Team GreenScape
Ethical and Legal assessment checklist of the project for Municipality of Breda

Team GreenScape - Kacper Janczyk, Simona Dimitrova, Matyas Tatar, Yuliia Bobrovytska, Collin Limoncelli-Buyskes

Breda University of Applied Sciences

Nitin Bhushan

June 23rd, 2023

Contents

Fundamental Rights Impact Assessment (FRIA)

1. Does the AI system potentially negatively discriminate against people on the basis of any of the following grounds (non-exhaustively): sex, race, colour, 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?

> The AI system is assembled using non-personal data, which means the negative affects it can cause should be non-existent. The use case for the presented system is merely a decision aid. It does not have the ability to make decisions, rather, to provide information and predictions.

Have you put in place processes to test and monitor for potential negative discrimination (bias) during the development, deployment and use phases of the AI system?

> Not needed as the AI system is not working with personal data that contains that kind of information, and it doesn't make judgements on it. Rather, it even highlights areas of segregation and ultimately motivates against it.

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?

> The system takes the child's best interests into account by aiming to create a more nature oriented city in which it can enjoy his/her childhood.

Have you put in place processes to address and rectify for potential harm to children by the AI system?

> The system is merely meant as a visualization and guiding tool which has no regard to age, gender, nationality, or people individually at all. All it is meant to do is provide a clear

understanding of the city's health and through this allow the Gemeente to make decisions that increase its wellbeing.

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?

The system is merely meant as a visualization and guiding tool which has no regard to age, gender, nationality, or people individually at all. As a result, it will not be in contact with children, or near them.

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

Yes, it does not work with personal data at all.

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?

Yes, we have created Legal and Ethical Checklists to ensure that the system cooperates with guidelines set forth by the European Union, and have come to the decision that data protection is an option, but not a necessity. Since all data used is publicly available, and has been provided by the municipality, the safeguards implemented need not be extensive, or cause concern if data is leaked, as it has no greater influence over the reputation of the city, or its municipality.

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?

Since the system does not work with any personal data whatsoever, there is no need of systems to safeguard it from being leaked. There are some measures to ensure that the publicly available, processed data we have used is not accessible to everyone.

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

Yes, it has no effect on these fundamental rights.

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?

No, since the project goal is not concerned with any fundamental rights, rather with the geographic wellbeing of the city.

Have you put in place processes to address and rectify for potential infringement on freedom of expression and information, and/or freedom of assembly and association, in the AI system?

An Ethical and Legal checklist has been assembled by each team member for their respective datasets, which will be aggreggated and the supervisor of the team will be consulted over this document in order to ensure its accuracy, and receive suggestions on further safeguards to implement.

Assessment List for Trustworthy AI (ALTAI):

Assessment 1: Human Oversight

Overview: The assessment primarily revolves around examining how the AI system affects users' autonomy and behaviour, particularly in the context of AI systems that provide

guidance, influence, and support for human decision-making. It specifically tackles concerns regarding the uncertainties surrounding algorithmic outcomes, user awareness and interaction with AI systems, overreliance on AI, potential interference with decision-making processes, and the risks of human manipulation. *(Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-assessment, 2020)*

a. Human Agency and Autonomy
    i. Ensure that the AI system supports human decision-making rather than replacing it entirely.
        1. The goal of the project is solely to support human decision making, and the system has no capabilities to make decisions by itself.
    ii. 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?
        1. No, the AI system component of the final deliverable is clearly labeled, and merely shows measurements resulting from a change in one of multiple variables.
    iii. Could the AI system affect human autonomy by generating over-reliance by end-users.
        1. Possibly, but the reliance effectively ends at gaining insight of what needs to change in order to increase the green score of the city.
    iv. Did you put in place any procedure to avoid the AI system inadvertantly affecting human autonomy?

1. Yes, the fact that the system only provides information, and is incapable of decision making by itself neutralizes the risk of affecting human autonomy.

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

1. The greatest measure of social interaction between the AI system and end-users is changing measurements of a few variables to see how it affects the green score. Otherwise, it is purely informative and does not converse or interact with humans directly.

Assessment 1.2 Ensuring Responsible AI Governance and Accountability

Overview: This assessment emphasises the integration of human involvement throughout the AI system's operation. It encompasses three key measures: human-in-the-loop, human-on-the-loop, and human-in-command. The human-in-the-loop measure ensures that humans have the ability to intervene in each decision cycle of the system. The human-on-the-loop measure focuses on monitoring the system's functioning. Lastly, the human-in-command measure guarantees that a human maintains oversight over the AI system's activities, including its societal impact, compliance with legal regulations, and adherence to ethical considerations. *(Data Ethics Decision Aid (DEDA), n.d.) (Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-assessment, 2020*

b. Human Oversight

i. Please determine whether the AI system (choose as many as appropriate):

1. Is overseen by a *Human*-in-Command

ii. Have the humans (human-in-the-loop, human-on-the-loop, human-in-command)
been given specific training on how to exercise oversight?
    1. Yes, through the instructions of the dashboard and the presentation to the municipality.

iii. Did you ensure a 'stop button' or procedure to safely abort an operation when needed?
    1. The only operations of the AI system are calculation of green score based on the change in value of a relating factor. A 'stop button' is not necessary.

iv. Did you take any specific oversight and control measures to reflect the self-learning or autonomous nature of the AI system?
    1. The AI system is not autonomous nor self-learning, so this ethical guideline can be excluded from the assessment.

Assessment 2: Technical Robustness and Safety

Assessment 2.1 Reliability

Overview: The assessment of reliability centres around the need for the AI model to consistently deliver reproducible results, ensuring reliability in its outcomes.(*Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-assessment*, 2020)

DISCLAIMER: Many of the assessment criteria in this section are not applicable due to the nature, and intended use of the AI system. Furthermore, the system is not advanced enough to be autonomous, or cause harm in any way.

c. Resilience to Attack and Security

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

1. The Model is not susceptible to Cyber Attacks, at it is meant to be an offline presentation which informs the decision makers of the Municipality of Breda to make decisions toward a better environment.

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

1. Integrity of the data largely relies on correct pre-processing, which has been done through continous peer-review and evidencing. What has been done to preprocess the data can easily be viewed through the project repository. System integrity and robustness has been ensured by unit testing.

Assessment 2.2 Safety

Overview: This assessment is responsible for assuring users regarding potential or/and existing risks and their identification, by implementation of adequate measures. In the end, it provides general safety of an AI system.(*Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-assessment*, 2020)

d. General Safety:

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

1. Risk levels are close to non-existent. We have clearly stated that the relationships and correlations derived by the system are not set in stone, since correlation does not equal causation.
2. Possible risks of the system end at decision making. If the municipality decides to install more solar panels, or try to reduce carbon emissions, the biggest threat is that it does not achieve the intended purposes of raising the green score.

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

1. Did you assess the risk of possible malicious use, misuse, or inappropriate use of the AI system?
    a. Yes, and the risk is close to zero. Malicious use is not a possibility since there is no information presented which can cause harm. Furthermore, it aims to neutralize environmental harm by making the city greener. Design faults and technical limitations have been highlighted in the ’Conclusion’ section of the dashboard.

Assessment 2.3 Accuracy

Overview: This assessment is aimed towards ensuring datasets used in AI systems, where they are of high quality, adequate to date and full. It also focuses on the matter of the data being nullified by how the AI model performs and whether there are actions to guarantee some level of accuracy of this model. This also connects to considerations of biases in the dataset, which can lead to inaccurate predictions. (*Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-assessment*, 2020) (*Data Ethics Decision Aid (DEDA)*, n.d.)

e. Accuracy

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

1. No, since the prediction of green score is also assessed by the end user, and the system itself does not make any decisions, only presents information. If incorrect correlations are drawn, the biggest threat is loss of money in a venture which does not yield a higher green score. However, all of the decisions made based on the predictions of the model have to go through multiple channels and checks to result in tangible action.

ii. Did you put in place measure to ensure that the data used to develop the AI system is up-to-date, of high quality, complete, and representative of the environment the sysem will be deployed in?

1. Yes, the goal of the Exploratory Data Analysis was to ensure the completeness and quality of the government provided data.

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

1. Yes, multiple models were trained for predicting the green score, and their accuracy was used to select the best performing model. Collins Models, Kacpers Model (final model used), Matyas's Models, Simonas Models, Yuliias Models

iv. 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?

1. Yes, which is why we included visualizations of all the datasets where changes in values can clearly be observed, and correlations can be drawn without a system.

v. 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?

1. Yes, the Municipality will be made aware that correlation does not equal causation, which is why we excluded CO2 emissions and renewables information from the prediction system, since correlation between those and the green score seem to be non-existent, however apparent it may be.

f. Reliability, Fall-back plans and reproducibility

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

1. Public safety is one of the measurements included in predicting the green score, which means that the system is specifically designed to increase safety as a means of increasing the green score. Furthermore, reproducibility is ensured through the use of a dependency management system, and clearly documented process of development in the project repository.

ii. 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?

1. Yes, unit testing, logging and linting, docstrings, and sphinx have been implemented.

2. Privacy and Data Governance
    a. 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?
        i. Yes, and the conclusion is that the AI system does not violate any of the aforementioned subjects. This is because of the didactic, and informational nature of the system.
    b. Depending on the use case, did you establish mechanisms that allow flagging issues related to privacy concerning the AI system.
        i. Since our use case does not include any privacy sensitive information, such a system is not necessary. However, a security page details contact forms in order to make complaints regarding any vulnerabilities of the project.
    c. Is your AI system being trained, or was it developed, by using or processing personal data (including special categories of personal data)?
        i. No, it was trained on GDPR compliant, publicly available, governmentally procured and provided data, which nullifies the need for extensive privacy related safeguards, or a Data Protection Officer.
    d. Did you implement the right to withdraw consent, to object, or to be forgotting into the development of the AI system?
        i. No, since the system does not actively collect data of its users, only presents them with information about how the city is doing, and how the green score would be affected, if one of the included values changed.

e. 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?
    i. Yes, if the project were to be released to the public, protective processes may be implemented to ensure the integrity of the data, however, if a leak were to happen, the police, green index, and liveability data are the only privacy sensitive data that may be at risk. However, their contents do not provide malicious opportunities to exploit since they are general in nature.
f. Did you consider the privacy and data protection implications of the AI system's non-personal training-data or other processed non-personal data?
    i. Yes, and at this stage, with its intended deployment level and scope of end-users, no further processes are required since the data worked with is purely statistical. The most sensitive data are police data, income data, green score, and liveability index which are generalized and have no privacy threatening effects. Their integrity, and accessiblity are ensured by the need for permission to access the private repository.

Assessment 3: Communication, Traceability & Explainability

Assessment 3.1 Traceability

Overview: This assessment focuses on adequate documentation process of development of AI systems, such as data preprocessing steps. This data should be accurately evaluated in terms of quality, based on the project; providing sources of the data is recommended. (*Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-assessment*, 2020) (*Data Ethics Decision Aid (DEDA)*, n.d.)

DISCLAIMER: Most questions regarding traceability and explainability refer to decision making, or recommendations by the AI system. In the case of GreenScape, the AI system does not make any decisions, or recommendations. It calculates an output based on an input, and the resulting green score is assessed by the end user. No decision, recommendation, or action is taken to any degree by the model itself, and questions relating to these processes are excluded.

g. Traceability
    i. Did you put in place measures that address the traceability of the AI system during its entire lifecyle?
        1. Did you put in place measures to continously assess the quality of the input data to the AI system?
            a. Yes, manual analysis of the data was done before training the model, and the data collected has been cleaned and curated by each team member. Unit testing also ensures the continuous functioning of the system.
    ii. Did you put in place measures to continously assess the quality of the outputs of the AI system?
        1. Yes, unit testing has been implemented in order to ensure integrity.

Assessment 3.2 Explainability

Overview: This assessment relates to the ability of explanation of processes regarding technique and decision-making of the AI models to the end-user. It's important to keep in mind to adequately inform the end-user about how this algorithm works and what results

were established afterwards. (*Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-assessment*, 2020) (*Data Ethics Decision Aid (DEDA)*, n.d.)

h. Explainability

i. Did you explain the decisions of the AI system to the users?

1. Yes, the use of a Random Forest Regressor is explained, and the variables used to predict the green score.

ii. Do you continously survey the users if they understand the decisions of the AI system?

1. Yes, if questions arise we are more than happy to answer them, but since the sytem makes predictions without decisions, it is more about visualizing and understanding how one variable affects the green score rather than anything else.

Assessment 3.3 Communication

Overview: This assessment's focus in terms of AI systems is to communicate and deliver information on goals, advantages, restrictions and means to the end-user.(*Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-assessment*, 2020)

i. Communication

i. In cases of interactive AI systems, do you communicate to users that they re interacting with an AI system instead of a Human?

1. Our system is minimally interactive, and the section of the dashboard where the machine learning algorithm is implemented is clearly labeled.

ii. Did you establish mechanisms to inform users about purpose, criteria, and limitations of the decisions generated by the AI system?

1. Did you communicate the benefits of the AI system to the users?
    a. Yes, we have clearly established the goal of the project through the proposed research questions. The approach to investigating the questions and building a model which correlates the different sectors is documented and can be reproduced. The ’Introduction’ and ’Conclusion’ page of the dashboard gives insights into the possible uses of the algorithm.
2. Did you communicate the technical limitations and potential risks of the AI system to users, such as its level of accuracy and/or error rates?
    a. Yes, evaluation metrics are presented and explained along with the algorithm which was used. The central pain-point of any project of this nature is also highlighted, correlation does not equal causation. Even if the model predicts an increase, it will be up to the human decision makers to figure out how to go about achieving it.
3. Did you provide appropriate training material and disclaimers to users on how to adequately use the AI system?
    a. Yes, a short explanation is included in the final deliverable, along with a live demo. However, interpreting how to change a sector in favor of increasing the green score accordingly to the model

requires further critical analysis. If the user changes the value in the 'Job Growth' section, decision makers still have to figure out how to achieve that change.

Assessment 4: Equity and Inclusivity

3. Diversity, Non Discrimination and Fairness
    a. Avoidance of Unfair Bias
        i. 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 algorithmic design?
            1. Yes, we clearly established a common consensus that our data was unlikely to contain biased data. This is due to the fact that it doesn't concern individuals, rather populations, neighborhoods, etc. The only possibly sensitive data to this criteria is income, which cannot be exploited to cause harm to citizens due to the design of the algorithm and final product. The algorithmic design is meant to be randomized, and every training batch receives an equally likely distribution of each data point. This further ensures that no bias is present in the algorithm's predictions.
        ii. Did you ensure a mechanism that allows for the flagging of issues related to bias, discrimination or poor performance of the AI system?
            1. We have open lines of communication through the school and included contacts throughout the project. A security page is also included in the repository, which details the contacts for reporting vulnerabilities.

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

1. Since our project is not concerned with distribution or acting autonomously, fairness implementation is not necessary. However, we have critically assessed the outcomes of the data and our model, and provided a 'Conclusion' section in our Dashboard which describes the limitations of the system.

b. Accessibility and Universal Design

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

1. Yes, we have made it intuitive, and easy to use, with little jargon.

ii. 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?

1. The target audience and our employer did not mention the need for such precautions, since the extent of use for the system is not meant for the long term, or for the wider public.

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

1. In almost all aspects yes, but this criteria is not entirely applicable if in the same context as the one for those at risk of exclusion.

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

1. Yes, the common agreement is it will only have positive impacts, as the aim of the project is a better, greener, and healthier Breda.

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

1. Yes, we have all discussed this topic thoroughly in the starting stages of development, where we clearly outlined the intentions and uses of the system.

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

1. Yes, which is partly why we included income as one of the correlated factors. However, the influence should be positive, as lower income areas should start to see more greenery if the overall green score of the city needs to be increased.

vii. Did you assess the risk of possible unfairness of the system onto the end-user's or subject's communities?

1. Yes, and we believe more greenery, more jobs, more income, and higher liveability for everyone are things that should not have an 'unfair' category.

c. Stakeholder Participation

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

1. Yes, through the Agile Scrum methodology we have been able to have a Product Owner who represents the views of the

municipality, and occasional conferences with the municipality itself in order to ensure that the project vision is in line with their requirements.

4. Societal and Environmental Well-being
    a. Environmental Well-being
        i. Are there potential negative impacts of the AI system on the environment?
            1. No, the goal of the project is to create a better, and healthier environment for Breda by taking a holistic approach to the factors which may influence the green score.
        ii. 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)?
            1. No, the environmental impact of development is miniscule (energy required for 5 laptops to operate about 8 hours a day for 8 weeks). The use of the system may result in possible positive changes for the environment, not stemming from operating the system itself.
    b. Impact on Work and Skills
        i. This section is excluded because the designed system is an easy to use tool which has little to no impact on the workforce itself, and may only generate new jobs as a result of decisions made based off of its predictive numbers.
    c. Impact on Society at Large or Democracy

i. Since the AI system is not meant to cause harm, but create progress towards a more sustainable future, this section is also excluded. Negative Impact possibilities have been assessed and no conceivable, or feasible idea has been found.

Assessment 5: Accountability

5. Accountability
    a. Auditability
        i. Did you establish mechanisms that facilitate the AI system's auditability (e.g. traceability of the development process, the sourcing of training data, and the logging of the AI system's processes, outcomes, positive and negative impact)?
            1. Yes, the entire development is documented, can be recreated and evaluated by a third party through the use of the Github repository
        ii. Did you ensure that the AI system can be audited by independent third parties?
            1. Yes, through the use of github anyone with the required permissions can audit the project and evaluate it's accuracy and use case.

Assessment 5.2 Risk Mitigation and Control

Overview: The risk mitigation and control assessment places significant importance on the involvement of an external ethics specialist who actively monitors and evaluates the project's

cohesion to the ALTAI and DEDA Framework throughout its entire lifecycle.(*Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-assessment*, 2020)

b. Risk Management

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

1. We consulted with an in-house ethical and legal counsel who assisted in the methodology of following legal and ethical guidelines during the development of the project. No third-party assistance was sought.

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

1. No, risk is minimal and are mentioned in the deliverable along with the presentation, legal frameworks are included alongside the final deliverable.

iii. 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?

1. We have conferred among the team and come to agreements, but since the project is not extensive and does not utilize personal data, most of these criteria apply very loosely, if at all. However, Zhanna Kozlova (Lecturer) (MA) has played a key role in establishing and ensuring the compliance to ethical and legal standards.

iv. 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)?

    1. Yes, through the Agile Scrum and use of Trello Board we have been able to keep adherence to ethical and legal frameworks at the forefront of our project.
    2. Identification and documentation of conflicts with these frameworks have been documented, exclusion of criteria has clearly been explained.
    3. Training was provided to those included in the process of creating the checklists for ethical and legal adherence.

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

    1. Does this process foster revision of the risk management process?

        a. We have open lines of communication through the GreenScape github page and multiple other channels through which the Municipality can contact the team. If the Project was to be made public, more systems could be put in place as needed. A Security.md file has been added to inform third parties of the channels in order to report vulnerabilities.

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

1. Not applicable since the application foreseeably has no chance of adversely affecting anyone. The project is also not intended for public release.

Data Ethics Decision Aid (DEDA)

1. Algorithms
    a. Does the project make use of an algorithm, or some form of machine learning or neural network?
        i. Yes, it uses a Random Forests Regressor algorithm to predict the change in Green Score based off of several factors.
    b. Is there someone withing the team who can provide an explanation that is accessible to the wider public?
        i. Yes, multiple people in the team can provide an explanation, and an informational document can be created to attach alongside the system for the public to understand if the need arises to deploy it to the wider public.
    c. Is there someone who can explain how the algorithm in question works?
        i. Yes, all team members are able to explain the functionality of the algorithm implemented (Random Forests for Regression).
2. Souce
    a. Where do the datasets come from?
        i. They are provided by the municipality, and are publicly available.
    b. In what ways have you checked the quality of the data?
        i. We have performed an Exploratory Data Analysis where we have cleaned the datasets to not contain any missing values, and have checked for consistency within the datasets, and externally

researching other measurements which may come up, these have been specified in the Data Quality Report document.

c. Does the team have an expiration date?
    i. Yes, the 23rd of June is the project deadline, however, contact can still be maintained in order to work on the system if needed.

3. Anonymization

a. Ensure that the datasets provided by the Municipality of Breda do not contain personal data or information that may cause bias.
    i. This has been ensured through the Exploratory Data Analysis process.

b. Should the data be anonymized, pseudonymized, or generalized?
    i. There is no need to do any of them, since the data doesn't contain personal information.

c. Verify that the data is collected and used in accordance with relevant data protection laws and privacy regulations.
    i. The municipality has provided the data on the premises of complying with their privacy regulations which detail not sharing data outside of the organization. Police data is protected under a Non-Disclosure Agreement, but still does not contain personal information through which individual officers could be identified. Green Score Index and Liveability Index are private datasets provided by the municipality, and are treated as such by privatizing the github repository.

4. Visualization

a. How will the result of the project be presented? Are the results suitable for visualization?

i. The final deliverable is a Streamlit dashboard which includes overview visualizations of the data, as well as the implemented Random Forests algorithm which showcases the change in green score based off of the selected criteria.

b. What alternative ways of visualizing the results are there?

i. A PowerBI dashboard is also feasible, but due to the easy implementation of the machine learning algorithm, Streamlit is more favorable. Furthermore, individual EDA visualizations are available for viewing in the github repository.

5. Access

a. Who has access to the data and under what condition?

i. The data is publicly available except for Green score, and liveability index. Access is granted under the condition of belonging to either the Buas or Gemeente Breda organization, and complying with privacy regulations.

b. How is access monitored?

i. To access the private repository, access needs to be granted by one of the team members. This ensures that the project and its components are only visible to people who have contacted us and are agreement with the privacy statements required by the municipality. People with access can be viewed in the collaborators and

6. Sharing, Reusing and Repurposing

a. Are any of the data suitable for reuse? If so, under what condition and for what new purposes could they be reused?

i. The datasets we have cleaned and uploaded can be reused for other purposes, but those purposes reach their full extent at gaining insight. There isn't much that can be done with the data otherwise. Drawing conclusions and making decisions based off of the data can only be done with human intervention, so there is no risk of losing control or negatively affecting people or the city.

b. Are there any obligations not to make the data publicly available? If you were to provide open access to parts of the data, what opportunities and risks might arise?

i. Yes, the municpality has directly requested to not share the data with people outside the scope of the project. If we were to release the data, more useful insights may be gained, but for the foreseeable future, no harm can come from it since the informations included are general measurements. The liveability and green score data have been provided from a private dataset, and with respect to this the entire project is made private on github.

7. Reponsibility

a. Which laws and regulations apply to your project?

i. Data Protection laws such as GDPR, Ethical and Regulatory frameworks regarding AI. However, since it won't be released to the wider public and will be kept within the project team, and due to the general nature of the data, these frameworks have no greater affect on the outcome of the project, other than being stored in a private repository.

b. Who is ultimately responsible for the project.

i. The GreenScape team, and in extension, the Product Owner.

c. Are the duties and responsibilites of that person clear, with regard to this project?

i. Yes, both the product owner, SCRUM master, and the rest of the team members have a solid grasp of Ethical and Legal components regarding the project. We are all responsible for ensuring that the project is inline with the frameworks that have been established.

d. Is the project suitable for cooperation with commercial partners? If so, which parties could that be?

i. Yes, construction companies, solar companies, and renewable production companies can all be included as cooperators in order to make Breda greener, healthier, and safer.

8. Communication

a. What is the communication strategy with regard to this project? Are all parties involved in agreement to this strategy?

i. The project team is using the Agile Scrum framework in order to cooperate, and utilizing Microsoft Teams, and Trello as its main form of communication. Along with github for the central project repository, and whatsapp for further communication outside the requirements of the Scrum framework.

b. What communication strategies are there for cases in which something oges wrong, and who is responsible for them.

i. Microsoft Teams is used as the main communication channel, and daily stand-up meetings ensure smooth progress of the project. If errors occur, github commits are the tool of choice to resolve issues with

code, and team members can all be contacted through whatsapp or teams. We are responsible as a team for the whole of the project, but if something goes wrong, the team member responsible for the component is the first line of defense. If they need help, the team is more than willing to cooperate to figure it out.

9. Transparency
    a. Does the project risk generating public concern or outrage?
        i. Minimally. Since the datasets include CO2 emissions, a miniscule portion of the population may be upset about the amount of CO2 produced, but visualizations show a clear, and substantial decrease. Otherwise, the project is good about minimizing the risk for public outrage since it doesn't contain inflammatory information, and only acts as an informational to assist in decision making for a healthier Breda.
    b. How transparent are you about this project towards citizens?
        i. Since they are not involved at any stage of the project, they don't require knowledge about it. If the municipality decides to release the project, it would be the equivalent of publishing their Breda in Cijfers website, which has the same objective as our project, without the added feature of predicting the green score based off of multiple factors. The streamlit dashboard is also intuitive and includes an introduction, a project vision, and explanations. If they require more transparency, they can request it through the municipality, who, along with Greenscape, are able to provide greater explanations to the data used, and the algorithms implemented. The only privacy concern raised is

regarding the liveability score and green index data, and the measures taken to ensure their security and integrity are a private github page which can only be accessed with permission.

c. Do citizens have the opportunity to raise objections to the results of the project?
    i. Since the project is not intended for public release, citizens objections are not considered. If the municipality decides to release it anyway, feedback from the public is taken with open arms. However, the Municipality itself, and the product owner are both representative of the citizens in the sense that they wish to make Breda a more pleasant place to live. Their views should be synonymous with the greater population.
d. Can citizens opt out of their involvement in the project? If so, when and how can they do this?
    i. Citizens are not included in the project, only material measurements. The furthest that citizen involvement goes is the Police Data, which they can opt out from if the project is released to the public, and they have an objection. However, since the Police Data does not include specific information about individual officers, only a quantitative report of how many officers are at a scene, it doesn't qualify as citizen involvement.

10. Privacy
    a. Is there a data protection officer or data privacy officer involved in this project?

    i. Since the project does not utilize personal data, or sensitive data, such an officer is not necessary.

b. Have you conducted a Privacy Impact Assessment or Data Protection Impact Assessment?

    i. No, a DPIA is required only if the project is likely to involve a high risk to people's personal information, and our project does not utilize data of this kind at all.

c. Does this project make use of personal data?

    i. No, all data provided is publicly available, and published by the government, so it complies with GDPR, and the data included in the project does not include any personal data, only quantitative measurements.

d. Does the data provide insight into the personal lives of citizens?

    i. No. The greatest extent of personal data is Average income, and segregated areas. However, these cannot be tied to any individual, neither can individual measurements be derived from them.

11. Bias

a. As a member of the project, what outcomes do you expect?

    i. To find factors that influence the green score, possibly determining carbon footprint, and ultimately, assisting the Municipality in making data driven decisions in order to make Breda a better place to live.

b. Is there anything about this project that makes you uneasy?

    i. No, since no personal data is used, and the goal of the project is informational, risks are minimized.

c. Will the results of the analysis be evaluated by a human before being implemented?
    i. Yes, EDAs are double checked, and calculated correlations can be checked against patterns in the data. If there is no correlation, the data points are not taken into account when predicting green score. Only strong correlations are considered as helpful in determining what the city can do to make Breda greener, and better.
d. Is there a risk that your project could contribute to discrimination against certain people or groups?
    i. No, since the project does not use personal data, and is aimed at making the city better for everyone.
e. Are all relevant citizens adequately represented within your data? Which oens are missing or under-represented?
    i. Does not apply, since citizens are not represented at all besides how average incomes, along with segregation, influence the green score. (Segregated areas have lower incomes, and as a result a lower green score since there is less development of green areas)
f. Are you gathering information that is appropriate for the purposes of your project?
    i. Yes, personal data is not used at all since individual people do not contribute to the green score of the city. Rather, it is a culmination of socioeconomic, geographic, and environmental factors which contribute to the green score of the city.
g. Is there a feedback loop in the model that might have negative consequences?

    i. There is no feedback loop inside the model, or deliverable. It is informational, and the feedback travels from the municipality/product owner to Greenscape. Through this simple pipeline we are able to make adjustments if needed, and risks are minimized.

h. Is there a risk that the project will unintentionally create incentives for undesirable behavior?
    i. Since the project is meant for the eyes of the municipality only, this risk is minimized. If released to the public, the appearance that segregation and income may be the cause of some areas having less greenery, and consequently, slightly lower quality of life, may result in uproar for equality. However, this happening is highly unlikely.

i. Function creep: can you imagine a future scenario in which the results of your project could be misused for alternative purposes?
    i. No, the results of the project and the possible intentions are very narrow. Making negative decisions based off of the green score predictions of our model is unsensible and highly unlikely. The team can not think of any adverse decisions drawn from the predictions.

j. Do your answers change when you consider possible long-term effects? Why?
    i. No, since in the long-term, the function of the model will remain the same. All of the variables and predictions contributing to the increase in green score should, result in a positive change for society. If negative consequences were to take place, the use of our model would not be of much help in masterminding a dystopian plan.

k. Openness, Respect for Autonomy, Efficiency, Integrity, Inclusion, Justice, Legitimacy.

i. All steps taken in assembling the project can be found in the public github repository (Openness). Like stated before, the algorithm and the final deliverable does not function autonomously, and only aids in decision making. It does not replace a human decision maker, but assists them. Efficiency is achieved by excluding irrelevant variables from the predictive algorithm (segregation variables, emissions and renewables). Integrity is ensured by the Data Quality Report and the collective completion of this document. Inclusion is ensured by excluding personal data, and including all data that is relevant to an environmentally healthier Breda. Justice is ensured by highlighting the unproportionate increase of the green score if an area has a higher income. Legitimacy of collected data is ensured by the Exploratory Data Analysis and the Data Quality Report

Exclusions

Overview: These questions have been excluded due to redundancy. Most topics proposed by the questions have been explored in one of the two frameworks, and ones that do not apply to the use case, deployment extent, and nature of the data and/or model have also been considered, but excluded from the final Legal and Ethical Assessment of the project.

Human Oversight and Autonomy

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

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?

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

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

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

Technical Robustness and Safety

How exposed is the AI system to cyber-attacks?

Did you red-team/pentest the system?

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

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

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

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

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?

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

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

Is your AI system using (online) continual learning?

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

Diversity, Non-discrimination and Fairness

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

References

Goodey, J., Heintz, F., Bonnet, Y., Chatila, R., Lucas, P., Renda, A., Sharkov, G., Tallinn, J., Wendling, C., Bonefeld-Dahl, C., Hidvegi, F., Bieliková, M., Höckner, K., & Kroplewski, R. (2019). *Assessment List for Trustworthy AI (ALTAI) HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE SET UP BY THE EUROPEAN COMMISSION THE ASSESSMENT LIST FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE (ALTAI) for self assessment Assessment List for Trustworthy AI (ALTAI)*.

Home page - ALTAI. (n.d.). Altai.insight-Centre.org. https://altai.insight-centre.org/

Schäfer, M., & Franzke, A. (2020, February). *Poster*. Data School. https://dataschool.nl/en/deda/poster/?lang=en