Province of Fryslân, Rijks ICT Gilde & the Z-Inspection® Initiative

Assessing the trustworthiness of an Al system in practice

Lessons Learned from the expert examination of the AI system "Monitoring grassification of heather fields"

Pilot Conditions Method Why What Z-Inspection The case study An assessment for trustworthy Al An international partnership The AI system A method for assessing responsible AI Assessment for trustworthy Al The government wants to seize the opportunities of AI, but the technology still raises many important An assessment for trustworthy AI involves a dialogue that Province of Fryslân, Al authority Prof. Dr. Zicari, his The Province of Fryslân is required by law to monitor Z-Inspection is a method for assessing trustworthiness of AI systems in practice and has the potential to play questions. How trustworthy are algorithms? Can an algorithm discriminate? And how transparent is the use of takes into account the values and interests of relevant international team of AI experts and Rijks ICT Gilde are biodiversity in nature areas. This is done by conducting a a key role in the context of the new EU Artificial Intelligence Act. Z-inspection is included in the OECD AI? The "Assessment for Trustworthy AI" pilot sought answers to the following questions: stakeholders to determine whether or not the AI system jointly investigating the trustworthiness of the AI system manual, visual inspection once every 10 years. There is a Catalogue of Tools & Metrics for Trustworthy Al. One of the key features of the Z-inspection method is its should be developed and used. The Z-inspection selfand its responsible use. need to monitor and map the nature areas more often and interdisciplinary and dynamic nature. The complexity of an AI system is reflected in the composition of the - As a government, how do you govern the development and use of responsible AI? faster. An AI system has been developed for this purpose. assessment method assists in developing, implementing team. The diversity of participants allows for a more inclusive assessment of the trustworthiness of an AI - What frameworks, laws and regulations are important, and how do we assess against them in the development and use of AI? and using the AI system in a trustworthy way. The AI system aims to monitor grassification of heathlands system. Using the Z-inspection method, potential ethical, technical and legal issues are identified. The - How do you analyze, assess and improve AI applications? through satellite imagery. Grassification disturbs the method is based on the European Commission's Ethics Guidelines for Trustworthy Artificial Intelligence (EU biodiversity in nature areas. The AI system helps ecologists - And are the applications consistent with public values and human rights? High-Level Expert Group on Artificial Intelligence). to quickly and frequently image the nature area so that it can be checked whether the intended nature quality In the pilot, an AI system of Province of Fryslân was assessed and analyzed. This assessment was conducted using the Z-inspection method of AI authority Prof. Dr. Roberto Zicari; a self-assessment in which participants objectives are being achieved, the right management discuss critical issues such as: the purpose of the algorithm, the development process, ethical dilemmas and measures are being taken and whether the approach to conflicts of interest. The Z-inspection method is a way of working to analyze, assess and improve AI systems, increasing biodiversity is working. sustainably, demonstrably and transparently. This enables organizations to develop and use responsible Al applications in a structured and accountable manner. Furthermore, it is very important to share the knowledge and experiences from the pilot. First, to stimulate digital awareness and dialogue about AI within the government. And second, to be able to confidently apply the technology to tomorrow's questions. **V**

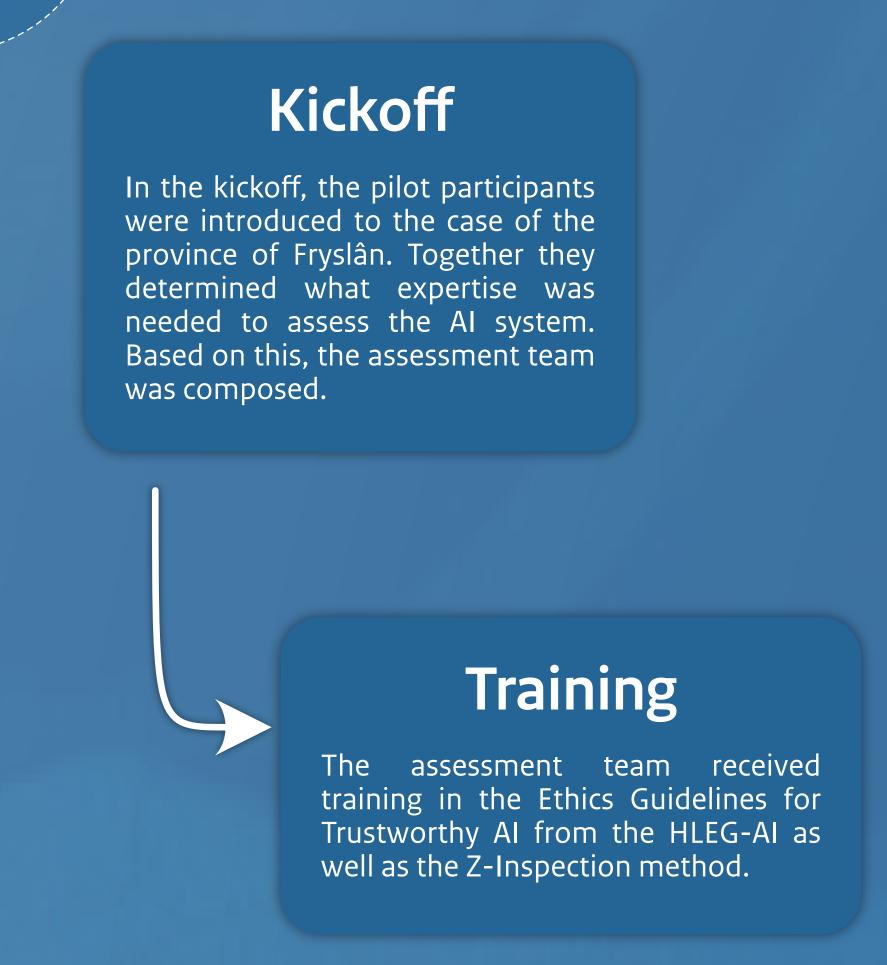
International

orthy

thrustw

Assessing Alin

organization



Scope The team determined the scope of the pilot before the assessment began: The Province of Fryslân has developed a new monitoring system based on remote sensing. Analyze whether this AI system is trustworthy and can be used responsibly in practice.

Approach The AI system was examined from three perspectives: technical, ethical and ecological. The stage in the life cycle of the AI system and the purpose of the system were taken into account. The assessment team was divided into three expert groups: a technical expert group, an ethical expert group and an ecological expert group. From each perspective, the risks, opportunities and recommendations were identified and linked to the European Commission's ethical principles. The different expert groups engaged in a structural dialogue with each other to sharpen the findings and recommendations. The conversation was led based on the claims-arguments-evidence framework. Expert groups Ecology Technology

Six months

lead time

Multidisciplinary

team

(+)

Ethical such as the right to a healthy environment. Technical findings were identified and recommendations made. Ecological of a fully automated field monitoring system? How is the algorithm used?

Open dialog

Dialogue & The ethical expert group used the Fundamental Rights and Algorithms Impact Assessment (FRAIA) to Presentation test the AI system against legal requirements regarding human rights. Human rights were then linked to ethical values and the system was tested from this broader, ethical perspective. Not only were human rights infringements considered, but also human rights that are actually protected or strengthened, The findings of the different expert groups were discussed by the full assessment team. This was done using the claims-arguments-evidence framework. Any ambiguities were refined and the The technical expert group assessed the technical robustness of the AI system. This took into account adjustments were the technical maturity of the AI system. Based on six categories - data management and processing, raw incorporated into the reports. data and data labeling, system architecture, robustness, explainability, and implementation and use -In addition to an ethical and technical assessment, domain experts - in this case ecologists - also reviewed the AI system. The ecologists looked at the AI system in two different ways: What is the impact

Report

Top 9 lessons learned

Clearly define the scope

It is important to establish a clear scope for the assessment in advance. What will the team assess? But also: what is the team not going to assess? It is tempting to zoom out, to look from a broad perspective and also to evaluate the government policy to which the algorithm relates. During the assessment, therefore, ensure a clear scope and a process supervisor who continuously monitors and defines the scope in advance.



Provide a common language

The assessment team is interdisciplinary with different backgrounds and skills. It takes time and patience to understand each other and develop a common language. Mapping findings to the ethical principles of the HLEG-AI framework helps create greater understanding and facilitates dialogue.



Z-inspection is more than a method

Z-inspection is more than a method. It is an international community in which there is joint learning on how to put ethics and responsible use of Al into practice.



Z-inspection is suitable for high-risk systems

The Z-inspection method is a good addition to the methods and tools already used by the Dutch government. Because of its depth and the knowledge and time required, it is especially suitable for high-risk algorithms.



Increasing digital awareness

The pilot and the Z-inspection method encourage dialogue about AI within the government, increasing digital awareness among civil servants. The benefit is that civil servants are more likely to understand the impact AI can have on their work, organization and society. It gives them more confidence to navigate the digital world and seize the opportunities of AI without losing sight of the risks. In this way, technology can be used with confidence for tomorrow's questions.



FRAIA and Z-Inspection strengthen each other

During the pilot, the assessment team used two different approaches: the Fundamental Rights and Algorithms Impact Assessment (FRAIA) and the Ethics Guidelines for Trustworthy AI. The two go hand in hand. Both approaches provide critical insights regarding the AI system. Both ethics and human rights are about norms and fundamental values in society. Since ethical reflection and ethical guidelines influence law, experts from both fields must work together when considering the design of AI systems and their societal implications. Ethics, a branch of philosophy, considers what is right and wrong. It seeks answers to questions such as "What should we do?" or "What is the right action? In the context of Al systems, an ethics-based approach focuses on questions such as "What is the right way to design, develop, deploy and use this type of technology so that it benefits individuals and society?

Such questions require reflection on the various courses of action around an AI system, on the different options and their implications. This reflection should not be limited to what legislation prescribes; a broader ethical perspective is needed. A human rights-based approach is closely linked to existing law and focuses primarily on aspects that are legally relevant and enforceable.



Organizations and project teams are looking for strong footing

Organizations and project teams are looking for strong footing. They often ask for one complete checklist so they can be sure that their Al system is reliable and compliant with all regulations. But do we want government to follow a checklist and check off all the boxes? No, we want government to be guided by public values and protect fundamental rights. Therefore, the need for a holistic, integrative and interdisciplinary method for responsible AI is great. Not just a static checklist, but a dynamic process for conducting assessments. An approach that ensures that the assessment team is representative, knowledgeable and independent. A method that facilitates meaningful consultation. A way to assess the reliability of AI systems, supported by arguments and evidence. The Z-inspection method meets this need.



Technical model validation is essential

Technical model validation is necessary to gain insight into potential ethical dilemmas. Also, an AI system that is not robust can never be trustworthy. Insight into how the algorithm performs is an important prerequisite for responsible use.



Courage

Working proactively, openly and transparently within government is new, and it causes quite a bit of cold feet. Conducting a self-assessment for responsible AI while providing full disclosure requires courage. Openness helps to learn from each other and steer the deployment of responsible AI in the right direction.



Summaries assessment

The AI system was examined on three components: technical, ethical and ecological. The findings were captured in three different reports. A summary has been prepared for each report. In addition to substantive reports, lessons learned from applying the Z-inspection method were also identified. This is summarized in the lessons learned overview.









ECOLOGY