AI4DI

Réponses aux questions reçues

QUESTION 1

Can we get the detailed process video mentioned page 7?

Réponse / Answer

FR :
https://otx.etat.lu/6fb86a5abdb084faa9285b701b8821e6cfb18cc31172518cb686048ab1ac0c0b

EN :
https://otx.etat.lu/6fb86a5abdb084faa9285b701b8821e6cfb18cc31172518cb686048ab1ac0c0b

QUESTION 2

Technical details of the computing stack the CTIE can allocate to this project: models and sizing of CPU, GPU, memory and storage

Réponse / Answer

FR : Il s'agit du modèle NVIDA GPU Tesla V100, retrouvez plus d'informations sous ce lien tesla-volta-v100-datasheet-letter-fnl-web.pdf (nvidia.com).

EN : The model is the NVIDA GPU Tesla V100, find more informations with the following link tesla-volta-v100-datasheet-letter-fnl-web.pdf (nvidia.com).

QUESTION 3

Does the GovCloud supports containers (Docker)? Does it support vitualization of resources (CPU & GPU mainly)?

Réponse / Answer

FR : Oui et Oui.

EN : Yes and Yes.

QUESTION 4

Should the Recueil*.docx files be kept in target workflow (meaning it will still be the reference of valid responses)? Is it kept in a specific place on the SP?

Réponse / Answer

FR : Le fichier docx n'est plus adapté pour conserver et classifier les réponses il faudra prévoir dans la solution un autre moyen (DB ou autre).

EN : The docx file is not suited anymore to store this kind of information, the solution should include a better way to store the information and keep it updated it.

QUESTION 5

Which frequency should the LLM "learn" about new validated entries in the Recueil.docx?

Réponse / Answer

FR : Le LLM devrait pouvoir être mis à jour tous les trimestres.

EN : The LLM should be able to be updated once per trimester.

QUESTION 6

Is there another anticipated usages of the target platform (different workflows, different documents, ...)?

Réponse / Answer

FR : Comme précisé dans le cahier des charges, le LLM devrait être capable de se référer à d'autres documents (lois, jurisprudence etc.. en pdf ou docx majoritairement), construire une réponse basée sur ces docs ou répondre à une question concernant ces docs pour que chaque agent de la CNPD puisse bénéficier de l'outil.

EN : As written in the specs, the LLM should be able to learn from other documents (laws, jurisprudence in pdf/docx mostly), build an answer based on them and

basically answer about their content so that every user in the CNPD can benefit ot it.

QUESTION 7

What are the evaluation criterias of the answers you'll get?

Réponse / Answer

FR : Le niveau d'expérience (en LLM, de travail avec une administration...), les langues maitrisees, le niveau de détai et la qualité de l'offre présentée et les CV's proposés.

EN : Experience (in LLM, in working with a government administration), languages mastered, level of details/quality off the offer and the suggested CV's.

QUESTION 8

Any specifics constraints or restrictions on the LLMs we might use on this project as long as they can be executed on the CTIE environments and as long as their licences are compatible with the approach?

Réponse / Answer

FR : Non.

EN : No.

QUESTION 9

Any existing vector database already deployed (FAISS, Annoy, Milvus, Weaviate, ...)? Any existing and usable ElasticSearch license?

Réponse / Answer

FR : Non, c'est notre 1er projet du genre.

EN : No, it is our first project of this kind.

QUESTION 10

P11: "The solution should provide user rights management and SSO authentication in the CNPD's windows environment" > does it mean Kerberos/SSO authentication? Can you provide more details about the

authentication requirements and the user referential we have to authenticate the user against? Supported protocols?

Réponse / Answer

FR : IAM-TAM – A definir lors du projet avec le CTIE.

EN : IAM-TAM – To be defined during the project with the Government IT Centre.

QUESTION 11

Any specific needs to log/audit the user interactions with the LLM?

Réponse / Answer

FR : Pas de besoin spécifique, mais un audit trail doit exister.

EN : Not really, the content managed by the LLM will not be highly sensitive but an audit log should exist.

QUESTION 12

How many users and interactions (queries/day) expected?

Réponse / Answer

FR : 100/jour.

EN : 100/day.

QUESTION 13

Can some parts of the offer be delivered in times and means, with the initial and maximum estimations (the others being fixed prices)?

Réponse / Answer

FR : Le projet est financé par le ministère de la Digitalisation ce qui impose une offre de prix fixe.

EN : We are funded through the Ministry of Digitalisation for this project, which requires a fixed price regime.

QUESTION 14

range of the budget for this call?

Réponse / Answer

FR : Maximum de 100k€ hors TVA.

EN : Maximum of 100k€ excluding VAT.

QUESTION 15

Which solution are you looking for:

- o Retrieval Augmented Generation (RAG) and standard LLM
- o Specifically trained (or fine-tuned) LLM (without RAG)
- o RAG which collaborates with the fine-tuned or specifically trained LLM?

Réponse / Answer

FR : Comme il s'agit de notre 1er projet IA, notre fournisseur de solutiuon devra nous conseiller sur la solution la plus adaptée. Cela dit la possibilité d'avoir un RAG semble intéressante pour ce projet car cela permettrait d'ajouter plus facilement des documents que la solution pourrait utiliser. En revanche le système n'aura pas d'accès en ligne.

EN : As this is our first AI project, our solution provider will have to advise us on the most suitable solution. That said, the possibility of having a RAG seems interesting for this project because it would make it easier to add documents the solution could use. No online access will be given to the solution.

QUESTION 16

If RAG, do you expect:

- o Proprietary Solution
- o Open Source Tools

Réponse / Answer

FR : L'open source est préférable mais pas obligatoire.

EN : Open source is preferable but not mandatory.

QUESTION 17

If RAG, regarding the 3 (FR, DE, EN) vectorization/embeddings:

- o Public commercial
- o Private commercial
- o Customized local

Réponse / Answer

FR : Pas de préférence.

EN : No preferences.

QUESTION 18

If fine-tuning or specific training of LLM:

The technical resources (servers, gpus ,etc.) will be made available by the CTIE?

Réponse / Answer

FR : Oui.

EN : Yes.

QUESTION 19

Which LLM do you want to use:

o Public commercial LLM (e.g. Chatgpt, Gemini)

o Private commercial LLM (e.g. Chatgpt on LU Gov Azure cloud)

o Open Source solutions installed & maintained locally at CTIE (e.g. Mistral, deepseek)

o Your own LLM installed & maintained locally at CTIE

Réponse / Answer

FR : La solution devra être hébergée sur le govcloud, quelques bonnes pratiques seront communiquées au début du projet.

EN : Solution should be hosted on govcloud, some guidelines will be provided at the project launch.

QUESTION 20

Regarding the sharepoint integrator, which connector is acceptable:

o Open Source

o Closed Source

o or will you provide the connector?

Réponse / Answer

FR : Pas de préférence, nous ne fournirons pas le connecteur.

EN : No preferences, we won't provide the connector.

QUESTION 21

- How many concurrent users do you expect?

Réponse / Answer

FR : 50.

EN : 50.

QUESTION 22

How many new documents to upload (day) do you expect?

Réponse / Answer

FR : Il y aura peu de documents au début, si le modele tient ses promesses des documents / infos seront ajoutés trimestriellement (20 / trimestre).

EN : There will be few documents at the beginning, if the model keeps its promises documents / information will be added quarterly (20 / quarter).

QUESTION 23

Have the documents been curated?

Réponse / Answer

FR : Non, c'est notre 1er projet du genre.

EN : No, it is our first project of this kind.

QUESTION 24

Are there any conflicting documents? different responses for similar questions?

Réponse / Answer

FR : Normalement au sein du « recueil » il a été vérifié de ne pas avoir de conflits mais une vérification de la cohérence du document pourrait être nécessaire. Entre les futurs documents c'est peu probable mais pas impossible, le système devra prévenir si c'est le cas.

EN : within the "collection" it has been checked not to have conflicts but a verification of the consistency of the document could be necessary. Among the

future documents it is unlikely but not impossible, the system will have to warn the user if this is the case.

QUESTION 25

Is there a guideline used to structure the responses (e.g. paragraph 1 describes the understanding of the question, paragraph 2 links with the relevant laws and articles, paragraph 3 provides details about similar cases)

Réponse / Answer

FR : Pas vraiment, la variété des types de questions/réponses ne permet pas de dégager un standard mais il y a des blocs standards en fonction de certains types de questions.

EN : Not really, the variety of question/answer types does not allow for a standard to be identified but there are standard blocks depending on certain types of questions.

QUESTION 26

Is the solution only used by trained teams ? (no outside requests - has an impact on security and safety, see below)

Réponse / Answer

FR : Yes, utilisation par les agents uniquement.

EN : Yes, internal users only.

QUESTION 27

Do the .msg files contain plain text questions or shall the system extract the attached documents, try to read the content and initiate OCR when needed?

Réponse / Answer

FR : le msg contiendra la demande en texte. Si il s'agit d'un scan le fichier d'entrée sera un pdf.

EN : .msg contain plain text. If we receive a letter it is scanned in pdf.

QUESTION 28

In terms of Security, which level do you expect?:

access management (with sso) is understood

1- specific access rights at document (or document chunk) level ? / Only some users have access to some documents or parts of documents.

2-anonymization of documents ? / Sensitive information (e.g. names) are only accessible to some users

3-Is db encryption required?

Réponse / Answer

FR : Non aux 3 questions. Les documents sources ne sont pas sensibles. En revanche les demandes ne doivent être accessibles qu'à la personne qui la faite et à un administrateur si nécessaire.

EN : No to the 3 questions. Referenced documents are not sensitive. Though the queries and answers of a user should remain accessible only to the user and an admin if necessary.

QUESTION 29

In terms of Safety, what are your requirements?:

Prompt Injections / Detect and block any user attempt of prompt injection or jailbreak.

Prompt Leakage / Block prompt leakage attempts before they reach the LLM.

SQL Enforcement / Detect and block attempts to use SQL operations that are not within the stated limits.

CNPD Policy / Block messages that do not uphold the terms of use.

Data Leakage / Block data leakage attempts before they reach the LLM.

Réponse / Answer

FR : Il sera du rôle du fournisseur de conseiller la CNPD sur ces points.

EN : The provider shall advise the CNPD about these features.

QUESTION 30

terms of AI reliability, what are your requirements?:

Toxicity / Detect and override prompts and responses containing profanity.

Allowed Topics / Ensure the conversation sticks to the stated topics.

Off-Topic Discussion / Block any prompt or response attempting to speak about restricted topics.

RAG Hallucinations / Detect and override any response that carries a high risk of hallucinations.

Réponse / Answer

FR : Il sera du rôle du fournisseur de conseiller la CNPD sur ces points.

EN : The provider shall advise the CNPD about these features.

QUESTION 31

Automation, would you need:

Automated document lifecycle (new, changed, replaced, outdated, deleted) management

Automated web retrieval for case laws and jurisprudence

Réponse / Answer

FR : Non.

EN : No.

QUESTION 32

Cost Tracking, do you expect:

Limits tracking and budget warnings.

Defined usage limits (per user or group)

Réponse / Answer

FR : Non.

EN : No.

QUESTION 33

AI Observability

Do you want the solution to provide reports about chain of thoughts and documents used to create responses

Réponse / Answer

FR : Oui.

EN : Yes.

QUESTION 34

In order to deliver the smart conversational AI system that goes through data and completes tasks, an LLM serving (and training) infrastructure / platform should be hosted and functional.
In the requirements it is stated that: "Integration into GovCloud managed by the CTIE is preferable to an on premises" solution". Do you already have such a system in place for GovCloud or serving LLMs is designed as part of AI4ID?
(We have our own proprietary solution platform for model serving and training.)

Réponse / Answer

FR : Non il faudra demander l'hébergement au CTIE.

EN : No, we don't have yet such an architecture it will be necessary to host it on the CTIE govcloud.

Question 35

Normally LLMs require considerable GPU resources to run. Does your cloud / on prem infrastructure provide such resources (e.g. Nvidia A100, H100, H200, H800 etc)

Réponse / Answer

FR : Le govcloud propose des GPU Nvidia Tesla V100.

EN : Govcloud offers Nvidia Tesla V100 GPUs.

Question 36

Does our company need to be registered before the application deadline or is there any chance to do it until the project starts?

Réponse / Answer

FR : Les marchés ne peuvent être attribués qu'à des opérateurs économiques légalement établis au jour de l'ouverture des offres (soit en l'occurrence le 21 février 2025). Il n'est pas obligatoire que cet établissement légal soit situé au Luxembourg.

EN : Contracts may only be awarded to economic operators who are legally established on the day the tender is opened (i.e. in this case February 21st, 2025). It is not a requirement that such legal establishment be in Luxembourg.

Question 37

Would a Microsoft Copilot-Based Solution Be Considered?

Réponse / Answer

FR : Oui, à ce stade du développement nous ne sommes pas fermés à une technologie.

EN : Yes, at this stage of the project we are not closed to any technology.

Question 38

Are there any specific anonymization or data protection rules we should follow when processing the documents ?

Réponse / Answer

FR : Non.

EN : No.

Question 39

What version of SharePoint is currently in use ?

Réponse / Answer

FR : SharePoint Server 2019.

EN : SharePoint Server 2019.

Question 40

Do you have an existing PDF template or mandatory elements (e.g., layout, references, dynamic elements) that must be included in the output ?

Réponse / Answer

FR : Pas encore.

EN : Not Yet.

Question 41

What are the key evaluation criteria for the PoC (e.g., response quality, response time, relevance rate) ?

Réponse / Answer

FR : Temps de réponse, pertinence de la réponse, qualité rédactionnelle, facilité d'intégrer de nouveaux documents.

EN : Response time, relevance of the response, editorial quality, ease of integrating new documents.

Question 42

Are there predefined test cases or scenarios to validate the solution performance and scalability ?

Réponse / Answer

FR : Pas encore.

EN : Not yet.

Question 43

Are there any constraints regarding the use of external APIs (such as OpenAI) versus open-source models?

Réponse / Answer

FR : Pas de contrainte mais une préférence pour des modeles open source hébergés sur un cloud. Quelques bonnes pratiques seront communiqués au début du projet.

EN : No constraints but cloud hosted open source models are preferred. Some guidelines will be provided at the project launch.

QUESTION 44

What security and confidentiality requirements apply to data processing via the language model, and do you have any preferences on the choice of model (e.g., GPT, Gemini, Llama, DeepSeek-R1)?

Réponse / Answer

FR : Pas de besoins spécifiques en terme de confidentialité, les documents de référence ne contiendront pas de données personnelles. Pas de préférence sur le choix du modèle en ce moment.

EN : No special requirements in terms of confidentiality, the reference documents will not contain personal data. No specific preference for the choice of the model at this stage.

QUESTION 45

Would it be acceptable for our PoC to include a fully functioning chatbot service that demonstrates the end-to-end solution ?

Réponse / Answer

FR : Oui.

EN : Yes.

QUESTION 46

Can we provide a detailed architecture and low-level design in our proposal that specifically addresses your use case ?

Réponse / Answer

FR : Oui.

EN : Yes.