Table of contents

Project goal

Developing a powerful tool to aid the municipality of Breda in analysing crime factors and implementing effective strategies for reducing crime rates. This innovative solution empowers policymakers to leverage valuable insights and make informed decisions in order to enhance public safety and facilitate city development.

By utilizing this tool, law enforcement agencies can optimize their efforts by pinpointing areas with high crime rates and utilizing this knowledge to their advantage. This proactive approach enables them to implement targeted measures and resources where they are most needed, ultimately making policing more effective and impactful.

Data

What kind of data will you be using?

- Demographic data

- Education level data: Education level code, neighborhood code, population

- Weather monthly data: count of rain days, rainfall, daylight hours

- Breda crime stats data: neighborhood code, year, month, registered crimes

- Emergency response times data

- Labor participation data: Gender, age, neighborhood code, contracts

- Country migration data: age code, migration background code, generation, population

Datasets come from many different sources:

- Municipality of Breda

- Police

- University-Database

- CBS (Centraal Bureau voor de Statistiek)

Effect of the project

1. Municipality officials and policymakers: The tool will directly impact the municipality officials and policymakers who are responsible for creating and implementing strategies to address crime rates. It will provide them with valuable insights and data-driven evidence to make informed decisions and develop effective policies for crime prevention.

2. Law enforcement agencies: The tool can greatly benefit law enforcement agencies operating in Breda. It will enable them to have access to detailed crime data, identify crime hotspots, and allocate resources more efficiently. This can lead to improved crime detection, prevention, and overall effectiveness of their operations.

3. Community and residents: The project will ultimately aim to create a safer environment for the community and residents of Breda. By reducing crime rates, the tool can contribute to enhancing public safety, fostering a sense of security, and improving the quality of life for the people living in the municipality.

4. Local businesses: Lowering crime rates can positively impact local businesses by creating a more secure and attractive environment for commerce. When businesses and potential investors perceive Breda as a safe city, it can lead to economic growth, increased business activities, and job opportunities.

Benefits of the project

1. Crime reduction: The primary benefit of the project is the potential reduction in crime rates within the municipality of Breda. By analyzing crime factors and implementing evidence-based strategies, the tool can contribute to creating a safer environment for residents and businesses.

2. Data-driven decision-making: The project enables policymakers and law enforcement agencies to make data-driven decisions. By providing comprehensive insights and analysis of crime data,

the tool empowers them to develop targeted interventions and allocate resources more effectively.

3. **Efficient resource allocation:** With insights gained from the tool, the municipality can optimize resource allocation for crime prevention and law enforcement efforts. By identifying high-crime areas and allocating resources accordingly, they can maximize the impact of their interventions and increase operational efficiency.

4. **Proactive crime prevention:** By leveraging the tool's capabilities, the municipality can adopt a proactive approach to crime prevention. Early identification of crime patterns and hotspots allows for timely interventions, preventing potential crimes and reducing the overall crime rate.

5. **Improved public safety:** As a result of reduced crime rates, the project enhances public safety in Breda. This benefits residents, local businesses, and visitors by creating a more secure environment that fosters a sense of well-being and encourages community engagement.

Fundamental Rights

The AI system could potentially discriminate against groups of people inhabiting certain neighbourhoods. The sensitive features can include the age of residents or their migration background.

If bias is identified, take appropriate steps to rectify it, such as adjusting the training data, retraining the system, or modifying the algorithms to ensure fairness and equality. During the development phase, measure the impact of the biases on the result of the model. Use diverse and representative datasets to train the AI system, ensuring that it covers a wide range of demographics and characteristics to minimise bias.

Laws that applied to the development of AI system include: AVG, WOO and WPG.

Human Agency and Autonomy

The AI system is designed to interact with human end-users and guide decision-making processes related to crime prevention. It aims to provide insights on the number of monthly crimes per neighbourhood and offers the opportunity to adjust features in order to lower crime rates.

The potential for confusion is rather low, as the system is built to be an informative tool that uses a dashboard to show insights to the end-users. However, it is still important to ensure that end-users and subjects are informed and made aware that the decisions, advice, or outcomes they receive from the AI system are generated by algorithmic processes.

There is a risk of over-reliance on the AI system, especially if end-users perceive it as infallible or authoritative. Therefore, it is necessary to implement procedures that encourage critical thinking and decision-making by end-users, ensuring they do not overly rely on the AI system's recommendations. Promoting a balanced approach that considers multiple factors and human judgement can help mitigate over-reliance.

The AI system's predictions may focus solely on numerical crime rates without considering the complex social, economic, and environmental factors that contribute to crime. Relying solely on the AI system's recommendations without considering the wider context can lead to ineffective or inappropriate interventions.

To avoid inadvertent effects on human autonomy, it is essential to develop and implement robust safeguards, such as ongoing monitoring and evaluation, bias detection and mitigation, transparency in decision-making, and providing end-users with the ability to understand, question, and challenge the system's recommendations.

Human Oversight

The AI system is self-learning that uses historical data to gather insights on how to make Breda a safer place. Humans engaged in development of the system (Human-on-the-Loop, Human-in-Command) were informed about possible disruptive technology risks and how to develop the system in a safe and ethical way taking into consideration biases. The results of the analysis will be evaluated by humans before being implemented. As the aim of our project is to create a dashboard/tool, the output generated by our AI project will always be used by humans and undergo evaluation before any decisions or actions are taken based on the results. The evaluation process ensures that the insights, predictions, and recommendations provided by our project align with the municipality's goals, policies, and existing legislation.

The detection and response mechanisms for undesirable adverse effects of the AI system for the end-user or subject that was applied in the AI system was: Bias Detection and Mitigation, Ongoing Risk Assessment, and Improvement and Ethical and Legal Compliance. The "stop button" was not applied to the AI system, because the system is an independent dashboard that does not involve any parts of the system within the organisation. This means that the system can be turned off at any moment. The impact of specific variables is examined to avoid bias and to control and explain self-learning of the model.

Technical Robustness and Safety

Resilience to Attack and Security

The AI system could have damaging effects such as design or technical faults that can lead to unreliable outcomes. This is caused by limited access to data used in development and overall character of the academic project.

The system is not certified for cybersecurity, however various actions to make it compliant with specific security standards were implemented. These actions include limited access to Github and vulnerability scanning.

General Safety

Use-cases of the AI system include only actions done by a data scientist team in Breda Municipality. Therefore, risks associated with malicious use are not likely to happen as the AI system will be used only for academic/research benefits.

However, flaws or errors in the design or implementation of the AI system can result in inaccurate predictions or unreliable outcomes. Relying heavily on the system's predictions for decision-making, it can lead to inappropriate actions to lower crime rates, potentially jeopardising human, or societal safety. However, end-users being conscious about possible flaws of the system or development process, will examine the decision-making process. Version control system (GitHub) was used as a form of a parallel system. Each branch can represent a different model or version of the AI system. This approach allowed us for the development, testing, and deployment of multiple models or variations in parallel. By switching between branches, we could compare performance and assess differences in outputs. Additionally, it provided the flexibility to iterate on individual models independently, making it easier to manage and track changes to the AI system over time.

Accuracy

Low accuracy of the model could lead to taking wrong actions to decrease crime rates. The critical consequences could be an increase in crimes and lowering safety levels in the city if the model is implemented in the wrong way. On the other hand, the system is an advising tool for improving city safety, low accuracy does not necessarily mean that the success of the project is not fulfilled as decision-making is done by independent human users.

All the data used in the development of the model is collected from the government. It is complete and representative of the Municipality of Breda, where the system is solely intended to be deployed. Data used is from years 2012-2022.

Accuracy of the model is documented, and different models were used to check correctness of the predictions. Model was evaluated to check the importance of different data points and variables. The system does not contain any operations that could invalidate the data it was trained on.

The end-user's expectation of the accuracy of the AI model was not high, as the access to data and sources for developing were limited.

Reliability, Fall-back plans and Reproducibility

The consequences of an unreliable or not reproductive system can include loss of trust, long adoption process and inaccurate predictions or recommendations. To avoid this, the development process was extended for testing models on different data and evaluating errors. Version control was implemented to make the system reproductible and error-free. Criteria of meeting the intended goals were not specifically distinguished, because the main goal of the system was helping with decision-making and allowing users to use the dashboard to get valuable insights. The AI system was tested on different devices and using different input data to evaluate reproducibility.

Unit tests was developed to check level of the code used during development of the model.

Privacy and Data Governance

Privacy

The AI system is not using personal or private data for its development. The data used was already processed to eliminate possible privacy concerns. Anonymization and de-identification techniques

were applied to all datasets. Process of collecting data was done by a trusted, government system, which can guarantee that data was collected obtaining informed consent from individuals.

The AI system does not impact fundamental rights of a person and only predicts overall crime predictions for the Neighbourhood, not for specific beings. Protected or private data is not included in training data or in development steps of the system.

Data Governance

The data used was not shared or leaked outside the project. The privacy impact assessment was conducted to evaluate the potential privacy risks associated with the non-personal training data or other processed non-personal data. The purposes for which the data is used were clearly communicated, the methods employed to protect privacy, and any rights individuals have regarding their data.

Pre-processing steps of non-personal data aimed to make all neighborhoods of Breda have representation. Pre-processing steps were clearly described in the documentation.

The AI system was not aligned with relevant standards (e.g., ISO25, IEEE26) or widely adopted protocols for (daily) data management and governance.

Transparency

Traceability

Traceability was measured at different steps of development of the AI system. From data collection to final product, the quality was assessed using different measures.

The input data was manually cleared, progressively transformed into a more performant version. The actions that were done to improve quality of the data were: checking duplicates and null values, adjusting date format, EDA of the key variables.

The model that was chosen was the best performing from 5 various models. The models tested in the development process are: XGBRegression, Gradient Boosting Regression, RandomForest Regression and Arima. Different target variables were chosen to assess prediction and vote for the best model. Each member of the development team was active in the model selection process, improving the objectivity of the decision.

The system uncovers meaningful correlations and data points of interest. Before choosing training data, correlations were investigated to ensure higher traceability of the final model. During the model developing process feature importance was analysed for possible bias and inaccuracies. This makes it easier to track which data influences the decision making.

Predictions of the model were evaluated using different measures, such as Mean Absolute Error, Mean Square error and Standard Deviation. Error analysis was implemented, however it did not highlight any specific error types. Logging practices were not implemented.

Explainability

Decisions of the AI system were based on historical data for crimes in the city of Breda. By plotting feature importance, the system is explainable. The other Explainable AI techniques were not applied, due to the Regression nature of the model, which can be easily understood by end-users.

Communication

The benefits are included in the DEDA framework and presented to the end-users using different business tools such as AI Canvas and Project Presentation. The communication strategy for this project aims to ensure effective and transparent information sharing among all relevant parties. Our strategy involves regular and clear communication channels to keep stakeholders informed about the project's progress, milestones, and outcomes.

The potential risks and technical limitations of the AI system are included in section REQUIREMENT #2 Technical Robustness and Safety of this report.

The users are informed on how to use the system by providing clear-designed and informative Dashboard.

Diversity, Non-discrimination and Fairness

Avoidance of Unfair Bias

By assessing quality of the data, analysing correlations and establishing feature importance of the model, the procedures to avoid unfair bias were applied. Pre-processing steps were included in Datasets documentation. In population data, which contained bias against people with migration background, the fairness method "Fairness through Unawareness" was applied to migrate biases.

Data of all Neighbourhoods of Breda was included and arranged to get a fair representation of these areas. Various features were used to develop the AI system in order to show diversity and structural differences within the city.

To get a better understanding of the data, the development team researched official police website's metadata and analyses provided by Municipality of Breda. To choose the best-performing model, different approaches were implemented and discussed within the team, but also with experts and stakeholders.

The potential biases were tested and monitored mostly during data-collection and processing phase. While implementing the model, steps to assess bias included feature importance analysis and performance evaluation for different targets.

The mechanism that allows for flagging issues related to bias or discriminations is a Feedback forms included in documentation of the project and in dashboard. This allows users to raise their concerns.

Possible subjects that could be indirectly affected by the AI system are inhabitants of the Breda as the AI system is designed for helping making decisions about urban adjustments.

Accessibility and Universal Design

The dashboard that is the final interface of the AI system is easy-to-read and accessible for end-users by using minimalist colours and features. The colourblind friendly design is implemented. The decision about implementing an accessible design dashboard was made by the development team. The end-users did not specify accessibility of the AI system.

The Universal Design principles were taken into account during the development process and designing phase. The main characteristics of Universal Design are implemented into the dashboard.

Stakeholder Participation

The stakeholders were participating and controlling development of the system. The mechanism implemented for stakeholders' engagement included: discussion of the development of the project once per week, discussion of possible project risks once a week.

The decision of the design and development process was taken by the development team as well as stakeholders.

Societal and Environmental Well-being

Impact on Work and Skills

The AI system aims to help the Municipality of Breda data scientist to make data-driven decisions about development of the city. Therefore, the system would impact human work as a tool to make it more efficient.

Further implementation of our AI system would be the decision of Breda governments. Actions Municipality would have to take before further actions in their organisation should include informing and consulting with impacted workers and their representatives, ensuring that workers understand how the AI system operates and testing impact of the AI system on the workforce.

The system is not likely to create deskilling risks as Human Autonomy is crucial for implementing the system. There are also no specific skills that could be replaced by an AI system; therefore the tool will not impact overall skilfulness of the workflow.

Impact on Society at large or Democracy

The system was not tested on the wider public. Therefore, it is not clear how the system would impact society. Minimising negative impact of the system included action such as analysing and removing bias in data, developing the system in an ethical, transparent way, and informing end-users about possible limitations and risks connected to the system.

The measures to evaluate AI system impact on democracy have not been taken, however further implementing the system could give insights about this risk.

Accountability

Auditability

There were several mechanisms used for facilitating the AI system's auditability. Methods such as Record Keeping and version control using Github are helpful for checking the progress of the system's development. The documentation of the system informs about data sources and formats. The documentation provides necessary information for project auditability and accountability.

Checklist

General Project Information

1. What is the project about and what is its goal?

2. What kind of data will you be using?

3. Who might be affected by the project?

4. What are the benefits of the project?

5. What are the problems or concerns that might arise in connection with this project?

Fundamental Rights

1. Does the AI system potentially negatively discriminate against people?

2. Which laws and regulations apply to your project?

Human Agency and Oversight

3. Could the AI system affect human autonomy?

4. Is the AI system designed to interact, guide or take decisions by human end-users that affect humans or society?

5. Could the AI system generate confusion for some or all end-users or subjects on whether they are interacting with a human or AI system?

6. Does the AI system risk creating human attachment, stimulating addictive behaviour, or manipulating user behaviour?

7. Did you take any specific oversight and control measures to reflect the self-learning or autonomous nature of the AI system?

8. Did you establish any detection and response mechanisms for undesirable adverse effects of the AI system for the end-user or subject?

9. Will the results of the analysis be evaluated by a human before being implemented?

Technical Robustness and Safety

10. Is the AI system certified for cybersecurity or is it compliant with specific security standards?

11. How exposed is the AI system to cyber-attacks?

12. Did you put measures in place to ensure the integrity, robustness and overall security of the AI system against potential attacks over its lifecycle?

13. Was risk assessment done for such as design or technical faults, defects, outages, attacks, misuse, inappropriate or malicious use?

14. Did you define risks, risk metrics and risk levels of the AI system in each specific use case?

15. Did you identify the possible threats to the AI system and the possible consequences?

16. Did you assess the dependency of a critical AI system's decisions on its stable and reliable behaviour?

17. Did you plan fault tolerance?

18. Could a low level of accuracy of the AI system result in critical, adversarial or damaging consequences?

19. In what ways have you checked the quality of the data?

20. Did you put in place a series of steps to monitor, and document the AI system's accuracy?

21. Did you put processes in place to ensure that the level of accuracy of the AI system to be expected by end-users and/or subjects is properly communicated?

22. Could the AI system cause critical, adversarial, or damaging consequences?

23. Did you consider whether the AI system's operation can invalidate the data?

24. Did you put in place verification and validation methods and documentation to evaluate and ensure different aspects of the AI system's reliability and reproducibility?

Privacy and Data Governance

25. Is your AI system being trained, or was it developed, by using or processing personal data (including special categories of personal data)?

26. Did you consider the impact of the AI system on the right to privacy?

27. Did you align the AI system with relevant standards or widely adopted protocols for (daily) data management and governance?

28. Did you consider the privacy and data protection implications of the AI system's non-personal training-data or other processed non-personal data?

Transparency

29. Did you establish mechanisms to inform users about the purpose, criteria and limitations of the decision(s) generated by the AI system?

30. Did you put in place measures that address the traceability of the AI system during its entire lifecycle?

31. Did you explain the decision(s) of the AI system to the users?

32. What is the communication strategy with regard to this project?

33. Are you gathering the information that is appropriate for the purposes of your project?

Diversity, Non-discrimination and Fairness

34. Did you establish a strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system?

35. Is there a risk that the project will unintentionally create incentives for undesirable behavior?

36. Did you consider diversity and representativeness of end-users and/or subjects in the data?

37. Is your definition of fairness commonly used and implemented in any phase of the process of setting up the AI system?

38. Did you ensure a mechanism that allows for the flagging of issues related to bias, discrimination or poor performance of the AI system?

39. Did you ensure that Universal Design principles are taken into account during every step of the planning and development process, if applicable?

40. Did you consider a mechanism to include the participation of the widest range of possible stakeholders in the AI system's design and development?

41. Did you take the impact of the AI system on the potential end-users and/or subjects into account?

Societal and Environmental Well-being

42. Could the AI system create the risk of de-skilling of the workforce?

43. Does the system promote or require new (digital) skills?

44. Does the AI system impact human work and work arrangements?

45. Could the AI system have a negative impact on society at large or democracy?

Accountability

46. Did you establish mechanisms that facilitate the AI system's auditability?

47. Did you ensure that the AI system can be audited by independent third parties?

Not Relevant

1. Does the AI system respect the rights of the child?

- The current system in question, as well as the data it operates on, are not directly linked or integrated with the implementation or enforcement of child rights. Furthermore, the system does not have any significant influence on sectors specifically involved in the care and well-being of children.

2. Does the AI system protect personal data relating to individuals in line with GDPR?

- Personal data was not used in any step of the development of the AI system. Therefore, protecting this data is not relevant.

3. Does the AI system respect the freedom of expression and information and/or freedom of assembly and association?

- The system in question is not designed or intended to be involved in any way with individual expressions or the freedom of assembly and association. It does not have any impact or influence on these fundamental aspects of individual rights and liberties.

4. Does the AI system simulate social interaction with or between end-users or subjects?

- The primary objective of the system is not focused on simulating social interactions. It does not aim to replicate or imitate human social behaviors or interactions in any significant capacity. The system's main purpose lies elsewhere and is not directly related to simulating social dynamics.

5. Have the humans (human-in-the-loop, human-on-the-loop, human-in-command) been given specific training on how to exercise oversight?

- For the academic purpose of the project, it is determined that conducting additional trainings is not necessary. The existing training and knowledge available are considered sufficient for the successful completion of the project.

6. Did you ensure a 'stop button' or procedure to safely abort an operation when needed?

- The "stop button" was not applied to the AI system, because the system is an independent dashboard that does not involve any parts of the system within the organisation. This means that the system can be turned off at any moment. The impact of specific variables is examined to avoid bias and to control and explain self-learning of the model.

7. Did you red team/pertest the system?

- The nature of the system is unlikely to attract attention or pose a significant risk that warrants red-teaming or penetration testing. The system's design, functionality, or potential vulnerabilities do not present a compelling need for conducting such security assessments.

8. Did you inform end-users of the duration of security coverage and updates?

- In the context of an academic project, the concept of security coverage may not be applicable or relevant. As the focus is primarily on academic research or learning objectives, the project may not involve practical implementation or real-world deployment that requires comprehensive security measures. Therefore, the concept of security coverage can be considered as a potential consideration rather than a practical requirement for the academic project.

9. Did you develop a mechanism to evaluate when the AI system has been changed to merit a new review of its technical robustness and safety?

- In the context of an academic project, the concept of security coverage may not be applicable or relevant. As the focus is primarily on academic research or learning objectives, the project may not involve practical implementation or real-world deployment that requires comprehensive security measures. Therefore, the concept of security coverage can be considered as a potential consideration rather than a practical requirement for the academic project.

10. Did you define tested failsafe fallback plans to address AI system errors of whatever origin and put governance procedures in place to trigger them?

- The AI system is designed and implemented in a way that minimizes the presence of system errors that could potentially disrupt its functionality or compromise its performance. As a result, the risk of significant system errors that could jeopardize the functioning of the system is low.

11. Did you put in place a proper procedure for handling the cases where the AI system yields results with a low confidence score?

- Since the AI system at its current state has only one use case, it may not be necessary to put in place a specific procedure for handling cases where the system yields results with a low confidence score. The use case and its associated tasks can be approached with a focus on ensuring accuracy and reliability within the system's capabilities.

12. Is your AI system using (online) continual learning?

- No, the model would not make use of online continual learning.

13. Did you put in place any of the following measures some of which are mandatory under the General Data Protection Regulation (GDPR), or a non-European equivalent?

- Since the AI system does not utilize personal data or sensitive information in its operations, it is not subject to the regulations imposed by the General Data Protection Regulation

(GDPR). As the GDPR primarily focuses on the protection of personal data and privacy rights, its applicability is contingent upon the handling and processing of such data. Therefore, in the absence of personal data usage, the AI system is not governed by GDPR policies.

14. Do you continuously survey the users if they understand the decision(s) of the AI system?

- As the focus of the project is on academic research and development rather than real-world deployment, there may not be a requirement or opportunity to engage with users for feedback or user interaction. Therefore, at this stage, user communication is not an integral part of the academic project's development process.

15. In cases of interactive AI systems (e.g., chatbots, robo-lawyers), do you communicate to users that they are interacting with an AI system instead of a human?

- Since the AI system is not designed to interact with users in a human-like manner, there is no need to explicitly communicate to users that they are engaging with an AI system rather than a human. The system's purpose and functionality do not rely on creating the illusion of human interaction, and users are aware that they are interacting with an automated system. As a result, there is no requirement to explicitly disclose the AI nature of the system during user interactions.

16. Did you put in place educational and awareness initiatives to help AI designers and AI developers be more aware of the possible bias they can inject in designing and developing the AI system?

- For the academic purpose of the project, it is determined that conducting additional educational and awareness initiatives is not necessary. The existing training and knowledge available are considered sufficient for the successful completion of the project.

17. Did you ensure that the AI system corresponds to the variety of preferences and abilities in society?

- The dashboard being developed is specifically intended as a tool for data scientists and researchers within the Municipality of Breda. As the target group is relatively small and

consists of a specific set of users with similar roles and needs, the need for accommodating a wide variety of preferences may not be applicable. Since the target audience shares similar requirements and expectations, the dashboard can be tailored to meet their specific needs without the need for extensive customization options to cater to diverse preferences.

18. Did you assess whether the AI system's user interface is usable by those with special needs or disabilities or those at risk of exclusion?

- The small group of users who will be utilizing the tool, comprising data scientists and researchers, has not expressed any concerns regarding the design of the dashboard leading to the exclusion of team members.

19. Are there potential negative impacts of the AI system on the environment?

- The system, in its current state, does not have a direct impact on the environment. While the question of environmental impact could be valuable for future development considerations, it is not applicable or relevant at this particular stage of system development.

20. Where possible, did you establish mechanisms to evaluate the environmental impact of the AI system's development, deployment and/or use (for example, the amount of energy used and carbon emissions)?

- To evaluate the environmental impact of the system, it would indeed be necessary to include specific features or considerations within the system. Incorporating environmental impact assessment as a step for further development would enable the evaluation of how the system interacts with the environment and its potential effects.

21. Did you pave the way for the introduction of the AI system in your organisation by informing and consulting with impacted workers and their representatives (trade unions, (European) work councils) in advance?

- As the development team for the AI system, our actions are limited and should not influence organization of Breda Municipality. If the AI system is intended for

implementation within the Municipality, it would be the responsibility of the organization to inform and consult with impacted workers and their representatives.

22. Did you adopt measures to ensure that the impacts of the AI system on human work are well understood?

- Due to the early stage of development, it may not be applicable or feasible to measure the impact of human work on the system. At this initial phase, the focus is primarily on building and refining the system itself rather than evaluating the impact of human efforts involved in its development. Measuring the impact of human work typically becomes more relevant as the system progresses towards a more advanced stage of development or deployment where the contributions of individuals can be assessed and quantified.

23. Did you foresee any kind of external guidance or third-party auditing processes to oversee ethical concerns and accountability measures?

- In the context of an academic project application, third-party auditing may not be applicable or necessary. The academic project is typically subject to internal evaluations, reviews, and assessments conducted by academic advisors or project supervisors to ensure its academic rigor and adherence to project goals. Therefore, third-party auditing is not typically a part of the academic project application process.

24. Did you organise risk training and, if so, does this also inform about the potential legal framework applicable to the AI system?

- The knowledge of the potential legal/ethical frameworks were included in a study material provided by academic experts. Therefore, for the project risk training was not planned as the team members were provided with necessary information to conduct individual ethical contribution.

25. Did you consider establishing an AI ethics review board or a similar mechanism to discuss the overall accountability and ethics practices, including potential unclear grey areas?

- Regarding the requirement for the ethical contribution of the project, there was no prior plan or intention to conduct an ethical review for the project.

26. Did you establish a process to discuss and continuously monitor and assess the AI system's adherence to this Assessment List for Trustworthy AI (ALTAI)?

    - In the case where the AI system development and testing do not extend beyond the scope of the academic project due to its academic structure, establishing a dedicated process to discuss, monitor, and assess the system's adherence to the Assessment List for Trustworthy AI (ALTAI) may not be applicable. The academic project's limited scope and duration might not warrant the establishment of an ongoing process specifically for ALTAI adherence.

27. Did you establish a process for third parties (e.g. suppliers, end-users, subjects, distributors/vendors or workers) to report potential vulnerabilities, risks or biases in the AI system?

    - If there is no involvement of third parties, such as suppliers, end-users, subjects, distributors/vendors, or workers, in the project, then establishing a process for them to report potential vulnerabilities, risks, or biases in the AI system may not be applicable. Since there are no third parties directly involved, there would be no need to establish a reporting process specifically for them.

28. For applications that can adversely affect individuals, have redress by design mechanisms been put in place?

    - If the project is to be implemented by the Municipality, it is important to develop and establish such redress mechanisms. This includes creating channels for individuals to report issues, providing processes for reviewing and addressing complaints, and implementing appropriate remedies or corrective actions when necessary. By incorporating redress by design mechanisms, the Municipality can enhance the fairness and accountability of the AI system's implementation, ultimately benefiting the individuals affected by it.