR-WS.org, 2022.

[3] E. Bellodi, A. Bertagnon, M. Gavanelli, Comparing emotion and sentiment analysis tools on italian anti-vaccination for covid-19 posts, in: D. Nozza, L. C. Passaro, M. Polignano (Eds.), 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), November 30, 2022, CEUR-WS.org, 2022.

[4] A. J. Lucassen, A. Testoni, R. Bernardi, Probability distributions as a litmus test to inspect nns grounding skills, in: D. Nozza, L. C. Passaro, M. Polignano (Eds.), 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), November 30, 2022, CEUR-WS.org, 2022.

[5] V. Basile, Is evalita done? on the impact of prompting on the italian nlp evaluation campaign., in: D. Nozza, L. C. Passaro, M. Polignano (Eds.), 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), November 30, 2022, CEUR-WS.org, 2022.

[6] M. S. Amin, A. Mazzei, L. Anselma, Towards data augmentation for drs-to-text generation, in: D. Nozza, L. C. Passaro, M. Polignano (Eds.), 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), November 30, 2022, CEUR-WS.org, 2022.

[7] I. Sucameli, A. Bondielli, L. Passaro, E. Annunziata, G. Lucherini, A. Romei, A. Lenci, Mate, a meta layer between natural language and database, in: D. Nozza, L. C. Passaro, M. Polignano (Eds.), 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), November 30, 2022, CEUR-WS.org, 2022.

[8] P. Bellan, M. Dragoni, S. P. Ponzetto, C. Ghidini, H. van der Aa, Process extraction from natural language text: the pet dataset and annotation guidelines, in: D. Nozza, L. C. Passaro, M. Polignano (Eds.), 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), November 30, 2022, CEUR-WS.org, 2022.

[9] T. Labruna, B. Magnini, Simulating domain changes in conversational agents through dialogue adaptation, in: D. Nozza, L. C. Passaro, M. Polignano (Eds.), 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), November 30, 2022, CEUR-WS.org, 2022.

[10] D. Borghesi, A. A. Ravelli, F. Dell'Orletta, What makes the audience engaged? engagement prediction exploiting multimodal features, in: D. Nozza, L. C. Passaro, M. Polignano (Eds.), 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), November 30, 2022, CEUR-WS.org, 2022.

[11] F. De Ponte, S. Rauchas, Grounding words in visual perceptions: Experiments in spoken language acquisition, in: D. Nozza, L. C. Passaro, M. Polignano (Eds.), 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), November 30, 2022, CEUR-WS.org, 2022.

[12] C. D. Hromei, D. Croce, R. Basili, Grounding end-to-end architectures for semantic role labeling in human robot interaction, in: D. Nozza, L. C. Passaro, M. Polignano (Eds.), 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), November 30, 2022, CEUR-WS.org, 2022.

[13] M. Papucci, C. De Nigris, A. Miaschi, F. Dell'Orletta, Evaluating text-to-text framework for topic and style classification of italian texts, in: D. Nozza, L. C. Passaro, M. Polignano (Eds.), 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), November 30, 2022, CEUR-WS.org, 2022.

[14] M. Dusi, N. Arici, A. E. Gerevini, L. Putelli, I. Serina, Graphical identification of gender bias in bert with a weakly supervised approach, in: D. Nozza, L. C. Passaro, M. Polignano (Eds.), 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), November 30, 2022, CEUR-WS.org, 2022.