

NL4AI 2024: Overview of the Eighth Workshop on Natural Language for Artificial Intelligence (NL4AI 2024)

Giovanni Bonetta¹, Claudiu Daniel Hromei², Lucia Siciliani³ and Marco Antonio Stranisci⁴

¹Fondazione Bruno Kessler, Italy

²Department of Enterprise Engineering, University of Rome Tor Vergata, Italy

³University of Bari Aldo Moro, Italy

⁴University of Turin, Italy

Abstract

The Natural Language for Artificial Intelligence (NL4AI) workshop serves as a platform to explore the area situated at the intersection between Natural Language Processing (NLP) and Artificial Intelligence (AI), with a special emphasis on recent activities carried out in both fields in Italy. The eighth edition of the workshop had 18 submissions, of which 16 were accepted. The submissions span a broad spectrum of topics, encompassing foundational NLP research, applied NLP, and works that bridge the realms of NLP and AI. This edition exhibited a strong international presence, featuring contributions from authors representing 6 countries. The submissions also reflect a diversity of languages (e.g., English, Italian) and modalities (e.g., text, vision), underscoring the workshop's commitment to inclusivity and comprehensive exploration.

The Natural Language for Artificial Intelligence (NL4AI) workshop is an annual initiative aimed at promoting a reflection and discussion about various interactions within the field of Artificial Intelligence (AI). The workshop specifically emphasizes the importance of Natural Language Processing (NLP) in AI research, highlighting its role in learning, knowledge representation, and cognitive modeling. Recent AI achievements demonstrate positive impact on complex inference tasks and offer extensive application possibilities in linguistic modeling, processing, and inferences. Nevertheless, Natural Language Understanding remains a rich research topic, whose cross-fertilization extends to diverse areas such as Cognitive Computing, Robotics, and Human-Computer Interaction. For AI, Natural languages serve not only as the central focus for paradigms and applications but also as fundamental elements, playing a crucial role in the automation, autonomy, and learnability of a broad spectrum of intelligent phenomena – from Vision to Planning and Social Behavior. Reflecting on these diverse and promising interactions constitutes a significant objective for ongoing AI studies, aligning seamlessly with the core mission of AIXIA. Specifically, the NL4AI workshop is endorsed by the Special Interest Group on NLP of the Italian Association for Artificial Intelligence (AIXIA) and by the Italian Association of Computational Linguistics (AILC). This year's edition attracted 18 submissions of high-quality papers by 64 distinct authors from Italy (51), United Kingdom (3), United States (3), Pakistan (3), Austria (3), Netherlands (1). After the peer-review process, 16 papers out of the initial 18 were accepted (Acceptance rate: 88.89%), which we believe provides a good balance between the different topics related to the workshop.

The contributions to the workshop cover a spectrum of topics, ranging from generative and conversational AI to applications of LLMs and Multimodal models in the NLP domain. In what follows, we provide a short overview of the accepted papers grouped by main topics.

In line with advancing the use of AI for questionnaire-related tasks, recent studies have leveraged large language models (LLMs) for both generating and filling questionnaires across different domains. Laraspata et al. [1] explored the use of GPT-3.5-Turbo and GPT-4-Turbo models to automate questionnaire generation for HR Management, releasing a novel dataset of HR survey questions. Their research show that, while AI-generated questionnaires are still distinguishable from human-authored ones, GPT-driven question generators are nonetheless a viable solution. On the other hand, Nardoni et

NL4AI 2024: Eighth Workshop on Natural Language for Artificial Intelligence, November 26-27th, 2024, Bolzano, Italy

✉ gbonetta@fbk.eu (G. Bonetta); hromei@ing.uniroma2.it (C. D. Hromei); lucia.siciliani@uniba.it (L. Siciliani); marcoantonio.stranisci@unito.it (M. A. Stranisci)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

al. [2] investigated LLMs' potential to automatically fill clinical questionnaires using patient records, achieving promising results in extracting relevant medical information.

Expanding the scope of AI multi-agent systems, Gosmar et al. [3] introduce an extension to the Multi-Agent Interoperability framework, improving the coordination of AI agents in multiparty conversations. This work introduces roles like Floor Manager and Convener Agent, along with mechanisms for handling interruptions and uninvited agents, which enhance agent collaboration and ensure efficient, structured multiparty exchanges. Brenna et al. [4] focused on proactivity in task-oriented dialogues, proposing the "last utterance proactivity prediction" task. Their research consists in instructing a model to detect when participants provide proactive, unrequested information, in dialogue snippets. This approach opens avenues for models capable of naturally generating proactive contributions, akin to human dialogue behavior.

Several authors have advanced domain-specific applications of AI, addressing key areas such as clinical data handling, legal text processing, educational tools, mental health support, and sign language generation. For instance, Styll et al. [5] introduced an NLP pipeline to automate the extraction of clinical data from free-text admission notes, using Named Entity Recognition (NER), for efficient integration into EHR systems, aimed at enhancing workflow and supporting healthcare management. In the legal domain, Valerio et al. [6] adapted a large language model to Italian legal texts, constructing a specialized corpus from public records and refining the model with Low-Rank Adaptation (LoRA), resulting in improved coherence and domain relevance across varying prompts and corpus sizes. In educational applications, Siragusa et al. [7] developed UniQA, a bilingual question-answering dataset focused on university course information, which includes 1k documents and 14k QA pairs. They assessed it with a Retrieval Augmented Generation model, making it suitable for both question-answering and translation tasks in Italian and English. For accessibility, Colonna et al. [8] introduced a model for generating Italian Sign Language (LIS) gestures for digital avatars, to enhance interaction for the deaf community, with potential applications in digital accessibility and education. Finally, Scozzaro et al. [9] conducted an interdisciplinary readability analysis of recent amendments to the Italian Constitution, incorporating readability metrics and language model evaluations to assess legislative clarity, contributing to the understanding of democratic document accessibility.

Multiple studies presented in this workshop focus on evaluating language models across diverse contexts, particularly on applications for Italian. The dissemination work presented by Seveso et al. [10] introduced a benchmark based on the INVALSI educational assessments to evaluate LLMs' proficiency in Italian, adapting the test format for automated scoring. Their findings highlight gaps in LLMs' performance relative to human standards and discuss educational implications. Scaiella et al. [11] evaluated a multimodal model, MiniCPM-V 2.6, on GQA-it, Italy's first large-scale VQA dataset, showing that fine-tuning improved its accuracy from 33.4% to 59.4%, underscoring the importance of language-specific adaptation for VQA tasks. Papucci et al. [12] addressed label selection in text-to-text classification, developing Value Zeroing, an attention-based method to optimize label representation for IT5, an Italian pre-trained T5 model. Their approach resulted in performance gains on the topic classification task. Lastly, Sartor et al. [13] examined coherence evaluation in small Italian language models, assessing 15 Transformer-based LLMs. They demonstrated that coherence modeling techniques, such as perplexity and semantic distance, show variable efficacy depending on text genre and data perturbations, revealing intricate dependencies that affect model performance on coherence tasks.

Di Quilio et al. [14] introduced a comprehensive framework for Aspect-Category Sentiment Analysis (ACSA), combining data conversion, semi-automatic annotation, and prediction-based reporting. They adapted an existing Aspect-Category-Opinion Sentiment (ACOS) tool to ACSA, developing a web application for annotating and enhancing their novel *beauty* dataset through manual or semi-automatic methods. Musacchio et al. [15] proposed LLaVA-NDiNO, a series of multimodal large language models tailored for the Italian language. By training these models on Italian-translated datasets derived from English vision-language resources, they address the gap in multimodal capabilities for non-English languages. Their work contributes to open science by releasing the models, data, and code, enabling further development in multimodal Italian LLMs.

References

- [1] L. Laraspata, F. Cardilli, G. Castellano, G. Vessio, Enhancing human capital management through gpt-driven questionnaire generation, in: G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci (Eds.), Proceedings of the Eighth 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), CEUR-WS.org, 2024.
- [2] V. Nardoni, M. Lippi, G. Hyeraci, M. Maccari, A. D. Tarazjani, G. Virgili, R. Gini, S. Marinai, Towards automatically filling questionnaires from clinical records with large language models, in: G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci (Eds.), Proceedings of the Eighth 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), CEUR-WS.org, 2024.
- [3] D. Gosmar, E. C. Deborah A. Dahl, D. Attwater, Ai multi-agent interoperability extension for managing multiparty conversations, in: G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci (Eds.), Proceedings of the Eighth 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), CEUR-WS.org, 2024.
- [4] S. Brenna, B. Magnini, Last utterance proactivity prediction in task-oriented dialogues, in: G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci (Eds.), Proceedings of the Eighth 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), CEUR-WS.org, 2024.
- [5] P. Styll, W. Kusa, A. Hanbury, Enhancing clinical data capture: Developing a natural language processing pipeline for converting free text admission notes to structured ehr data, in: G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci (Eds.), Proceedings of the Eighth 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), CEUR-WS.org, 2024.
- [6] F. Valerio, P. Basile, M. de Gemmis, Adapting a large language model to the legal domain: A case study in italian, in: G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci (Eds.), Proceedings of the Eighth 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), CEUR-WS.org, 2024.
- [7] I. Siragusa, R. Pirrone, Uniqa: an italian and english question-answering data set based on educational documents, in: G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci (Eds.), Proceedings of the Eighth 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), CEUR-WS.org, 2024.
- [8] E. Colonna, A. Arezzo, D. Roberto, D. Landi, F. Vitulano, G. Vessio, G. Castellano, Towards italian sign language generation for digital humans, in: G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci (Eds.), Proceedings of the Eighth 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), CEUR-WS.org, 2024.
- [9] C. J. Scozzaro, M. Delsanto, A. Mastropaolo, E. Mensa, L. Revelli, D. P. Radicioni, On the reform of the italian constitution: an interdisciplinary text readability analysis, in: G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci (Eds.), Proceedings of the Eighth 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), CEUR-WS.org, 2024.
- [10] A. Seveso, F. Mercurio, M. Mezzanzanica, D. Potertì, A. Serino, Disce aut deficere: Evaluating llms proficiency on the invalsi italian benchmark, in: G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci (Eds.), Proceedings of the Eighth 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), CEUR-WS.org, 2024.
- [11] A. Scaiella, D. Margiotta, C. D. Hromei, D. Croce, R. Basili, Evaluating multimodal large language

- models for visual question-answering in italian, in: G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci (Eds.), Proceedings of the Eighth 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), CEUR-WS.org, 2024.
- [12] M. Papucci, A. Miaschi, F. Dell’Orletta, Fantastic labels and where to find them: Attention-based label selection for text-to-text classification, in: G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci (Eds.), Proceedings of the Eighth 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), CEUR-WS.org, 2024.
- [13] M. Sartor, F. Dell’Orletta, G. Venturi, Coherence evaluation in italian language models, in: G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci (Eds.), Proceedings of the Eighth 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), CEUR-WS.org, 2024.
- [14] L. Di Quilio, F. Fioravanti, A comprehensive framework for aspect-category sentiment analysis, in: G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci (Eds.), Proceedings of the Eighth 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), CEUR-WS.org, 2024.
- [15] E. Musacchio, L. Siciliani, P. Basile, G. Semeraro, Llava-ndino: Empowering llms with multimodality for the italian language, in: G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci (Eds.), Proceedings of the Eighth 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), CEUR-WS.org, 2024.