Project specifications:

AI4DI

Analysing requests for information against a knowledge base

Version	Date	Author	Modifications
1.0	09/12/2024	VLE	Initial version + review VBE
1.1	19/12/2024	AHA	Modified
1.2	31/12/2024	VLE	After feedback Ministry of Digitalization
1.3	06/01/2025	MLE	minor modifications for clarification
1.4	07/01/2025	VLE	Final considerations
1.5	20/02/2025	CZO	Minor modification 2.3

Contents

1. Context

1.1 Introduction

The CNPD[1] (Commission Nationale pour la Protection des Données) is the Luxemburgish Data Protection Authority and the contracting authority of the project. One of the CNPD's mission is to provide guidance to organizations and citizens interested in the subject of personal data protection. To this end, it operates a service called "Demandes d'informations" (DI in Short) and receives various questions via different communication channels (email, mail, phone).

Since 2018 and the entry into application of the GDPR, the CNPD has built up a rich collection of questions/answers documented and structured in a Word document.

When a DI is received, the procedure is therefore to search through this collection to see if the question has not already been dealt with in an identical or similar way and then to construct a personalized answer for the new request from the existing text blocks. If it is a new type of DI (new theme, specific variation on a recurring theme, etc.), this DI will be added to the collection after legal analysis.

The collection has in the meantime become so big that the search carried out manually by the agent in charge takes a lot of time. The CNPD therefore wishes to automate this search as well as the production of an initial draft response proposal based on the existing collection. The type of solution required is the development of an IT solution which should be based on an Artificial Intelligence Large Language Model.

1.2 Project Team

Name	Organization	Division	Title / Role
Marc Lemmer	CNPD	Guidance	Commissionner / Sponsor
Vincent Legeleux	CNPD	Guidance	IT Expert / Project Manager
Vanessa Bemtgen	CNPD	Guidance	Acting Head of DI / Legal Expert - User POC
Aurélie Lagoutte	CNPD	Guidance	Legal Expert / User POC
Jan Kuffer	CNPD	Administration	Head of Internal IT / IT Support
Arnaud Habran	CNPD	Guidance	Head of Division / Legal Expert
	Ministry for Digitalisation		Support in Project Management / Financing

---

1.3 Target Audience

The target audience is the CNPD. The project is specifically directed to the CNPD agents answering to the requests for information ("Demandes d'Informations", in short DI). The typical users are legal advisors with no particular IT knowledge. Once the initial goal of searching and writing the response is achieved the tool may be used by all the agents of the CNPD.

1.4 Deliverables - Risks & preventive measures

1.4.1 Expected deliverables

It is expected, within a period of 6 months from kick-off, to complete the deliverables below:

- IT solution that can be hosted and maintained by the CTIE (Centre des Technologies de l'Information de l'Etat)
- IT solution compatible with current IT environment (Sharepoint + Office)
- IT solution able to give an answer under 30s
- IT solution able to learn new reference documents, new entries via supervised learning
- IT solution able to handle multi-language (FR, DE, EN)
- Deliver a solution that is easily useable by a person with no programming knowledge
- A final report and supporting documentation

For more in depth features specifications see 2.1.2.

1.4.2 Risks & mitigating measures

Risk	Probability	Impact	Mitigating action	Responsibility
Inability to use the solution	Low	High	Good knowledge transfer process and training (hands-on workshops and dedicated documentation)	Project manager, experts, consultants
Poisoning of the LLM	Low	High	Create a test set with known results and a procedure to test the model after each Supervised learning.	Experts and consultants
Delayed project results	Medium	Medium	Targeted update meetings, timeline of milestones	All involved parties
Inadequate hardware/software	Low	High	Provide technical requirements	Consultants and CTIE (State technology center)

2. Project Goal

2.1 Work Description

DI: Current Workflow
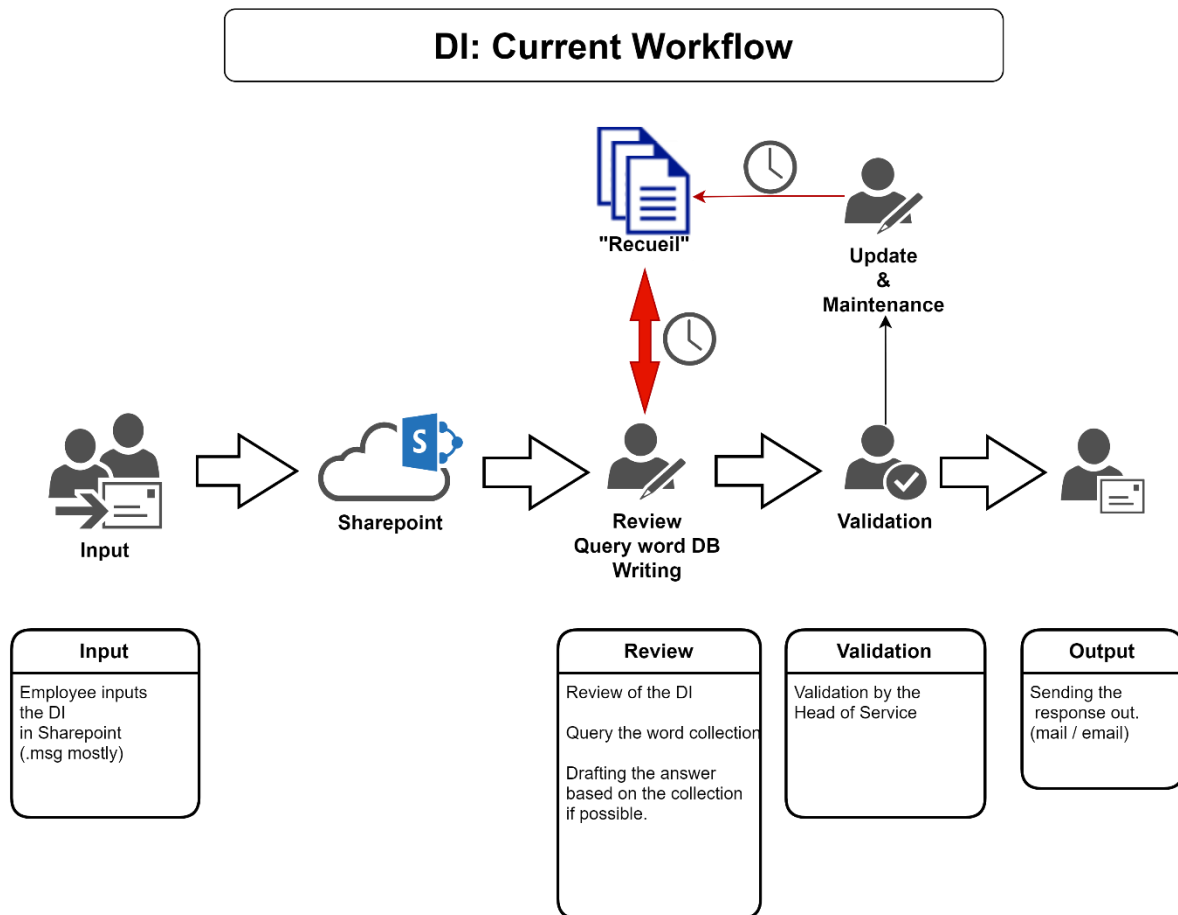
Input:

A user creates a new case in the internal Sharepoint case management system and stores the initial request for information received by the CNPD. The file is mostly in a .msg format because it is an email transferred from outlook. If the initial request is a letter, it is scanned and stored in pdf in the Sharepoint case.

Redaction / Query:

A user of the DI service which has been assigned the current DI case will read the mail/email and search through the word document "Recueil" (Recueil Light.docx) if a similar issue has already been raised and answered. The user copies the question and the identified answer(s) to a draft word document to be transferred in the Sharepoint case. Based on this copy the user writes the draft answer to be submitted for internal validation

If the issue is not found in the "Recueil" a legal analysis is performed by the user to prepare the answer.

Validation:

The final answer is reviewed in Sharepoint by the Head of the Service and marked as ready to be sent or goes back for editing.

A video explaining the process above in detail is available upon request at vincent.legeleux@cnpd.lu

Update:

Currently all the new "DI" of the running year are stored after anonymisation in a new word document (2024_Recueil des réponses DI.docx). The main document (Recueil Light.docx) is regularly updated by a legal expert after duplicates removal.
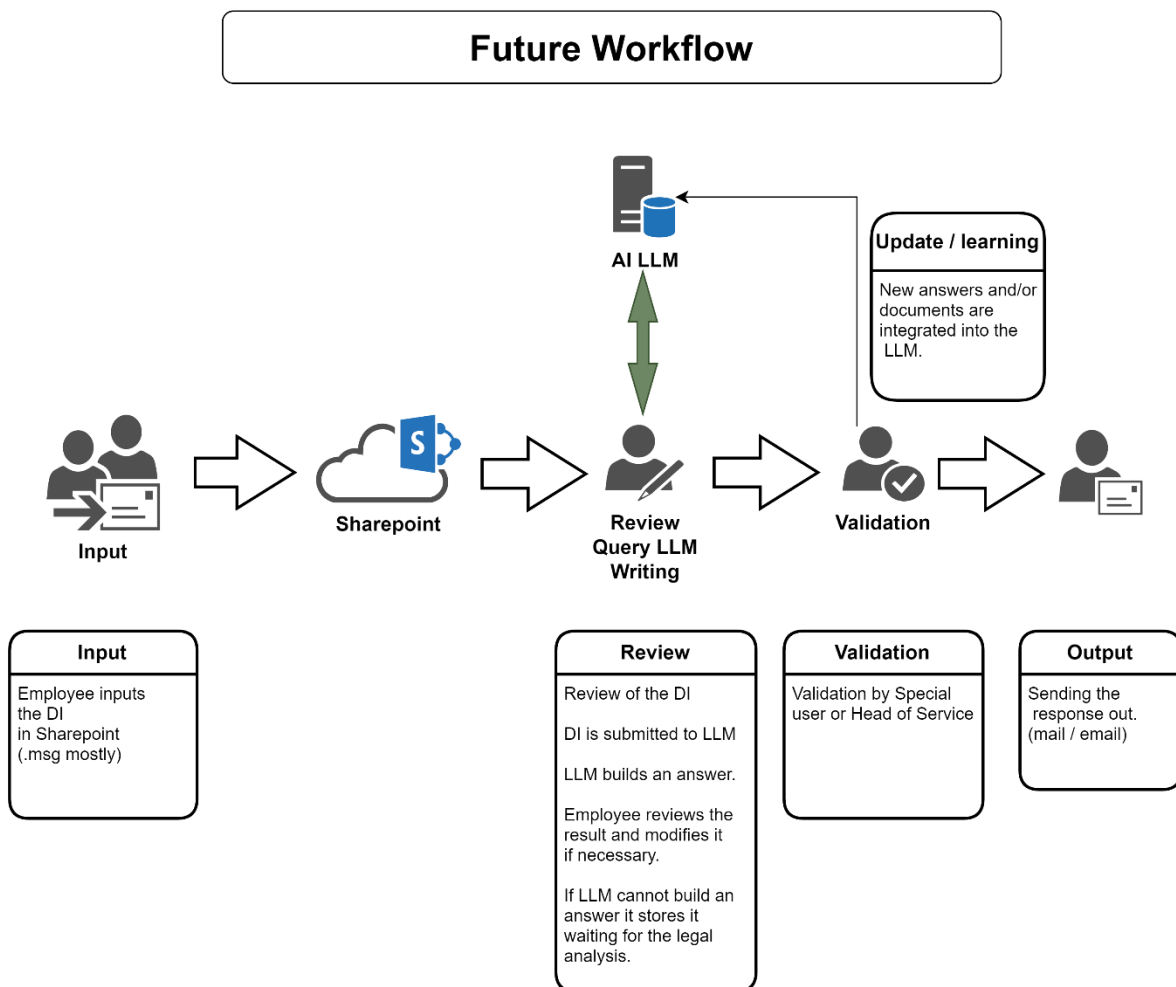
Input (same as previous workflow):

A user creates a new case in the internal Sharepoint case management system and stores the initial request for information received by the CNPD. The file is mostly in a .msg format because it is an email

transferred from outlook. If the initial request is a letter, it is scanned and stored in pdf in the Sharepoint case.

Redaction / Query:

A user of the DI service which has been assigned the current DI will read the mail/email and if necessary, submit it to the LLM by copying the text into the GUI (Graphical User Interface) or importing the pdf. The LLM will - if the issue has already been handled - suggest an answer in a formal way based and all the examples it learned and give as reference the date of the similar cases. If the issue is in relationship with a guidance that has been imported in the LLM, the LLM shall make a reference to it or even suggest building the answer upon it (the LLM should inform the user).

The user interacts with the LLM to adapt the answer (asking to change the tone or asking for more details and/or references). The user copies the answer directly from the GUI to a word document or generates an export pdf to be transferred in the Sharepoint case.

If the LLM has not enough references to build the answer it informs the user that the corresponding answer will have to be submitted to the LLM in a near future (by the agent in charge of the supervised learning).

The LLM should be able to answer all the questions about the content it handles and be specific about the source of the information. Thus, any query can be also made outside of a specific DI workflow.

E.g.:

"According to the EDPB what is needed to consider a data anonymized? Please give the reference of the document."

"What are the articles and recitals talking about third country transfers in the GDPR?"

Validation:

The final answer is reviewed in Sharepoint by the Head of the DI Service and marked as ready to be sent or goes back for editing.

Update:

Each time a new request for information that needs a full legal analysis is received it is processed outside the LLM and integrated during a specific learning process. The special user in charge of this operation will upload a document containing several couples of question & answers (specs to be defined) to be integrated into the LLM.

Anytime the special user may upload new documents (guidance produced by the EDPB or other authorities, laws etc.) to the LLM.

2.1.1 Data Specifications

The available data consists of a word document entitled "Recueil Light.docx" currently consisting of 1396 pages divided in 5 sections depending on the applicable law. The document contains consolidated answers to questions in a form as presented in the following example:

Title: [Year-Month] [Language] [Elements of the Request in french]

2021-10 EN Champ d'application territorial du RGPD dans le cas de traitement des données des résidents et citoyens non-UE par une entreprise européenne

In your message you seem to imply that Articles 3.1 and 3.2 may apply cumulatively. This is not the EDPB' and the CNPD's interpretation: in our view Article 3.1 refers to EU "established" controller and processors, whereas Article 3.2 of the GDPR applies to non-EU established organisations who are targeting data subjects in the European Union. Therefore, Articles 3.1 and 3.2 apply alternatively.

Regarding Article 3.1 of the GDPR, the decisive criterion for its application is whether an organisation has an "establishment" in the EU. There is no definition of establishment in the regulation, however, recital 22 provides information on the meaning to this notion: "establishment implies the effective and real exercise of activity through stable arrangements. The legal form of such arrangements, whether through a branch or a subsidiary with legal personality, is not the determining factor in that respect." In Case C-230/14 Weltimmo, the Court of Justice of the European Union (CJEU) considered that an organisation may be established where it exercises "through stable arrangements in the territory of that member state, a real and effective activity even a minimal one" (see page 6 and reference 7 of the Guidelines 3/2018 on territorial scope). The test to conclude whether a company/organisation has an establishment in the EU is whether necessary technical and human resources are available.

In order to evaluate the notion of "in the context of the activities" of an establishment, we kindly refer you to Case C-131/12 Google Spain v AEPD of the CJEU (paras 47 to 60).

Example 4 of the Guidelines 3/2018 on territorial scope to which you refer, states that the company's main establishment is in France, therefore it refers to a controller established in the European Union. In this case, the company falls under the conditions of Article 3.1 and the criterion of Article 3.2 (i.e. data subjects in the Union) is not applicable.

As your organisation is based in the EU through stable arrangements, Article 3.1 of the GDPR applies. Even if your customers are based outside the EU, you are still required to comply with the obligations of the GDPR, including the principles of Article 5, the requirement to maintain a record of processing activities (Article 30), the obligations relating to security (Article 32), as well as the obligation to cooperate with the EU data protection authorities.

Formatting and links are important, keywords are used. The content is anonymized. More information about the document and a bigger extract can be requested per email (vincent.legeleux@cnpd.lu).

Currently all the new "DI" of the current year are stored in another word document "2024_Recueil des réponses DI.docx" and the main document is regularly updated by a legal expert after anonymization and duplicates removal.

In addition, other kind of documents should be progressively integrated to the LLM, for instance:

- guidance (see example[2]) and opinions of the EDPB (European Data protection Board)
- EU laws like the GDPR[3], ePrivacy, DGA, DSA or AI Act
- relevant national Luxembourg laws [4]
- specific case law such as from CJUE, TA
- Jurisprudence

These documents should be used to build the answers and/or cited as reference in the answers. The users can query those documents through the LLM.

2.1.2 Features specifications & Technology requirements

Multilingual:

The solution should be able to display, understand and output in French, English and German (Luxembourgish is optional but would be a plus). This means that a query can be made in those languages, the resulting answer can be produced by request in any of those languages and the LLM can learn from documents in any of those languages.

Inputs:

For the queries and the supervised learning the solution should be able to accept direct input from its user interface for text and accept files input for usual text based documents including pdf, .msg, docx..

Ability to process scanned documents in pdf it would be a plus.

User-friendly graphical interface and process for uploading new documents and advanced / continued training of the LLM, including manual correction input.

Outputs:

The solution will output directly in the UI like a classic chatbot but should be able to produce a standardised pdf (with date, references, tbd.) also.

Statistics:

The solution should be able to produce statistics about its usage (number of queries / days, frequency of recuring topics, duration of sessions, tbd.)

---

[2] https://www.edpb.europa.eu/system/files/2024-10/edpb_guidelines_202302_technical_scope_art_53_eprivacydirective_v2_en_0.pdf

[3] https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R0679

[4] https://legilux.public.lu/eli/etat/leg/loi/2018/08/01/a686/jo

User management and authentication:

The solution should provide user rights management and SSO authentication in the CNPD's windows environment.

Environments: At least 2 environments should be set up TEST and PROD with a procedure to refresh the TEST.

Hosting:

Integration into GovCloud managed by the CTIE is preferable to an "on premises" solution.

Modularity:

The solution should be modular, based on standards and apis that can be adapted for instance to integrate new kind of input documents.

Ability to change the underlying base LLM to use a different one.

2.2 Rights of inspection

To facilitate both the transfer of knowledge to the contracting authority and quality control, the contracting authority authorizes itself a right of inspection over all deliverables, methods and practices used by the economic operator during the production of the results inherent in these project specifications. The proposed solution and approach must take into account the fact that at the end of the contract the contracting authority wishes to be able to guarantee the maintenance of the project itself. During the project implementation phase, the economic operator grants a right of inspection to the contracting authority in order to allow it to monitor the quality and control the progress of the work.

2.3 Company profile and profile of consultants

Profile of company:

Requirements:

- Should be able to meet and work physically in Luxembourg.
- Can provide support on short notice (within a couple days or less)
- Previous experience with a Luxembourgish state administration is a plus
- Provide a track record of the number of ML-related projects completed, with a short explanation and including contact names for reference

Profile of consultants:

Requirements:

- Communicate (writing, speaking) either in English or French
- Provide an extract of criminal record (Casier judiciaire record 3)
- Has to work at least partly on site.
- Provide a CV demonstrating years of expertise in ML-related analysis

2.4 Security, confidentiality and protection of personal data

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.

A non-disclosure agreement, a personal data processing agreement and a non-competition clause need to be signed by the company selected as well as any person working on this project. Technologies implying a transfer of the contracting authority's data to private cloud services are not allowed in the project. Government cloud service (govCloud) will be used. The source code and all project deliverables will be made available to the contracting authority.

In the event of the processing of personal data on behalf of the contracting authority, the economic operator undertakes to conclude with the contracting authority a data processing agreement compliant with the provisions of Article 28 of GDPR. In addition, the contracting company must provide evidence that it provides sufficient guarantees to implement appropriate technical and organizational measures in such a manner that processing of personal data covered by this project will meet GDPR requirements, that data-protection by design is respected and that the protection of the rights of the data subjects in accordance with the GDPR are ensured.

2.5 Intellectual property rights

The results inherent in these specifications, algorithm, solution, results as well as the data generated are protected by the relevant intellectual property and copyright laws and will remain the exclusive property of the ordering party CNPD for the duration of the project and beyond.

The contracting authority grants no license or authorization regarding the intellectual property rights, which it holds in respect of the project. The economic operator does not have, without the prior written consent of the contracting authority, the right to use the results to market them.

2.6 Communication about the Project

The economic operator does not have, without the prior written consent of the contracting authority, the right to advertise the services provided within the framework of this project. In the event of

communication about the project, the economic operator shall mention in addition to the contracting party, also the Ministry for Digitalisation and the Tech-in-Gov Call for Projects.

3. Execution

3.1 Timeframe

The timeframe is up to six months from kick-off to proof of concept. Therefore, the end of the project is latest at the end of 2025.

3.2 Work Environment

The IT solution should be coded by the company. However, a close collaboration with the project team is required to ensure meeting the objectives. As such, regular progress meetings and workshops will be organized on site.

As the data processed to build the LLM is anonymized and / or public there are no particular limitations in sending them to the IT provider but confidentiality must be still guaranteed. As the CNPD works with the CTIE as IT provider it is crucial that the solution is compatible with the CTIE's technological requirements and availabilities. The point of contact will be Jan Kuffer from the CNPD or directly the CTIE depending on the situation.

The consultants will be able to work hybrid (onsite + remote) but must be present if specifically asked by the CNPD. Management and milestone meetings can be held online (each 2 weeks) via Microsoft Teams but regularly some update and expert meetings should be held on site (at least once per month). The Ministry for Digitalisation will also be involved in giving feedback.

3.3 Phasing of the Project

After the selection process, the next step will be to sign the necessary contracts. The kickoff session will mark the start of this project. Regular support and management meetings will help with the progression of the project.

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.

The following phases are expected in terms of project management by the external party.

3.3.1   Requirements Analysis Phase

Objective: Understand the project's goals, user needs, and technical constraints.

Activities:

- o   Conduct stakeholder interviews and workshops.
- o   Gather and document functional and non-functional requirements.
- o   Analyze existing systems and workflows.
- o   Develop a requirements traceability matrix.

Deliverables:

- o   Requirements document.
- o   Stakeholder approval on requirements.
- o   Initial project risk assessment.

3.3.2   Research and Evaluation Phase

Objective: Explore potential solutions, evaluate options, and define the project's scope and approach.

Activities:

- o   Perform market research or technology evaluation.
- o   Benchmark potential solutions against requirements.
- o   Conduct feasibility studies or proofs of concept.
- o   Assess compliance with industry standards or regulations.

Deliverables:

- o   Technology evaluation report.
- o   Recommendation of technical solutions.
- o   High-level project roadmap.

3.3.3   Definition of Technical Infrastructure

Objective: Design and plan the technical foundation for the project.

Activities:

- o   Create the system architecture and data flow diagrams.
- o   Define hardware, software, and network requirements.
- o   Plan integration with existing systems.

Deliverables:

- o   Technical design document.
- o   Approved system architecture.

3.3.4   Iterative Implementation and Testing

Objective: Build, deploy, and test the system incrementally to meet requirements.

Activities:

- o   Develop features in iterative cycles.
- o   Conduct unit, integration, and user acceptance testing.
- o   Incorporate stakeholder feedback after each iteration.
- o   Resolve identified defects and optimize performance.

Deliverables:

- o   Functional system increments.
- o   Testing reports and defect logs.
- o   Updated documentation reflecting changes.

3.3.5   Final Evaluation Against Acceptance Criteria

Objective: Validate that the final product meets all agreed requirements and is ready for deployment.

Activities:

- o   Conduct a comprehensive review against acceptance criteria.
- o   Perform final user acceptance testing (UAT).
- o   Obtain formal stakeholder sign-off.

o Develop a transition plan to operational teams.

Deliverables:

o Final acceptance report.

o System handover and training materials.

o Post-project review documentation.

3.4 Knowledge transfer

A final knowledge transfer session (minimum 2hrs for the users and 2h for the admin), before the end of the project, should enable the consultants to provide the necessary knowledge to use, understand and maintain the solution to the stakeholder. This session can be held with all involved parties.

3.5 Applicable laws and regulations

Unless otherwise provided herein, the following laws and regulations, as amended, shall apply:

- la loi modifiée du 8 avril 2018 sur les marchés publics (L MP) ;
- le règlement grand-ducal d'exécution du 8 avril 2018 de la loi modifiée du 8 avril 2018 sur les marchés publics (RGD MP) ;
- le règlement grand-ducal modifié du 27 août 2013 relatif à l'utilisation des moyens électroniques dans les procédures des marchés publics et les procédures d'attribution de contrats de concession ;
- la loi du 23 juillet 1991 ayant pour objet de réglementer la sous-traitance (L ST) ;
- la loi du 10 novembre 2010 instituant les recours en matière de marchés publics (L RMP) ;
- la loi modifiée du 8 juin 1999 sur le budget, la comptabilité et la trésorerie de l'Etat (L COMP);
- les prescriptions du Code Civil (C CIV) ;
- le règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (RGPD) ;
- la loi du 1er août 2018 portant organisation de la Commission nationale pour la protection des données et mise en œuvre du règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données), portant modification du Code du travail et de la loi modifiée du 25 mars 2015 fixant le régime des traitements et les conditions et modalités d'avancement des fonctionnaires de l'État (L CNPD) ;
- Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828 (règlement sur l'intelligence artificielle) (R IA) ;

- la loi modifiée du 31 juillet 2006 portant introduction d'un code du travail (C TRAV) ;

3.6 Content of the offer

Must be included in the offer:

- a description of the economic operator,
- CVs detailing the background and relevant experience for each person who will work on the project
- a detailed description of the proposed approach, detailing all stages of the project, the delivering and a reasoning and justification for the chosen approach.
- a summary project plan proposal including milestones, deliverables and timing,
- a financial proposal including all billed costs and hourly/daily fees of involved persons

3.7 Price Regime

The contract falls under the fixed global price regime.

3.8 Submitting an offer

The offer must be sent via email to vincent.legeleux@cnpd.lu with a copy to nt@digital.etat.lu.