GENerative, Explainable and Reasonable Artificial Learning Workshop 2023

Federico Torrielli, Luigi Di Caro and Amon Rapp

University of Torino, Department of Computer Science, Corso Svizzera 185 - 10149 Torino

Abstract

The GENERAL (GENerative, Explainable and Reasonable Artificial Learning) workshop, held at CHITALY 2023, delves into the advancements in General and Generative Artificial Intelligence (GGAI), with a focus on breakthroughs in natural language processing (NLP) and computer vision (CV). The workshop highlights the capabilities of Large Language Models (LLMs) and Latent Diffusion Models (LDMs) in generating human-like content across text and images. It emphasizes the importance of AI explainability, aiming to understand, explain, and control the complexities of these AI systems in terms of fairness, accountability, and transparency. The workshop encourages interdisciplinary collaboration across fields like HCI, psychology, social studies, and the arts to better understand AI's societal and cultural impacts. Topics of interest include user perceptions of generative AIs, machine psychology, AI assistants, ethical issues in Generative AI, and safety and control mechanisms for large language models.

1. Introduction

General and Generative Artificial Intelligence (GGAI) is rapidly advancing, with significant breakthroughs in areas such as natural language processing (NLP), computer vision (CV), and interdisciplinary interests spanning from psychology and Human-Computer Interaction (HCI), to social studies and the arts. The GENERAL workshop aims to provide a platform for exploring these frontiers, discussing emerging properties of human-like intelligence in Large Language Models (LLMs) and Latent Diffusion Models (LDMs), and fostering interdisciplinary collaboration to discuss better ways for interacting with this technology.

A prominent theme in the GENERAL workshop is the exploration and development of Human-Centered AIs, given their increasing popularity even among the general population, due to the rapid diffusion of applications like ChatGPT and Midjourney. The workshop aims to delve into cutting-edge techniques, applications, and challenges in generating human-like (and likable) content across text and images, as well as discuss interfaces and techniques to interact with these technologies. This includes understanding the emergence of creativity, coherence, and unexpected properties of these AI models.

Moreover, in light of the growing complexity and scale of AI models, the GENERAL workshop advocates for a renewed commitment to AI explainability. As models become larger and more

powerful, their inner workings become increasingly opaque, making them difficult to analyse and comprehend. This presents challenges in areas such as fairness, accountability, and transparency and how to communicate them to the users. The workshop will explore novel approaches to understanding, explaining, and controlling these sophisticated systems.

Finally, the workshop also places emphasis on interdisciplinary interests, linking together HCI, psychology, social studies, AI NLP, and the arts. By fostering collaboration and dialogue between these fields, we aim to unlock new insights into the nature of intelligence, the human mind, and the impact of AI on society and culture.

2. Workshop Contributions

This section provides brief descriptions of the six selected contributions included in the workshop. Vladimiro Lovera Rulfi from the University of Bologna and Ivan Spada from the University of Torino demonstrate the remarkable ability of ChatGPT to generate a complete academic article autonomously. Laura Ventrice and Giovanni Siragusa, both from the University of Torino, present their work on enhancing knowledge bases like "Semagram" using open-source Large Language Models, yielding intriguing results. Christian Morbidoni and Annalina Sarra from D'Annunzio University of Chieti–Pescara illustrate how prompt engineering techniques on LLMs outperform human-created supervised baselines in detecting online misogyny. Federico Torrielli from the University of Torino introduces "BLACK", an innovative methodology for prompting in Latent Diffusion Models, which produces high-quality images with minimal effort. Luigi Di Caro, Laura Ventrice, Rachele Mignone, and Stefano Locci, all from the University of Torino, explore the use of RLHF-augmented LLMs like ChatGPT to mimic lexical resources using only contextual information. Lastly, Elena Callegari from the University of Iceland, Desara Xhura from SageWrite ehf., and Peter Vajdecka from the University of Prague propose a method for controlled text generation and style modification using a model trained on text+integer data.

3. Conclusions

In wrapping up this first workshop, we can't help but be optimistic about the future. Seeing six high-quality submissions in our very first event was truly inspiring and has set a strong foundation for our upcoming workshops. We couldn't have asked for a better start. We're at a special moment in the field of Generative AI. The technologies are maturing, and the ideas that came out of this workshop show that we're just scratching the surface of what's possible.

Looking forward, we're excited to see what the future holds. We believe this workshop series will continue to bring forward cutting-edge ideas, deepening our understanding and pushing the boundaries of Generative AI. This is just the beginning, and the best is yet to come.