DEDA CHECKLIST

Privacy and consent:

- ☐ Determine if the data being gathered contains personally identifiable information (PII) and whether consent is required.
- ☐ When required, get informed consent from individuals and explicitly express the purpose and scope of data collection.

Fairness and bias:

- ☐ Identify and mitigate any data biases to ensure fairness in algorithm development and execution.
- ☐ Assess the algorithm's influence on various demographic groups and correct any potential biases or prejudice.

Legal Obligation:

- ☐ Ensure that appropriate rules, regulations, and ethical norms, such as data protection and privacy legislation, are followed.
- ☐ Consult with legal specialists to understand the legal consequences of building and employing the prediction algorithm.

Transparency and comprehensibility:

- ☐ Document the data sources, methods of collection, and any data pretreatment or cleaning activities that were used.
- ☐ Explain to stakeholders the algorithm's approach, variables, and decision-making processes in detail.

Data security

- ☐ Encrypt sensitive data while it is being sent and stored.
- ☐ To safeguard confidentiality, establish standards for data exchange both inside the project team and with external stakeholders.

Risk assessment:

- ☐ Conduct a thorough risk assessment to detect any negative repercussions or unforeseen outcomes of the algorithm.
- ☐ Implement risk-mitigation strategies, such as monitoring for algorithmic bias and assessing and updating the model on a regular basis.

Communication:

- ☐ Communicate the algorithm's aim, advantages, and restrictions to stakeholders, including the municipality and the general public.
- ☐ Provide transparency about the algorithm's intended application and answer any concerns or inquiries from stakeholders.

Evaluation:

- ☐ Monitor the algorithm's performance on a regular basis to guarantee its correctness, fairness, and efficiency in anticipating and distributing resources.
- ☐ Evaluate the algorithm's societal impact and make any required improvements to mitigate any unforeseen repercussions.

ALTAI Checklist

1. Human Agency and Oversight:

- ☐ Could the AI system generate confusion for some or all end-users or subjects regarding the origin of algorithmic decisions?
- ☐ Are end-users or subjects adequately informed that a decision, content, advice, or outcome is the result of an algorithmic decision?
- ☐ Could the AI system lead to over-reliance by end-users, potentially affecting human autonomy?
- ☐ Have procedures been established to prevent unintended interference with human autonomy?
- ☐ Are human stakeholders involved in the decision-making process of the project?
- ☐ Is there a clear understanding of the roles and responsibilities of human stakeholders?
- ☐ Are there mechanisms in place for human oversight and intervention in the prediction process?
- ☐ Is there a process to review and override automated decisions if necessary?
- ☐ Are there provisions for accountability in case of errors or negative impacts?

2. Technical Robustness and Safety:

- ☐ Could the AI system pose risks or threats due to design or technical faults, defects, outages, attacks, misuse, inappropriate or malicious use?
- ☐ Have potential forms of attacks and vulnerabilities been assessed?
- ☐ Are measures in place to ensure the integrity, robustness, and overall security of the AI system against potential attacks throughout its lifecycle?
- ☐ Is there a mechanism to evaluate when changes to the AI system warrant a new review of its technical robustness and safety?
- ☐ Is the prediction model robust and reliable, producing accurate and consistent results?
- ☐ Are there measures in place to ensure the safety and integrity of the data used for predictions?
- ☐ Has a risk assessment been conducted to identify and mitigate potential risks and hazards associated with the prediction model?

3. Privacy and Data Governance:

- ☐ Have you considered the impact of the AI system on the right to privacy?
- ☐ Have mechanisms been established to flag privacy-related issues related to the AI system?
- ☐ Have you implemented measures mandated by data protection regulations?
- ☐ Have you implemented the right to withdraw consent, the right to object, and the right to be forgotten into the development of the AI system?
- ☐ Have you considered the privacy and data protection implications of data collected, generated, or processed throughout the AI system's life cycle?
- ☐ Are there measures in place to protect personal data and comply with relevant data protection regulations?
- ☐ Has a data governance framework been established to ensure responsible data handling practices?

4. Transparency:

- ☐ Have you implemented measures to address the traceability of the AI system throughout its entire lifecycle?
- ☐ Can you trace back which data was used by the AI system to make specific decisions or recommendations?
- ☐ Can you trace back which AI model or rules led to the decisions or recommendations of the AI system?

- ☐ Have you put in place measures to continuously assess the quality of the AI system's outputs?
- ☐ Have you explained the decisions of the AI system to the users?
- ☐ Have you provided appropriate training materials and disclaimers to users on how to adequately use the AI system?
- ☐ Is the prediction model transparent and explainable?
- ☐ Are explanations provided to stakeholders on how their data is used and how predictions are made?

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?
- ☐ Have you considered diversity and representativeness of end-users and/or subjects in the data?
- ☐ 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?
- ☐ Have you assessed whether the AI system's user interface is usable by those with special needs or disabilities or those at risk of exclusion?
- ☐ Have you ensured that information about the AI system and its user interface is accessible and usable to users of assistive technologies?
- ☐ Have you involved or consulted with end-users or subjects in need of assistive technology during the planning and development phase of the AI system?
- ☐ Have you considered the impact of the AI system on potential end-users and/or subjects?
- ☐ Have you assessed whether the team involved in building the AI system engaged with possible target end-users and/or subjects?
- ☐ Have you assessed whether there could be groups who might be disproportionately affected by the outcomes of the AI system?
- ☐ Have you assessed the risk of possible unfairness of the system on the end-users' or subjects' communities?
- ☐ Have you considered a mechanism to include the participation of a wide range of stakeholders in the AI system's design and development process?

6. Societal and Environmental Well-being:

- ☐ Are there potential negative impacts of the AI system on the environment?
- ☐ 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?
- ☐ Did you ensure that workers understand how the AI system operates and its capabilities?
- ☐ Could the AI system create the risk of de-skilling the workforce?
- ☐ Have you taken measures to counteract de-skilling risks?
- ☐ Does the system promote or require new (digital) skills?
- ☐ Have you provided training opportunities and materials for re- and up-skilling?
- ☐ Could the AI system have a negative impact on society at large or democracy?
- ☐ Have you assessed the societal impact of the AI system's use beyond the end-users and subjects, considering potentially indirectly affected stakeholders or society at large?
- ☐ Have you taken action to minimize potential societal harm caused by the AI system?
- ☐ Have you taken measures to ensure that the AI system does not negatively impact democracy?
- ☐ Has the potential impact of the prediction on society and the environment been considered?
- ☐ Are there measures in place to mitigate any negative consequences or externalities?
- ☐ Does the project align with sustainable and ethical practices in urban planning and neighborhood development?
- ☐ Has the project considered the potential social and environmental impacts of predicting public nuisance?
- ☐ Are there mechanisms in place to address and mitigate any negative social or environmental consequences?
- ☐ Has the project considered the broader well-being and quality of life of the neighborhoods and communities involved?

7. Accountability:

- ☐ Did you establish mechanisms that facilitate the AI system's auditability (e.g., traceability of the development process, sourcing of training data, logging of processes, outcomes, positive and negative impact)?
- ☐ Did you foresee any external guidance or third-party auditing processes to oversee ethical concerns and accountability measures?
- ☐ Did you consider establishing an AI ethics review board or a similar mechanism to discuss overall accountability and ethics practices, including potential unclear areas?
- ☐ Did you establish a process to discuss, continuously monitor, and assess the AI system's adherence to this Assessment List for Trustworthy AI (ALTAI)?
- ☐ Does this process include identification and documentation of conflicts between the requirements or between different ethical principles, and explanation of the "trade-off" decisions made?
- ☐ Did you provide appropriate training to those involved in the process, including coverage of the legal framework applicable to the AI system?
- ☐ For applications that can adversely affect individuals, have redress-by-design mechanisms been put in place?
- ☐ Is there a clear framework for accountability and responsibility within the project?
- ☐ Are there mechanisms in place for addressing any concerns or issues that may arise?
- ☐ Is there a process for addressing complaints, concerns, or disputes related to the predictions made?
- ☐ Are there mechanisms in place to track and evaluate the impact and effectiveness of the prediction model over time?