

Is your LLM really clever? Can it mark its own homework?

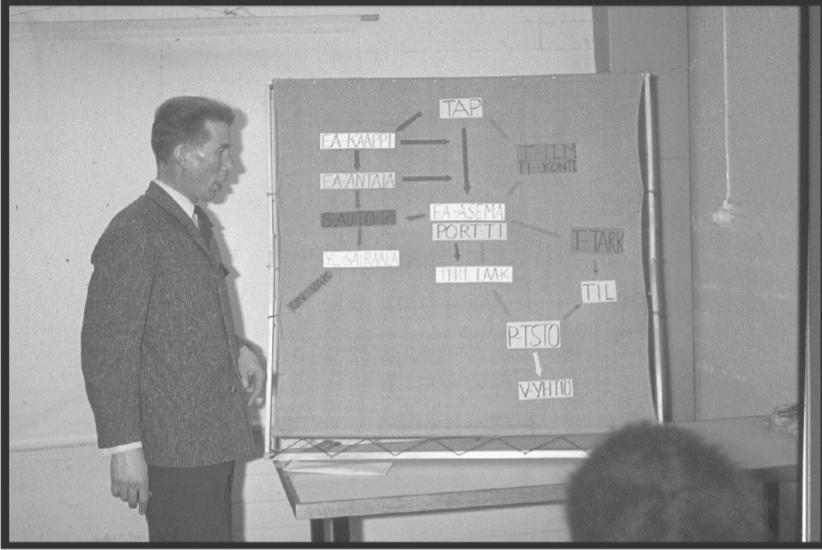
ELOQUENT

The first ELOQUENT Workshop, Grenoble,
September 9-12 2024. 

ELOQUENT - EVALUATION OF GENERATIVE LANGUAGE MODEL QUALITY AND CHARACTERISTICS

TASK 1: *Topical Competence*

Does your LLM know what it is talking about?

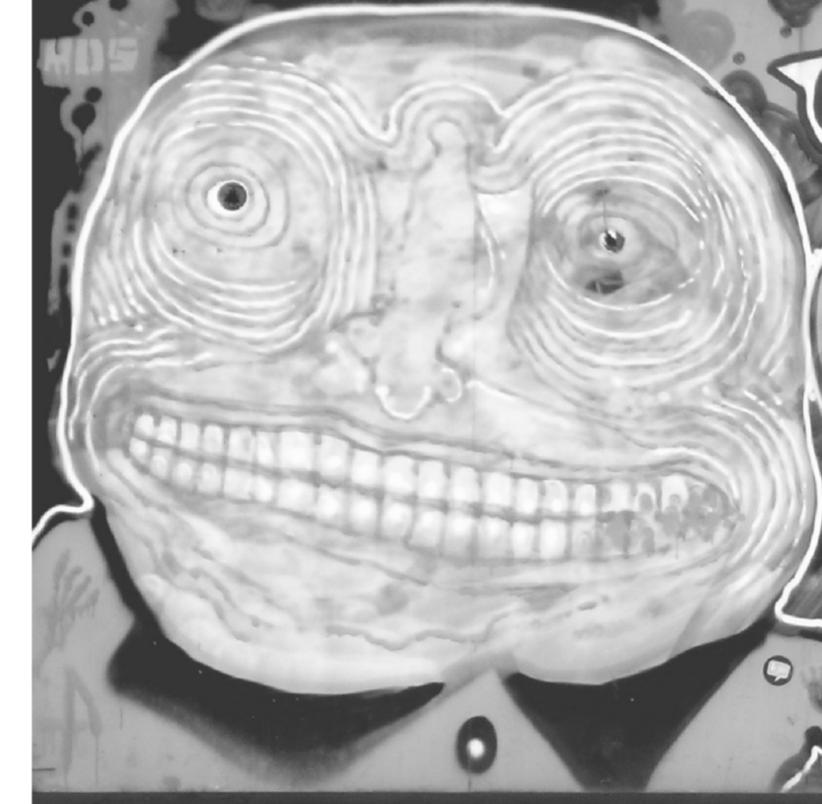


This task will test and verify that a system based on a generative language model is able to handle material from some given topical domain of interest, by having systems automatically generate tests of domain knowledge.

TASK 2: *Veracity and Hallucination*

Is it true? Or make-believe?

This task will test how the truthfulness or veracity of automatically generated text can be assessed.



TASK 3: *Robustness*

Will it give the same content to all of us?



This task will test the capability of a model to handle input variation — e.g. dialectal, sociolectal, and cross-cultural — as represented by human-generated varieties of input prompts. The results will be assessed by how variation in output is conditioned on variation of equivalent but non-identical input prompts.

TASK 4: *Voight-Kampff*

Has a machine written this? Or has a human author put together these words?

This task will explore whether automatically-generated text can be distinguished from human-authored text. This task will be organised in collaboration with the [PAN lab](#) at CLEF.



How to participate

To join the conversation, the mailing list, to get the detailed task descriptions, and to participate in all or some of the tasks: sign up through the CLEF web site.

Timeline

Spring 2024 Registration
22 April 2024 last registration for participation
May 2024 submission of experiments
June 2024 participant report submission
July 2024 camera ready report submission
September 2024 workshop at CLEF

Committee

Silo AI Jussi Karlgren
Helsinki University Aarne Talman
Edinburgh University Liane Guillou
Rise ICT Luise Dürlich, Evangelia Gogoulou, Joakim Nivre
AI Sweden Magnus Sahlgren

... there is more!

We plan to organise ELOQUENT in 2025 as well. Do you have an idea for a new task? Would you like to help organise? Contact the organisers!