Project specifications

Study LLM4Gov

Contents

1. INTRODUCTION

The Ministry for Digitalisation has received of a multitude of requests through the AI4Gov call for projects. The focus of these requests was the provision of a 'Large Language Model' (LLM) dedicated to the public sector's needs, the main use cases revolving around chat and mailbots, some covering specific business needs. The main objective behind the integration of such a model is to automate a variety of tasks. These tasks predominantly include the generation of text-based content and summaries. The model is for example expected to be capable of generating responses to questions raised by citizens and help in the preparation phase of harmonized answers to them.

Automating these tasks can significantly enhance the efficiency of public sector organisations. An LLM solution can streamline their operations and provide quicker, responses to citizen inquiries. A study to identify the model(s) best suited to the needs of the Luxembourgish government is the first step prior to a potential implementation of one or more LLM's. To identify the needs of public sector organisations regarding the use of Large language models, a survey has been sent out to gather their input. The results of the survey will be shared with the selected economic operator. Use cases such as analysing documents, an alternative research tool and content creation will be in the focus.

In addition to the survey results, the chosen economic partner will also be informed about several use cases (some mentioned below in section 2.1) of similar complexity that have already been identified for an LLM model within the government.

This project aims at producing a study for the Ministry for Digitalisation ranking existing LLMs on the following relevant criteria, along with a plan to implement one or more LLM tools.

2. ELEMENTS TO BE CONSIDERED FOR THE DELIVERABLES

2.1 CRITERIA

The following list of criteria are a non-exhaustive list of criteria that must at a minimum be considered in the study, and therefore also included in the final report:

Models' selection

- The leading state-of-the-art LLMs should at minima be considered (both proprietary and Open-Source).

- LLMs with different numbers of parameters should be considered (lightweight to flagship).

- LLMs using different technologies must be considered (for instance transformers, decoder-only, encoder only, mixture of experts)

Capabilities and features

- General capabilities: the overall capabilities of candidate LLMs should be provided.

- Use-case specific capabilities: the ability for each candidate LLM to provide a relevant answer to each use case should be evaluated.

- Versatility:

    o The ability for each LLM to cover multiple use case with high-quality solutions should be outlined.

    o The ability for each LLM to onboard new use cases easily. More information on possible use cases is going to be provided at a later stage (survey results).

- Language Support:

    o It should be evaluated whether the model has to support English (EN), French (FR), German (DE) and Luxembourgish (LU).

    o LLMs should also be capable of understanding and generating code in various coding languages.

- Benchmarking:

    o The capabilities and versatility of different LLMs should be supported by relevant benchmarks, as described above.

    o LLMs should be benchmarked on various use cases, including risk of hallucination, speed, and accuracy.

    o Characteristics of the benchmarking environment should be given, along with all relevant metrics to help ensuring that the obtained results are reproducible.

Integration

- Integration with existing Models:

    o The feasibility of integrating existing state models (like speech to texts models, an employment specific chatbot etc.) should be analysed. More information on those models will be provided to the chosen economic partner.

- API Availability: The model should provide an accessible and reliable API for integration with existing tools.

- Hosting and Infrastructure:

    o The pros and cons of cloud vs on-premises hosting, licensed vs non-licensed solutions, and open-source vs proprietary solutions should be evaluated.

    o The ability to deal with sensitive data (personal or others) in a sovereign and secured environment should be evaluated.

    o The ability for the LLM to be deployed on an OpenShift container platform should be evaluated.

    o The ability for the LLM to be deployed on air-gapped platforms such as sovereign disconnected cloud platforms should be evaluated.

- The Integration in a complete environment should be evaluated. This includes:

- o The ease of integration with tools such as, but not limited to:
  - vector DBs,
  - caches,
  - monitoring, logging, and tracing tools
  - orchestration tools
  - agents.
  - o The model's REST API should support OAuth2.0 and OIDC
  - o the possibility to implement the LLMs on different environments (Build and run environment)
  - o the existence of a fully-fledged suite integrating the LLM.
- Maintenance and Integration: Considerations should include versioning, model updates, ensuring reproducibility of generated answers, maintenance efforts, hardware requirements (CPU, GPU, memory, storage), and energy consumption.
  - o This should be done in general and specifically for the provided use cases.

Cost

- Cost Analysis: A comprehensive cost analysis should be conducted, including the estimate cost of hosting, maintenance, potential license fees or human resources. In order to compare the LLMs the costs should be indicated in TCO (total cost of ownership) for a period of 5 years.

Other criteria might arise from answers given to a survey that will be shared at project launch.

Compliance

- Data Protection:
  - o The model should be GDPR compliant.
  - o The exposure of the LLM to extraction attacks should be provided.

2.2 EXPECTED DELIVERABLES

It is expected, within a period of approximately 6 months, to complete the deliverables below:

- Analysis of survey results and priorization of different use cases.

- Perform an analysis on existing LLMs with respect to their suitability for the public sector's needs, considering the criteria mentioned under section 2.1 of this document.

- Provide a report describing the methodology of the study.

- Provide a report on the findings as well as an overall and use-case specific rankings with a maximum of 8 use cases of similar complexity and of suitable models prioritizing the use cases out of the survey such as analysing documents, an alternative research tool and content creation.

- Suggest a future implementation plan, making sure that implementation in the IT infrastructure of the State or a sovereign disconnected cloud is possible. More information can be provided to the chosen economic partner.

- A cost comparison among the models (TCO based), encompassing aspects such as licensing costs, hosting costs, and the costs associated with machine/server availability for model implementation. Identification of the model that is most suitable for a typical use case, but also demonstrates versatility in handling other use cases. This includes an evaluation of its capabilities in areas such as content creation, code generation, and information research.

2.3 TYPE OF OFFER

The offer should be a financial and time proposal and a CV for each person who will participate in the project, whether it is a single person or a dedicated team, detailing their background and relevant experience, must also be provided.

The offer must include a concept that outlines the proposed approach, detailing the various stages, deliverables, and providing the reasoning and justification for the chosen approach.

Since the contract falls under the fixed global price regime, the economic operator must therefore provide a fixed price for all elements of the contract. The contract should be considered as a fixed-price service contract, with an obligation to achieve specific results and including all requested services. To enable the contracting authority to assess the validity of the proposed prices, the economic operator must provide a detailed breakdown of their calculations for each element of the contract.

The documents must be sent via email to datascience@digital.etat.lu. The deadline for submitting an offer is **Tuesday, November 19th 2024**.

Profile of consultants as subject matter experts

Requirements:

- Fluency the following languages English, French (C1 level at minima for one language)

- Provide an extract of criminal record (record 3)

Criteria:

- Provide a CV demonstrating some prior experience in data science with focus on LLM's

- Provide a track record of the number of projects implementing LLM's or create Benchmarks on LLM's

- Bibliography of recent published articles and/or publicly available contributions (such as GitHub / Hugging Face accounts).

3. EXECUTION

3.1 TIMEFRAME

The timeframe is 6 months from kick-off to deliverables.

3.2 WORK ENVIRONMENT

Management and milestone meetings can be held online via Microsoft Teams or on site.

4. PROCEDURE

After the selection process, the next step will be to sign the necessary contracts. The kick-off session will mark the start of this project. Regular support and management meetings will help with the progression of the project. Weekly progress will be assessed and provided to management. The project will finish when the deliverables have been achieved within their timeframe and when the solution is given during a final knowledge transfer session. A final presentation of results to higher management will mark the end of the project.

4.1 SECURITY AND CONFIDENTIALITY

The security and confidentiality aspects are extremely important considering the sensitive nature of the data. A background check and an extract from the criminal record of each employee working on this project is required. A non-disclosure agreement and a non-compete clause need to be signed by any person working on this project as well as for the representative of the company.

Throughout the whole project and beyond, all employees involved in the project shall agree to maintain the confidentiality of all information and data in relation to the project and shall not use, disclose, furnish or make it accessible to anyone other than authorized employees.

4.2 INTELLECTUAL PROPERTY RIGHTS

The deliverables are protected by the relevant intellectual property and copyright laws and will remain the exclusive property of the Ministry for Digitalisation for the duration of the project and beyond.
All disputes concerning this project shall be governed by Luxembourg law, and the courts of the Grand Duchy of Luxembourg shall have exclusive jurisdiction to hear and settle such disputes.

4.3 KNOWLEDGE TRANSFER

A final knowledge transfer or presentation session, at the end of the project to provide the necessary knowledge is required. This session should be held with all involved parties.

4.4 EVALUATION

The project will be evaluated based on the deliverables given in this document and their timeliness, but also on the candidate's ability to come with their own initiatives to achieve the deliverables.

The offers will be evaluated based on the following criteria:

- Quality of the offer submitted (approach, structure, level of detail, completeness,) – 40%
- Proposed schedule and pricing – 30%
- Profiles of consultants (relevant experience) – 30%