

Pre-deployment Questionnaire

Phase 1 - Context and purpose

A. General information:

- a. *Describe the context and the specific purpose of the AI system.*

The Bank plans to enhance its existing infrastructure (traditional MLS) by integrating a GenAI-powered conversational DSS. LoanLens will combine the computed credit score with unstructured data (e.g., documents collected during loan processing and historical customer data with 10-year historical depth). LoanLens will be a chatbot interface, with which humans will be able to dialogue in a simple and fast manner. The specific purpose of the AI system is then to generate a credit score of natural persons to be used as “suggestions” for loan decisions, and to facilitate data extraction and information synthesis about customers, through a friendly interaction in natural language between humans and LoanLens.

- b. *When (dd/mm/yyyy) will the AI system begin operation?*

The deployment of the AI system will occur on 03/06/2025.

- c. *Specify the end date if known or estimate the duration in months/years.*

We estimate that the AI system will be used for at least three years.

- d. *Usage Frequency: How many times per day/week/month will the AI system be used? Provide an exact count.*

The AI system will be used for all the loan applications that the Bank receives from retail customers (natural persons). It is not possible to give an exact number, but based on the history, it is expected that around one thousand requests per week will be processed.

- e. *What are the expected outputs of the AI system and how will they be used in decision-making processes?*

The expected output will be a numerical score ranging from 1 to 100, representing the creditworthiness of customers. This credit score will be available through LoanLens interface, together with the possibility of asking for additional information and pieces of supporting documents. The final decision remains human, and operators can decide autonomously whether to ‘follow’ the AI system's suggestion, or to deviate by providing an explanation.

f. *Is the AI system classified as High-risk under article 6 of the AI Act and why?*

Yes, the LoanLens has been classified as a high-risk AI system since it is used to produce a credit score of natural persons and therefore determining access to and enjoyment of essential private services as stated in Annex III, 5 (b) of the AI Act.

g. *Does the AI system have a GenAI component?*

Yes, LoanLens includes a GenAI component, which allows the answers to be generated. We specify that the credit score remains calculated by the traditional MLS and only 'reported' in natural language by the chatbot. .

h. *Does the GenAI component present systemic risks as defined by the AI Act, and if so, what are the specific factors contributing to this assessment?*

No, it does not fall within Art. 3, par. 65 of the AI Act.

B. Categories of natural persons and groups

a. *Who are the affected persons?*

The affected persons are all loan applicants, belonging to the 'retail' segment of customers (natural persons).

b. *Are there any categories of natural persons that can be considered as more vulnerable groups? What are their specific risks?*

The disproportionate vulnerability between different subgroups of people in AI credit scoring application is a phenomenon widely analysed in the literature, with many points of attention raised. AI-driven credit scoring poses significant risks of exacerbating existing inequalities and creating new ones, particularly for the following subgroups of people:

- Youngest People. Young individuals, especially those just entering the workforce or with limited credit histories, often possess "thin files." AI systems, reliant on extensive data, may struggle to accurately assess their creditworthiness, leading to unfair denials or inflated interest rates.
- Elderly People. Elderly individuals may face unique challenges, such as declining income due to retirement, increased healthcare expenses, or denials of opportunities due to their life expectancy. AI systems, may not adequately account for these life-stage shifts, resulting in an unfair risk assessment. Also, older people may not have as much digital credit-related footprint as younger generations, that could result in less data for the AI to work with.

- People with Disabilities. Individuals with disabilities may experience employment instability, fluctuating income, increased healthcare costs, or denials of opportunities due to their life expectancy. AI systems, lacking nuanced understanding of these factors, could misinterpret these patterns as indicators of higher credit risk, leading to discriminatory outcomes. Moreover, AI systems could also factor in health data, if it is available, in a way that could be discriminating. The instability of employment, and health conditions, can cause patterns in financial history that are not a sign of poor credit worthiness, but a sign of societal inequality.
- Subgroups subject to biases due to sociodemographic characteristics. AI systems are susceptible to inheriting and amplifying biases present in training data, which can disproportionately affect subgroups based on race, ethnicity, gender, or socioeconomic status. Therefore, historical data may reflect discriminatory lending practices, leading to biased AI-driven credit scoring. It is critical to be aware that historical data often contains societal bias, and AI systems may perpetrate or exacerbate those biases in a systematic way.

The groups outlined above are potentially more vulnerable to adverse outcomes in AI-driven credit scoring applications such as LoanLens, due to factors such as data scarcity, life-stage transitions, disability-related challenges, and the multi-faceted risk of algorithmic bias.

c. *Who will directly interact with the AI system?*

The AI system will only interact with the Bank's operators, credit specialists, through a conversational mode. There will be no direct interaction with the Bank's customers.

Phase 2.1 - Deployment process section

- a. What type of algorithm(s) or model will be used? Please provide detailed examples and context for each risk level.
 - i. a non-self-learning algorithm in which humans specify the rules the computer must comply to (risk level 1 to 3)
 - ii. a System that is entirely or partially based on a self-learning algorithm, where the machine itself is finding patterns in the data (choose a risk level from 4 to 10)

RISK LEVEL: 8

COMMENT: The risk level has been chosen considering that the system is composed of a Gen AI component and a narrow AI component. The risk level chosen is 8 since the Gen AI components find patterns in the data.

- b. Among different alternatives suitable to achieve similar goals and performances, do you plan to use a simpler and more explainable algorithm?
 - i. Yes (choose a risk level between 1 and 5)
 - ii. No. Please explain your choice (choose a risk level between 6 and 10)

RISK LEVEL: 7

COMMENT: In this use case, we opted against using a simpler, more explainable algorithm due to performance considerations. While simpler methods may offer greater transparency, they typically cannot match the higher accuracy and efficiency provided by the more complex models.

Deployment process section risk level: 7.5

Phase 2.1 - Input data and Fairness section

- a. Are all input data (from internal and external sources), including third-party training data and data added by the Deployer, governed by data governance and data quality processes?
 - i. Yes, all, in compliance with GDPR, AIA and IP requirements (risk level 1)
 - ii. Only partially (e.g., in case a model has been pre-trained by an external provider and there are no warranties about data used). Please specify (choose a risk level between 2 and 10)
 - iii. No¹

RISK LEVEL: 4

COMMENT: The Narrow AI component is fully compliant with the relevant regulations and IP requirements, since it is designed and developed internally, so, minimizing risks. The GenAI component, being developed externally, introduces some risks related to compliance (which are mitigated by legal clauses) and intellectual property (which cannot be fully guaranteed). Thus, we have assigned a "risk level 4" based on the combination of these factors.

- b. If the AI system includes a GenAI component and it uses personal data to produce the output, could this lead to the generation of unexpected and/or undesired content? Do not count this question for the Risk Indicator calculation

¹ If the Team selected this answer, the Questionnaire shall be suspended.

in case the System has no GenAI components or if it does not use personal data to produce the output.

- i. Yes (risk level 10)
- ii. No (risk level 1)

RISK LEVEL: 1

COMMENT: The GenAI component is designed to respond strictly based on the documents it has received as input. Consequently, it does not generate new or unexpected content; rather, it accelerates the user's access to relevant information without producing any automated decisions or recommendations.

- c. Have you or your Provider analyzed the input data to assess possible biases that could lead to a negative impact on fundamental rights? Please specify your choice and provide detailed examples.

- i. Yes, please specify (e.g. in Fairness Assessment or data exploration, documentation provided by the Provider) (choose a risk level between 1 and 10).
- ii. No²

RISK LEVEL: 3

COMMENT: We conducted a direct assessment of the ML component's input data to identify and address any potential biases that could negatively affect fundamental rights. Additionally, the provider of the gen AI component has assured us that they perform a similar bias analysis on their input data, but we do not have direct access to them.

- d. Have you or your Provider collected all the relevant and available data needed to assess if the algorithm may discriminate against or disadvantage specific population subgroups (e.g., nationality, gender, age, etc.)? Please specify your choice and provide detailed examples.

- i. Yes, please specify (e.g. in Fairness Assessment, data exploration, or in order to do this assessment; choose a risk level between 1 and 10)
- ii. No³

RISK LEVEL: 3

COMMENT: Regarding the ML component, we collected all the relevant and available data to assess if the algorithm may discriminate

² If the Team selected this answer, the Questionnaire shall be suspended.

³ If the Team selected this answer, the Questionnaire shall be suspended.

against or disadvantage specific population subgroups. Additionally, the provider of the Gen AI component has assured us that they perform a similar analysis on their input data, but we do not have direct access to them.

- e. Do you or your provider plan to monitor the trend over time of fairness metrics and/or AI System performance referring to specific population subgroups? Please specify your choice and provide detailed examples.

- i. Yes (choose a risk level between 1 and 5)
- ii. No (choose a risk level between 6 and 10)

RISK LEVEL: 1

COMMENT: We plan a monthly monitoring.

Input data and Fairness section risk level: 2.4

Phase 2.1 - Transparency section

- a. Are the components of the AI System and their outputs explainable, interpretable and/or verifiable ?

- i. Yes, all the components are designed to be explainable, interpretable and/or verifiable. Describe what techniques are employed (Choose a risk level from 1 to 3)
- ii. Not all the components of the AI System are explainable, interpretable and/or verifiable. Describe what techniques are employed (Choose a risk level from 4 to 10)
- iii. No⁴

RISK LEVEL: 3

COMMENT: The ML component is explainable and interpretable. We chose risk level 3 due to the black-box nature of the Gen AI component, even though its outputs are verifiable.

- c. Have you identified the subjects (Provider, Deployer, and Affected persons) for which the output of the AI System shall be made sufficiently understandable?

- i. Yes (risk level 1)
- ii. No⁵

⁴ If the Team selected this answer, the Questionnaire shall be suspended

⁵ If the Team selected this answer, the Questionnaire shall be suspended

RISK LEVEL: 1

d. Are the AI System outputs designed as sufficiently understandable for the Deployer?

- i. Yes (choose a risk between 1 and 10)
- ii. No⁶

RISK LEVEL: 1

e. If applicable, are the AI System outputs designed as sufficiently understandable for the staff and other persons dealing with the operation and use of AI system on behalf of the Deployer? If not applicable, do not consider this question within the Risk indicator calculation.

- i. Yes (choose a risk between 1 and 10)
- ii. No⁷

RISK LEVEL: 1

f. If applicable, are the AI System outputs designed as sufficiently understandable for the affected person? If not applicable, do not consider this question within the Risk indicator calculation.

- i. Yes (choose a risk level between 1 and 10)
- ii. No⁸

RISK LEVEL: n/a

Transparency section risk level: 1.2

Phase 2.2 - Performance section

a. Have you planned to test the AI system performance and proper functioning?

- i. Yes (choose a risk level between 1 and 10)
- ii. No⁹

If yes, specify what is the metric chosen for that purpose and what is the frequency of the testing. On the basis of which evaluation criteria is a metric value considered acceptable (ex. acceptable thresholds)?

⁶ If the Team selected this answer, the Questionnaire shall be suspended

⁷ If the Team selected this answer, the Questionnaire shall be suspended

⁸ If the Team selected this answer, the Questionnaire shall be suspended

⁹ If the Team selected this answer, the Questionnaire shall be suspended

RISK LEVEL: 4

COMMENT: The ML component performance is assessed using accuracy metric as a reference. Gen AI component performance is assessed looking at the rate of correct responses or location of relevant documents at manager-posed queries. We plan to test the proper functioning on a monthly basis.

Performance section risk level: 4

Phase 2.3 - Human Oversight section

a. What is the degree of automated decision-making in the AI system?

- i. The decision is taken by a human being and the AI System provides only an additional layer of information (Choose a risk level between 1 and 2)
- ii. The decision is made by the application of AI System are only executed after human review or approval (Choose a risk level between 3 and 4)
- iii. The decision relies on the AI System, but it is possible for a human to override the outcomes (Choose a risk level between 5 and 10)
- iv. The decision-making process is completely reliant on the AI System, without possibility for overrides¹⁰

RISK LEVEL: 1

b. Can the operation of the AI system be interrupted through a 'stop' button or similar procedure?

- i. Yes, intervention possible (choose a risk level between 1 and 10)
- ii. No, intervention is not possible. Provide details¹¹

RISK LEVEL: 1

COMMENT: The human can interrupt the AI system in each moment of its interaction.

c. Is it ensured that the staff and other persons dealing with the operation and use of AI systems on behalf of the Deployer and the Provider have the means to interact with the AI system and use it in an informed and conscious manner?

¹⁰ If the Team selected this answer, the Questionnaire shall be suspended

¹¹ If the Team selected this answer, the Questionnaire shall be suspended

- i. Yes, those subjects have received a specific training and they can consult technical guidelines or support. Please specify. (choose a risk level from 1 to 4)
- ii. Yes, those subjects have received a specific training but they can not consult technical guidelines or support. Please specify (choose a risk level from 5 to 10)
- iii. No, those subjects have not been trained for the specific purpose¹²

RISK LEVEL: 1

COMMENT: We provide specialized training for managers working with the AI system, offering intensive and mandatory sessions before they can use it. Additionally, we have a detailed set of guidelines for handling AI systems and offer technical support to managers in case of any doubts or malfunctions.

- d. Are there support mechanisms for the subjects mentioned above designed to address issues or concerns with AI system operation?
 - i. Yes (risk level 1)
 - ii. No (risk level 10)

RISK LEVEL: 1

Human Oversight section risk level: 1

Phase 2.4 - AI system Monitoring and Maintenance

- a. If unfair behavior emerges, will it be possible to intervene to correct it?
 - i. Yes, it will be possible to intervene (risk level 1)
 - ii. No, it has been not possible to correct the unfair behavior¹³

RISK LEVEL: 1

- b. Do you plan to adopt monitoring techniques and related countermeasures in order to intercept anomalous behavior or performance deterioration?
 - i. Yes (risk level between 1 and 10)
 - ii. No¹⁴

¹² If the Team selected this answer, the Questionnaire shall be suspended

¹³ If the Team selected this answer, the Questionnaire shall be suspended

¹⁴ If the Team selected this answer, the Questionnaire shall be suspended

RISK LEVEL: 4

COMMENT: We plan to implement monitoring techniques across the entire AI system to detect any anomalous behavior or performance drops. While we can directly apply countermeasures to the ML component, we lack direct access to the Generative AI model provided by an external party. To address this limitation, we will employ a guardrail technique to ensure controlled and secure outputs.

- c. How do you plan to update the AI System (retrain, tune, knowledge source update, etc)?
- i. Update not needed given the AI system's technical features (rule-based system without variations over time) (risk level 1)
 - ii. Not applicable, as in the case of GenAI Systems, either the AI system has been integrated by external providers (i.e. "proprietary models") or its update is economically too burdensome and operationally unnecessary (risk level between 2 and 9).
 - iii. Manual or automatic update planned (e.g., using new data, re-estimation of the hyperparameters of the model, vector store update in RAG applications.) (choose a risk level between 1 and 9)
 - iv. No update planned (risk level 10)

RISK LEVEL: 4

COMMENT: We chose risk level 4 because the gen AI component is periodically updated by the external party and not directly by us.

- d. Have you planned procedures for emergency/urgent updates?
- i. Yes, please describe (choose a risk level between 1 and 5)
 - ii. No (choose a risk level between 6 and 10)

RISK LEVEL: 4

COMMENT: We chose risk level 4 because we can immediately intervene on the ML component in case of an emergency, while we can't immediately intervene on the gen AI component.

- e. Have you planned any processes in order to ensure the ongoing and continued availability of data, coherently to the business needs?

- i. Not needed since not foreseen by business requirements (risk level 1)
- ii. Yes (choose a risk level between 1 and 5)
- iii. No (choose a risk level between 6 and 10)

RISK LEVEL: 1

f. Do you plan to oversee the stability of input data/features used by the AI system?

- i. Yes, overseeing the stability of input data and features has been planned with one or more of these options (choose a risk level between 1 and 5):
 - 1. General and specific quality controls of input data
 - 2. Trend analysis of input data
 - 3. Trend analysis of Model features
 - 4. Specific Data Drift synthetic measures
 - 5. Evaluation of suitability/toxicity of prompts
- ii. Input data quality controls already in place are deemed as effective safeguards and no other actions are planned (choose a risk level between 6 and 9)
- iii. No action planned (risk level 10)

RISK LEVEL: 1

AI System Monitoring and Maintenance section risk level: 2.5

QRI: 3.1

Post-deployment Questionnaire

In this section we describe the post-deployment Questionnaire results. In order to avoid repetition, we will only present the additional question to those already present in the pre-deployment Questionnaire. Please note that in a real scenario the post-deployment Questionnaire is completed only after the AI system has been deployed, incorporating any new information that arises post-deployment. This post-deployment Questionnaire serves a different purpose than the one used during the design phase, so it must be filled out from scratch rather than relying on responses previously provided during the design stage.

Phase 2.1 - Ownership and Control

- a. Does the AI System use software libraries, and/or modules that have been developed by third parties (e.g., pretrained models, outsourced code, ...)? Please specify your choice and provide detailed examples.
 - i. No, AI System has been developed internally and/or using documented and accessible open-source libraries (e.g., pandas, scikit-learn) (risk level 1)
 - ii. Yes (risk level 10)

RISK LEVEL: 10

COMMENT: The AI system includes a gen AI component pre-trained and developed by an external party.

Ownership and control section risk level: 10

QRI: 4.1