**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**

**Checklist**

**Assessment 1: Human Oversight**

- Have steps been taken to ensure that end-users receive comprehensive information regarding algorithmic decisions, possess a clear understanding of their interactions with an AI system, and are protected against negative consequences such as manipulation and undue reliance?

- How is oversight of the AI model's actions maintained, including its impact on society, compliance with legal regulations, and ethical considerations?

**Assessment 2: Technical Robustness and Safety**

- Has the comprehensive safety of the AI system been assured through the assessment of potential risks, such as misuse or technical malfunctions? Have these risks been effectively communicated to users?

- Are there established measures to guarantee that the data utilised for AI system development, including training data, remains current and of exceptional quality? How is the accuracy of the model continuously upheld?

- Are there documented protocols for testing and verifying the reliability and reproducibility of the AI system? How are outcomes with low confidence scores handled?

**Assessment 3: Communication, Traceability & Explainability**

- Has proper documentation been created for each phase of the AI system's life cycle, including evaluation of data sources, assessment of data quality, selection of AI models, and assurance of output quality?

- Have users been provided with explanations of predictions and insights into the AI model's process, including comprehensive information about the underlying algorithm employed?

- Have users been effectively informed about the purpose, advantages, limitations, and capabilities of the AI model?

**Assessment 4: Accountability**

-   Has the capability for auditing and promoting transparency been guaranteed through comprehensive documentation? How can users voice objections or concerns about the outcomes?

-   Is there an independent third party engaged in performing audits to supervise ethical considerations and accountability measures? Are they actively involved in evaluating the compliance of the AI system with the ALTAI and DEDA framework?

-   Are the roles and responsibilities of project personnel explicitly defined, and who holds the final accountability for the project's outcomes?

-   Which entities are authorised to access the data, under what circumstances, and how is access monitored?

-   What approaches are being employed for communication within the team and throughout the project? Who bears responsibility in the event of issues or failures?

**Assessment 5: Equity and Inclusivity**

-   Has the user interface of the AI system been intentionally designed to cater to the needs of users with disabilities or special requirements, integrating technologies that follow universal design principles? How are the findings presented visually?

-   Were stakeholders actively engaged in the decision-making process throughout the entire development lifecycle?

-   Is the data capable of being reused, and if so, what are the terms and purposes for its reuse? Are there any considerations for facilitating open access to the data?

**Included Ethical and Legal assessments**

**Assessment 1: Human Oversight**

***Assessment 1.1 Awareness and Protection of the user***

Have steps been taken to ensure that end-users receive comprehensive information regarding algorithmic decisions, possess a clear understanding of their interactions with an AI system, and are protected against negative consequences such as manipulation and undue reliance?

**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)*

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Assessment 1.2 Ensuring Responsible AI Governance and Accountability**

How is oversight of the AI model's actions maintained, including its impact on society, compliance with legal regulations, and ethical considerations?

**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)*

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Assessment 2: Technical Robustness and Safety**

**Assessment 2.1 Reliability**

Are there documented protocols for testing and verifying the reliability and reproducibility of the AI system? How are outcomes with low confidence scores handled?

**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)

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Assessment 2.2 Safety**

Has the comprehensive safety of the AI system been assured through the assessment of potential risks, such as misuse or technical malfunctions? Have these risks been effectively communicated to users?

**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)

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Assessment 2.3 Accuracy**

Are there established measures to guarantee that the data utilised for AI system development, including training data, remains current and of exceptional quality? How is the accuracy of the model continuously upheld?

**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.)

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Assessment 3: Communication, Traceability & Explainability**

**Assessment 3.1 Communication**

Have users been effectively informed about the purpose, advantages, limitations, and capabilities of the AI model?

It is crucial to have effectively informed users about the purpose, advantages, limitations, and capabilities of the AI model in the project aiming to make Breda greener. By providing transparent and comprehensive information, users can better understand and engage with the project, contribute their insights, and help ensure the responsible and beneficial use of AI technology for the benefit of the city and its residents.

**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)

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Assessment 3.2 Traceability**

Has proper documentation been created for each phase of the AI system's life cycle, including evaluation of data sources, assessment of data quality, selection of AI models, and assurance of output quality?

**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.)

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Assessment 3.3 Explainability**

Have users been provided with explanations of predictions and insights into the AI model's process, including comprehensive information about the underlying algorithm employed?

**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.)

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Assessment 4: Accountability**

**Assessment 4.1 Communication**

What approaches are being employed for communication within the team and throughout the project? Who bears responsibility in the event of issues or failures?

**Overview:** This assessment refers to good communication within the team. It takes into account all perspectives of team members and makes sure that the decision-making process is excellent and takes all team's considerations into account. (*Data Ethics Decision Aid (DEDA)*, n.d.)

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Assessment 4.2 Risk Mitigation and Control**

Is there an independent third party engaged in performing audits to supervise ethical considerations and accountability measures? Are they actively involved in evaluating the compliance of the AI system with the ALTAI and DEDA framework?

**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)

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Assessment 4.3 Personal Accountability**

Are the roles and responsibilities of project personnel explicitly defined, and who holds the final accountability for the project's outcomes?

**Overview:** This assessment revolves around classifying responsibilities of each part of the project to different team members. (*Data Ethics Decision Aid (DEDA)*, n.d.)

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Assessment 4.4 Responsibility and Transparency**

Has the capability for auditing and promoting transparency been guaranteed through comprehensive documentation? How can users voice objections or concerns about the outcomes?

**Overview:** The assessment of responsibility in this assessment puts emphasis on the AI system's documentation and comments, so at the end independent third parties are able to revise this system without incurring any obstacles. This assessment puts emphasis as well on transparency of the project, so the wider population can revise the project themselves and explain their concerns regarding the final product. .(*Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-assessment*, 2020)  (*Data Ethics Decision Aid (DEDA)*, n.d.)

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Assessment 4.5 Access**

Which entities are authorised to access the data, under what circumstances, and how is access monitored?

**Overview:** This assessment makes sure that private data of clients is accessible by organisation only and also that it is properly secured and monitored.  *(Data Ethics Decision Aid (DEDA), n.d.)*

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Assessment 5: Equity and Inclusivity**

**Assessment 5.1 Sharing, Reusing and repurposing**

Is the data capable of being reused, and if so, what are the terms and purposes for its reuse and are there any considerations for facilitating open access to the data?

**Overview:** This assessment is related to the data used in the project and its ways of reusing, sharing and repurposing it to other projects. What is worth mentioning is if this data is adequate to different projects, containing different goals set. It also connects with consideration of providing open access to it, because not all of the data used is relevant for such an access. (*Data Ethics Decision Aid (DEDA)*, n.d.)

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Assessment 5.2 Inclusivity and Design for All**

Has the user interface of the AI system been intentionally designed to cater to the needs of users with disabilities or special requirements, integrating technologies that follow universal design principles? How are the findings presented visually?

**Overview:** This assessment refers to the importance and specification of how the data will be visualised, in order to create a universal access to the findings for all end-users. This also applies to the ability for end-users with special needs, disabilities or ones who can potentially be excluded, when it comes to an AI system's UI. .(*Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-assessment*, 2020)  (*Data Ethics Decision Aid (DEDA)*, n.d.)

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Assessment 5.3 Engagement of Stakeholders**

Were stakeholders actively engaged in the decision-making process throughout the entire development lifecycle?

**Overview:** This assessment refers to engagement of stakeholders throughout the lifecycle of the development of the project. It is crucial to experience such an engagement, because

this factor provides useful feedback throughout the whole development process.

(*Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-assessment*, 2020)

**Why and evidence (where possible):** HERE WE NEED TO PROVE WHY WE INCLUDED/EXCLUDED THIS ETHICAL ASSESSMENT FOR THE PROJECT. EVIDENCE IT BASED ON DATA/OTHER SOURCES

**Excluded Ethical and Legal assessments:**

**DEDA Framework**

10. Privacy

11. Bias

3. Anonymization

**ALTAI Framework**

Requirement number 6 - Societal and Environmental Well-being I

Environmental Well-being

Impact on Work and Skills

Impact on Society at Large or Democracy

Requirement number 3 - Privacy and Data Governance

Privacy

Data Governance

Requirement number 2 - Technical Robustness and Safety

Resilience to Attack and Security

Requirement number 5  -Diversity, Non-discrimination and Fairness

Avoidance of Unfair Bias