

Ethical framework

While working on this project we are of course keeping the legal and ethical framework in mind. This includes answering the DEDA and AITAI frameworks in this document as well. The checklists for the same have been answered below. I will also provide a link to the download path for the whole document here:

REQUIREMENT #1 Human Agency and Oversight

1. Does the AI system potentially negatively discriminate against people on the basis of any of the following grounds (non-exhaustively): sex, race, color, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation?

When making a model like this, you always keep these things in mind. It is, however, almost impossible. There are sadly almost always innate biases when working with machine learning models. Moreover, we are building a model that uses data that might be sensitive and we make an effort into securing the data in a proper manner.

2. Does the AI system respect the rights of the child, for example with respect to child protection and taking the child's best interests into account? Have you put in place processes to address and rectify potential harm to children by the AI system? Have you put in place processes to test and monitor for potential harm to children during the development, deployment and use phases of the AI system?

We are making our model with the child's best interest at heart, and making sure to protect not only the child but everyone's best interests with our model.

3. Does the AI system protect personal data relating to individuals in line with GDPR?¹⁶ Have you put in place processes to assess in detail the need for a data protection impact assessment, including an assessment of the necessity and proportionality of the processing operations in relation to their purpose, with respect to the development, deployment and use phases of the AI system? Have you put in place measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data with respect to the development, deployment and use phases of the AI system? See the section on Privacy and Data Governance in this Assessment List, and available guidance from the European Data Protection Supervisor.

We keep in mind the GDPR. We do this by making it mandatory to log in to the website. This way we lock any data that might unintentionally harm anyone behind a password. Furthermore, this log-in code will only be provided to the “product owner” and the municipality of Breda. We do this to keep our model secure.

4. Does the AI system respect freedom of expression, information, and/or freedom of assembly and association? Have you put in place processes to test and monitor for potential infringement on freedom of expression and information, and/or freedom of assembly and association, during the development, deployment and use phases of the AI system? Have you put in place processes to address and rectify potential infringement on freedom of expression and information, and/or freedom of assembly and association, in the AI system?

Yes, Our AI system respects freedom of expression, information, and/or freedom of assembly and association. We are making a very basic model that does not infringe on any of these basic rights.

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

No, our model only gives outputs based on the information it has been given. Then it predicts how much the other variable needs to be in order to keep the correlation intact. What is done with these predictions is up to the end user. Our model predicts the crime rate for the asked year, what is done with that information is down to the end user.

- Could the AI system generate confusion for some or all end-users or subjects on whether a decision, content, advice or outcome is the result of an algorithmic decision?

No. Our AI system gives you one value at the end of its prediction: The number of policemen/BOAs that will be necessary for a year that has been requested.

- Are end-users or other subjects adequately made aware that a decision, content, advice or outcome is the result of an algorithmic decision?

Yes, in our dashboard the inner workings of the model will be explained and made easy to understand for the end user. Furthermore, we document all the code in this document as well. This will make it easier for the end user to understand and use our model.

- 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?

No In our dashboard, we will make it very clear that they are working with a machine-learning model. We do this by only making it give you outputs

based on what basic information you give it. This will make it so there is no other interaction that could resemble that of a human.

- Are end-users or subjects informed that they are interacting with an AI system?

Yes. The path that leads towards the model outputs will be named in such a way that it's clear that the end user is interacting with a machine learning model.

- Could the AI system affect human autonomy by generating over-reliance by end-users?

No. Our model is only a tool that will be very useful in one singular situation. This makes it so it has no additional value besides that one specific situation.

- Did you put in place procedures to avoid end-users over-rely on the AI system?

Yes. As stated before, our model has only a simple goal and that goal is just an output of numbers. It will be very difficult for end-users to become over-reliant on the model.

- Could the AI system affect human autonomy by interfering with the end-user's decision-making process in any other unintended and undesirable way?

No. As stated before, our model functions to assist the end-user by giving them suggestions for how we might need to improve our

police force with the number of officers. How many BOAs will be necessary for a year will be the main metric.

- Did you put in place any procedure to avoid that the AI system from inadvertently affecting human autonomy?

The way we avoid affecting human autonomy is by way of not making our model too complex. We have it calculate a simple metric based on the model's inputs and have that as our output. These outputs are merely a suggestion.

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

Not necessarily. The model calculates a simple metric based on the info it has been given and then gives an output.

- Does the AI system risk creating human attachment, stimulating addictive behavior, or manipulating user behavior? Depending on which risks are possible or likely, please answer the questions below:

Our model does not create human attachment or anything in that regard since it just calculates a simple metric based on the given inputs and has it as its output.

- Did you take measures to deal with possible negative consequences for end-users or subjects if they develop a disproportionate attachment to the AI System?

Our model does not invoke such things since it's so simple and does not resemble any kind of human interaction.

- Did you take measures to minimize the risk of addiction?

Our model does not invoke such things since it's so simple and does not resemble any kind of human interaction.

- Did you take measures to mitigate the risk of manipulation?

Our model does not invoke such things since it's so simple and does not resemble any kind of human interaction.

- **Please determine whether the AI system (choose as many as appropriate):**

- **Is a self-learning or autonomous system;**
- **Is overseen by a Human-in-the-Loop;**
- **Is overseen by a Human-on-the-Loop;**
- **Is overseen by a Human-in-Command.**

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

Yes. We are currently doing a study on it. We have been well-informed on how to make and maintain these models.

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

We are overseeing every moment the model is running, we do this to catch any undesirable effects and patch them out immediately so we can deliver a proper product.

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

Yes. There is a built-in stop button in the environments we use.

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

No. The model is not complicated enough to require this.

REQUIREMENT #2 Technical Robustness and Safety

Resilience to Attack and Security

- Could the AI system have adversarial, critical, or damaging effects (e.g. to human or societal safety) in case of risks or threats such as design or technical faults, defects, outages, attacks, misuse, inappropriate or malicious use?

No. it will not. It's a very simple model.

- Is the AI system certified for cybersecurity (e.g. the certification scheme created by the Cybersecurity Act in Europe) or is it compliant with specific security standards? • How exposed is the AI system to cyber-attacks?

No. it doesn't need to be.

- Did you assess potential forms of attacks to which the AI system could be vulnerable?

No. since it's not that advanced.

- Did you consider different types of vulnerabilities and potential entry points for attacks such as:

- Data poisoning (i.e. manipulation of training data);
- Model evasion (i.e. classifying the data according to the attacker's will);
- Model inversion (i.e. inferring the model parameters)

Yes. but it does not apply to our model.

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

○

No, our model is not complex enough to make it necessary.

○ Did you red-team/pentest the system?

Not applicable.

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

Not applicable.

- What length is the expected timeframe within which you provide security updates for the AI system?

Not applicable.

General Safety

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

Yes. We have brainstormed and looked at our potential risks.

○ Did you put in place a process to continuously measure and assess risks?

Yes. The model gives constant output, making it unnecessary.

○ Did you inform end-users and subjects of existing or potential risks?

Yes, we will be informing the end-users how to use our model. But the model comes without risks seeing as it's not a complex model.

- Did you identify the possible threats to the AI system (design faults, technical faults, environmental threats) and the possible consequences?

Yes. We have brainstormed and looked at our potential risks.

- o Did you assess the risk of possible malicious use, misuse, or inappropriate use of the AI system?

Yes. We have brainstormed and looked at our potential risks.

- o Did you define safety criticality levels (e.g. related to human integrity) of the possible consequences of faults or misuse of the AI system?

Yes. We have brainstormed and looked at our potential risks.

- Did you assess the dependency of a critical AI system's decisions on its stable and reliable behavior?

Yes. We have brainstormed and looked at our potential risks.

- o Did you align the reliability/testing requirements to the appropriate levels of stability and reliability?

Yes. We have brainstormed and looked at our potential risks.

- Did you plan fault tolerance via, e.g. a duplicated system or another parallel system (AI-based or 'conventional')?

No. We did not, seeing as our model is not complex enough

- 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?

We are evaluating our model constantly throughout the project lifecycle. Updating it constantly.

Accuracy

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

No, it can not.

- Did you put in place measures to ensure that the data (including training data) used to develop the AI system is up-to-date, of high quality, complete, and representative of the environment in the system will be deployed?

Yes. We have, seeing as we are using data from the municipality of Breda. Furthermore, we clean all the data before using it.

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

Yes.

- Did you consider whether the AI system's operation can invalidate the data or assumptions it was trained on, and how this might lead to adversarial effects?

Yes. This is why it needs to be fed new and updated information if it wants to stay operational.

- 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?

Yes.

- Could the AI system cause critical, adversarial, or damaging consequences (e.g. pertaining to human safety) in case of low reliability and/or reproducibility?

No.

o Did you put in place a well-defined process to monitor if the AI system is meeting the intended goals?

Yes. We have been documenting all our progress when working with the model so far. We do this to see if the model has the capabilities to reach our end goal.

o Did you test whether specific contexts or conditions need to be taken into account to ensure reproducibility?

Yes. We are using datasets from different years to validate the reproducibility of our model.

- Did you put in place verification and validation methods and documentation (e.g. logging) to evaluate and ensure different aspects of the AI system's reliability and reproducibility?

Yes. We are logging everything we are working on in this document as a form of code documentation.

o Did you clearly document and operationalize processes for the testing and verification of the reliability and reproducibility of the AI system?

Yes. We are logging everything we are working on in this document as a form of code documentation.

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

We do not need a fallback plan since our model isn't advanced enough for that.

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

Yes. We will be giving it different data to improve its confidence score.

- Is your AI system using (online) continual learning?

No.

- o Did you consider the potential negative consequences of the AI system learning novel or unusual methods to score well on its objective function?

Our model does not have these learning methods so we do not need something like that.

REQUIREMENT #3 Privacy and Data Governance

Privacy

- Did you consider the impact of the AI system on the right to privacy, the right to physical, mental, and/or moral integrity, and the right to data protection?

Yes. We always try to keep the legal and ethical framework in mind when working on our AI-centered project.

- Depending on the use case, did you establish mechanisms that allow flagging issues related to privacy concerning the AI system? Data Governance This subsection helps to self-assess the adherence of the AI system('s use) to various elements concerning data protection.

Yes.

Data Governance

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

No.

- 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?

1. Data Protection Impact Assessment (DPIA)²³;
2. Designate a Data Protection Officer (DPO)²⁴ and include them at an early stage in the development, procurement, or use phase of the AI system;
3. Oversight mechanisms for data processing (including limiting access to qualified personnel, mechanisms for logging data access, and making modifications);
4. Measures to achieve privacy by design and default (e.g. encryption, pseudonymization, aggregation, anonymization);
5. Data minimization, in particular, personal data (including special categories of data);

Our project does not mention any personal data of any kind. These are not applicable.

o Did you implement the right to withdraw consent, the right to object, and the right to be forgotten into the development of the AI system? o Did you consider the privacy and data protection implications of data collected, generated, or processed over the course of the AI system's life cycle?

Our project does not mention any personal data of any kind. This is not applicable.

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

Yes.

• Did you align the AI system with relevant standards (e.g. ISO25, IEEE26) or widely adopted protocols for (daily) data management and governance?

Our project does not mention any personal data of any kind. This is not applicable.

REQUIREMENT #4 - Diversity, Non-Discrimination, And Fairness

Traceability

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

Yes. This document will serve as part of that.

o Did you put in place measures to continuously assess the quality of the input data to the AI system?

No. Data that is used needs to be assessed beforehand.

o Can you trace back which data was used by the AI system to make a certain decision(s) or recommendation(s)?

Yes. Since you know exactly what info you're feeding the model. You can trace back what information it used to make the decisions.

o Can you trace back which AI model or rules led to the decision(s) or recommendation(s) of the AI system?

Yes. Through the same method.

o Did you put in place measures to continuously assess the quality of the output(s) of the AI system?

Yes.

o Did you put adequate logging practices in place to record the decision(s) or recommendation(s) of the AI system?

Yes.

REQUIREMENT #5 - Diversity, Non-Discrimination, And Fairness

Avoidance of Unfair Bias

1. Did you establish a strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design?

Yes. We are following the ethical guidelines mentioned in the GDPR framework.

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

We are working on the crime statistics related to different Breda neighborhoods for our project. The average annual income of Breda residents is also shown. Therefore, the likelihood of bias or discrimination in these data is extremely low. So I can confirm that sure, we have thought about the effects of our data not being sufficiently representative.

- a) Did you test for specific target groups or problematic use cases?

No. It was not necessary.

- b) Did you research and use publicly available technical tools, that are state-of-the-art, to improve your understanding of the data, model, and performance?

Yes

- c) Did you assess and put in place processes to test and monitor for potential biases during the entire lifecycle of the AI system (e.g. biases due to possible limitations stemming from the composition of the used data sets (lack of diversity, non-representativeness)?

Yes. This is because we are following the GDPR conventions.

- d) Where relevant, did you consider the diversity and representativeness of end-users and or subjects in the data?

Yes.

3. 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?

Yes. We first ensured that everyone knew what each other was doing and thus followed all the right practices and educational and awareness initiatives. We filled the DEDA framework to ensure that we were ethical and considerate.

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

Yes. We have given different members of the group responsibility for detecting bias, discrimination, or poor performance.

a) Did you establish clear steps and ways of communicating on how and to whom such issues can be raised?

Yes. We made sure to communicate all the requirements when it comes to ethics at the beginning and there is also constant checking for this as well.

b) Did you identify the subjects that could potentially be (in)directly affected by the AI system, in addition to the (end-)users and/or subjects?

Yes, and we also talked about the consequences that the people who will be affected by this will face i.e. criminals. And since we are also taking the income and crime rates per neighborhood, we won't be focusing on any individual group/person.

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

Yes. We are following different ethical guidelines to ensure this as well. It was implemented during the process of collecting and cleaning the data.

a) Did you consider other definitions of fairness before choosing this one?
No. This was something I always intended on using.

b) Did you consult with the impacted communities about the correct definition of fairness, i.e. representatives of elderly persons or persons with disabilities?
No, since it is required for our dataset as mentioned above.

6. Did you ensure a quantitative analysis or metrics to measure and test the applied definition of fairness?

We had considered this, however, the data set does not contain the vulnerable data and hence fairness metrics are not required in our case.

Accessibility and Universal Design

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

The project our group is working on focuses on the crime rates per neighborhood and we also take into consideration other factors such as education and income. Since this is mainly being worked on to improve Breda, it is quite evident that the AI system corresponds to preferences in society. What we are building is just a basic idea that the municipality can improve based on their needs. We are also constantly incorporating user feedback into the development process to help identify areas where the AI system may not adequately correspond to the preferences and abilities of users.

2. 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?

This interface is not accessible to citizens. It is going to be used by the municipality to improve the quality of life of the citizens of Breda.

a) Did you ensure that information about, and the AI system's user interface of, the AI system is accessible and usable also to users of assistive technologies (such as screen readers)?

No. Since it is going to be used by the municipality to improve the quality of life of the citizens of Breda.

b) Did you involve or consult with end-users or subjects in need for assistive technology during the planning and development phase of the AI system?

It was never mentioned during our consults with the end users.

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

This is applicable only to a very small extent because our final project is going to be a website and it is often very easy to use.

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

The primary focus of the project is to ensure that consequences are appropriately applied to individuals who engage in criminal activities. The project aims to maintain a professional approach by adhering to legal frameworks, due process, and the principles of justice. The objective is to hold criminals accountable for their actions through lawful means, promoting a safe and secure environment for all members of the community.

a) Did you assess whether the team involved in building the AI system engaged with the possible target end-users and/or subjects?

We maintain ongoing consultations with the end users to ensure their active participation and engagement in the project. By involving them in the process, we aim to enhance their understanding of the potential outcomes and gather valuable insights into their specific needs and requirements. These consultations serve as a means to foster effective collaboration, promote transparency, and align our efforts with the expectations and aspirations of the end users.

b) Did you assess whether there could be groups who might be disproportionately affected by the outcomes of the AI system?

Yes. There aren't any groups.

c) Did you assess the risk of the possible unfairness of the system onto the end-users or subject's communities?

Yes.

Stakeholder Participation

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

No. We have not come up with a specific mechanism to include the participation of the stakeholders. However, we have frequent meetings with the involved parties to ensure that the progress of the group is up to their expectations.

REQUIREMENT #6 Societal And Environmental Well-Being

Environmental Well-being

1. Are there potential negative impacts of the AI system on the environment?
Which potential impact(s) do you identify?

There can be some negative impacts on the environment when it comes to the AI model such as increased energy consumption, environmental consequences of data center infrastructure, e-waste from improper disposal of AI hardware, biases in decision-making leading to discriminatory environmental outcomes, etc.

2. 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)?

We have not come up with a mechanism yet.

a) Did you define measures to reduce the environmental impact of the AI system throughout its lifecycle?

No. Mainly because what we are doing right now particularly does not produce a lot of environmental impacts

Impact on Work and Skills

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

Yes. It may change certain work arrangements because the algorithm might try to predict areas with higher crimes and hence the municipality might have to make better arrangements to take care of that.

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

No, it didn't seem necessary.

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

Yes. By doing the DEDA framework and completing this checklist we are getting a better idea of the ethical considerations when it comes to the project.

a) Did you ensure that workers understand how the AI system operates, and which capabilities it has and which it does not have?

Yes.

4. Could the AI system create the risk of deskilling the workforce?

No one is responsible for the task or outcome of our model.

5. Does the system promote or require new (digital) skills? Did you provide training opportunities and materials for re- and up-skilling?

No, the model does not require new digital skills.

Impact on Society at Large or Democracy

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

No. It could not have a negative impact. In fact, it can only have a positive impact since our project aims on reducing crime rates in different neighborhoods of Breda.

a) Did you assess the societal impact of the AI system's use beyond the (end-)user and subject, such as potentially indirectly affecting stakeholders or society at large?

Yes. The only people this project is going to affect are those committing crimes.

b) Did you take action to minimize the potential societal harm of the AI system?

No, as this is not applicable to us.

REQUIREMENT #7 Accountability

Auditability

1. Did you establish mechanisms that facilitate the AI system's Auditability (e.g. tractability of the development process, the sourcing of training data and the logging of the AI system's processes, outcomes, positive and negative impact)?

Yes. We have ensured to make a report on everything. The product underwent various stages, including conceptualization, design, and development. The data for the report was gathered from open-source data provided by the Gemeente. We had weekly consultations with the people involved in the project.

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

Yes, since this project is a part of our university curriculum, we have ensured to document everything and make everything easy for understanding.

Risk Management

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

No. Not yet. We are still in the process of it. However, our mentor will make sure that we have followed all the ethical guidelines.

a) Does the involvement of these third parties go beyond the development phase?

Yes.

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

The DEDA framework covers data management, ethics, design considerations, and regular audits, while the ALTAI framework emphasizes accountability, liability, transparency, auditability, and interpretability. These frameworks serve as comprehensive methodologies that guide our ethical decision-making throughout the product lifecycle, promoting trust and integrity in our operations.

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

Yes.

4. 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)?

Yes. We are ensuring this by following the Altai handbook and completing all the checklists.

a) Does this process include identification and documentation of conflicts between the 6 aforementioned requirements or between different ethical principles and explanation of the 'trade-off' decisions made?

Yes. We have made sure to complete all of the aforementioned requirements.

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

Yes, redress-by-design feedback mechanisms have been diligently implemented for applications that can potentially have adverse effects on individuals.