SpeedUP

# Study LLM4Gov

| | |
|---|---|
| **Ordering party** | Ministry for Digitalisation |
| **Objectives of the call for solutions** | The Ministry for Digitalisation has received of a multitude of requests on the provision of a 'Large Language Model' (LLM) dedicated to the public sector's needs. So, the main objective behind the integration of such a model is to automate a variety of tasks.<br>These tasks predominantly include the generation of text-based content and summaries. The model is also expected to be capable of generating coherent responses to questions regarding specific business needs.<br><br>As a first step prior to a potential implementation of one or more LLM's, a study to identify the best solution is to be done with this call for solution. |
| **Type of solution required** | Study of feasibility, benchmark of the existing, proposition of a concept |
| **Selection criteria** | • Quality of the offer submitted (approach, structure, level of detail, completeness) – 40%<br>• Proposed schedule and pricing – 30%<br>• Profiles of consultants (relevant experience) – 30% |

| | |
|---|---|
| **Standards to be met** | NA |
| **IP and other details** | 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.<br>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. |
| **Deadlines for submission of the offer** | November 19th 2024 |
| **Schedule** | The length of the project should not exceed 6 months. The Kick-off is planned for 01/2025 at the latest. |
| **Contact for questions** | datascience@digital.etat.lu |
| **Deadline for questions** | November 6th 2024 |

| | |
|---|---|
| **Description of the problem / challenge** | 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.<br><br>In addition to the survey results, the chosen economic partner will also be informed about several use cases 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 relevant criteria, along with a plan to implement the identified use cases based on selected LLMs.<br><br>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:<br><br>***Models' selection:***<br><br>• The leading state-of-the-art LLMs should at minima be considered (both proprietary and Open-Source).<br>• LLMs with different numbers of parameters should be considered (lightweight to flagship).<br>• LLMs using different technologies must be considered (for instance transformers, decoder-only, encoder only, mixture of experts)<br><br>***Capabilities and features:***<br><br>• General capabilities: the overall capabilities of candidate LLMs should be provided.<br>• Use-case specific capabilities: the ability for each candidate LLM to provide a relevant answer to each use case should be evaluated.<br><br>**Versatility:**<br><br>• The ability for each LLM to cover multiple use case with high-quality solutions should be outlined.<br>• The ability for each LLM to onboard new use cases easily.<br>More information on possible use cases is going to be provided at a later stage (survey results).<br><br>**Language Support:**<br><br>• It should be evaluated whether the model has to support English (EN), French (FR), German (DE) and Luxembourgish (LU).<br>• LLMs should also be capable of understanding and generating code in various coding languages.<br><br>**Benchmarking:**<br><br>• The capabilities and versatility of different LLMs should be supported by relevant benchmarks, as described above.<br>• LLMs should be benchmarked on various use cases, including risk of hallucination, speed, and accuracy.<br>• 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:**

- 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:**

- The pros and cons of cloud vs on-premises hosting, licensed vs non-licensed solutions, and open-source vs proprietary solutions should be evaluated.
- The ability to deal with sensitive data (personal or others) in a sovereign and secured environment should be evaluated.
- The ability for the LLM to be deployed on an OpenShift container platform should be evaluated.
- 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:**

- The ease of integration with tools such as, but not limited to:
    - vector DBs,
    - caches,
    - monitoring, logging, and tracing tools
    - orchestration tools,
    - agents.
- The model's REST API should support OAuth2.0 and OIDC.
- the possibility to implement the LLMs on different environments (Build and run environment)
- 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.

- 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***<br>• Data Protection:<br>  ○ The model should be GDPR compliant.<br>  ◦ The exposure of the LLM to extraction attacks should be provided.<br><br>**For complete information regarding the requirements, the deliverables, and the type of offer to be submitted, please refer to the project specifications linked on the website.** |