THE COMPASS FRAMEWORK RESULTS: STRENGTHS AND WEAKNESSES.

The results are intended to direct actions. Those serve as a navigation to turn complex details into tailored manageable guides for actions and decisions.

Evaluator(s) name and role: João Gonçalves

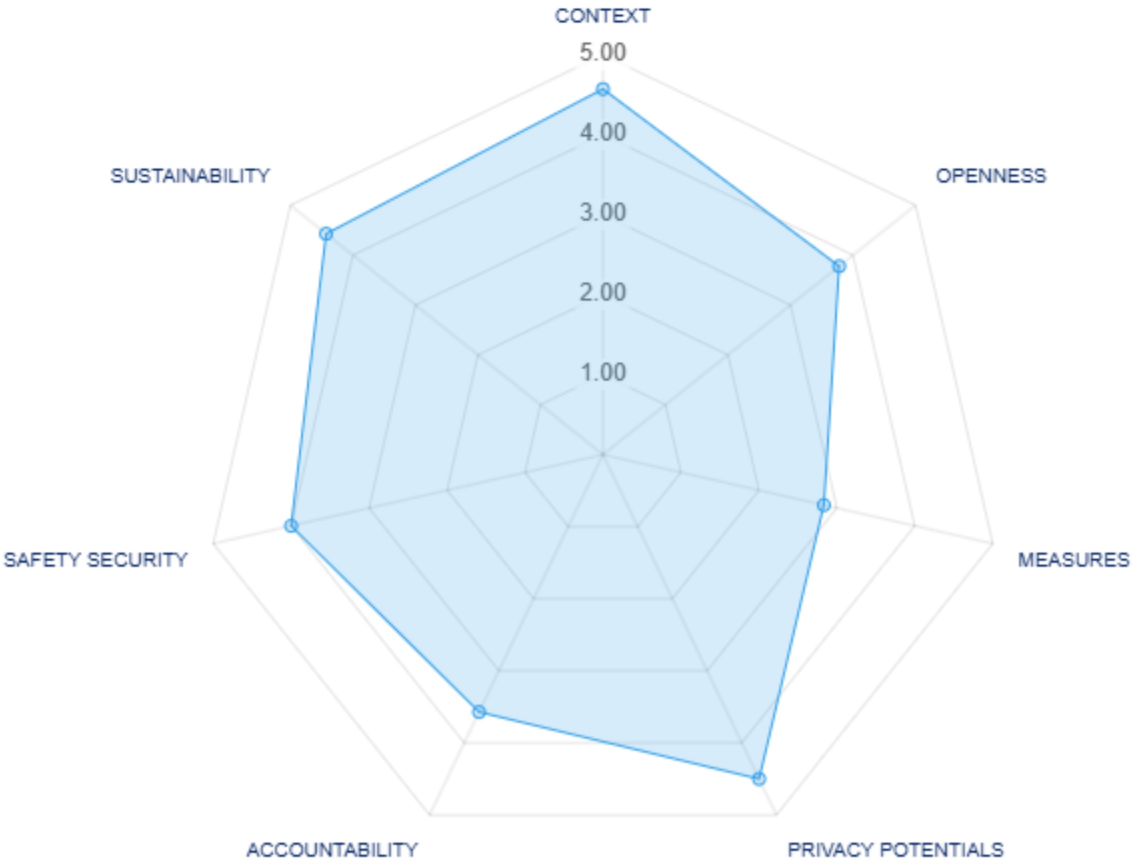
Date of evaluation: 03/07/2024

Number of evaluation: 1

Name/working title of AI system: Erasmian Language Model

Stage of development: The Erasmian Language Model ELM is a community driven large language model tailored to the research and education needs of Erasmus University (EUR, Netherlands) students and staff. The model draws inspiration from ChatGPT in terms of architecture, but it aims to be privacy sensitive, environmentally conscious, and from and for the Erasmus community.

Description: The Erasmian Language Model ELM is a community driven large language model tailored to the research and education needs of Erasmus University (EUR, Netherlands) students and staff. The model draws inspiration from ChatGPT in terms of architecture, but it aims to be privacy sensitive, environmentally conscious, and from and for the Erasmus community.



CLOSE-ENDED QUESTIONS	QUESTION SCORE
CONTEXT •	**4.57**
To what extent have you described the problem definition and use case for your AI system?	5
To what extent have you determined what machine learning domain the AI system will fall under and what consequences this may bring about in its application(s)?	5
To what extent AI system is required for the job or task in question and whether it should be used instead of traditional engineering approaches?	4

To what extent have you comprehensively outlined the roles and responsibilities of all current or potential stakeholders?	5
To what extent have you identified who will be impacted by the AI system and in what ways, throughout its whole life cycle?	3
To what extent have you specified selection criteria for the data(sets) used in the training of your model?	5
To what extent have you included considerations for disadvantages, inaccuracies, potential biases, and other concerns related to including certain data(sets) in the definitions of these selection criteria?	5
OPENNESS	**3.78**
To what extent is it possible to obtain information about the basic structure and nature of the data used?	5
To what extent have you designed the AI system documentation to be accessible and usable?	4
How well are stakeholders informed about data collection, processing, and storage during the AI lifecycle?	5
To what extent does the AI system provide meaningful and understandable explanations and justifications regarding its output?	0
Is there a mechanism for human intervention when it is necessary such as the operator taking control in situations where artificial intelligence encounters uncertainties?	1
To what extent is the AI system transparent about human intervention (e.g., HITL, HOTL, Human-in-Command)?	5
To what extent have you shared relevant performance indicators (e.g. accuracy, precision,...) for your models with your stakeholders?	4
To what extent do you make research findings open and do you interact with expert communities about the AI system?	5
To what extent can you rate the suitability and openness of the AI system for external review and auditing?	5
MEASURES	**2.83**
To what extent have you established a method for assessing how well the AI system operates, including which aspects you'll evaluate and the metrics you'll use?	3
To what extent have you defined the criteria for success and failure of the AI system?	3

To what extent have you considered standardizing these criteria?	2
To what extent have you determined if the AI system will function differently in other contexts, for various (groups of) people or different applications, and whether this different functioning is appropriate or acceptable?	4
To what extent have you integrated measures to prevent overdependence on the AI system and deskilling of the workforce?	1
To what extent have you integrated (potentials for) human intervention (e.g., HiTL, HoTL, Human in Command)?	4
PRIVACY POTENTIALS	**4.50**
To what extent does the AI system align with current privacy regulations or standards such as GDPR, AI ACT, ISO, IEC standards, or other relevant regulations?	5
How thoroughly do you review third-party data sources, tools, libraries, or otherwise for potential privacy risks before integrating them into your AI system?	5
To what extent does the system allow stakeholders to control their data (e.g., data review request, data deletion request)?	2
To what extent are privacy protection techniques incorporated in the AI system's design?	5
To what extent have you identified risk scenarios that could harm privacy, including direct and indirect causes?	4
To what extent have you ensured that the AI system notifies relevant parties of changes and adjustments, especially users?	5
To what extent does the AI system prioritize data minimization and anonymization practices?	5
How effective is the level of the AI system at obtaining necessary permissions and preventing unauthorized access to sensitive information?	NA
To what extent have you considered the possibility that the output of the AI system could lead to privacy violations?	5
To what extent do you ensure privacy (preservation) beyond regulatory compliance?	5
Have you tested the resilience of your system against state of the art privacy attacks such as membership inference attacks? Is your system yielding high-enough robustness?	4

ACCOUNTABILITY	**3.57**
To what extent do you have methods in place to ensure traceability and/or explainability of the AI system throughout its life cycle?	1
To what extent can you explain and justify the operations and decision-making processes of your AI system to different stakeholder groups?	2
To what extent can you justify the responsible use of your training/testing/validation datasets?	5
To what extent is your development process auditable; how open/prepared is your development process for external auditing?	5
To what extent have you established internal structures to ensure ethical and responsible development and deployment of the AI system?	5
How well can you ensure that the AI-system is accountable in the event of failures or undesirable consequences?	5
To what extent have you discussed the cost of predictive errors against different groups? Are they asymmetric (i.e. certain groups can suffer more from specific errors?)?	2
SAFETY SECURITY	**4.00**
To what extent is security considered throughout the life cycle of the AI system, e.g., access and control, resilience, data security/protection and breach management?	5
To what extent is responsibility for security of the AI system determined, documented, and communicated, considering the complete life cycle and all stakeholders involved?	2
To what extent have you defined how security is balanced with other characteristics of the AI system, e.g., openness, transparency, and performance?	5
To what extent have you recognized and defined the (safety) risks with the current and potential future use cases of the AI system?	5
To what extent have you determined potential individual and societal harms stemming from use or misuse of the AI system?	4
To what extent have you established a procedure for managing security or safety incidents?	4
To what extent have you identified and analyzed potential threats to your AI-based system and defined corresponding mitigation strategies?	2

To what extent have you evaluated the robustness of your AI systems against state-of-the-art adversarial ML attacks (e.g. data poisoning attacks, evasion attacks)?	5
SUSTAINABILITY	**4.43**
To what extent does your system development actively acquire and use feedback from different stakeholder groups?	4
To what extent is your organization responsible for quality assurance and maintenance of the system during deployment and beyond?	5
To what extent do you consider futureproofing of the AI system in your development process?	5
To what extent do you remain updated on future developments and changes in standards relevant to your AI system?	4
To what extent have you formulated detailed short-term plans for updates to the AI system?	5
To what extent have you developed detailed long-term plans for updates to the AI system?	3
To what extent do you consider the environmental impact of the development and deployment of your AI system?	5