COMPUTER AND CONTROL ENGINEERING

DAUIN - Preference models for multimodal annotations

Funded By	Dipartimento DAUIN
Supervisor	CAGLIERO LUCA - luca.cagliero@polito.it
Contact	PASTOR ELIANA - eliana.pastor@polito.it BARALIS ELENA MARIA - elena.baralis@polito.it
Context of the research activity	Data sources are commonly enriched with multimodal annotations, e.g., a video can be annotated with visual tags, textual summaries, audio excerpts, and OCR text. The choice of the modality and style of the data annotations is often arbitrary and independent of the downstream models and tasks. The research aims to define automatic preference models for Multimodal LLMs for annotations that automatically recommend the right modality, format, and type according to the task, context, and model.
Objectives	Objectives Data sources are commonly enriched with multimodal annotations, e.g., a video can be annotated with visual tags, textual summaries, audio excerpts, and OCR text. The choice of the modality and style of the data annotations is often arbitrary and independent of the downstream models and tasks. The research aims to (1) Study the correlation between annotation modality and style and the performance of Multimodal LLMs on the annotated data. (2) Explore the use of modality transfer techniques (e.g., audio transcription, text-to-speech, visual-to-text) to boost the performance of Multimodal LLMs on complex multimodal tasks (e.g., video summarization, action recognition, multimodal sentiment analysis). (3) Design and test autonomous agents capable of detecting, recommending, or adopting the most appropriate annotation modality and style based on the target task, LLM, and application context. Tentative work plan During the first year the PhD student will benchmark Multimodal LLMs on established tasks (VQA, Summarization, Entity Extraction, Segmentation) by prompting them with annotation of different types and modalities. It not only studies the effect of annotation types and modality but also the ways to transfer modalities and input formats effectively and efficiently. During the second year, the research will focus on designing a modality preference model able to recommend the right modality and format of the input annotations according to the model, context, and task. Finally, in the third year the research will extend the modality preference model and test in different real-world use cases.

	List of possible publication venues - Conferences: ACL, EMNLP, ACM Multimedia, KDD, ACL, COLING, IEEE ICDM, ECML PKDD, ACM CIKM - Journals: IEEE TKDE, ACM TKDD, IEEE TAI, ACM TIST, IEEE/ACM TASLP, ACL TACL

Skills and competencies for the development of the activity	The PhD candidate is expected to - Have the ability to critically analyze complex systems, model them and identify weaknesses; - be proficient in Python programming; - know data science fundamentals; - have a solid background on machine learning and deep learning; - have natural inclination for teamwork; - be proficient in English speaking, reading, and writing; - proficiency with Docker and Kubernetes software is a plus.