Proposal Evaluation Form

EUROPEAN COMMISSION

Horizon Europe (HORIZON)

Evaluation Summary Report - Research and innovation actions

Call: HORIZON-CL4-2022-DATA-01
Type of action: HORIZON-RIA
Proposal number: 101092975
Proposal acronym: CLOUDMINER
Duration (months): 36
Proposal title: Cloud-Native and Intelligent Polyglot Data Mining and Analytics Pipelines
Activity: HORIZON-CL4-2022-DATA-01-05

N.	Proposer name	Country	Total Cost	%	Grant Requested	%
1	The Open Group Limited	UK	373,327.5	8.01%	373,328	8.01%
2	INSTITUT FÜR ANGEWANDTE SYSTEMTECHNIK BREMEN GMBH	DE	471,772.5	10.12%	471,773	10.12%
3	ZORGIOS IOANNIS	EL	389,470	8.35%	389,470	8.35%
4	UNIVERSITY OF YORK	UK	572,386.25	12.27%	572,386	12.27%
5	UNIVERSITA DEGLI STUDI DELL'AQUILA	IT	367,451.25	7.88%	367,451	7.88%
6	UNIVERSITEIT MAASTRICHT	NL	464,560	9.96%	464,560	9.96%
7	EDGE HILL UNIVERSITY	UK	476,628.75	10.22%	476,629	10.22%
8	COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES	FR	565,333.75	12.12%	565,334	12.12%
9	INFOTRIPLA OY	FI	320,286.25	6.87%	320,286	6.87%
10	KACHELMANN GMBH	CH	0	0.00%	0	0.00%
11	CNET CENTRE FOR NEW ENERGY TECHNOLOGIES SA	PT	142,498.75	3.06%	142,499	3.06%
12	MAIEUTICA COOPERATIVA DE ENSINO SUPERIOR CRL	PT	119,446.25	2.56%	119,446	2.56%
13	MUNICIPIO DA MAIA	PT	0	0.00%	0	0.00%
14	Continental Automotive Technologies GmbH	DE	400,366.25	8.59%	400,366	8.59%
	Total:		4,663,527.5		4,663,528	

Abstract:
The increasing volume, heterogeneity and importance of data, has created a strong need for intelligent and scalable data mining and analytics technologies. At the same time, the need for data-based decision making and the shortage of skilled software developers, has boosted the demand for data mining solutions that can be used by domain experts with no computer programming background. Several low-code data mining and analytics tools are currently available and are widely used across different scientific and industrial domains. Their main appeal, compared to code-based data science toolkits such as Jupyter Notebooks, is that they enable an increasingly digital-native workforce who may lack a computer programming background, to develop and execute pipelines for scientific and decision-making purposes.

A major shortcoming of existing tools is that they require all the components of a pipeline to be implemented in the same language instead supporting mixing and matching polyglot best-of-breed components. Also, they provide distribution mechanisms that are significantly underdeveloped compared to container orchestration technologies such as Kubernetes, and offer little support for intelligent recommendations that can help users select and configure pipeline components.

The aim of CLOUDMINER is to deliver an open-source platform to support the development and execution of extreme data mining pipelines by providing automations i) for assisting users with relevant recommendations while specifying data pipelines ii) for optimizing the execution of pipelines by efficiently distributing their components, and iii) for supporting the exploration of analyzed data through smart and multidimensional visualizations. The developed solutions will be integrated in a scalable and dynamically reconfigurable micro-service architecture and a low-code development environment that will complement and enhance the work of domain experts while performing extreme data mining activities.

Evaluation Summary Report

Evaluation Result

Total score: 12.50 (Threshold: 10)

Criterion 1 - Excellence

Score: **4.00** (Threshold: 3/5.00 , Weight: -)

The following aspects will be taken into account, to the extent that the proposed work corresponds to the description in the work programme:
- Clarity and pertinence of the project's objectives, and the extent to which the proposed work is ambitious and goes beyond the state of the art.
- Soundness of the proposed methodology, including the underlying concepts, models, assumptions, inter-disciplinary approaches, appropriate consideration of the gender dimension in research and innovation content, and the quality of open science practices, including sharing and management of research outputs and engagement of citizens, civil society and end users where appropriate.

The proposal CLOUDMINER is well in-line with the topic and work program, as it aims to develop a low code, polyglot extreme datamining, aggregation, and analytics open-source platform, with special focus on easing the work of users and data scientists in creating adequate and configurable pipelines for a large variety of data science tasks.

While the objectives are generally well described the measures of success in achieving these objectives are not sufficiently defined. For example, objectives STO2 “Design and implementation of a polyglot data pipeline development environment” and “STO4: Development of intelligent recommenders for data mining assistance”.
The proposal clearly describes its beyond state-of-the-art contribution and innovation. The current state of the art for the different components of the platform is well described and is defined as a baseline to evaluate the beyond state-of-the-art achievements targets. The innovation potential is good with the proposal expecting to achieve TRL 6 or TRL 7. A minor issue is that the proposal does not clearly describe which TRL starting point the background technologies are already at. The proposal describes appropriate links with some national and international research activities and standardisation efforts.
Overall, the provided approach is well detailed, and its soundness is well established in terms of how the effectiveness, efficiency, performance, and business applicability of the platform would be measured during the project. However, the proposal has inconsistent levels of detail regarding how the planned research and technical activities are structured and will work together.
Furthermore, the proposal does not convincingly describe how knowledge gained from use cases in unrelated domains can then be to adapt and generalise the framework across domains.
A welcome addition is that the proposal points out common pitfalls when evaluating improvements and gives reasonable ways of handling them.
The soundness and credibility of the concept is very good. However, it is not sufficiently clear from the proposal which activities are being delivered by the partners themselves and which are research activities developed by the community at large.
The proposal lacks a clear description of the evaluation of the AI approaches to be used in the project; specifically for the recommenders’ design and knowledge base creation. In general, the proposal does not present sufficient information to assess the technical robustness of the AI and whether it may pose concerns.
The consideration of interdisciplinary approaches is good. It is mainly focused on computer science (ML, AI, NLP, etc.), although use cases are applied to different domains like meteorological forecasting or industry. Stakeholder knowledge is well integrated within the proposal.
The proposed open science and open data approaches are well in-line with the work program requirements and the FAIR principles.

Criterion 2 - Impact

Score: **4.00** (Threshold: 3/5.00 , Weight: -)

The following aspects will be taken into account, to the extent that the proposed work corresponds to the description in the work programme:
- Credibility of the pathways to achieve the expected outcomes and impacts specified in the work programme, and the likely scale and significance of the contributions from the project.
- Suitability and quality of the measures to maximise expected outcomes and impacts, as set out in the dissemination and exploitation plan, including communication activities.

Overall, the target impact for the project is highly ambitious. The proposed contribution to improve the accessibility of the industry to advanced Extreme AI platforms and knowledge, at reduced cost and with easier implementation, is significant and can attain large scale in the key use cases presented in the proposal. However, the proposal lacks a sufficiently clear description of how it would contribute to improving European leadership in global data economy, maximising social/economical benefits from wider use of data or reinforcing Europe's ability to manage urgent societal challenges.
Furthermore, the potential barriers to impact are not sufficiently addressed in the proposal.
The proposal convincingly details the approach and activities for dissemination, communication and exploitation. These also include general justification and targets for open science and open data plans.
A positive aspect of the proposal is the aim to involve domain experts and to create low-code solutions which could be used by people with no computer-science background. The exploitation plans indicate that they will make the proposed project's results broadly available as open source, embedding results in commercial IT products.

Criterion 3 - Quality and efficiency of the implementation

Score: **4.50** (Threshold: 3/5.00 , Weight: -)

The following aspects will be taken into account, to the extent that the proposed work corresponds to the description in the work programme:
- Quality and effectiveness of the work plan, assessment of risks, and appropriateness of the effort assigned to work packages, and the resources overall.
- Capacity and role of each participant, and the extent to which the consortium as a whole brings together the necessary expertise.

The proposed management structure and procedures are adequate. The overall work-packages structure, organisation, milestones and deliverables are clearly defined and well in-line with the project objectives and ambitions. The consortium presented well organized quality and performance monitoring and evaluation procedures for the major activities. While the proposal provides a comprehensive risk list, the presented mitigation plans are not sufficiently detailed. Risks, with regard to, project timings and dependencies are not sufficiently addressed.
The Consortium composition demonstrates it includes adequate expertise and resources in the key areas required for the project execution. The consortium has good complementarity. The allocation of tasks is good. All partners seem to have a valid role and adequate resources to fulfil it. The roles of the 14 participants (partners and associated) are well described and their contribution and expected value generation to the plans are convincingly detailed. A key strength is in the involvement of a renowned standards organisation (TOG) in order to attain the necessary linkage, involvement and participation in the generation of de-facto standard in cloud aware, low-code development, Big data AI systems.

Scope of the application

Status: **Yes**

Comments (in case the proposal is out of scope)

Not provided

Exceptional funding

A third country participant/international organisation not listed in the General Annex to the Main Work Programme may exceptionally receive funding if their participation is essential for carrying out the project (for instance due to outstanding

expertise, access to unique know-how, access to research infrastructure, access to particular geographical environments, possibility to involve key partners in emerging markets, access to data, etc.). (For more information, see the HE programme guide)

Please list the concerned applicants and requested grant amount and explain the reasons why.

Based on the information provided, the following participants should receive exceptional funding:

Not provided

Based on the information provided, the following participants should NOT receive exceptional funding:

Not provided

Use of human embryonic stem cells (hESC)

Status: **No**

If YES, please state whether the use of hESC is, or is not, in your opinion, necessary to achieve the scientific objectives of the proposal and the reasons why. Alternatively, please state if it cannot be assessed whether the use of hESC is necessary or not, because of a lack of information.

Not provided

Use of human embryos

Status: **No**

If YES, please explain how the human embryos will be used in the project.

Not provided

Activities excluded from funding

Status: **No**

If YES, please explain.

Not provided

Do no significant harm principle

Status: **Yes**

If Partially/No/Cannot be assessed please explain

Not provided

Exclusive focus on civil applications

Status: **Yes**

If NO, please explain.

Not provided

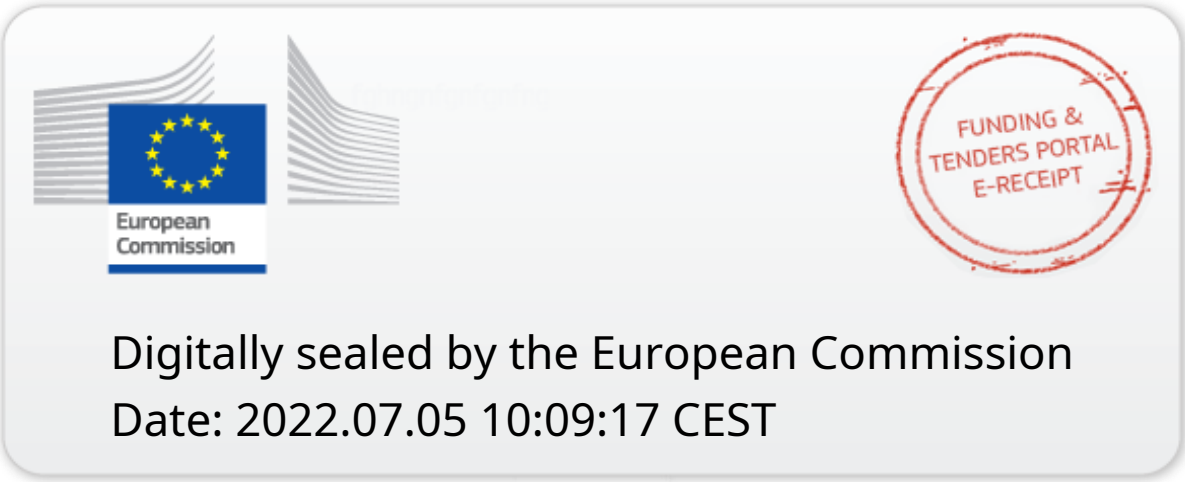
Artificial Intelligence

Status: **Yes**

If YES, the technical robustness of the proposed system must be evaluated under the appropriate criterion.

Overall comments

Not provided



This electronic receipt is a digitally signed version of the document submitted by your organisation. Both the content of the document and a set of metadata have been digitally sealed.

This digital signature mechanism, using a public-private key pair mechanism, uniquely binds this eReceipt to the modules of the Funding & Tenders Portal of the European Commission, to the transaction for which it was generated and ensures its full integrity. Therefore a complete digitally signed trail of the transaction is available both for your organisation and for the issuer of the eReceipt.

Any attempt to modify the content will lead to a break of the integrity of the electronic signature, which can be verified at any time by clicking on the eReceipt validation symbol.

More info about eReceipts can be found in the FAQ page of the Funding & Tenders Portal.

(https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/faq)