

Legal and normative text analyser

Réponses aux questions reçues

QUESTION 1

Scope: It is not totally clear to us if we are supposed to “develop” the actual solution, or “define” a solution (supported by PoC) for you to then deliver a public RFP. Per section 2.1.1: “The first step of the project is to define specifications in order to submit a public contract to identify an economic operator which will be responsible for the development of the AI tool-based application capable to meet the OLAS needs”.

Réponse / Answer

FR : L’objectif du projet est de développer un proof of concept qui fera par la suite l’objet d’un déploiement sur l’infrastructure informatique du CTIE.

EN : The objective of the project is to develop a proof of concept which will subsequently be deployed on the CTIE IT infrastructure.

QUESTION 2

Location: We understand the necessity to have on-site presence to facilitate the project, which is obviously feasible with our local Luxemburgish entity (that I represent and manage). However, we do have a dedicated entity within our group (Smart 4 Engineering group), focused on IA & Data projects (350 people, in Portugal), that we aim to work with for this project. Thus, do you have a minimum required time of presence expected in Luxembourg maybe?

Réponse / Answer

FR : A ma connaissance, il n'existe pas d'exigence spécifique concernant le temps minimum de présence attendu du prestataire au Luxembourg.

EN : To my knowledge there is no specific requirement regarding the minimum required time of presence expected for the provider in Luxembourg.

QUESTION 3

Duration: We would need to confirm if you really expect a 6 months project (to align with your plans per your involvement) or if we could define a faster timeline, potentially?

Réponse / Answer

FR : L’un des critères d’attribution associé à la proposition de projet définit que les délais à respecter pour disposer d’un « proof of concept » est évalué à 6 mois après le « kick of meeting ». Rien ne dit que le projet doit obligatoirement durer précisément 6 mois, ce qui laisse la porte ouverte à un délai de réalisation potentiellement plus court.

EN : One of the award criteria associated to the project proposal defines that the deadline for obtaining a "proof of concept" is estimated at 6 months after the "kick-off meeting." There is no



requirement that the project must last exactly 6 months, which open the door to a potentially shorter completion time.

QUESTION 4

Legal status: Could you please specify what legal status is required for candidates who wish to respond to this call for projects?

Réponse / Answer

FR : Les marchés ne peuvent être adjugés qu'aux opérateurs économiques qui sont juridiquement établis au jour de l'ouverture de la soumission.

EN : Contracts may only be awarded to economic operators who are legally established on the day the tender opens.

QUESTION 5

Power computing: Are there any limitations regarding the computing power in GovCloud? Does it support all possible LLMs?

Réponse / Answer

FR : Oui, il y a des limitations.

EN : Yes, there are limitations in terms of computing power.

QUESTION 6

Perating system: What operating system is expected to be used on GovCloud?

Réponse / Answer

FR : Il existe deux types de services GovCloud : 1) GovCloud où vous aurez à votre disposition des machines virtuelles Windows ou Linux (mode IaaS) 2) GovCloud GCCS qui permet de gérer des namespaces OpenShift (mode PaaS).

EN : It exists two types of GovCloud services: 1) GovCloud where you will have at your disposal some Virtual Machines Windows or Linux (mode IaaS) 2) GovCloud GCCS which allows to manage OpenShift namespaces (mode PaaS).

QUESTION 7

Technical question: Is containerization a viable option for GovCloud?

Réponse / Answer

FR : Oui.

EN : Yes



QUESTION 8

Non-competition clause: We did not fully understand the non-compete clause. As we are a software company applying to build tools for a government entity, how can we ever be in competition? Can the restriction of this clause be explained?

Réponse / Answer

FR : La signature d'une clause non-compete n'est pas nécessaire pour ce projet.

EN : The signing of a non-compete clause by the economic operator will not be required by the contracting authority for this project.

QUESTION 9

Infrastructure IT: Est-ce que l'infrastructure IT GovCloud est gérée par le CTIE ?

Réponse / Answer

FR : Oui.

EN : Yes.

QUESTION 10

Infrastructure IT: L'infrastructure inclut-elle des GPU, S3 or Blob Storage, VMs... ?

Réponse / Answer

FR : Actuellement, l'infrastructure dispose de S3 pour le stockage et de machines virtuelles. Les ressources GPU sont limitées et sont exposées aux machines virtuelles GovCloud ou au namespace OpenShift de manière limitée (pas de GPU dédiés,).

EN : Currently, the infrastructure has S3 for storage, and VMs. There are limited GPU resources, which are exposed to the GovCloud VMs or OpenShift namespaces in limited way (no dedicated GPUs).

QUESTION 11

Cluster git: A-t-on un cluster git, K8S et image registry dédiés pour le projet ?

Réponse / Answer

FR : Il est possible d'avoir un namespace OpenShift. Nous avons un dépôt d'images centralisé avec le CI/CD à partir de projets git.

EN : It is possible to have an OpenShift namespace. We have a centralized image repository with the the CI/CD from git projects.



QUESTION 12

Solutions VectorDB: Toutes les solutions VectorDB sont-elles acceptables ? Nous prévoyons d'utiliser ChromaDB.

Réponse / Answer

FR : Actuellement, le seul VectorDB pris en charge au niveau de l'infrastructure est PGVector. Cependant, vous pouvez gérer une instance de ChromaDB dans votre propre namespace. Cela signifie que les opérations de sauvegarde/restauration doivent être documentées et planifiées ; les mises à jour de ChromaDB doivent être gérées par l'application. Par conséquent, tous ces aspects doivent être pris en compte lors de la conception, du développement et du déploiement de l'application.

EN : Currently, the only supported VectorDB at infrastructure level is PGVector. However, you could manage an instance of ChromaDB in your own namespace. This means, that backups/restores operations must be documented and scheduled; updates of ChromaDB should be managed by the application. Therefore, all these aspects need to be considered when designing, developing and deploying the application.

QUESTION 13

Update des fichiers: Est-ce qu'un update des fichiers tous les mois est acceptable ?

Réponse / Answer

FR : Une mise à jour une fois par jour est nécessaire avec la possibilité de lancer une mise à jour manuellement au besoin.

EN : An update once a day is required with the ability to manually initiate an update if needed.

QUESTION 14

Data sources: A-t-on besoin de toutes les datas des différents data sources, ou faut-il filtrer au préalable (ex : Seulement plusieurs corpus de lois Legilux, et non l'ensemble) ?

Réponse / Answer

FR : Comme indiqué au point 1.3 « type of Solution Required » du cahier des charges « Extract from the identified documents the relevant information », l'application doit identifier les documents pertinents. Il est clair que l'ensemble des documents de 'legilux.public.lu', par exemple, ne sont pas concernés. Nous n'excluons pas la possibilité d'avoir le choix dans l'application de pouvoir choisir les data sources que nous souhaitons utiliser. De plus, comme indiqué au point 2.2.1 « Data Specifications » du cahier des charges, il sera possible de sélectionner des fichiers stockés localement (data source : « Locally stored files (working documents - not yet published) ») afin de lancer une analyse sur des fichiers préalablement sélectionnés.



EN : As indicated in point 1.3 "Type of Solution Required" of the specifications "Extract from the identified documents the relevant information", the application must identify the relevant documents. It is clear that all documents from 'legilux.public.lu', for example, are not concerned. We do not exclude the possibility of having the choice in the application to be able to choose the data sources that we wish to use. In addition, as indicated in point 2.2.1 "Data Specifications" of the specifications, it will be possible to select locally stored files (data source : "Locally stored files (working documents - not yet published)") to launch an analysis on previously selected files.

QUESTION 15

Update des fichiers: Est-ce que l'update des fichiers est entièrement automatisé, ou actions manuelles à prévoir (User Interface required) ?

Réponse / Answer

FR : L'update des fichiers doit être automatisé tant que possible afin de limiter au stricte minimum l'interaction de l'utilisateur.

EN : File updates should be automated as much as possible to keep user interaction to a minimum.

QUESTION 16

Gestion de l'authentification: L'authentification doit-elle être gérée de notre côté ?

Réponse / Answer

FR : L'application finale doit être intégrée au système IAM du CTIE. Cela signifie que l'application doit implémenter un protocole Open ID Connect afin de permettre l'authentification à partir de l'IAM du CTIE.

EN : The final application needs to be integrated with CTIE's IAM system. This means that the application needs to implement an Open ID Connect protocol in order to allow authentication from CTIE's IAM.

QUESTION 17

Responsible for the coordination: Who is responsible for the coordination with entities like Legilux, EUR-Lex etc. (ex: getting accesses, API info, etc.)?

Réponse / Answer

FR : L'opérateur économique est responsable d'obtenir les détails techniques sur la manière d'intégrer les sources de données et de les mettre en œuvre, et l'ILNAS est responsable de fournir les informations d'accès, les clés d'authentification API...

EN : The economic operator is responsible to get technical details on how to integrate data sources and to implement it, and ILNAS is responsible to provide access information, API authentication keys...



QUESTION 18

List of connection modes: Can we have an exhaustive list of connection modes for the data sources?

Réponse / Answer

FR : Les modes de connection seront à élaborer au début de la phase projet ensemble avec l'opérateur économique sélectionné.

EN : The connection modes will be defined at the beginning of the project phase together with the selected economic operator.

QUESTION 19

Technologie stack: What is the required technological stack?

Réponse / Answer

FR : Une solution basée sur Python ou JavaScript. Pour le langage Python: Python, FastAPI, Flask ou DJANGO. Pour JavaScript: TypeScript, React, Next.js, Node.js, Prisma. Pour la base de données PostgreSQL et PGVector pour le DB Vector.

EN : We welcome solution based on both Python or JavaScript language and frameworks. For Python Language: Python, FastAPI, Flask or DJANGO. For JavaScript: TypeScript, React, Next.js, Node.js, Prisma. For the Database we manage PostgreSQL and PGVector for the DB Vector.

QUESTION 20

Programation language: Do you impose the usage of specific programming languages (ex: Java or Python)?

Réponse / Answer

FR : Il n'y aura pas d'imposition sur les langues de programmation mais elles seront décidées ensemble, voir également question 19.

EN : There are no specific languages imposed, but there should be an exchange on what language to use, also see question 19.

QUESTION 21

Libraries: Do you impose the usage of specific libraries for RAG and/or LLMs?

Réponse / Answer

FR : Non.

EN : No.



QUESTION 22

Cloud provider: Are there any requirements or preferences for specific cloud provider technologies (Azure, AWS, GCP)?

Réponse / Answer

FR : Les points Cloud/Govcloud devront être vus avec le CTIE.

EN : The Cloud/Govcloud issues will have to be discussed with the CTIE.

QUESTION 23

Vector database: Do you impose a specific vector database?

Réponse / Answer

FR : Regardez la réponse à la question 12.

EN : View answer on question 12.

QUESTION 24

CTIE documentation: Can you provide the documentation from CTIE regarding guidelines for AI solution development?

Réponse / Answer

FR : Il n'est pas prévu de partager de la documentation du CTIE à ce stade. Des informations / documents supplémentaires pourront être partagés avec l'opérateur économique choisi à un stade ultérieur dans le projet.

EN : At this stage, it is not foreseen to share CTIE documentation. Additional information/documents will be shared with the chosen economic operator at a later stage in the project.

QUESTION 25

Evaluation criteria: What are the key evaluation criteria for the PoC?

Réponse / Answer

FR : Les principaux critères d'évaluation sont :

- Couverture du cahier des charges
- Interface utilisateur conviviale
- Qualité du rapport en termes de contenu
- Rapidité d'exécution



- Exhaustivité et niveau de détail du rapport détaillé sur la solution privilégiée
- Qualité de l'accompagnement en formation
- Meilleur rapport qualité-prix du projet par rapport aux coûts, comme indiqué au point 2.3 « Livrables attendus » du cahier des charges

EN : Key evaluation criteria are:

- Coverage of the specifications
- User-friendly application interface
- Quality of the report in term of content
- Speed of the application
- Completeness and level of details for the detailed preferred solution report
- Quality of the training support
- Best value for money of the project related to costs as stated in point 2.3 "Expected Deliverables" of the specifications

QUESTION 26

Quality criteria: How do you quantify the expected quality of the PoC?

Réponse / Answer

FR : Les principaux critères d'évaluation sont :

- Couverture du cahier des charges
- Interface utilisateur conviviale
- Qualité du rapport en termes de contenu
- Rapidité d'exécution
- Exhaustivité et niveau de détail du rapport détaillé sur la solution privilégiée
- Qualité de l'accompagnement en formation
- Meilleur rapport qualité-prix du projet par rapport aux coûts, comme indiqué au point 2.3 « Livrables attendus » du cahier des charges

EN : Key evaluation criteria are:

- Coverage of the specifications
- User-friendly application interface



- Quality of the report in term of content
- Speed of the application
- Completeness and level of details for the detailed preferred solution report
- Quality of the training support
- Best value for money of the project related to costs as stated in point 2.3 "Expected Deliverables" of the specifications

QUESTION 27

Curriculum Vitae: Are the CV criteria required only on one CV or multiple CVs?

Réponse / Answer

FR : Ces critères sont requis pour un seul CV.

EN : These criteria are only required for one CV.

QUESTION 28

Document integration: Do all documents need to be integrated within 6 months of the PoC?

Réponse / Answer

FR : Oui

EN : Yes

QUESTION 29

Languages of documents: Besides English and French, will the system need to process documents in other languages - even if final reports are only in English and French?

Réponse / Answer

FR : Non

EN : No

QUESTION 30

Interfaces: Apart from the administration interface for imports and launching reports, are other interfaces expected?

Réponse / Answer

FR : Comme indiqué au point 2.3 « Livrables attendus » des spécifications, il est mentionné « Guides utilisateur (pour l'application standard, les utilisateurs à privilèges élevés et



l'administrateur du déploiement, de la maintenance et des mises à niveau de l'application) ». Nous attendons donc des « utilisateurs standard » et des « utilisateurs à privilèges élevés », qu'ils utilisent ou non la même interface utilisateur.

EN : As indicated in point 2.3 "Expected Deliverables" of the specifications, it is mentioned "User guides (for application standard, application high privilege users and application deployer/maintenance/upgrade administrator)". So, we expect "standard user" and "high-privilege users" whether or not they use the same user interface.

QUESTION 31

Is it a problem if some unspecified elements, such as user interface, detailed metadata available, source connexions are time-boxed ?

Réponse / Answer

FR : A reasonable timeframe to define these elements can be proposed by the economic operator.

EN : Un délai raisonnable pour définir ces éléments peut être proposé par l'opérateur économique.

QUESTION 32

The actual software is not mentioned anywhere. Furthermore, the costs for developing the prototype are listed as deliverables in this phase, which appears to contradict the requirement to deliver the actual prototype within the same phase. Would you please clarify further on these points?

Réponse / Answer

FR : Le logiciel lui-même (partie du PoC) n'est pas spécifiquement identifié parmi les livrables. Veuillez considérer la livraison du PoC telle que décrite dans les spécifications comme faisant partie des livrables attendus.

Par ailleurs, le « coût lié à la mise en œuvre du PoC dans des environnements standardisés (développement, test et production) » ne correspond pas au coût du PoC, mais au coût de sa mise en œuvre dans les trois environnements (développement, test et production).

EN : The actual software (part of the PoC) is not specifically identified from the deliverables. Please consider the delivery of the PoC as described in the specifications as part of the expected deliverables.

Otherwise, the "cost related to the implementation of the PoC in standardized environments (Dev, Test and Prod)" is not the cost of the PoC but the cost to implement the PoC in the 3 environments (Dev, Test and Prod).



QUESTION 33

Profile of the economic operator: Would the candidacy of an economic operator can be considered if all of the criteria defined in the paragraph 2.4 of the project specifications have not been satisfied? Additional context - (i) In case of a start-up, having produced similar projects to those requested but have not already work for a public administration. (ii) In case it is not possible to identify a firm that want to remain anonymous.

Réponse / Answer

FR : Applicants are required to fulfill the requirements listed under 2.4 to be eligible.

EN : Les candidats doivent remplir les conditions indiquées au point 2.4 pour être éligibles.

QUESTION 34

Profile of the economic operator: Is it possible to include CVs of persons with whom a firm has contractual agreements in place (service providers / partners), but are not part of this firm?

Réponse / Answer

FR : En cas de sous-traitance ou de groupements d'opérateurs économiques, cela doit être clairement indiqué dans l'offre. Le CV de chaque personne intervenant dans l'exécution du marché doit être inclus dans l'offre.

EN : In the case of subcontracting or groups of economic operators, this must be clearly indicated in the offer. The CV of each person involved in the execution of the contract must be included in the offer.

QUESTION 35

Profile of the economic operator: Are there any restrictions on data on the GovCloud being accessed by persons outside of Luxembourg (e.g. Europe) or in third countries (e.g. India) for the development of the intended application ? Additional context - Assuming that this is done in line with robust security measures (including encryption in transit and rest, least privilege, access control, authentication, logs), and in line with the GDPR - where applicable (Satisfying art. 28, and Chapter V - e.g. putting in place standard contractual clauses, and Transfer Impact Assessments).

Réponse / Answer



FR : Les points Govcloud devront être vus avec le CTIE.

EN : These Govcloud issues will have to be discussed with the CTIE.

QUESTION 36

Scope and outputs: Do you expect the six-month project to deliver a proof-of-concept or a production ready application? If a PoC, what success criteria will decide whether it is turned into production ?

Réponse / Answer

FR : L'objectif est de fournir un PoC pouvant être utilisé pour les tests et toutes les informations nécessaires pour déployer l'application dans un environnement de production tel que décrit dans les spécifications.

Les principaux critères d'évaluation sont :

- Couverture du cahier des charges
- Interface utilisateur conviviale
- Qualité du rapport en termes de contenu
- Rapidité d'exécution
- Exhaustivité et niveau de détail du rapport détaillé sur la solution privilégiée
- Qualité de l'accompagnement en formation
- Meilleur rapport qualité-prix du projet par rapport aux coûts, comme indiqué au point 2.3 « Livrables attendus » du cahier des charges

EN : The objective is to provide a PoC that can be used for testing and all information required in order to deploy the application in a production environment such as described in the specifications.

Key evaluation criteria are:

- Coverage of the specifications
- User-friendly application interface
- Quality of the report in term of content
- Speed of the application
- Completeness and level of details for the detailed preferred solution report
- Quality of the training support



- Best value for money of the project related to costs as stated in point 2.3 "Expected Deliverables" of the specifications

QUESTION 37

Handling of activities: Which “activities” must be fully supported, and which can be handled in a scaled-up phase?

Réponse / Answer

FR : Toutes les activités doivent être gérées pendant le projet.

EN : All activities have to be handled during the project.

QUESTION 38

Handling of activities: What’s the desired level of detail in identifying “requirements” and “guidance”?

Réponse / Answer

FR : Le niveau d’exigences et d’orientations est celui qui est disponible dans la réglementation (nationale ou européenne) et/ou les documents normatifs (normes nationales, européennes ou internationales).

EN : The level of requirements and guidance is the one that is available in the regulation (national or European) and/or normative documents (national, European or international standards).

QUESTION 39

Model of report: What should the reports look like ? Do you have an existing template or mandatory elements ? Should there be metadata?

Réponse / Answer

FR : La composition du rapport sera discutée lors de la phase de définition de l'architecture fonctionnelle. Nous ne disposons pas encore de modèle.

EN : The composition of the report will be discussed during the functional architecture definition phase. We don’t have any existing template.

QUESTION 40

Additional fields: Besides the five items listed (document ref, section, activity type, requirement, guidance), are any additional fields mandatory ?

Réponse / Answer

FR : Aucun champ supplémentaire n'est identifié mais il pourrait évoluer lors de la phase de définition de l'architecture fonctionnelle.



EN : There is no additional field identified but it could evolve during the functional architecture definition phase.

QUESTION 41

Application: Should the application expose an API for downstream systems, or is GUI + file export sufficient?

Réponse / Answer

FR : Aucun champ supplémentaire n'est identifié mais il pourrait évoluer lors de la phase de définition de l'architecture fonctionnelle.

EN : There is no additional field identified but it could evolve during the functional architecture definition phase.

QUESTION 42

Preference for a specific model: Is there a preference for using open-source models versus hosted APIs ?

Réponse / Answer

FR : Pas de préférence.

EN : No preference.

QUESTION 43

Vector source: Any preference for a particular commercial or open-source vector store ?

Réponse / Answer

FR : Voir question 12, des discussions à ce sujet sont à mener après le choix de l'offre.

EN : See question 12, additional discussions will be held after choosing an offer.

QUESTION 44

Vector database: Any existing vector database already deployed?

Réponse / Answer

FR : Non

EN : No

QUESTION 45

Report generation: Is there a maximum acceptable latency from query to report generation ?



Réponse / Answer

FR : Non mais ce point sera abordé lors de la phase de définition de l'architecture fonctionnelle.

EN : No but this point will be addressed during the functional architecture definition phase.

QUESTION 46

Specific requirements: Is a pure RAG pipeline acceptable, or do you expect additional fine-tuning on accreditation-specific corpora?

Réponse / Answer

FR : Nous n'avons pas d'exigences spécifiques concernant le cadre d'IA à utiliser.

EN : We don't have specific requirements on the AI framework to be used.

QUESTION 47

LLMs: Any specific constraints or restrictions on the LLMs we might use on this project ?

Réponse / Answer

FR : Non.

EN : No.

QUESTION 48

Format of documents: Do the source documents have consistent formatting (e.g., headings, structure), or are they highly varied?

Réponse / Answer

FR : Cela devra être analysé lors de la phase de définition de l'architecture fonctionnelle.

EN : It will have be analysed during the functional architecture definition phase.

QUESTION 49

Language of documents: What proportion of source documents is expected in French vs. English vs. other languages?

Réponse / Answer

FR : Voir question 29

EN : See question 29



QUESTION 50

Official API: For each source (Legilux, EUR-Lex, ISO, CEN, EA, ILAC, IAF) can you confirm whether an official API is available and may be used, or whether scraping is required?

Réponse / Answer

FR : Cela devra être analysé lors de la phase de définition de l'architecture fonctionnelle.

EN : It will have be analysed during the functional architecture definition phase.

QUESTION 51

Conflicting documents: Are there any conflicting documents?

Réponse / Answer

FR : Cela ne devrait normalement pas se produire, mais nous ne pouvons pas l'affirmer à ce stade.

EN : It should normally not occur but we cannot confirm at this stage.

QUESTION 52

Contradicting requirements: If the application encounters accreditation requirements or guidance that contradict one another across different legal or normative sources, how should it behave? Is there a predefined hierarchy or reliability ranking among sources (e.g., EU regulations > national laws > ISO/CEN standards , etc).

Réponse / Answer

FR : En cas de contradiction, elle doit être mentionnée dans le rapport. Aucune hiérarchie ni aucun classement de fiabilité ne doivent être couverts.

EN : If a contradiction occurs, it shall appear in the report. No hierarchy or reliability ranking has to be covered.

QUESTION 53

Test dataset: Will a formal test dataset or gold-standard annotations be provided?

Réponse / Answer

FR : Un ensemble de données de test doit être identifié lors de la phase de définition de l'architecture fonctionnelle.

EN : A test dataset should be identified during the functional architecture definition phase.

QUESTION 54

Test dataset: Are there test cases or scenarios to validate the solution performance?



Réponse / Answer

FR : Les cas de test ne sont pas encore construits.

EN : Test cases are not yet built.

QUESTION 55

Computing resources: Could you detail the GovCloud computing resources that will be made available for this project (CPU model and core count, GPU model and number of cards, RAM per node, total and per-volume storage capacity, as well as any relevant performance or quota limits)?

Réponse / Answer

FR : Le CTIE propose des machines virtuelles, dont certaines avec des capacités GPU (NVidia "Tesla V100" GPU), spécifiquement dédiées aux applications de machine learning. Les environnements sont évolutifs en termes de RAM et de vCPUs. Des informations détaillées sur le type ou les modèles de matériel ne sont pas disponibles pour le moment. Le système d'exploitation peut être Linux Ubuntu, Redhat ou Windows Server.

EN : The CTIE proposes virtual machines , some with GPU capabilities (NVidia "Tesla V100" GPU), specifically dedicated to ML applications. The environments are scalable in terms of RAM and vCPUs. Detailed information on hardware type or models is not available at this time. OS can be Linux Ubuntu or Redhat or Windows Server environment.

Limits of storage: Any hard limits on storage, GPU/CPU quotas or outbound traffic from GovCloud?

Réponse / Answer

FR : Le CTIE propose des machines virtuelles, dont certaines avec des capacités GPU (NVidia "Tesla V100" GPU), spécifiquement dédiées aux applications de machine learning. Les environnements sont évolutifs en termes de RAM et de vCPUs. Des informations détaillées sur le type ou les modèles de matériel ne sont pas disponibles pour le moment. Le système d'exploitation peut être Linux Ubuntu, Redhat ou Windows Server. EN : The CTIE proposes virtual machines , some with GPU capabilities (NVidia "Tesla V100" GPU), specifically dedicated to ML applications. The environments are scalable in terms of RAM and vCPUs. Detailed information on hardware type or models is not available at this time. OS can be Linux Ubuntu or Redhat or Windows Server environment.

OLAS licences: Does OLAS already own licences for NLP/LLM frameworks, OCR engines ?

Réponse / Answer

FR : Non.

EN : No.



QUESTION 58

Number of users: Expected number of concurrent users and peak document ingests per day?

Réponse / Answer

FR : Le nombre d'utilisateurs simultanés à court terme est de 6 (10 à long terme). Le pic d'ingestion de documents ne peut être anticipé, car il dépendra des nouvelles publications.

EN : Concurrent users in short term is 6 (long term is 10). The peak document ingestion cannot be anticipated as it will depend on new publications.

QUESTION 59

Ajout de documents: Faut-il ajouter la possibilité de rajouter manuellement des documents ? (auquel cas il faut prévoir un front-end)

Réponse / Answer

FR : Comme indiqué dans le cahier des charges au points 2.2.1 "Data Specifications", la possibilité de "Locally stored files (working documents - not yet published)" doit faire partie de la solution.

EN : As stated in the specifications in point 2.2.1 "Data Specifications", the possibility of "Locally stored files (working documents - not yet published)" must be part of the solution.

QUESTION 60

Vectorisation de documents: Faut-il ajouter la possibilité de vectoriser des documents ? (auquel cas il faut prévoir une vectorDB). On pourrait s'en sortir sans, à partir du moment où on pourra se baser sur un file storage et un LLM

Réponse / Answer

FR : Non.

EN : No.

QUESTION 61

Volume of documents: What is the estimated volume and frequency of legal/normative documents that the system is expected to process?

Réponse / Answer

FR : Le volume et la fréquence des documents à gérer ne peut être anticipé, car il dépendra des nouvelles publications. La récupération des documents dans l'application d'IA doit pouvoir être lancée manuellement par l'utilisateur final ; sinon, elle doit être effectuée au moins une fois par jour.



EN : The volume and frequency of documents cannot be anticipated as it will depend on new publications. The document retrieval in the AI application shall be able to be started manually by the end user, otherwise it shall be performed at least once a day.

QUESTION 62

Format of reports: Should the detailed report be generated in all three formats (PDF, Microsoft Word, and Excel), or is it acceptable to provide it in just one of these formats?

Réponse / Answer

FR : Il peut être généré dans les trois formats

EN : It can be generated in all three formats.

QUESTION 63

Template of reports: Do you already have a predefined report template, and how many different templates are expected for the output reports?

Réponse / Answer

FR : Aucun modèle de rapport n'est prédéfini, il devra être identifié lors de la phase d'analyse des besoins métier.

EN : No report template are predefined, it will have to be identified during the business requirements analysis phase.

QUESTION 64

Authetification/access control: Is an authentication or user access control module required for the Proof of Concept (PoC) phase, or only in the production stage?

Réponse / Answer

FR : Oui, le module d'authentification ou de contrôle d'accès utilisateur doit faire partie du PoC.

EN : Yes, authentication or user access control module has to be part of the PoC.

QUESTION 65

Password protection: Will access to password-protected files and archives be required during the Proof of Concept (PoC) phase or only in the production stage?

Réponse / Answer

FR : Oui, les fichiers et archives protégés par mot de passe doivent faire partie du PoC.

EN : Yes, password-protected files and archives have to be part of the PoC.



QUESTION 66

AI models: Are there specific frameworks, or AI models preferred or mandated for this project?

Réponse / Answer

FR : Non.

EN : No.

QUESTION 67

IT Infrastructure: Could you please provide more details about "environment within the GovCloud"? What infrastructure resources will be provided for development and testing (e.g., VM specs, GPU availability)?

Réponse / Answer

FR : Il s'agit du modèle NVIDIA 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 NVIDIA GPU Tesla V100, find more informations with the following link [tesla-volta-v100-datasheet-letter-fnl-web.pdf](#) (nvidia.com).

QUESTION 68

GDPR compliance: What specific measures should be taken to ensure GDPR compliance throughout the document processing cycle?

Réponse / Answer

FR : Il n'y a pas de mesures spécifiques à prendre pour assurer la conformité au RGPD tout au long du cycle de traitement des documents.

EN : There is no specific measures to be taken to ensure GDPR compliance throughout the document processing cycle.

QUESTION 69

Budget flexibility: Is there flexibility in the budget concerning the procurement of additional licenses or resources if needed?

Réponse / Answer

FR : Non.

EN : No.



QUESTION 70

Internal team: Is there an internal team designated to take over the system post-contract, and will they be available for collaboration during development?

Réponse / Answer

FR : Oui.

EN : Yes.

QUESTION 71

Partnerships: Would the organization be open to proposals that include alliances or partnerships?

Réponse / Answer

FR : Des groupes d'opérateurs économiques, y compris lorsqu'ils se sont regroupés sous la forme d'une association temporaire, peuvent participer sans qu'il soit nécessaire qu'ils prennent une forme juridique spécifique. Toutefois, les groupes d'opérateurs économiques présentant une offre collective seront tenus solidairement responsables de l'exécution du marché si leur offre est retenue. En cas d'offre collective, l'offre doit donc être accompagnée d'une déclaration de responsabilité solidaire, datée et signée par chaque opérateur économique participant au consortium, dans laquelle les opérateurs économiques désignent un représentant parmi eux. L'offre doit indiquer soit la part assumée dans l'exécution du marché et, le cas échéant, dans chacun de ses éléments, par chacun des opérateurs, soit la contribution proportionnelle de chacun d'eux à l'exécution du marché dans son ensemble ou dans ses différents éléments. L'offre d'un consortium doit également respecter pleinement le droit de la concurrence.

EN : Groups of economic operators, including where they have come together in the form of a temporary association, may participate without it being necessary for them to take on a specific legal form. However, groups of economic operators submitting a collective offer will be held jointly and severally liable for the execution of the contract should their offer be selected. In the event of a collective offer, the offer must therefore be accompanied by a declaration of joint and several liability, dated and signed by each economic operator participating in the consortium, in which the economic operators designate a representative from among themselves. The offer must indicate either the proportion assumed in the performance of the contract, and, where applicable, in each of its elements, by each of the operators, or the proportional contribution made by each of them in the performance of the contract as a whole or in that of its various elements. A consortium bid must also fully respect competition law.

QUESTION 72

Format of data: What format will the system support for ingesting data from sources like Legi-lux, EUR-Lex, ISO etc.? Please explain the Data normalization process? - The system can fail to parse essential clauses, especially if formats (PDF/HTML/DOCX/XML) are not normalized.



Réponse / Answer

FR : Comme indiqué au chapitre 2.2.1 « Spécification des données », toutes les sources et tous les formats de données sont répertoriés. Nous ne fournissons pas d'informations supplémentaires concernant la normalisation des formats spécifiés. Il appartient à l'opérateur économique de déterminer comment les gérer.

EN : As indicated in section chapter 2.2.1 "Data Specification", all data sources and data format are listed. We don't provide additional information related to the normalisation of specified formats. It is up to the economic operator to determine how to manage it.

QUESTION 73

Password protection: Documents may be password protected or require API Authentication. What strategies will be used to manage credentials and access controls effectively? - Missing restricted documents (e.g., draft standards) due to poor credential management.

Réponse / Answer

FR : La gestion des événements (c'est-à-dire des erreurs) lors du téléchargement et du traitement des documents fait partie des bonnes pratiques en matière de développement d'applications informatiques et doit donc faire partie de la solution.

EN : The event handling (i.e. errors) during upload and document processing is part of the good practices in IT applications development and it shall therefore be part of the solution.

QUESTION 74

Key attributes: How will key attributes be captured and stored within the system. Attributes such as Metadata, document source, publication date, jurisdiction etc... - Losing document provenance invalidates audits.

Réponse / Answer

FR : Il appartient à l'opérateur économique de proposer des solutions pour collecter et stocker ces informations. Ces solutions pourraient être intégrées au « Rapport de proposition de solutions (avec avantages et inconvénients) », comme indiqué au chapitre 2.3 « Livrables attendus ».

EN : This is up to the economic operator to propose solutions to capture and store those information. It could be integrated in the "Solutions proposal report (including pros and cons" as indicated in chapter 2.3 "Expected Deliverables".

QUESTION 75

GovCloud data storage: How will the Data be stored and organized within GovCloud after processing?

1) What are the types of Databases that will be utilized (SQL, Non-SQL)



2) What type of indexing or search technologies will be applied (e.g., ElasticSearch)

- Slow retrieval for queries like "Show all CAB requirements from 2023."

Réponse / Answer

FR : Il appartient à l'opérateur économique de proposer des solutions en matière de technologies de recherche. Cela nécessitera au préalable une analyse des besoins métier.

EN : It is up to the economic operator to propose solutions regarding search's technologies. It will require first the analysis of business requirements.

QUESTION 76

AI outputs: What strategies will be implemented or utilized to ensure that AI outputs are unbiased and traceable? Which methods, if any, such as SHAP, LIME or others, will be used for transparency and uphold interests of OLAS stakeholders? - Bias in interpreting "shall" vs. "should" in regulations.

Réponse / Answer

FR : Il appartient à l'opérateur économique de proposer des solutions pour garantir la traçabilité, la cohérence et l'homogénéité des données dans les sorties générées par l'application d'IA.

EN : This is up to the economic operator to propose solutions to guaranty the tracability, consistency and consistency of the data in the output generated by the AI application.