TLDR of ATLAI (Assessment list for Trustworthy Artificial Intelligence)

ATLAI was developed by the High-Level Expert Group on Artificial Intelligence (A group of experts to provide advice to the European Commission on its artificial intelligence strategy.) set up by the European Commission to help assess whether the AI system that is being developed deployed procured or used complies with the seven requirements of Trustworthy AI, as specified in these 7 key Ethics Guidelines for Trustworthy AI.:

1. Human Agency and Oversight
2. Technical Robustness and Safety.
3. Privacy and Data Governance.
4. Transparency.
5. Diversity, Non-Discrimination, and Fairness.
6. Societal and Environmental well-being.
7. Accountability.

The primary goal and purpose of ALTAI is to establish a framework for evaluating the trustworthiness of AI through self-evaluation. By utilizing ALTAI, organizations can select relevant elements from the framework or add their own elements to address the specific considerations of their respective sectors. This enables organizations to comprehend the concept of trustworthy AI, including the potential risks associated with AI systems. Additionally, ALTAI aims to raise awareness about the societal, environmental, and ethical impacts of AI, emphasizing the well-being of consumers, workers, citizens (especially children and marginalized groups), and other stakeholders both within and outside organizations.

ALTAI serves as a tool to assess whether adequate measures and processes, such as internal guidelines and governance procedures, are already in place to ensure compliance with trustworthy AI requirements or if further actions need to be implemented. The overarching objective is to foster a responsible competitive environment, where all users and those affected by AI systems can have confidence that the design, development, and usage of such systems are lawful, ethical, and robust. This approach paves the way for responsible and sustainable AI innovation in Europe, establishing a distinctive approach that benefits and safeguards individuals and society. By embracing ALTAI, European organizations can position themselves as global leaders in innovative AI that deserves the trust and confidence of individuals and communities. ALTAI was developed over a two-year period, from June 2018 to June 2020, to provide a comprehensive and reliable framework for assessing trustworthy AI.

Checklist

1. Human Agency and Oversight:

Have you defined a clear purpose and objective for the AI system?

Yes, a clear purpose and objectives have been defined for the AI system.

Have you considered the roles and responsibilities of humans in the AI system's development, deployment, and use?

Yes, the roles and responsibilities of humans have been carefully considered throughout the AI system's lifecycle.

Have you established mechanisms for human oversight and control of the AI system?

Yes, mechanisms for human oversight and control have been implemented to ensure responsible use of the AI system.

Have you ensured that humans can understand and explain the AI system's decisions and behavior?

Yes, efforts have been made to ensure transparency and interpretability of the AI system's decisions and behavior for humans.

Evidence: The model created has been clearly defined by the members of the group. Input and output selections have been discussed and the best ones have been made into the final model. The model has also been uploaded and incorporated into streamlit, allowing users who have no knowledge of AI to use the model and understand what it does,

2. Technical Robustness and Safety:

Have you conducted a comprehensive risk assessment of the AI system's development, deployment, and use?

Yes, a comprehensive risk assessment has been conducted to identify and mitigate potential risks associated with the AI system's development, deployment, and use.

Have you implemented measures to ensure the reliability, accuracy, and security of the AI system?

Yes, measures have been implemented to enhance the reliability, accuracy, and security of the AI system, following best practices and industry standards.

Have you assessed and mitigated risks related to data quality, integrity, and security?

Yes, risks related to data quality, integrity, and security have been assessed, and appropriate measures have been taken to mitigate those risks.

Have you considered the potential impact of adversarial attacks or manipulation of the AI system?

Yes, the potential impact of adversarial attacks or manipulation of the AI system has been considered, and measures have been implemented to enhance the system's resilience against such threats.

Evidence: A risk assessment with mitigation plans has been made. A data quality report has been written to discuss the limitations and quality of the data we selected and appropriate measures have been taken to ensure these risks are eliminated.

3. Privacy and Data Governance:

Have you conducted a thorough assessment of the AI system's data needs and data sources?

Yes, a thorough assessment of the AI system's data needs and data sources has been conducted to ensure compliance with privacy regulations and ethical considerations.

Have you considered the principles of data minimization and purpose limitation in the collection and use of data?

Yes, the principles of data minimization and purpose limitation have been considered, and data collection and use have been limited to what is necessary and aligned with the defined purposes of the AI system.

Have you obtained informed consent or implemented appropriate legal bases for processing personal data?

Yes, informed consent or appropriate legal bases for processing personal data have been obtained in compliance with relevant data protection laws and regulations.

Evidence: Almost all of the data collected are from public sources and informed consent has been given for data protection.

4. Transparency:

Have you ensured transparency in the AI system's capabilities, limitations, and potential impact?

Yes, efforts have been made to ensure transparency regarding the AI system's capabilities, limitations, and potential impact. Clear documentation or explanations have been provided where necessary.

Have you documented and made available information about the AI system's development, deployment, and use?

Yes, documentation about the AI system's development, deployment, and use has been created and made available to relevant stakeholders as appropriate.

Evidence: Unit tests have been conducted on the model, both of which are uploaded to the shared group repository.

5. Diversity, Non-discrimination, and Fairness:

Have you established a strategy or procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data and algorithm design?

Yes, strategies or procedures have been established to avoid creating or reinforcing unfair bias in the AI system, including careful consideration of input data and algorithm design.

Have you considered diversity and representativeness of end-users and/or subjects in the data?

Yes, diversity and representativeness of end-users and/or subjects in the data have been considered to minimize bias and ensure fairness in the AI system's outcomes.

Evidence: A data quality report was written to assess the quality of the data. Personal or sensitive data was avoided to make sure the model could not be biased, so that certain groups would not be discriminated against and that the model would be fair.

6. Societal and Environmental Well-being:

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

No, there are no potential negative impacts of the AI system on the environment.

Evidence: The model created does not have any potential negative impacts, as it only predicts the amount of nuisance reports.

7. Accountability

Did you ensure that workers understand how the AI system operates and its capabilities?

Yes, efforts have been made to ensure that workers understand how the AI system operates and its capabilities, enabling them to effectively collaborate with and adapt to the system.

Evidence: A streamlit page was made where the model was present and ready to use by anyone, including all stakeholders.

Relevance

Compliance with regulations: Considering legal aspects ensures that your research and the implementation of the prediction model adhere to relevant laws and regulations. This is important to avoid any legal issues or non-compliance with data protection, privacy, or discrimination laws.

Protection of individuals' rights: Ethical considerations play a significant role in safeguarding the rights and well-being of individuals. By addressing ethical aspects, you can ensure that the prediction model and its use in addressing public nuisance do not infringe upon individuals' rights, discriminate against specific groups, or reinforce biases.

Accountability and transparency: Including legal and ethical aspects in your checklist promotes accountability and transparency in the development and use of the AI system. It establishes mechanisms to trace the development process, document decisions, and facilitate third-party auditing. This helps build trust among stakeholders, as they can be assured that the system is being developed and used in a responsible and accountable manner.

Mitigation of biases and unfair outcomes: Legal and ethical considerations guide you in identifying potential biases and discriminatory outcomes in the prediction model. By implementing measures to minimize biases and ensure fairness, you can reduce the risk of perpetuating or exacerbating social inequalities in neighborhoods.

Social acceptance and trust: Addressing legal and ethical aspects contributes to the social acceptance and trustworthiness of the prediction model. When stakeholders, including residents, community organizations, and local authorities, perceive that the AI system is developed and used in an ethical and legally compliant manner, it enhances their confidence in the system and its outcomes.

Justification for excluding certain questions

Here are the questions we did not use along with a short explanation.

Technical Robustness and Safety:

Have you conducted comprehensive risk assessments for the AI system to identify potential safety hazards?

Justification: This question focuses on safety hazards, which is not relevant to the project. The project at hand does not require comprehensive risk assessments for safety hazards. Therefore, this question does not apply in the context of the project.

Privacy and Data Governance:

Have you implemented measures to ensure the privacy and confidentiality of personal data used by the AI system?

Have you implemented measures to protect personal data and privacy throughout the AI system's lifecycle?

Justification: These questions primarily address data privacy, which is not related to the project. The project does not involve the collection, storage, or processing of personal data. Therefore, ensuring privacy and implementing data protection measures throughout the AI system's lifecycle is not applicable in this case.

Transparency:

Have you ensured that information about the AI system and its capabilities is accessible and understandable to users?

Have you provided explanations for the AI system's decisions or outcomes, particularly in cases where it affects individuals?

Have you considered the requirements and expectations of transparency from end-users or subjects?

Justification: These questions relate to ensuring accessibility and understanding of the AI system, which is not related to the project. The project does not involve user interaction, decision-making, or the need to provide explanations for the AI system's behavior. Therefore, ensuring transparency and meeting user expectations regarding the AI system's capabilities and decision-making processes are not applicable in this case.

Diversity, Non-Discrimination, and Fairness:

Have you tested for specific target groups or problematic use cases?

Have you assessed and implemented processes to test and monitor for potential biases throughout the AI system's lifecycle?

Have you implemented educational and awareness initiatives to help AI designers and developers be more aware of possible biases in the AI system?

Have you ensured a mechanism for flagging issues related to bias, discrimination, or poor performance of the AI system?

Have you established clear steps and communication channels for raising such issues?

Justification: These questions focus on the impact on diversity, non-discrimination, and fairness, which is not related to the project. The project does not involve the use of AI systems in a manner that directly affects different target groups, introduces biases, or requires specific monitoring and reporting mechanisms. Therefore, these questions do not apply in the context of the project.

Societal and Environmental Well-being:

Which potential impacts do you identify?

Where possible, have you established mechanisms to evaluate the environmental impact of the AI system's development, deployment, and/or use (e.g., energy consumption, carbon emissions)?

Have you defined measures to reduce the environmental impact of the AI system throughout its lifecycle?

Does the AI system impact human work and work arrangements?

Did you inform and consult with impacted workers and their representatives (e.g., trade unions, work councils) in advance of introducing the AI system in your organization?

Have you adopted measures to ensure a clear understanding of the impacts of the AI system on human work?

Does the AI system impact human work and work arrangements?

Justification: These questions focus on the impact on societal and environmental well-being, which is not related to the project. The project does not have significant implications for the environment, workforce, or societal well-being. Therefore, assessing potential impacts, evaluating environmental factors, or addressing impacts on human work arrangements are not applicable in this case.

Accountability:

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

Justification: This question pertains to establishing an AI ethics review board, which is not related to the project. The project does not require or involve the establishment of a dedicated ethics review board or a similar mechanism. Therefore, this question does not apply in the context of the project.

References

Data.breda.nl. (n.d.). Open Data Breda. Retrieved from https://data.breda.nl/

European Commission. (2020). The Assessment List for Trustworthy Artificial Intelligence (ALTAI) (ISBN: 978-92-76-20008-6). doi:10.2759/002360 KK-02-20-479-EN-N.

Politie Nederland. (n.d.). Politie Open Data. Retrieved from https://data.politie.nl/#/Politie/nl/

Statistics Netherlands. (n.d.). CBS - Statistics Netherlands. Retrieved from https://www.cbs.nl/