Ref. Ares(2026)4470211 - 30/04/2026



D8.2

Report initial pilot - wastewater processing pilot

Related Work Package	**WP8 – Implementation of industrial processes LSP**
Related Task	Task 8.2 – Initial Pilot
Lead Beneficiary	SINTEF
Contributing Beneficiaries	Veas
Document version	v.1.0
Deliverable Type	R — Document, report
Distribution level	PU — Public
Contractual Date of Delivery	30/04/2026
Actual Date of Delivery	30/04/2026

Authors	**Asbjørn Følstad (SINTEF), Hilde Johansen (VEAS), Sølve Eidnes (SINTEF)**
Contributors	Patrick Nitschke (SINTEF), Pål Johnsen (SINTEF), Anne-Kari Marsteng (Veas), Yves Le Naour (Veas), Theresa Liedke (Veas), Kjell Rune Jonassen (Veas), Torbjørn Mørk Madsen (Veas)
Reviewers	V. Pezoulas (FORTH), Dania Cioni (UNIPI), Francesca Pia Caputo (UNIPI)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Directorate-General for Communications Networks, Content and Technology. Neither the European Union nor the granting authority can be held responsible for them.

Version history

Version	Description	Date completed
0.1	Outline	26/02/2026
0.2	Early draft	02/04/2026
0.3	Intermediate draft	10/04/2026
0.4	Draft for internal review	17/04/2026
0.5	Draft with input from review	28/04/2026
1.0	Final version	30/04/2026

Statement of Originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Disclaimer

This document contains material, which is the copyright of one or more FAITH consortium parties, and may not be reproduced or copied without permission.

All FAITH consortium parties have agreed to full publication of this document.

The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the FAITH consortium as a whole, nor individual FAITH consortium parties, warrant that the information contained in this document is capable of use, nor that use of the information is free from risk, accepting no liability for loss or damage suffered by any person using this information.

Executive summary

This deliverable documents the implementation, evaluation, and outcomes of the initial pilot of Large Scale Pilot 5 (LSP5) in the FAITH project. LSP5 focuses on applying the FAITH Artificial Intelligence Trustworthiness Assessment Framework (FAITH AI TAF) to AI-based prediction and control models in the context of industrial wastewater treatment. The pilot is a collaboration between SINTEF as pilot organizer and Veas, operator of Norway's largest wastewater treatment plant, as host.

The primary objective of the initial pilot has been to develop, validate, and assess trustworthy AI prediction models for nitrate levels in treated wastewater. These models will in turn serve as basis for model predictive control (MPC) in the replication pilot. The work addresses technical performance and broader trustworthiness concerns, including explainability and interpretability as well as transparency and accountability.

Significant progress has been made in advancing AI models for time-series prediction in wastewater treatment. Multiple modelling approaches have been explored, including regression-based, tree-based, deep learning, and transformer-based methods. In particular, the Recency-Weighted Temporally-Segmented Ensemble for Time Series Modeling (ReWTS) approach has shown promise in both accuracy and robustness under varying seasonal and operational conditions. Two ReWTS-based models have been validated with domain experts.

A key achievement of the pilot is the integration of explainability and human-centred evaluation into the AI development lifecycle. Through bi-weekly validation meetings and evaluation workshops, engineers and operators at Veas have been actively involved in assessing model predictions, interpreting errors, and refining evaluation metrics and visualisations of model predictions. This strengthens interpretability and transparency, as well as a shared understanding of AI capabilities and limitations.

The FAITH AI TAF and associated tools—TrustGuard, TrustSense, and the AI Model Hub—have been applied in the pilot to structure trustworthiness assessments, identify risks and vulnerabilities, and establish appropriate controls. These tools have supported systematic analysis of technical, organisational, and human-centred risks, while also generating feedback for future framework and tool improvements.

Due to the promising results of the initial pilot, the project has also progressed well toward readiness for AI-based predictive control for the forthcoming replication pilot.

Table of Contents

1 Introduction 8
1.1 Scope and objectives of the initial pilot in LSP5 - the industrial processes LSP 8
1.2 Evolution of the AI system under test (SUT) from initial setup 10
1.2.1 Evolution of initial AI prediction models – nitrate levels in outflowing wastewater 10
1.2.2 Evolution of interim AI prediction models – other aspects of denitrification 10
1.2.3 Evolution of advanced AI prediction models – process control 11
1.3 Structure of the document 11
2 Progress of the AI system deployment and execution of the pilots 12
2.1 Status of the initial pilot execution 12
2.1.1 Deployment context and domain specifics 12
2.1.2 From AI model validation to AI for predictive control 13
2.1.3 AI prediction models advanced in the initial pilot 15
2.1.4 Explorations of increased data capture – sensor trial 18
2.1.5 User and stakeholder involvement in the initial pilot 18
2.1.6 Evaluation process in LSP5 19
2.2 Status of the replication pilot preparation and initialization 20
2.3 Deviations from the original deployment plan 20
2.4 Data collection throughout the initial pilot execution 20
2.5 Integration into LSP5 workflows: Lessons learned 21
3 Application of FAITH Tools 23
3.1 Application of the FAITH AI TAF V1 methodology 23
3.1.1 Cartography 23
3.1.2 Threat analysis 24
3.1.3 Impact assessment 25
3.1.4 Vulnerability assessment 26
3.1.5 Risk analysis 27
3.1.6 Countermeasures 27
3.2 Tool validation planning & training processes 29
3.2.1 Dry-run 29
3.2.2 Planning the validation and training 29

3.2.3 Tool training 30
3.3 FAITH tool validation activities 30
3.3.1 TrustSense 30
3.3.2 AI Model Hub - Data and Model Passports 31
3.3.3 TrustGuard 33
3.4 Threat analysis 34
3.5 Initial risk assessment results 35
3.6 Countermeasures 35
3.7 Formal tool feedback 37
3.7.1 TrustSense tool feedback 37
3.7.2 AI Model Hub tool feedback 37
3.7.3 TrustGuard tool feedback 37
4 AI system evaluation 38
4.1 Evaluation Approach 38
4.2 Trustworthiness assessment 38
4.2.1 Validity & Reliability 39
4.2.2 Explainability & Interpretability 41
4.2.3 Transparency & Accountability 43
4.2.4 Security & Resilience 44
4.2.5 Safety 44
4.3 Preliminary impact evaluation 45
5 Regulatory and ethical assessment 47
5.1 Alignment with the AI Act 47
5.2 Alignment with NIS2 47
6 Conclusions and roadmap for the replication pilot 49
6.1 Summary of achievements of the initial pilot 49
6.2 Plan for replication pilot 50
6.3 Identified risks for the replication pilot 52
7 References 54

List of Abbreviations

Abbreviation	Explanation
LSP	Large scale pilot
TRL	Technology readiness level
FAITH AI TAF	FAITH artificial intelligence trustworthiness assessment framework
SUT	System under test
PID	Proportional-integral-derivative controller
MPC	Model predictive control
ReWTS	Recency-weighted temporally-segmented ensemble for time series modeling
LSTN	Long short-term memory
XGBoost	Extreme gradient boosting
TCN	Temporal convolutional network
TFT	Temporal fusion transformer
DCS	Decision control system
SCADA	Supervisory control and data acquisition
OPC-UA	Open platform communications unified architecture
MAE	Mean absolute error
MSE	Mean squared error
RMSE	Root mean square error
MAPE	Mean absolute percentage error

List of Tables

Table 1. Assets in LSP5 initial pilot 24
Table 2. Users in LSP5 initial pilot 24
Table 3. Threats identified during threat analysis in LSP5 25
Table 4. Impact assessment – business consequences for compromised trustworthiness characteristics 26
Table 5. Risk levels associated with trustworthiness characteristics for LSP5 27
Table 6. Vulnerabilities and countermeasures for identified threats in LSP5 28
Table 7. Controls and corresponding threats in TrustGuard results for LSP5, initial pilot 33
Table 8. Countermeasures and associated controls from example risks in TrustGuard 36
Table 9. Evaluation approaches to address relevant trustworthiness characteristics 38
Table 10. Experimental results, nowcasting, TFT compared to other model families 40

List of Figures

Figure 1: Overview of setup for AI prediction model development and validation 14
Figure 2: Overview of setup for trial of AI-based predictive control 14
Figure 3. Stylized overview of the ReWTS approach. 16
Figure 4. Photos from Veas denitrification facility showing trial low-cost chemical sensor 18

Figure 5. FAITH AI TAF V1 23
Figure 6. Excerpt of validation and training plan for LSP5 29
Figure 7. Screenshots of TrustSense as applied in LSP5 30
Figure 8. Screenshot of AI Model Hub as applied in LSP5 32
Figure 9. Screenshots of TrustGuard as applied in LSP5 33
Figure 10. Table overview of mean absolute error of predictions by time-segmented ReWTS models and non-time-segmented global models 39
Figure 11. ReWTS vs. non-time segmented (global) TCN prediction models. ReWTS with lower MSE for the vast majority of time chunks 40
Figure 12. Covariate impact of prediction accuracy for LSTM nowcast model 41
Figure 13. Impact of previous time-chunk model weights on current ReWTS prediction model. 42
Figure 14. Visualizations of actual and predicted nitrate levels, error rate, methanol levels 43
Figure 15. Screenshot of prediction model outputs and actual nitrate levels 44
Figure 16. Gantt chart of MPC trial in LSP5 replication pilot. 52
Figure 17. Known risks for LSP5 53

1 Introduction

This document reports on the initial pilot of the FAITH Large Scale Pilot 5 (LSP5). LSP5 is one of seven large scale pilots in FAITH and enables the trial and validation of the FAITH artificial intelligence trustworthiness assessment framework (FAITH AI TAF) on a specific industrial process – wastewater treatment.

LSP5 involves two of the FAITH project partners: SINTEF and Veas. The organizer of LSP5 is SINTEF, an independent research organization for technology research with AI and industrial applications as specific research priorities. The host of LSP5 is project partner Veas, who runs the largest wastewater treatment plant of Norway.

The Veas wastewater treatment plant has a well-established digital infrastructure for control and data management. In LSP5 we draw on this digital infrastructure to leverage AI for improvement of process control in denitrification.

The objective of this deliverable is to present the initial pilot of LSP5. Here, SINTEF and Veas has collaboratively advanced AI models for prediction of nitrate levels in treated wastewater and started the work to apply these models in predictive process control. Furthermore, we have made use of the FAITH AI TAF through TrustGuard, TrustSense, the AI Model Hub.

In the remainder of this first section we will provide an overview of the scope and objective of the initial pilot in LSP5. Then we will present the system under test (SUT), that is AI prediction models assessed by application of FAITH AI TAF, before providing an overview of the entire deliverable.

1.1 Scope and objectives of the initial pilot in LSP5 - the industrial processes LSP

The initial pilot concerns the application of the FAITH AI TAF and associated tool support for assessing and supporting trustworthiness of the AI prediction models for use in wastewater treatment.

Within this scope, the pilot has included research to advance the AI prediction models to be assessed, and to ensure that these advancement are optimized with respect to key trustworthy AI attributes as identified in FAITH AI TAF.

The following objectives were outlined for the initial pilot, as detailed in the plan for LSP presented in D8.1: (a) the development and validation of AI prediction models, (b) ensuring AI model explainability, (c) involving stakeholders in assessments, (d) applying pilot outcomes for improved decision making, and (e) applying the FAITH AI TAF for support of the pilot. The full detail of the objectives is provided in Table 1.

Objective	**Detailing**
Development and validation of AI prediction models	Development and validation of models to predict nitrate levels in treated wastewater, from baseline models available as a starting point in FAITH. Particular attention will be given to models accounting for variation across seasons and weather conditions.

Ensuring model explainability	The explainability of AI prediction models in LSP5 is critical for models to satisfy quality requirements and be trusted as part of the critical infrastructure of wastewater processing. SINTEF and Veas collaborate to identify explanations and information needed to satisfy the requirements of the LSP stakeholders.
Involving stakeholders in assessments	To facilitate and guide stakeholder involvement in assessments, the main activities include involvement via routine and ad-hoc meetings, access to AI prediction models through a Veas-controlled user interface, and the planning and execution of evaluation workshops with stakeholders. The main focus in the initial pilot has been on stakeholders at the level of process owners and engineers. LSP5 also presents AI models and initial outcomes to at least three wastewater treatment plants at a national level.
Applying pilot outcomes for improved decision-making.	To enable improved decision-making through the use of the initial pilot outcomes, the pilot outcomes will be applied to identify both improvement potential in the current process, and to identify needs and requirements for additional AI prediction models to be taken forth in the replication pilot.
Apply FAITH AI TAF to support initial pilot in LSP5	The assessments of the initial pilot will be supported by the FAITH AI TAF, in part for the assessment and risk modelling concerning the system under test, in part to assess the trustworthy AI maturity of the involved team including SINTEF and Veas participants.

As will be seen in this report, we have delivered on all objectives presented in D8.1. We have developed a range of AI model types, using tree-based models, linear regression, and deep learning, including transformer-based approaches. Furthermore, promising prediction models have then been validated over time on current data.

Model explainability has been ensured through the development of relevant metrics and metric visualizations in collaboration between AI developers and domain experts, and further enhanced through ongoing work on a dashboard for prediction model overview.

Stakeholders and domain experts have been involved in assessments through several collaborative meeting places including bi-weekly stakeholder meetings, bi-weekly model validation meetings, and routine evaluation workshops.

Applying model outcomes for improved decision making is a key objective for Veas. Consequently, in the initial pilot we have conducted research to establish the basis for using AI prediction models for proactive MPC to enhance control of methanol dosage in denitrification.

Applying the FAITH AI TAF and associated tools to support the trustworthiness of the SUT has been key to the research and development in LSP5. Furthermore, this application has been important to provide feedback for further enhancement of the framework and tools.

1.2 Evolution of the AI system under test (SUT) from initial setup

The SUT of LSP5 is the AI prediction models developed within the project to enhance wastewater treatment. In the initial pilot, models have been developed to enable forecasting of nitrate levels in treated wastewater. Then, in the replication pilot, these models will serve as basis for process control, specifically the control of methanol in denitrification.

During the initial pilot, important advances of the SUT have been made in order to enable the trustworthy application of the AI prediction models in wastewater treatment. As described in D8.1, the SUT includes a range of AI prediction models at different levels of maturity, including prediction models of nitrate levels in wastewater treatment (initial versions, TRL4 at project start), interim prediction models of other aspects of denitrification (interim versions, TRL2 at project start), and advanced prediction models to control methanol dosage in denitrification (advanced versions, TRL2 at project start). In the initial pilot, our main concern has been with the initial and interim versions of the SUT. In the replication pilot, our focus will also include the advanced versions.

1.2.1 Evolution of initial AI prediction models – nitrate levels in outflowing wastewater

To enhance the trustworthiness of initial AI prediction models, key trustworthiness attributes include validity and reliability as well as explainability and interpretability. In response to this, we have developed AI prediction models with enhanced validity and reliability, and established processes, metrics, and visualizations to strengthen explainability and interpretability.

The evolution of AI prediction models to strengthen validity and reliability has included explorations of tree-based models, linear regression, and neural network models including recurrent neural networks and transformer-based approaches. In addition, we have continued advancing a temporally segmented ensemble model for enhanced predictions of timeseries with substantial periodic variation (the ReWTS approach). We are also exploring opportunities for enhanced data access through additional sensors in the wastewater treatment. The work on initial AI prediction models is detailed in Section 2.1.

1.2.2 Evolution of interim AI prediction models – other aspects of denitrification

The initial AI prediction models concern the prediction of nitrate levels in treated wastewater. However, to enable process control based on these prediction models also other aspects of denitrification need to be modelled.

In this reporting period, as part of the work towards the replication pilot, we have addressed this need through developing an environment for simulating denitrification. In this simulation environment key aspects of the denitrification process are represented which enables initial testing of AI-based process control prior to testing in the real wastewater treatment process.

Additionally, during the reporting period, we have explored AI models to predict time periods with relatively low efficiency in wastewater treatment or conditions for which it is challenging to make predictions. This work is further detailed in Section 2.1.3.

1.2.3 Evolution of advanced AI prediction models – process control

The advanced AI prediction models concern the application of these models to process control in wastewater treatment. This work is conducted within the replication pilot, but we will also provide a status update on this in the present report.

The proactive process control is planned as an MPC for signals to methanol dosage in denitrification. The planning and development of the MPC are conducted in close collaboration with AI developers at SINTEF and stakeholders, domain experts and technical expertise at Veas. Details on the plans for advancing MPC are provided in Section 6.2.

For the MPC trial in the operative environment of Veas, key trustworthy AI attributes include not only validity and reliability, explainability and interpretability, and accountability and transparency, but also safety, as well as security and resilience. As such, the evolution of advanced AI prediction models draws on five of the seven trustworthy AI attributes of FAITH AI TAF.

1.3 Structure of the document

The remainder of this report is structured into five main sections, each reflecting a core element of the FAITH pilot activities and the ongoing maturation of the AI system within the LSP context.

Section 2 provides an overview of the progress of the AI system to date, including its deployment status, user engagement, training activities, and the initial evaluation processes undertaken during the first pilot phase.

Section 3 describes the application of the FAITH AI TAF and associated tools. It details how each component of the TAF has been applied and how framework and tool feedback has been provided. This section also reports on tool-specific validation, and documents the outcomes of initial threat identification, risk assessment, and countermeasure analysis.

Section 4 presents the evaluation of the AI system. It reports on the quantitative performance of the model using pilot data and provides a trustworthiness assessment. The section concludes with a preliminary impact evaluation, examining both the positive and negative effects of the AI system within the specific operational context of LSP5.

Section 5 addresses regulatory and ethical considerations relevant to the deployment and future scaling of the AI system. This includes an assessment of the system's trajectory toward potential certification pathways and its alignment with applicable regulatory frameworks, such as the requirements of the EU AI Act for high-risk AI systems.

Finally, Section 6 summarises the key achievements of the second reporting period and outlines the roadmap for the replication pilot. It provides a high-level plan for the replication phase and identifies the principal risks that must be monitored and managed as the system progresses toward broader deployment.

2 Progress of the AI system deployment and execution of the pilots

2.1 Status of the initial pilot execution

During the initial pilot, there has been extensive advancement of AI prediction models for nitrate levels in treated wastewater as well as other aspects of denitrification. Selected models have been validated with domain experts and stakeholders at Veas through routine validation meetings, and a broader involvement of users and stakeholders has been enabled through evaluation workshops. Outcomes of these evaluations have been used to improve the prediction models and establish confidence in them, allowing for the initiation of a trial of AI-powered process control. In this Section 2.1, we provide an overview of these advances, following an initial overview of the deployment context and domain specifics.

2.1.1 Deployment context and domain specifics

The deployment context of LSP5 is wastewater treatment at Veas, the largest wastewater treatment plant in Norway. The Veas plant is key to the treatment of wastewater from Oslo and neighbouring municipalities in the inner part of the Oslo fjord. The marine environment of the Oslo fjord is under threat, in particular from spill of nitrogen to the fjord with implications for algae growth and subsequently the marine environment at large. Municipal wastewater is a potential source of nitrogen spill to the fjord.

To mitigate the spill of nitrogen, Veas runs an advanced facility for denitrification as part of their wastewater treatment process. We here provide a brief summary of the process. For details see the deliverable D8.1.

Denitrification is part of the biological process at Veas, conducted as a final treatment step after mechanical and chemical treatment of the wastewater. In denitrification, bacteria convert nitrate in the wastewater to nitrogen gas released into the atmosphere. The internal goal of Veas at project onset was 80% nitrogen removal through the entire course of the wastewater treatment process, with a current adjustment to 78%.

The process of denitrification follows a preceding sub-step of nitrification, and is driven by adding methanol to the wastewater. Denitrification at Veas requires nearly 4500 metric tons of methanol per year. A reduction in methanol consumption by transitioning from reactive control to AI-based predictive process control represents substantial efficiency potential.

Veas has implemented automated process control on the basis of an advanced data infrastructure. However, Veas has not yet taken up AI for process control. Uptake of AI would enable predictive process control through MPC, and estimates prior to the project has indicated substantial potential to process effectiveness and efficiency from such predictive control. As detailed in D8.1, key business and sustainability goals for Veas for the uptake of AI-powered MPC in denitrification are:

- **Effectiveness improvements**, in compliance with Veas' internal goal of 80% nitrogen removal by the end of the FAITH project.

- **Efficiency improvements**, with significant reduction in methanol spent for removal of nitrogen in wastewater by project end.

2.1.2 From AI model validation to AI for predictive control

In LSP5, AI-based decision support and control will advance from explorations of AI models on the basis of historical data (TRL2-4) to AI models demonstrated in the operational environment at Veas (TRL7).

As outlined in D8.1, LSP5 was foreseen to move through three scenarios: AI model development and validation (scenario 1), model predictions used by operators (scenario 2), and AI models for predictive control through MPC (scenario 3).

In the initial pilot, AI model development and validation have been sufficiently successful for Veas to advance to early trials of AI-based MPC. This has happened earlier than anticipated. Because of this, we in this reporting period have prioritized early preparation for scenario 3, rather than the intermediated scenario 2. This allows us to run the planned trial of MPC already in the period M32-M39, leaving room for a second trial of MPC in the period M40-M51 provided the first trial is successful.

As such, the updated aims for the initial pilot and advanced pilot include the following:

- **Initial pilot (M3-M28):** AI model development and validation.
- **Replication pilot (M20-M51):** Trials of predictive control (MPC) in two iterations, in parallel with continued AI model development and validation.

During the initial and replication pilots, we make use of two different setups at Veas. The first setup is for model development and validation. The second setup is for trials of predictive control.

Setup for model development and validation: During development and validation, AI prediction models are developed and run by researchers at SINTEF on historical and real time data from the Veas control system. For model development, data are made available to researchers at SINTEF via the ABB Edge Insight edge interface. For model validation, AI prediction models run at a SINTEF server and are made available to personnel at Veas through the third-party platform Intelecy. An overview of the setup for AI prediction model validation is presented in Figure 1. Details on ABB Edge Insight and Intelecy are provided in D8.1.

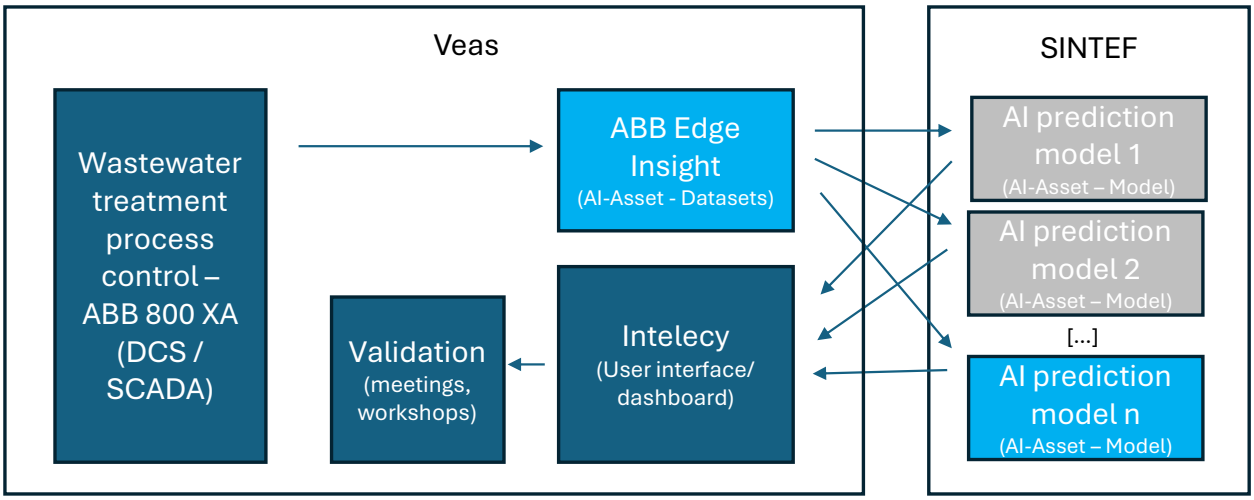


Figure 1: Overview of setup for AI prediction model development and validation

Setup for trial of AI-based predictive control (MPC): During the trials the MPC will be made available on a dedicated PC at Veas running an OPC-UA server. The server sends a control signal from the MPC to the Veas process control system ABB 800 XA (DCS / SCADA). The control signal is sent so that it can be utilized by the Veas control system, but without the MPC or the OPC-UA server having access to the control system. In addition to the process control system, Veas operators are also provided access to information on the MPC via a dedicated dashboard developed by SINTEF. AI developers at SINTEF can monitor the process and MPC performance. Process data are provided through ABB Edge Insight. An overview of the setup is provided in Figure 2.

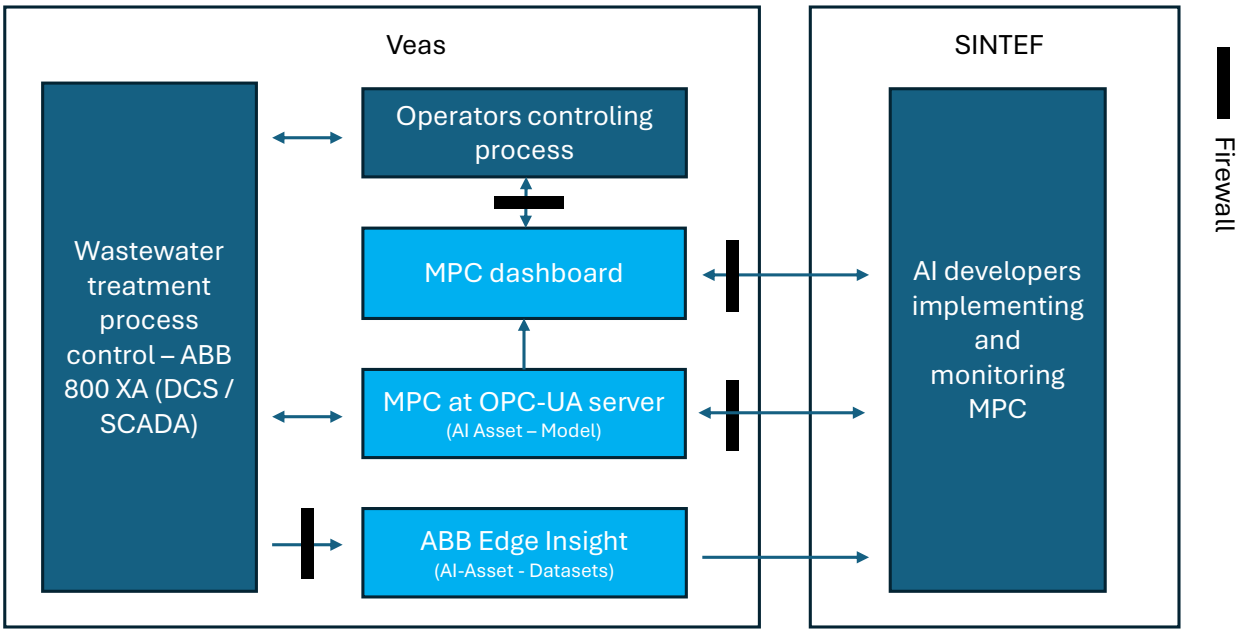


Figure 2: Overview of setup for trial of AI-based predictive control

2.1.3 AI prediction models advanced in the initial pilot

During the initial pilot, we have explored and advanced a range of AI-models to predict key aspects of the denitrification process. We have specifically focussed on the prediction of nitrate levels in treated wastewater, but also other aspects such as predicting time periods with relatively low efficiency in wastewater treatment. In this section we present the explored approaches and advances made.

2.1.3.1 Initial nowcasting and forecasting models

In an initial exploration of AI models for predicting time series data of relevance for denitrification in wastewater treatment, a range of approaches was applied during the first project year. This exploration was grounded in work preceding the FAITH project. In these initial explorations, AI prediction models were developed on data from a small separate testing facility at Veas – the P*ilot Reactor*. Timeseries for the Veas Pilot Reactor are assumed to be less challenging to model than timeseries from the main reactors, as the Pilot does not apply feedback from the process outcome to update the control of methanol.

For initial nowcasting and forecasting models, three main types of approaches were explored: linear regression, neural networks, and tree-based methods.

- **Linear regression – ElasticNet:** ElasticNet is a regularized regression method that utilizes the L_1 and L_2 penalties from the Lasso and Ridge methods to avoid overfitting. As a regression method, ElasticNet is simple and interpretable, but at the cost of expressiveness as it can only learn functions with linear inputs.
- **Neural networks – LSTM and TCN**: For initial explorations of neural networks, long short-term memory (LSTM) and temporal convolutional networks (TCN) were applied. These methods have higher expressiveness than regression models as they can approximate any function of inputs, but at the cost of computational complexity and interpretability.
- **Tree-based methods – XGBoost:** For tree-based methods, we explored gradient-boosted decision trees with XGBoost. Tree-based methods typically have lower expressivity than neural networks, but higher than regression-based approaches, and may be preferable in cases of limited data for training.

The initial explorations provided indications of the strengths and limitations of the different approaches. The initial explorations are described in detail in a conference paper by Böhn et al. (2026) which is currently under review.

2.1.3.2 Chunk-based forecasting models

Timeseries data from wastewater treatment vary across hours of the day, seasons, and changes in weather such as heavy rainfall or snow-melting. Following from this, robust AI prediction models need to be capable of accurate predictions under substantial seasonal and weather change.

To allow for robust predictions, we have in the initial pilot advanced a chunk-based approach to forecasting. This approach is referred to as the Recency-Weighted Temporally-Segmented (ReWTS) ensemble method, and is presented in an article published in the Journal of Artificial Intelligence Research (Johnsen et al., 2025).

The ReWTS approach trains specialized models on data from segmented time periods, so-called chunks. In forecasting it then optimizes performance by selecting the best-suited models based on recent patterns in a recent look-back period. Tested on process data from Veas, ReWTS outperformed global models where one model is trained on the entire dataset, reducing forecasting errors by 10–70% while showing significantly higher resilience to outliers. A stylized overview of the ReWTS approach is provided in Figure 3. For detailed findings on the accuracy and robustness of the ReWTS approach, see Section 4.2.1 below.

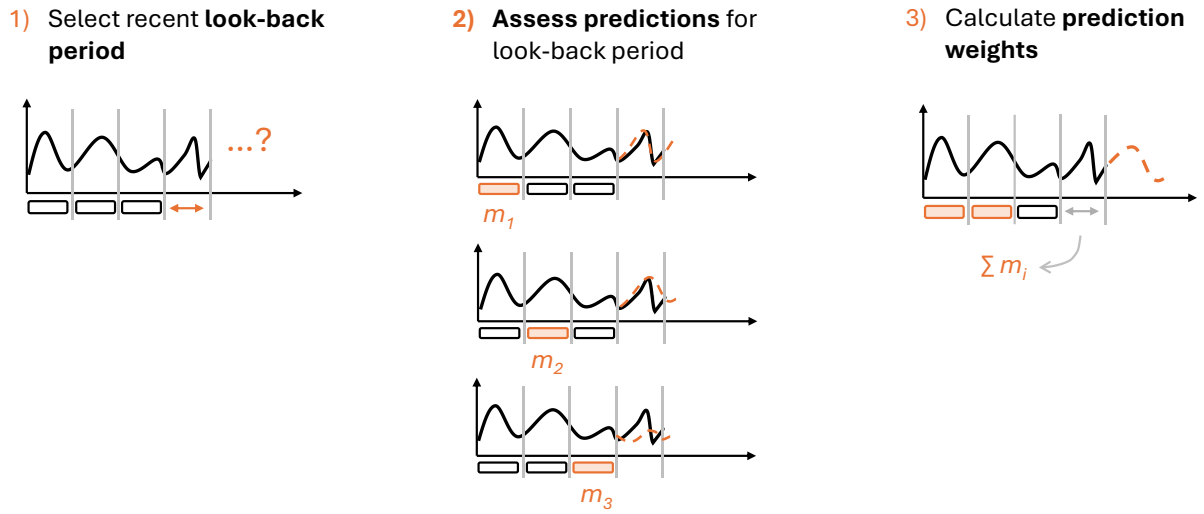


Figure 3. Stylized overview of the ReWTS approach where prediction model weights are calculated on the basis of assessments of models from previous limited time chunks.

2.1.3.3 Forecasting models advanced to validation with domain experts

Based on initial explorations and research on the ReWTS approach, two chunk-based forecasting models have been advanced in LSP5 for validation with domain experts: The regression-based ElasticNet and the neural network TCN. Both models were set up in line with the ReWTS approach, with model weights continuously updated based on the prediction performance of the available time-segmented models in the recent look-back period.

While initial explorations were conducted on data from the Veas Pilot Reactor, the forecasting models advanced for validation with domain experts were trained on data from one of the eight Veas main reactors, Veas *Main Reactor 1*. The eight reactors run in parallel, and a prediction model suitable for Veas Main Reactor 1 will be suitable also for the remaining seven reactors.

Real time predictions from the two forecasting models that were advanced to validation with domain experts, have been made available to Veas personnel through the Intelecy platform. Furthermore, weekly metrics and graphical representations of prediction performance and explanations are provided for validation meetings. See details in Section 4.2.2 below.

2.1.3.4 Models for efficiency and effectiveness predictions

In addition to the AI model advances described above, we have in the initial pilot included two exploratory approaches to AI modelling. The first of these, described in this section, is explorations of AI models to predict the efficiency and effectiveness of nitrate removal. Specifically, we have explored models to predict periods where methanol added to the denitrification process has relatively low returns.

For this purpose, regression-based models were developed to estimate nitrate removal performance as a function of process covariates such as methanol dosage, inflowing nitrate concentration, oxygen levels, and temperature. These models aimed to quantify the expected nitrate reduction under given operating conditions and thereby identify situations where increased methanol dosage does not yield proportional improvement.

However, the results highlighted several limitations, including sensitivity to noise in sensor measurements, limited explanatory power due to missing or unobserved variables, and challenges in sufficiently capturing the dynamic nature of the biological process. Consequently, the models provided useful directional insights but are not yet sufficiently robust for direct operational decision-making. As such, we instead reoriented our exploratory work towards transformer-based approaches presented in the following section.

2.1.3.5 Transformer-based nowcasting and forecasting models

The second exploratory approach in the initial pilot, has been to advance transformer-based nowcasting and forecasting models. The benefit of a transformer-based approach in wastewater treatment is its ability to consider substantial complexity over a relatively large time span, however at the cost of computational resources.

In the initial pilot, a temporal fusion transformer (TFT) has been adapted into a regression setting to estimate nitrate concentration based on historical multivariate sensor data. The architecture supports context-aware feature weighting, meaning that the importance of variables such as methanol dosing, oxygen levels, or flow conditions can shift depending on the current process state, providing a more flexible and realistic representation of system dynamics.

The result from this exploratory work suggests that the transformer-based approach may provide a marked improvement over classical regression and tree-based models, as well as traditional deep learning methods such as LSTM and TCN. The transformer-based approach demonstrates stronger ability to represent the nonlinear and time-dependent relationships, which may lead to more stable and accurate estimation of nitrate levels across varying operational conditions. Furthermore, the model benefits from its capacity to integrate information across both short-term fluctuations and longer-term trends, which are often difficult to capture simultaneously using conventional approaches.

The initial explorations of the transformer-based approach has been conducted on data from the Veas Pilot and is presented in a paper accepted for an academic conference (Uddin et al., 2026). Selected findings from this work, on model performance in terms of accuracy, is provided below in Section 4.2.1.

The next step in our transformer-based explorations, conducted as part of the replication pilot, will be to continue model advancement on data from one of the main process reactors. If successful, we may consider a transformer-based approach also for predictive control.

2.1.4 Explorations of increased data capture – sensor trial

Denitrification at Veas is conducted in eight parallel main reactors. Each main reactor includes four biological filters where bacteria transform nitrate to nitrogen gas and released from the wastewater. Currently, each main reactor includes only one sensor for measurement of nitrate levels in treated wastewater. For strengthening accuracy and robustness in AI prediction models, each reactor would ideally have one such sensor in each of its four biological filters – that is four sensors in each main reactor. However, the purchase cost of current sensors is prohibitive.

As an alternative to the current cost intensive optical sensors, we in LSP5 are exploring whether the current optical sensors can be augmented with low-cost chemical sensors. For this, in the initial pilot, we have installed one low-cost chemical sensor next to the current optical sensor in a test tub in the Veas Main Reactor 1. The low-cost chemical sensor trial has been running since M18. We are currently reaching a volume of data where it is meaningful to explore the performance of the chemical sensor relative to the optical sensor – given the substantial variation in wastewater characteristics over longer time spans, including temperature changes and changes in the composition of ions in the wastewater.

To explore the performance of the low-cost chemical sensor and potentially benefit from its data we will, during the replication pilot, seek to calibrate its outputs to those of the optical sensor. For the calibration, we will apply regression-based and machine learning approaches set up as a nowcasting of optical sensor data values from the low-cost sensor data values and process covariates.



Figure 4. Photos from Veas denitrification facility showing trial low-cost chemical sensor mounted in test tub next to optical sensor

2.1.5 User and stakeholder involvement in the initial pilot

Users and stakeholders in LSP5 encompass three key groups: process owners, engineers, and operators. Process owners hold high-level responsibility at Veas and include the production leader and process leader. Engineers have responsibility for specific aspects of the wastewater treatment process or related functions and includes process engineers, system engineers, operations engineers, test engineers, and development engineers. Operators have responsibility for the operative control of the wastewater treatment process.

Users and stakeholders have been involved through different means in the initial pilot. In particular, three forms of involvement have been important throughout the initial pilot: Stakeholder meetings, validation meetings, and evaluation workshops.

Stakeholder meetings are meeting points for SINTEF and Veas for status reporting. These are conducted biweekly throughout the project period. Here, Veas is represented with several engineers to ensure sufficient coverage of the denitrification process – including development engineer, process engineer, system engineer, and operations engineer. At the stakeholder meetings we report on status and advances for ongoing tasks with the aim of guiding the research process according to the needs and requirements of Veas.

Validation meetings have been conducted as biweekly meeting points between AI researchers at SINTEF and process engineers at Veas. Validation meetings have been conducted since real time AI model predictions were made available to Veas personnel – from M18 and onward. In preparation for the meetings a range of metrics and visualizations are provided for the AI model predictions of the last two weeks, with the aim of scrutinizing model prediction quality and contrasting this to known events in the weather or at the plant – such as heavy rainfall or filter cleaning. Hence, the validation meetings enables insights into events that may impact model prediction quality. Details on the measures and visualizations used during the validation meetings are provided in section 4.2.2 below.

Evaluation workshops are the arenas for broader involvement within Veas and cover all user and stakeholder groups. Here, key advances and plans are presented to production and process leaders, engineers, and operators for discussion and guidance on future directions. A series of eight evaluation workshops has been planned – four with a strategic orientation and four with an operational orientation. At the time of writing of this report, three of the eight workshops have been conducted – two with a strategic orientation and one with an operational orientation. A fourth workshop with operational orientation is planned before the onset of the MPC trial (M32).

2.1.6 Evaluation process in LSP5

The evaluation process in LSP5 has been organized to address AI trustworthiness characteristics of key importance to wastewater treatment. As detailed in D8.1, trustworthiness characteristics of particular relevance concern validity and reliability (particularly the accuracy and robustness of AI model predictions), explainability and interpretability (with particular concern for providing engineers and operators needed information to adequately make use of and monitor AI models for process control), transparency and accountability (with particular relevance for operational use of AI models), security and resilience (with particular concern for protecting the Veas control system during trials and subsequent operation), and safety (with particular concern for safeguarding against risk towards the marine environment in the Oslofjord during trials or subsequent operation)

To evaluate the AI prediction models of LSP5 on key trustworthiness characteristics, the evaluation process has involved assessments of models according to established metrics as well as analyses for explanations and insight into the robustness of model predictions. In addition,

human-centred evaluations has been conducted in the form of validation meetings and evaluation workshops.

Details on the evaluation process are provided in Section 4.1. Evaluation results for the relevant trustworthiness characteristics are provided in Section 4.2. Furthermore, a preliminary impact evaluation is provided in Section 4.3.

2.2 Status of the replication pilot preparation and initialization

Due to promising results of the AI prediction models in the initial pilot, we have been able to move towards trials of MPC early in the replication pilot. The replication pilot runs in the months M20-M51. Within this period we are currently planning for a first trial of MPC in the period M32-M39. Following this, provided the first trial is successful, we foresee a second trial of MPC in the period M40-M51.

An overview of the detailed plan for the replication pilot is provided in Section 6.2.

2.3 Deviations from the original deployment plan

The initial pilot has closely followed the DoA description of T8.2, and also largely adhered to the plan for LSP5 as detailed in D8.1.

The initial pilot has, however, deviated from the plan in D8.1 on one point. In D8.1 we envisioned three scenarios of use for AI prediction models in the Veas wastewater treatment process: (1) AI model development and validation, (2) model predictions used by operators, and (3) AI models for predictive control through MPC.

Due to successful advancement of AI prediction models, we have moved to the third scenario – predictive control through MPC – more rapidly than envisioned. This entails that the interim scenario we initially foresaw, with the use of AI prediction models to guide manual control by operators, is not required. Instead, operators are involved in the trial of predictive control through MPC.

While this is a deviation from the plan as presented in D8.1, we see this deviation as a benefit and a reflection of the results achieved in LSP5 so far. In particular, we see it as beneficial that the relatively early onset of MPC trial potentially allow us to include also a trial of a second version MPC, allowing for more comprehensive outcomes of the replication pilot.

2.4 Data collection throughout the initial pilot execution

Throughout the initial pilot, data has been collected on the following:

- **Process data from Veas Pilot Reactor:** Data on all key process variables and co-variates from the Pilot Reactor, available through ABB Edge Insight at Veas. Data availability until March 2025, after which the Pilot Reactor has had a period of non-use.
- **Process data from Veas Main Reactor 1:** Data on all key process variables and co-variates from Main Reactor 1, available through ABB Edge Insight at Veas. Data availability from several years back and throughout the project period. Missing data period from mid-November 2025 to mid-February 2026 due to reactor updates.

- **Prediction output data from ElasticNet and TCN ReWTS models:** Predictions of nitrate levels in treated wastewater, provided at 10 minute intervals for 5 to 60 minute prediction horizons, for both ElasticNet and TCN ReWTS models. Includes data on previous time-chunk models contributing to model weights. Data stored at SINTEF server. Data availability from June 2025. Missing data period from mid-November 2025 to mid-February 2026 due to reactor updates.
- **Sensor data from low-cost chemical sensor:** Data from low-cost chemical sensor gathered through a local computer (Raspberry Pi) and transmitted over a wireless connection. Data availability from July 2025 and onwards.
- **Data and summaries from validation with domain expert:** Weekly summarization of metrics and visual representations for consideration of accuracy, robustness and potential sources of disruption to model predictions, as well as explanations of sources of ReWTS model weights. Summaries serve as a log of validation with domain experts.

Data and models from the initial explorations of prediction models for the Veas Pilot Reactor have been made openly available (https://data.sintef.no/feature/fe-63c812ef-13fe-4b0c-936e-44da3303a781; https://github.com/SINTEF/veas-denit-pilot).

2.5 Integration into LSP5 workflows: Lessons learned

The integration of AI prediction models into the workflows of LSP5 in the initial pilot has been achieved through model advancement, validation of model predictions with domain experts, and preparations for predictive control with stakeholders and users.

Through this work, several lessons have been learnt. In the following we summarize those we perceive as key:

- **Close collaboration between AI developers and domain experts is essential:** Regular validation meetings and iterative discussions significantly improves efficiency and effectiveness of model development with implications for performance, interpretability, and practical relevance.
- **Stakeholder and user involvement should be structured and frequent.** Biweekly status and validation meetings, combined with strategically oriented workshops, enable alignment, identify operational issues early, and facilitate organizational acceptance.
- **Infrastructure for real-time validation of AI prediction models accelerates progress and builds trust.** The opportunity for live predictions and weekly summaries provided domain experts with direct exposure to model behavior, enabling meaningful, data-driven assessments.
- **Explainability and interpretability must be developed in parallel with predictive performance.** Metrics and visualizations established in collaboration with AI developers and domain experts were critical for Veas engineers to trust, understand, and evaluate

the AI models. Furthermore, such metrics and visualizations in turn are critical for advancing model performance.

- **Applying the FAITH AI TAF and related tools provides systematic structure and sensitizes AI developers and domain experts to improvement needs:** Using TrustGuard, TrustSense, and the AI Model Hub supported risk identification, explainability improvements, and maturity assessment – thereby supporting and guiding the process.

3 Application of FAITH Tools

The criticality of wastewater treatment for the sustainability of the Oslofjord marine environment accentuates the importance of ensuring trustworthiness for AI models in support of this process. In LSP5, the FAITH framework for trustworthy AI (FAITH AI TAF) and associated tool support (TrustGuard, TrustSense, AI Model Hub) are deployed in the lifecycle of AI model development and deployment at Veas.

In this chapter, we report on the application of the FAITH TAF V1 methodology, the tool validation and planning process, and the validation of each of the FAITH tools.

3.1 Application of the FAITH AI TAF V1 methodology

The FAITH AI TAF V1 methodology has been used to support the development and validation of the AI-based prediction models in the initial pilot.

The FAITH AI TAF V1 consists of six different phases (see Figure 5). Each phase has been conducted for LSP5 through collaboration with SINTEF and Veas.

Below we provide accounts for each phase.



Figure 5. FAITH AI TAF V1

3.1.1 Cartography

The cartography phase concerns the identification of the AI system under test. Specifically, it is important to identify assets and users of relevance to the AI system.

In LSP5, two assets and two main user groups have been identified. The assets and user groups are presented in the tables below.

Table 1. Assets in LSP5 initial pilot

Asset	Description
Data from the Veas process	The data for the AI prediction models are provided from the Veas control system through ABB Edge Insight. Datasets for training, testing, and validation of AI prediction models are exported from ABB Edge Insight to a local server at the AI research partner SINTEF. Data used when running the prediction models are made available from ABB Edge Insight.
AI prediction models	AI prediction models have been trained, tested, and validated on the basis of data from ABB Edge Insight. For a detailed overview of the AI prediction models, see Section 2.1.3. Two AI prediction models, one regression model (ElasticNet) and one deep learning model (TCN) has been made available to users at Veas through a third party platform Intelecy.

Table 2. Users in LSP5 initial pilot

Users	Description
Veas engineers and operators	Engineers and operators at Veas with responsibilities for maintaining, enhancing, and running the wastewater treatment process. This user group has expertise in wastewater treatment and well established competence in digital process control.
SINTEF AI developers	AI developers at SINTEF with responsibilities for development of AI prediction models for use in the Veas wastewater treatment process. This user group has expertise in AI research and development, but limited knowledge of wastewater treatment.

As part of the cartography phase, an initial assessment of the team AI maturity through TrustSense was conducted (see Section 3.3.1 below), gathering further insight into the users. Furthermore, an initial application of the AI Model Hub (see Section 3.3.2 below) was conducted to gather further insights on the model and data assets .

As part of the cartography phase, an initial criticality assessment was conducted. Here we examined the AI system relative to the AI Act and NIS2 directive. Given the current application of the AI system for internal validation purposes only, it is uncertain whether the system in the initial pilot should be characterized as high-risk under the AI Act or as an essential or important service under NIS2. This will be reconsidered for the replication pilot, when the aim is to include AI-based predictive control in the wastewater treatment process.

3.1.2 Threat analysis

During the treat analysis, technical threats and social threats were assessed with support from the list of potential threats provided in FAITH TAF v1. Here, 13 threats that may affect the AI system assets (model and data) were assessed with regards to likelihood and contextual considerations.

In addition, we identified key threats corresponding to the specific AI system in LSP5. In table 3 we list identified threats corresponding to the five AI trustworthiness characteristics of particular relevance to LSP5. These threat has been included in the FAITH Knowledge Base.

Table 3. Threats identified during threat analysis in LSP5

Threat title	Threat description	AI life-cycle stage	Affected AI Asset	Affected TAI characteristic
Inaccurate predictions	Inaccurate predictions from model will impact negatively its applicability and relevance for the use case.	Testing	Models	Valid & reliable
Non-robust predictions	Predictions that are not robust over seasonal variation will negatively impact model applicability	Testing	Models	Valid & reliable
Insufficient data quality	Data quality for given time periods may be reduced due to issues with sensors, or lack of sufficiently fine-grained sensors	Testing	Data	Valid & reliable
Reduced operator accountability due to lack of interpretability	Operator responsibility requires explainability. If operators cannot interpret easily the AI model output, their accountability is at risk.	Testing, Design	Models	Explainable & Interpretable
Reduced ability of operators to identify prediction errors	Lack of interpretability may challenge operators ability to identify errors in the AI model predictions	Testing, Design	Models	Explainable & Interpretable
Loss of trust due to lack of interpretability	Lack of interpretability may reduce operator trust, even for predictions of adequate accuracy.	Testing, Design	Models	Explainable & Interpretable
Over-reliance in model predictions	Lack of interpretability may unduly increase operator trust, even for predictions of inadequate accuracy	Testing, Design	Models	Explainable & Interpretable
Deskilling from lack of transparency	Lack of transparency in models may over time threaten operator skill	Testing	Models	Accountable & transparent
Diffusion of accountability due to lack of transparency	Lack of transparency may make it challenging to correctly assign accountability and responsibility in use of AI prediction model	Testing, Design	Models	Accountable & transparent
Security and resilience threats	Malfunction of ML model, e.g. following from lack of input data, is assumed to be most frequently encountered threat.	Testing, Design	Models	Secure & Resilient
Environmental risk from AI model error	Risk of insufficient wastewater treatment following AI model prediction error	Design	Models	Safe

3.1.3 Impact assessment

Following treat assessment, we conducted an impact assessment related to the different trustworthiness characteristics. Here, with basis in the analysed threats, we assessed the business level consequences of compromise to each of the seven trustworthiness characteristics considered in FAITH AI TAF.

The impact assessment served to clarify that in LSP5, the trustworthiness characteristic of being Valid & Reliable is of particular relevance due to the substantial business consequences that may result from a breach. Furthermore, also the characteristics of Explainable & Interpretable and Accountable & Transparent are of high importance in this regard.

As such the impact assessment was instrumental for subsequent vulnerability and risk analysis, as well as the identification of countermeasures.

An overview of the impact assessment for the AI trustworthiness characteristics is provided in Table 4.

Table 4. Impact assessment – business consequences for compromised trustworthiness characteristics

Trustworthiness characteristic	**Business consequence if compromised**	**Details on business consequence**
Valid & reliable	Very high	Value of AI prediction model depends on its performance. Without sufficient reliability it may not be relevant for wastewater treatment.
Explainable & Interpretable	High	Important to maintain operator skill over time. Lack of E&I may lead to reduced trust and hamper uptake. May also impact operator accountability, as operator cannot be accountable for something they cannot interpret.
Accountable & transparent	High	Compromised A&T may entail loss of competence, skill-degradation. Important that operators also understand rationale for model output.
Secure & Resilient	Low	Poor data, e.g. from sensor or system failure, could have some business consequences. But are manageable within current regime.
Safe	Low	In the case of failure, wastewater treatment will default to current non-AI-based procedures. Hence, low business consequence.
Fair with harmful bias managed	N/A	Fairness not relevant in AI prediction model as wastewater treatment does not entail differentiation between groups but treats water from a larger region.
Privacy-enhanced	N/A	Person data is not handled in the AI prediction model.

3.1.4 Vulnerability assessment

Following the assessment of the business impacts of compromised trustworthiness characteristics, we conducted a vulnerability assessment related to all threats identified in the threat analysis above.

The vulnerability assessment was conducted both for the threats indicated in FAITH AI TAF as well as the threats identified as specific to LSP5. A full overview of the considered vulnerabilities are presented in Table 6 at the end of Section 3.1. In the following, we discuss in more detail two examples of the vulnerabilities considered.

3.1.4.1 Example vulnerability assessment – model accuracy

The assessment noted that 'inaccurate predictions' is a threat. This accentuates that the AI model itself may be considered in terms of vulnerability. In LSP5, the accuracy of the AI model depends on available training data being sufficiently representative of the current situation. This may represent a challenge due to the changing conditions of wastewater treatment across weather and seasons, and has been addressed through the ReWTS approach with time-segmenting and a

recent look-back period for identification of the most accurate models in the current time window. However, some weather situations may be exceptional, and it may be difficult to find adequate models for predicting time series changes in these conditions. In such situations, model accuracy is a vulnerability to the system.

3.1.4.2 Example vulnerability assessment – model interpretability

The assessments also noted that 'reduced ability of operators to identify prediction errors' is a threat. This accentuates the importance of interpretable model outputs. Without such interpretability, operator difficulties in identifying error becomes a vulnerability.

3.1.5 Risk analysis

The vulnerability analysis indicated three trustworthiness characteristics as having particularly high business consequences in cases of being compromised: Valid & Reliable, Explainable & Interpretable, and Accountable & Transparent.

These three trustworthiness characteristics are also characterized as being likely of very likely compromised provided that no mitigation actions are in place. In consequence, risk levels for these three trustworthiness characteristics are high or very high, and these should be particularly addressed through controls in the form of countermeasures.

For the other trustworthiness characteristics in FAITH AI TAF, business consequences are low or not applicable. For these characteristics, the likelihood of being compromised is also relatively lower than for the three characteristics with higher risk levels. An overview is provided in Table 5.

Table 5. Risk levels associated with trustworthiness characteristics for LSP5

Trustworthiness characteristic	**Business consequence if compromised**	**Estimated likelihood of compromise if not controlled**	**Risk if not controlled**
Valid & Reliable	Very high	Very high	Very high
Explainable & Interpretable	High	High	High
Accountable & Transparent	High	High	High
Secure & Resilient	Low	Medium	Medium-Low
Safe	Low	Medium	Medium-Low
Fair with harmful bias managed	N/A	N/A	N/A
Privacy-enhanced	N/A	N/A	N/A

3.1.6 Countermeasures

Following the risk analysis, countermeasures for specific threats and vulnerabilities were identified. The countermeasures serve as controls for managing risks and vulnerabilities related to the different trustworthiness characteristics of specific relevance to LSP5.

An overview of the countermeasures for the identified risks and vulnerabilities are provided in Table 6.

Table 6. Vulnerabilities and countermeasures for identified threats in LSP5

Threat title	Affected AI Asset	Affected TAI characteristic	Vulnerabilities	Countermeasures
Inaccurate predictions	Models	Valid & reliable	Model accuracy	Iterative training and assessment of AI prediction models. Trial of different modelling approaches.
Non-robust predictions	Models	Valid & reliable	Model robustness	Iterative training and assessment of AI prediction models. Trial of different modelling approaches.
Insufficient data quality	Data	Valid & reliable	Data quality	Improve data quality through sensors for added data collection. Identify and mitigate missing data in dataset.
Reduced operator accountability due to lack of interpretability	Models	Explainable & Interpretable	Operator responsibility requires explainability	Human-centred evaluations with operators for improved interpretability
Reduced ability of operators to identify prediction errors	Models	Explainable & Interpretable	Difficult to identify prediction errors if poor interpretability	Human-centred evaluations with operators with focus on identification of prediction errors
Loss of trust due to lack of interpretability	Models	Explainable & Interpretable	Loss of operator trust	Human-centred evaluations with operators with focus on trust-enhancing measures.
Over-reliance in model predictions	Models	Explainable & Interpretable	Over-reliance in model predictions	Training and AI skill development in process operators and engineers.
Deskilling from lack of transparency	Models	Accountable & transparent	De-skilling of operators	Transparent models, assessed with experienced operators and engineers. Insight into complex relationships between factors
Diffusion of accountability due to lack of transparency	Models	Accountable & transparent	Diffusion of stakeholder accountability	Accountability strengthened through transparency and interpretability, evaluated with stakeholders.
Security and resilience threats	Models	Secure & Resilient	Resilience issues due to data fallout	Improved data capture through additional sensors, and monitoring of data quality.
Environmental risk from AI model error	Models	Safe	Potential environmental risk from erroneous AI model predictions, leading to insufficient wastewater treatment	Risk is low as the AI prediction models are planned deployed in an existing process with continuous monitoring of process output, and where the established control system will serve as fallback in case of issues.

3.2 Tool validation planning & training processes

The tool validation planning and training process has spanned the period of the initial pilot. Key activities included an initial trial of the precursor to TrustGuard – a so-called 'dry-run'. In addition, we have engaged in plan development and tool training.

3.2.1 Dry-run

The dry run of the precursor to TrustGuard for LSP5 was conducted in collaboration between SINTEF and Veas. Through the dry-run, the first four phases of the FAITH AI TAF V1 were addressed in order to provide guidance to the initial pilot.

The dry-run helped establish initial versions of the assessments presented in Section 3.1. The assessments in turn provided basis for the AI system evaluations presented in Section 4.

3.2.2 Planning the validation and training

To coordinate the tool validation and training in LSP5 with the development schedules of the FAITH tools and the schedules of the other pilots, a plan for training and validation was established under the guidance of WP2.

Here, SINTEF and Veas collaboratively planned training and support activities as well as feedback activities for the FAITH tools and coordinated this with the technology advances to be made in the AI technology to be deployed in LSP5.

The planning has been critical for the coordination of LSP5 activities relative to the method and tool development in WP2. An illustrative excerpt of the plan is provided in Figure 6.

Veas & SINTEF				
	Project Start:		M23	
	Display Month:		1	
TASK	ASSIGNED TO	PROGRESS	START	END
Planning				
Define plan	Veas / SINTEF	100%	M25	M25
Agree plan	Veas / SINTEF	100%	M25	M25
Training and support				
Attend training for AI Model Passport (pythonic library)	SINTEF	100%	M26	M26
Attend training for AI Data Passport	SINTEF	100%	M26	M26
Attend training for trustSENSE	Veas & SINTEF	100%	M23	M23
Attend training for trustGUARD	Veas & SINTEF	100%	M26	M26
Attend training for SSM	Veas & SINTEF	0%	M32	M33
Attend introduction to FAITH STM Platform registration and workflows	Veas & SINTEF	0%	M33	M33
FAITH Tools Implementation and Iteration 2026				
Deploy SUT as initial trial MPC in Veas Pilot	Veas & SINTEF	30%	M26	M33
Use trustSENSE for team maturity assessment (early deployment period) - as part of evaluation workshop(operative focus)	Risk Assessor + End Users	100%	M27	M27
Provide formal feedback on Usability and Usefulness of trustSENSE	Risk Assessor	100%	M28	M28
Implement trustSENSE activities	SINTEF & Veas	0%	M33	M33

Figure 6. Excerpt of validation and training plan for LSP5

3.2.3 Tool training

In addition to the training provided on the FAITH AI TAF, the partners involved in LSP5 have received training in the FAITH tools TrustSense, AI Model Hub, and TrustGuard.

The training for TrustSense was delivered through plenary workshops that included detailed presentations and discussions of key aspects of the tool.

The training for AI Model Hub and TrustGuard was provided as dedicated partner-specific meetings – one for each tool – where the tools were walked through and explained with particular concern for the LSP-specific aspects of the tool.

In addition to the training sessions, we have reached out to the tool developers for explanation and support when needed, and has through this been able to do a thorough initial validation activities.

3.3 FAITH tool validation activities

In LSP5, the three FAITH tools TrustSense, AI Model Hub, and TrustGuard have been validated through use on the AI models advanced over the course of the initial pilot. In the following, we report on these validation activities for each of the tools separately.

3.3.1 TrustSense

TrustSense is a tool for assessing team maturity levels concerning trustworthy AI. Following a completed assessment, the tool provides a maturity level and recommended mitigation actions across eight categories of team proficiency.

TrustSense is sensitive to team expertise and distinguishes between AI experts and domain experts in addition to cybersecurity experts and facilities for adversary profiling.



Figure 7. Screenshots of TrustSense as applied in LSP5

3.3.1.1 Application of TrustSense in the initial pilot

In LSP5, two categories of users are relevant for TrustSense: Domain experts (Domain AI User Team) and AI developers (AI Technical Team). We have applied TrustSense for both groups and on two different occasions. First in M24, and then in M27.

The application of TrustSense has involved AI developer representatives from SINTEF and domain experts from Veas. TrustSense was also presented in an evaluation workshop involving 14 representatives from Veas and SINTEF. Here, TrustSense questions of particular importance were discussed with regards to strengthening team maturity.

3.3.1.2 Implications of TrustSense in the initial pilot

The application of TrustSense in the initial pilot has enabled a discussion within the LSP5 team on how to establish the needed maturity in domain experts when AI tools are introduced in wastewater treatment. Key questions in the discussion on team maturity, driven by the application of TrustSense, are:

- How to ensure that everyone on the team knows the purpose of the AI systems, their use, and their expected output?
- How to ensure that everyone on the team can critically evaluate AI predictions and is capable of identifying any errors or deviations in those predictions?
- How should the team handle deviations related to AI incidents (or threats)?

In response to these questions, plans have been established for further involvement of users and stakeholders in the implementation of AI to the wastewater treatment process, the provision of training of operators prior to AI system use, written procedures for monitoring AI systems during trials, and written procedures for the handling of unforeseen events associated with the AI systems.

3.3.1.3 Validation feedback on TrustSense in the initial pilot

SINTEF and Veas have provided validation feedback on TrustSense following both occasions of use – in M24 and M28. The validation feedback has addressed points of potential improvement, for further strengthening of the tool.

The feedback has concerned tool presentation and reworking of questionnaire items in the tool for general alignment with existing quality processes in the industry. The adaptations following the feedback has improved the fit of the tool to the context of LSP5.

Examples of items with improved fit of the tool include:

- The team collectively ensures that all members understand their roles in maintaining AI trustworthiness, holding **routine** review sessions. (boldened words replacing ‘monthly’)
- The team adheres to policies by maintaining a compliance score on Trustworthy AI **in accordance with established requirements** during regular audits. (boldened words replacing ‘above 95%’)

3.3.2 AI Model Hub - Data and Model Passports

AI Model Hub provides a digital record for an AI system’s lifecycle, supporting transparency and traceability. It can document datasets and model runs, including configurations and performance metrics, to help ensure that the model's origins and behaviours are verifiable.

The AI Model Hub includes two components of particular relevance to LSP5, a model passport and a data passport. The model passport enables the digital record for the AI model, and the complementary data passport enables a digital record for the data on which the model is trained, tested, and validated.

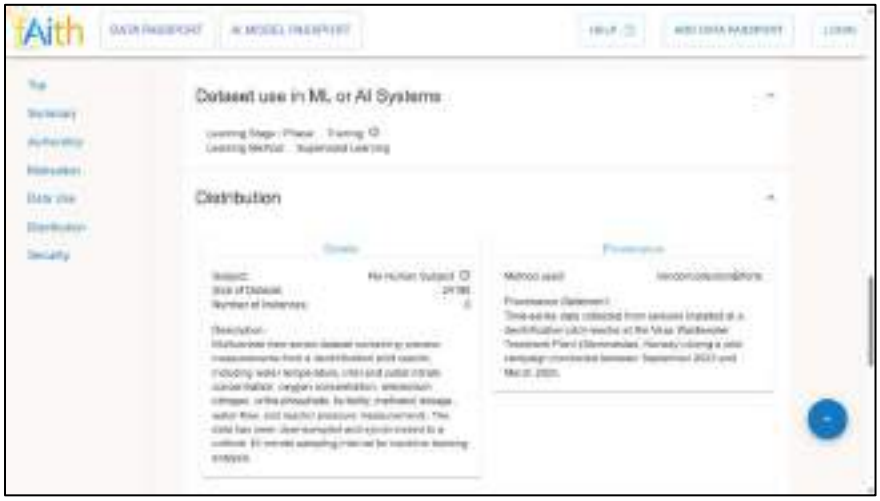


Figure 8. Screenshot of AI Model Hub as applied in LSP5

3.3.2.1 Application of AI Model Hub in the initial pilot

In LSP5, AI Model Hub has been used to implement a model passport for an AI prediction model for nitrate levels in wastewater from the Pilot reactor of Veas. The model runs on a SINTEF server and the model passport was configured following training provided by project partner FORTH. As a complement to the model passport, also a data passport has been registered for the data used for training, testing, and validating this model.

3.3.2.2 Implications of AI Model Hub in the initial pilot

AI Model Hub allows AI developers inLSP5 to provide logs of model background, runs, and metadata in a systematic manner. This is potentially valuable for transparency and accountability and is hence of relevance for risk management.

The systematic approach provided by the AI Model Hub may be particularly relevant when a larger number of models is implemented in the LSP5 context of use, as it potentially facilitates model reviews as part of the routine quality reporting.

The background and log information provided by AI Model Hub may also facilitate handover of responsibilities for model maintenance and development between AI Developers or across partners.

3.3.2.3 Validation feedback on AI Model Hub in the initial pilot

The AI development team in LSP5 has provided feedback on the AI Model Hub following implementation and trial in the initial pilot. During implementation, the AI Development team was in close interaction with the partner providing the AI Model Hub for details on configuration. Furthermore, structured feedback has been given in a questionnaire form provide by leader of WP2.

3.3.3 TrustGuard

TrustGuard implements the FAITH AI TAF. It provides an environment for running assessments covering the six phases of this framework. Here, users provide details for the cartography, including asset and user identification, before running assessments of threats, consequences, and vulnerabilities.

TrustGuard enables AI teams to systematically consider relevant risks, consequences and vulnerabilities of relevance to key trustworthiness characteristics, as well as support the identification of relevant assets and users.

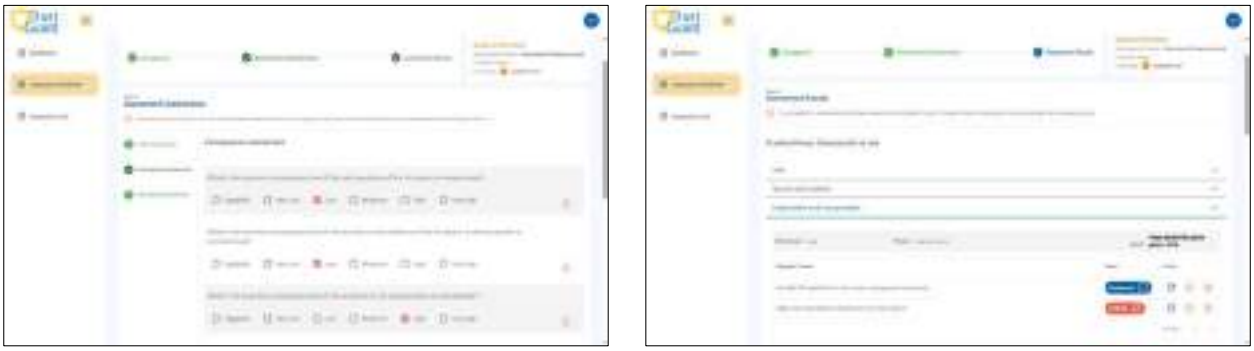


Figure 9. Screenshots of TrustGuard as applied in LSP5

3.3.3.1 Application of TrustGuard in the initial pilot

In LSP5, TrustGuard has been applied in collaboration between SINTEF and Veas. The application was in part conducted through work meetings and in part through asynchronous discussion to adequately complete the six assessment phases.

The users and assets were identified at the onset of the process, whereas some of the threats, consequences, and vulnerabilities were not previously considered and needed particular consideration.

As a result of the assessment of threats, consequences, and vulnerabilities, an overall risk analysis was provided for the trustworthiness characteristics addressed in FAITH and associated controls.

3.3.3.2 Implications of TrustGuard in the initial pilot

A number of controls for mitigating risk were proposed following the TrustGuard assessment. Several proposed controls were of particular relevance for LSP5. An overview of the most relevant controls and corresponding threats are presented in Table 7.

Table 7. Controls and corresponding threats in TrustGuard results for LSP5, initial pilot

Trustworthiness characteristic	**Threat**	**Control**
Explainable & Interpretable	Human error	Apply documentation requirements to AI projects.
Explainable & Interpretable	Over-reliance in model predictions	Provide training and AI skill development to stakeholders, including operators.
Accountable & Transparent	Deskilling from lack of transparency	Transparent models, assessed with experienced operators.

Accountable & Transparent	Diffusion of accountability due to lack of transparency	Accountability strengthened through transparency and interpretability, evaluated with stakeholders
Secure & Resilient	Compromise of ML application components	Define and monitor indicators for proper functioning of the model

The identified controls are further discussed in Section 3.6 on countermeasures below.

3.3.3.3 Validation feedback on AI TrustGuard in the initial pilot

The team in LSP5 has provided feedback on the framework underpinning TrustGuard during its development. The aim of the feedback has been to ensure the relevance of the framework for application in industrial processes such as wastewater treatment.

Following implementation, structured feedback has been given in a questionnaire form provided by leader of WP2.

3.4 Threat analysis

The FAITH tools, and in particular TrustGuard, helped structure the threat analysis that was already initiated in LSP5 with initial applications of the FAITH AI TAF. In line with the framework, a broad range of threats corresponding to the different trustworthiness characteristics had been identified in collaboration between the LSP5 partners. On this basis, the tool-supported analysis provided a more systematic perspective, explicitly structuring the threats according to specific trustworthiness characteristics and supporting the identification of controls corresponding to the characteristics of the users and assets. This helped ensure completeness and traceability in the identification of threats relevant to the LSP5 context.

Through the use of FAITH AI TAF and TrustGuard, threats related to the AI prediction models were identified as the most critical. Specifically threats concerning the trustworthiness characteristics Valid & Reliable, but also Explainable & Interpretable, Accountable & Transparent, Secure & Resilient, and Safe. These threats pertained to inaccurate or non-robust predictions under changing seasonal or weather conditions, insufficient interpretability of model outputs, and potential over-reliance on model predictions by operators. The tools highlighted how such threats affect not only model performance, but also operator decision-making, accountability, and trust. For example, threats related to explainability and transparency were closely linked to human–AI interaction risks, such as reduced ability to detect prediction errors or gradual deskilling of operators.

For the data assets, the primary threats identified through FAITH AI TAF concerned potential degradation in data quality due to sensor failure which may entail erroneous or missing data. TrustGuard helped clarify how data issues could result in reduced model accuracy and robustness with misleading predictions as a result.

Threats related to the maturity of the AI teams were also assessed through TrustSense, providing insight into threats related to human-centred aspects of the implementation of AI in wastewater treatment – in particular corresponding to the trustworthiness characteristics Explainable & Interpretable and, Accountable & Transparent.

3.5 Initial risk assessment results

The risk assessment in the initial pilot, supported by FAITH AI TAF and TrustGuard, integrated the identified threats, estimated likelihoods, and assessed business consequences to produce an overall risk profile for the AI system in LSP5.

The risk assessment shows that risks are not evenly distributed across trustworthiness characteristics. Instead, high risk is found mainly for the trustworthiness characteristics of Valid & Reliable, Explainable & Interpretable, and Accountable & Transparent – as shown in Table 5 above. These risks are closely tied to the accuracy and robustness of AI prediction models, as well as to the human-centred aspects of applying such models in the context of wastewater treatment.

The highest risk levels were identified for threats affecting the trustworthiness characteristic Valid & Reliable. Here, risks were assessed as very high due to both the severe consequences of inaccurate predictions and the realistic likelihood of performance degradation under changing operational conditions, such as seasonal effects or atypical weather events. For the trustworthiness characteristics Explainable & Interpretable and Accountable & Transparent, risk levels were high driven, e.g., by the possibility that insufficiently interpretable model outputs could impair operators' ability to identify errors or assess model suitability, in addition to long term implications for operator skill.

Risks related to human-centred aspects of AI implementation, such as over-reliance on model predictions, diffusion of accountability, and long-term deskilling, may appear secondary to model performance risks. However, the assessment emphasized their role as amplifying factors: weaknesses in team preparedness or AI literacy can increase the likelihood and impact of technical failures.

Risks related to the other trustworthiness characteristics, including Safe and Secure & Resilient were assessed as medium to low in the initial pilot, largely because the AI system does not yet directly control the process and existing fallback mechanisms remain in place – in addition to safety and security being managed through established quality processes for the digital infrastructure and overall wastewater treatment process at Veas.

3.6 Countermeasures

Based on the risk assessment results produced through TrustGuard and the FAITH AI TAF process, a set of countermeasures has been identified in collaboration between SINTEF and Veas to mitigate the highest-priority risks in LSP5. The countermeasures consist of controls primarily addressing risks related to model validity and reliability, explainability and interpretability, and accountability and transparency, as these were assessed to have the highest combined likelihood and business impact.

For risks related to validity and reliability, the main countermeasures concern continuous performance monitoring and iterative model development. These include systematic evaluation of prediction accuracy and robustness over time, comparison of alternative modelling approaches, and explicit handling of missing or erroneous input data. The adoption of the time-

segmented ReWTS approach is further intended to strengthen validity and reliability through directly enhancing AI model performance under changing conditions of wastewater treatment. In addition, the exploratory deployment of supplementary sensors is treated as a longer-term countermeasure to strengthen data quality and resilience, with calibration and validation steps planned before broader adoption.

For risks related to explainability, accountability, and human–AI interaction, the FAITH tools emphasised organisational and procedural countermeasures. These include human-centred validation of models with engineers and operators, provision of interpretable metrics and visualisations alongside predictions, and targeted training to prevent both under-trust and over-reliance on AI outputs. Clear procedures for monitoring AI behaviour and responding to anomalies are planned to be established prior to trials of AI-based process control. Responsibility for implementing these countermeasures is shared: SINTEF leads technical controls related to model development and monitoring, while Veas leads organisational measures such as operator training, procedural safeguards, and integration with existing quality and safety routines. Together, these countermeasures provide a structured foundation for risk mitigation as LSP5 progresses from model validation toward live trials of predictive control.

FAITH AI TAF and TrustGuard support the structuring of trustworthiness threats, vulnerabilities and controls, facilitating an overview of relevant countermeasures and corresponding controls. Controls and corresponding countermeasures in LSP5 from TrustGuard of particular relevance are presented in Table 8.

Table 8. Countermeasures and associated controls from example risks in TrustGuard

Trustworthiness characteristic	**Threat**	**Control**	**Countermeasures in LSP5**
Explainable & Interpretable	Human error	Apply documentation requirements to AI projects.	Documentation for operators to be develop as part of future trials of AI models in operational context.
Explainable & Interpretable	Over-reliance in model predictions	Provide training and AI skill development to stakeholders, including operators.	Training for operators to be developed and conducted prior to future trials of AI models in operational context.
Accountable & Transparent	Deskilling from lack of transparency	Transparent models, assessed with experienced operators.	Development of dashboards to provide required information for operator transparency requirements
Accountable & Transparent	Diffusion of accountability due to lack of transparency	Accountability strengthened through transparency and interpretability, evaluated with stakeholders	Routine validation meetings as part of AI prediction model development strengthen transparency and interpretability
Secure & Reliable	Compromise of ML application components	Define and monitor indicators for proper functioning of the model	Notifications to be implemented for ML application failure or unexpected behaviour as part of future trials of AI models in operational context.

3.7 Formal tool feedback

Formal tool feedback has been provided from LSP5 at different points in time during the initial pilot.

3.7.1 TrustSense tool feedback

LSP5 has provided feedback on TrustSense in two iterations during the initial pilot. First in M23/M24 following initial trial use, then in M27/28 following completed use in the initial pilot. The feedback was provided through structured evaluation questionnaires. In addition, a note with detailed suggestions for further strengthening the tool was provided on the initial trial use. For both feedback sessions, the feedback was provided by representatives of AI developers, domain experts, and risk manager at SINTEF and Veas.

In the feedback for TrustSense, it was noted that the assessment of team trustworthy AI maturity is highly relevant and that the tool may be supportive of this objective. In the feedback, the LSP5 team also provided input on potential improvements in the tool to further strengthen its usefulness and ease of use.

3.7.2 AI Model Hub tool feedback

LSP5 has provided feedback on the AI Model Hub upon completing the initial implementation and application of the model and data passports in M27/M28. The feedback was provided in a structured questionnaire by representatives of the AI developers and risk manager in LSP5.

In the feedback on the AI Model Hub, it was noted that the tool provides structure and transparency for assessment of AI models and datasets as part of risk management procedures. The feedback also included points on which AI developers in LSP5 needed support to complete implementation which indicates opportunities for further strengthening of the tool.

3.7.3 TrustGuard tool feedback

Finally, LSP5 has provided feedback on the TrustGuard tool. This tool embodies the FAITH AI TAF methodology. Feedback was provided in M17, following the so-called dry-run of the FAITH AI TAF tooling, and in M27/28 following use of the tool in the initial pilot. The feedback was provided by representatives of AI developers, domain experts, and risk manager at SINTEF and Veas.

In the feedback on TrustGuard it was noted that the tool provides structure and support to trustworthy AI risk assessment and management. Specifically, it was noted that the tool may enable comprehensive coverage of risks pertaining to the different trustworthiness characteristics. In the feedback, also some points for further strengthening the tool was pointed out. These notes, along with similar notes for the other tools will be summarized in WP2.

4 AI system evaluation

In this section, we report on the evaluation of the AI prediction models in the initial pilot. First we provide an overview of the evaluation approach. Then we report on the assessment for the trustworthiness characteristics of relevance to LSP5. Finally, we reflect on the general impact of the AI system in the pilot context.

4.1 Evaluation Approach

The evaluation approach was initially detailed in D8.1. Here, different evaluation methods were planned for coverage of the relevant trustworthiness characteristics for LSP5, and enable thorough involvement of users and stakeholders in the evaluation process.

This entails that the AI prediction models in LSP5 have been evaluated both through application of metrics for assessing the reliability of model predictions, as well as human-centred evaluations involving users and stakeholders.

Key metrics involved MAE, MSE, RMSE, and MAPE for assessing discrepancies between model predictions and measurements of wastewater nitrate levels. Furthermore, metrics were provided for trend agreement and ReWTS-based model weights to overview impact of previous time periods in AI prediction model.

Human-centred evaluations has involved validation meetings with domain experts and evaluation workshops with a broader set of users and stakeholders. Validation meetings have particularly addressed validity and reliability, as well as explainability and interpretability. Evaluation workshops have addressed a broader range of trustworthiness characteristics.

An overview of the evaluation approach corresponding to the different trustworthiness characteristics under assessment is provided in Table 9.

Table 9. Evaluation approaches to address relevant trustworthiness characteristics

Evaluation approach	**Trustworthiness characteristics**				
	Valid & Reliable	**Explainable & Interpretable**	**Accountable & Transparent**	**Secure & Resilient**	**Safe**
Metrics and analysis	X	X	X		
Validation meetings	X	X	X		
Evaluation workshops		X	X	X	X

4.2 Trustworthiness assessment

Assessments of trustworthiness in line with the recommended controls in FAITH AI TAF and TrustGuard, have provided insight into the key trustworthiness characteristics for LSP5. Specifically, Valid & Reliable is key for the application of AI prediction models in wastewater treatment, but also Explainable & Interpretable and Accountable & Transparent, as well as Secure & Resilient and Safe.

In the following, we provide an overview of assessments in the initial pilot corresponding to the relevant trustworthiness characteristics.

4.2.1 Validity & Reliability

Valid & Reliable has been identified as the most important trustworthiness characteristic for AI prediction models in wastewater treatment. In particular, it has been seen as critical to assess prediction accuracy as well as robustness over time, as the performance of AI prediction models here will be decisive for whether they can replace the existing PID-based automated process controls and help Veas reach their objectives of treatment effectiveness (80% of wastewater nitrogen removed) and efficiency (significant reduction in methanol spent for nitrogen removal).

Prediction accuracy has been analysed for all models advanced in the initial pilot, including the initial nowcasting and forecasting models (ElasticNet, TCN, LSTM, XGBoost), the chunk-based forecasting models (ReWTS ensemble models), the models advanced to validation with domain experts (ReWTS, ElasticNet and TCN), and transformer-based nowcasting and forecasting models (TFT).

4.2.1.1 ReWTS models for strengthened accuracy

Accuracy was analysed through established metrics. Assessments of **initial nowcasting and forecasting models** showed better accuracy for regression and LSTM models for nowcasting and comparable performance of all regression, tree-based, and deep learning models for forecasting.

Assessments of temporally segmented **ReWTS models** showed improved accuracy compared to initial models. A comparison of ReWTS models and initial global models in terms of accuracy is presented by Johnsen et al (2025). Their findings on the comparison of ReWTS models with similar non-time segmented (global) models are provided in Figure 10.

		Architecture			
Dataset		Elastic net	XGBoost	TCN	LSTM
Veas	ReWTS	0.73	0.92	1.01	0.99
	Global	0.81	1.15	1.33	1.16
	% Difference	10.1	22.5	28.0	15.9

Figure 10. Overview table of the mean absolute error of predictions by time-segmented ReWTS models and non-time-segmented global models. From Johnsen et al., (2025)

Following from their promising prediction quality, two ReWTS models have been advanced for validation with domain experts at Veas. See the section 4.2.2 on explainability and interpretability below.

4.2.1.2 Transformer models promising for further explorations

In parallel with the ReWTS approach, we have explored transformer models for enhanced prediction capabilities. As a first step, the temporal fusion transformer model TFT has been run on the Veas Pilot data, corresponding to the dataset used for initial nowcasting and forecasting.

TFT models are more resource demanding, but initial explorations on nowcasting suggests that they hold potential for timeseries predictions in wastewater treatment. A comparison of TFT

models with regression, tree-based and machine learning models in terms of nowcasting accuracy is presented by Uddin et al. (2026) in an accepted conference paper, showing superior nowcasting accuracy for TFT models (see Table 10).

Table 10. Experimental results, nowcasting, TFT compared to other model families, presented in Uddin et al. (2026)

Model Family	Best Configuration Hyperparameters	Test MAE	Test RMSE
XGBoost	400 trees, LR=0.05, depth=4, subsample=0.8, colsample=0.8	1.9246	2.6388
Linear Regression	positive=True, fit_intercept=True	2.2162	2.9756
TCN	3 blocks, 64 channels, k=3, dropout=0.1, lr=0.0005	2.0310	3.1473
LSTM	2 layers, 32 hidden units, dropout=0.2, lr=0.001	2.3477	3.5156
TFT (Run 1)	hidden_size=16, attention_heads=2, hidden_cont=8, dropout=0.1, lr=0.001	0.7415	1.6097
TFT (Run 2)	hidden_size=32, attention_heads=4, hidden_cont=16, dropout=0.2, lr=0.0005	0.8295	1.6343
TFT (Run 3)	hidden_size=64, attention_heads=4, hidden_cont=32, dropout=0.3, lr=0.0001	0.6041	1.4783

Following the promising results of transformer models, in the replication pilot we will advance such models for consideration in process control.

4.2.1.3 ReWTS enabling insight on robustness

The ReWTS approach to AI prediction models for time series data has also provided a basis for considering prediction robustness. In ReWTS, time series are divided into temporal segments or chunks. For each chunk, a separate prediction model is provided. This allows forecasting to be done based on those time-segment models that best predict the immediate history – assuming that these models will also be best for predicting the immediate near future.

As such, ReWTS should enhance prediction robustness as accuracy should be improved for each given time interval. Johnsen et al. (2025) investigated this assumption by analysing accuracy for different time chunks in the dataset. The analysis shows that for the vast majority of chunks, the ReWTS approach increases accuracy compared to non-time-segmenting (global) approaches – suggesting substantial enhancement of robustness in predictions. An example for the deep learning TCN is provided in Figure 11.

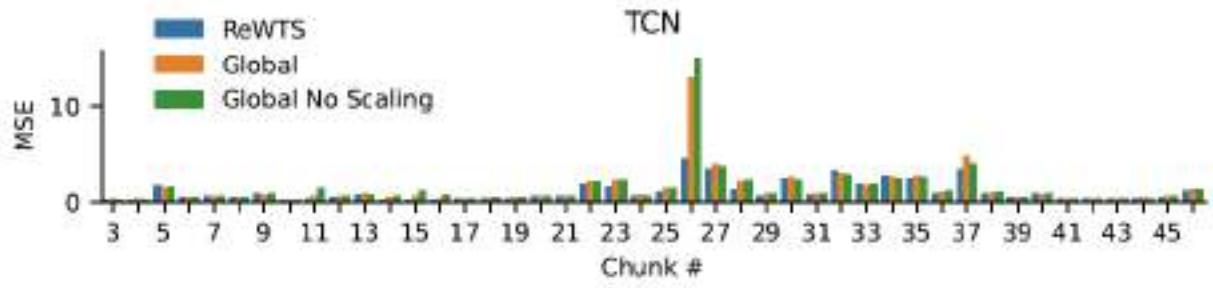


Figure 11. ReWTS vs. non-time segmented (global) TCN prediction models. ReWTS with lower MSE for the vast majority of time chunks. From Johnsen et al. (2025).

4.2.2 Explainability & Interpretability

In the initial pilot, substantial efforts have been placed in making AI prediction models explainable and interpretable. In particular, interpretability for operators has been considered highly important.

Explainability and interpretability have been sought through (a) analyses of covariate impact on model predictions, (b) analysis of time-chunk impact on ReWTS model weights, and (c) visualizations of covariates and prediction accuracy for human-centred validation with domain experts.

4.2.2.1 Analysis of covariate impact on model predictions

AI prediction models for nitrate levels in treated wastewater depend on a broad range of covariates, including levels of nitrate, ammonium, and oxygen in untreated wastewater, as well as water temperature, turbidity, water pressure, and added methanol. To enhance explainability of the models, it is helpful to consider the impact of the different covariates on model predictions.

Analysis of covariate impact has been conducted for both nowcasting and forecasting models, as presented by Böhn et al. (2026) in a paper currently under review. An overview of the impact of covariates for an LSTM machine learning nowcast model is provided in Figure 12.

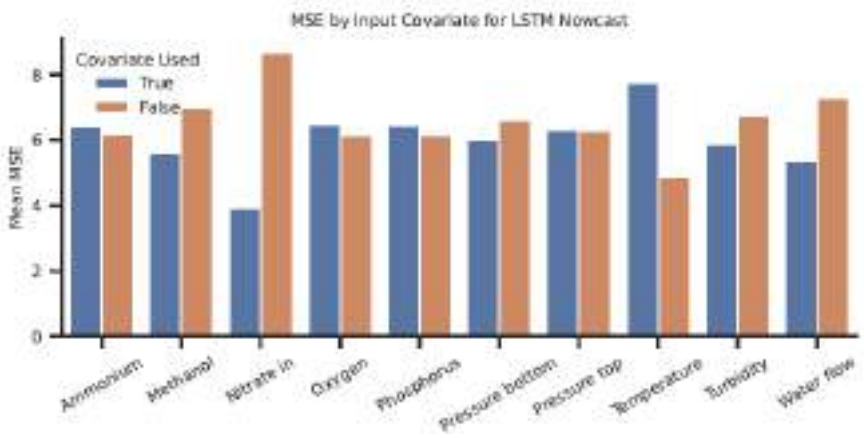


Figure 12. Covariate impact of prediction accuracy for LSTM nowcast model. From Böhn et al. (2026)

4.2.2.2 Analysis of time-chunk impact on ReWTS model weights

For ReWTS prediction models, an important aspect of explainability concerns providing information on the time-chunk models that have contributed to the model weights. This is done to enable the identification of time periods with similar characteristics with regards to key aspects of wastewater treatment.

For the prediction models advanced to human-centred validation, information on time-chunk impact on ReWTS model weights is provided through visualizations of time-chunk share of model weights. Figure 13 shows how the share of weights from prediction models is distributed across different historical time-chunks. This figure shows share of weights for a ReWTS model of April 10, 2026. Time-chunks with much weight in this example include early April 2023, early April 2024, and late November 2022 – all periods for which the relevant prediction models have been among the best in predicting nitrate levels in treated wastewater in the very recent past. While

this example intuitively shows two April months as lending weights to current predictions, this is not always the case but depends solely on previous models' prediction capabilities for the present situation.

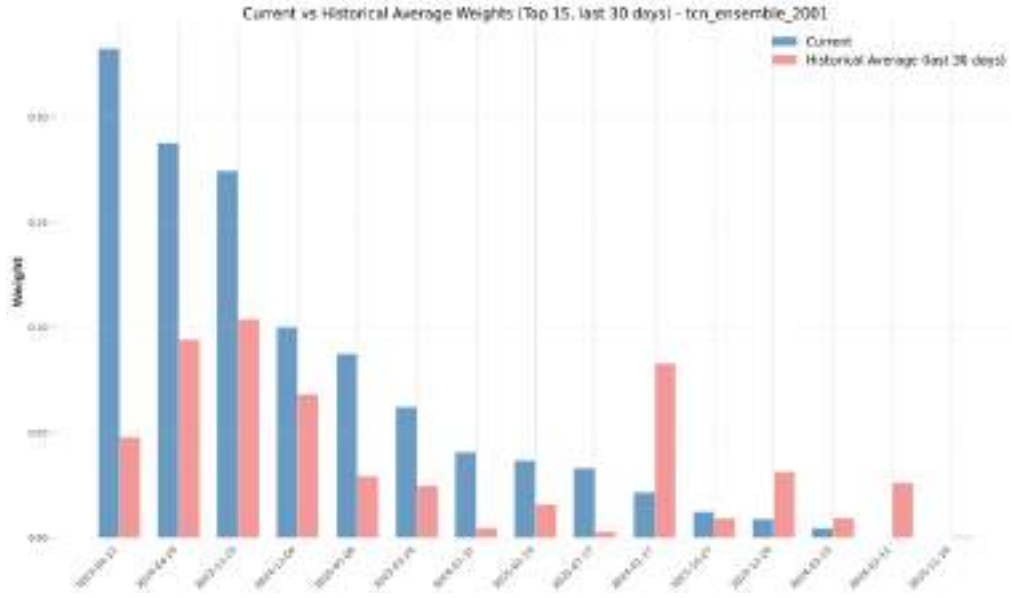


Figure 13. Impact of previous time-chunk model weights on current ReWTS prediction model, April 10, 2026. Figure shown in human-centred validations with domain experts.

4.2.2.3 Summaries and visualizations for human-centred validation

Process engineers from Veas, holding the role of domain experts, join AI researchers at SINTEF for biweekly validation meetings. In these meetings, the group go through model predictions of the last two weeks, with particular regard to deviations in model predictions.

For these meetings to be productive, a range of metric summaries and visualizations are provided. The set of summaries and visualizations has evolved over the course of the validation meetings to reflect the needs of domain experts.

Currently the summaries and visualizations include the following for both the past week and historical periods:

- Actual values: Nitrate level, methanol level, and covariates
- Predicted values: Nitrate level (Mean, SD)
- Accuracy metrics for 5 min to 1 hour prediction horizon: MAE, MSE, RMSE, MAPE
- Historical accuracy metrics for 1 hour prediction horizon: MAE, MSE, RMSE, MAPE
- Time series visualizations: Actual values, predictions, error
- Trend visualizations: Filtered time series, rate of change, and trend agreement
- Time-chunk visualizations: Share of weights for models of historical time-chunks

To exemplify visualizations, time series visualizations of actual and predicted nitrate levels, error, and methanol levels are provided in Figure 14.



Figure 14. Visualizations of actual and predicted nitrate levels, error rate, and methanol levels, at April 10 2026.

The provided summaries and visualizations have during the initial pilot been provided in a spreadsheet. For the replication pilot, an interactive web-based dashboard is developed for presenting summaries and metrics to domain experts.

4.2.3 Transparency & Accountability

Transparency and accountability has been identified as important for AI prediction models used in wastewater treatment. This is particularly relevant given the criticality of wastewater treatment, as well as the treatment is provided as a municipal service with regulatory requirements on treatment effectiveness. As of today, there is a requirement for removal of 70% of nitrogen from municipal wastewater in Oslo. Veas internal aim is an 80% removal.

4.2.3.1 Transparency through predication access

In the initial pilot, transparency and accountability have been sought by providing direct access to real time prediction model outputs to Veas personnel. This has been done through the third party platform Intelecy. Here, the predictions of the two models advanced for human-centred validation can be scrutinized in terms of current predictions and for historical periods. A screenshot of the Intelecy platform with the two prediction model outputs along with actual nitrate levels is presented in Figure 15.

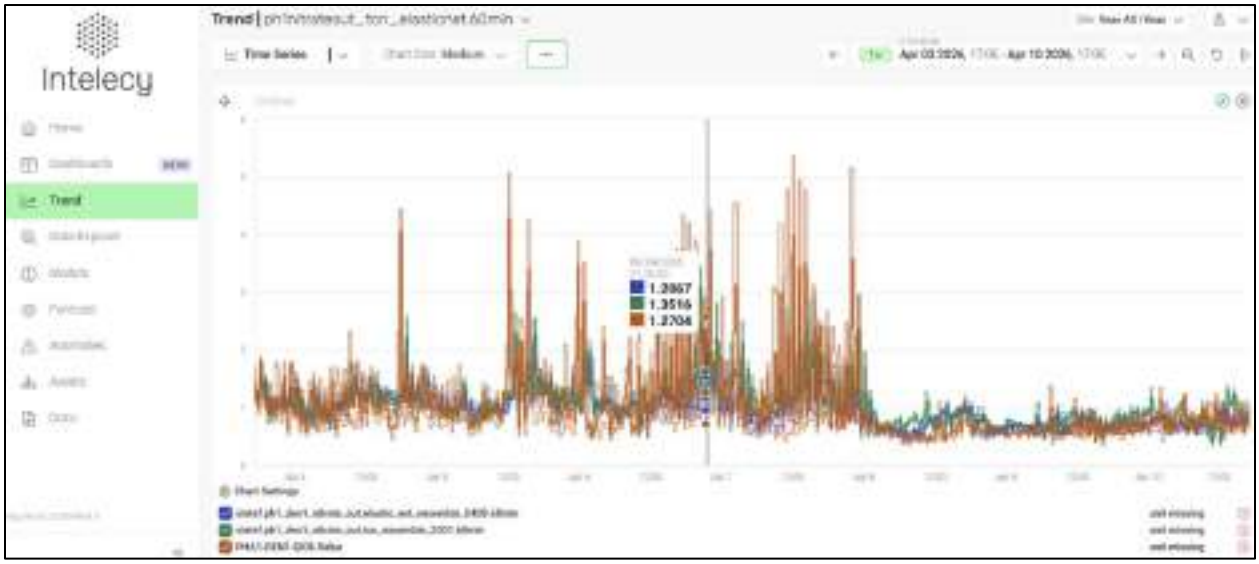


Figure 15. Screenshot of prediction model outputs (blue, green) and actual nitrate levels, shown in the Intelecy platform, April 10, 2026.

4.2.3.2 Transparency through dashboard for summaries and analyses

Based on the experiences from the validation meetings, with the evolving set of metric summaries and visualizations, we are in the replication pilot developing a dashboard for near real-time overview and analyses of predictions.

The dashboard will strengthen transparency through access to the current and historical analyses relevant to model predictions. The transparency provided by the dashboard should also enhance Veas accountability in adopting AI prediction models for process control.

4.2.4 Security & Resilience

Security and resilience have been identified as important for the use of AI predictions for process control in wastewater treatment. As Veas has a mature digital platform for process control and well established procedures for quality control and security, security and resilience in the initial pilot has concerned complying with existing requirements of Veas policy.

One example of security measures in place concerns access to wastewater treatment process data. Here, process data are available through ABB Edge Insight, which allows verified third parties access to read and export data. Verification is conducted through Veas procedure for confidentiality agreements.

In the replication pilot, security will be given more comprehensive concern as the setup for AI-based predictive control requires the Veas control system to access control signals without risk of being accessed or compromised. This is currently managed by the Veas setup corresponding to Figure 2 higher up in the report.

4.2.5 Safety

Safety is the final trustworthiness characteristic identified as relevant for LSP5. Safety is important due to the environmental implications of wastewater treatment. Hence, it is a priority

to avoid breaches that may lead to negative consequences for the marine environment, such as reduced wastewater treatment quality or excess use of methanol to drive denitrification.

In the initial pilot, safety has not yet been given high priority, as we have only advanced prediction models and validated them without deploying them in live trials.

In the replication pilot, AI models will form the basis of predictive control with the aim of trials in the live wastewater treatment process. Before such trials can be conducted, the team will establish safety precautions to reduce risk of safety breaches including (a) fallback procedures in the case of subpar nitrate removal and (b) safeguards against excessive methanol use.

4.3 Preliminary impact evaluation

The initial pilot of LSP5 has already had a positive impact on both technical practice and organisational understanding of AI-supported process control at Veas. Through regular validation meetings and more broad-reaching evaluation workshops, user and stakeholders have gained insight into the performance of AI prediction models and their opportunities and limitations. This has contributed to increased confidence in the feasibility of AI-based predictive control and has enabled us to prepare for MPC trials earlier than initially planned. In D8.1, we foresaw to start preparing such trials only in year 3. As such, being able to start preparing for these already in at the onset of the replication pilot, in M20, represents a substantial advance.

From a development perspective, the application of FAITH AI TAF and associated tools has positively impacted the project by providing a shared vocabulary and systematic process for discussing trustworthiness, risks, and mitigation measures.

At the same time, the initial pilot has revealed challenges that should be managed as the system moves closer to trials in an operational context. The reliance on historical and near-real-time process data highlights that AI models may be sensitive to data quality issues, sensor outages, and operational changes at the plant, as demonstrated by periods of missing data during reactor upgrades. The ReWTS approach strengthens robustness and as a consequence reduces this operational risk, but it will be important, in a live trial, to stress test AI models with regard to how they perform in case of data quality issues or sensor outages. It will also be important to enable easy fallback to conventional control systems to reduce operational safety risk. Furthermore, facilities for monitoring predictive control through easy-to-use dashboards for human oversight along with adequate training and procedures will be highly important for operational trials.

Looking forward, the potential impacts of the AI system in LSP5 are significant, particularly if the planned MPC trials confirm expected gains in effectiveness and efficiency. Positive potential impacts include contribution towards Veas' goal of nitrogen achieving 80% nitrate reduction and reduced methanol consumption. Reduced nitrogen levels helps achieve sustainability objectives for the marine ecosystem of the inner Oslofjord. Reduced methanol consumption help reduce operational cost at Veas. The advances made in prediction of nitrogen removal as detailed in Sections 4.2-4.3 are indicative of the intended contributions concerning nitrogen reduction and wastewater treatment efficiency are within reach. We foresee the replication pilot to enable us to provide precise quantitative estimates of effectiveness and efficiency gains.

At the same time, users, stakeholders, and AI developers involved in LSP5 are aware of the substantial challenges involved in predictive control. We therefore plan on a series of shorter trials instead of proceeding directly to the replication pilot. This will enable the necessary changes and updates prior to longer trials.

5 Regulatory and ethical assessment

The application of AI in wastewater treatment should be considered from regulatory and ethical perspectives. In particular, it is important to ensure compliance with regulation. Motivated by the cartography-phase of FAITH AI TAF, we particularly discuss the relevance of aligning LSP5 with the AI Act and with NIS2.

5.1 Alignment with the AI Act

The EU Artificial Intelligence Act (AI Act) (EU, 2024) is relevant to LSP5 due to the intended use of AI-based prediction models as part of process control in a critical infrastructure context, namely municipal wastewater treatment.

While the AI system in the initial pilot phase has been limited to validation and decision-support activities and has not yet exercised direct control over the treatment process, the planned replication pilot includes trials of AI-based MPC. Such use brings the system closer to applications that may be considered high-risk under the AI Act, given its potential impact on environmental protection, public health, and the continuity of essential services.

In TrustGuard, the cartography phase requires an assessment of whether the system under test is high-risk under the AI Act. As part of the replication phase it will be valuable to clarify the status of AI systems for predictive control in wastewater treatment are to be considered high-risk.

A preliminary consideration may suggest that the status of AI-systems in wastewater treatment may be high-risk as wastewater treatment is considered a societally critical service. However, as stated in the AI Act article 6, AI-systems “intended to perform a narrow procedural task” are except from being considered high risk where they do not “pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making” (EU, 2024). As such, the AI system under trial in the replication pilot may potentially be considered to be of limited risk given its narrow procedural focus and processes in place for mitigating risk of harm. This consideration will be further developed in the replication pilot.

5.2 Alignment with NIS2

The NIS2 Directive (EU, 2022) is an EU regulation that mandates stricter cybersecurity requirements, incident reporting, and management accountability for organizations in critical sectors. It is relevant to LSP5 due to the role of municipal wastewater treatment as a potentially essential service and its dependence on digital systems for safe and reliable operation.

The Veas wastewater treatment plant relies on a complex digital infrastructure for monitoring and controlling critical processes, and the introduction of AI-based prediction models and predictive control adds a new digital component to this environment. While the AI system in the initial pilot has operated in parallel with existing control mechanisms and without direct influence on the control system, the planned replication pilot involves tighter integration with operational decision-making, thereby increasing its relevance for cybersecurity and resilience considerations under NIS2.

In TrustGuard, the cartography phase requires an assessment of whether the system under test is used by an operator of essential services, and whether it is used for the provision of an essential or important service. As part of the replication phase, it will be important to clarify how AI-based predictive control systems should be treated within the scope of NIS2, especially with regard to requirements for risk management measures, incident handling, and continuity of essential services. Here it will be important for the LSP5 partners to clarify whether the implementation of AI-based predictive control systems require particular attention with regards to NIS2, or whether this should be handled at the level of the entire digital control system at Veas.

6 Conclusions and roadmap for the replication pilot

To conclude this report on the initial pilot of LSP5, we will provide a summary of achievements in the initial pilot, our plan for the replication pilot, and an overview of relevant risks for the replication pilot.

6.1 Summary of achievements of the initial pilot

The initial pilot of LSP5 has successfully met its objectives and established a solid foundation for the replication pilot with trials of AI-based predictive control in wastewater treatment. The pilot has delivered substantial achievements across technical development, trustworthiness assessment, stakeholder involvement, and operational readiness.

First, the initial pilot achieved significant progress in advancing and validating AI prediction models for nitrate levels in treated wastewater. A range of modelling approaches were explored, including regression, tree-based methods, deep learning models, and transformer-based architectures. Building on this exploration, the time-segmented ReWTS approach was further advanced and has shown promising results for accuracy and robustness under varying seasonal and operational conditions. Two ReWTS-based models (ElasticNet and TCN) has been validated with domain experts. Furthermore, transformer-based models have shown promising results in exploratory studies positioning these as candidates for further development in the replication pilot.

Second, the initial pilot made advances in model explainability and interpretability, which is critical for trust in AI systems applied to industrial process control. Explainability was strengthened through analyses of covariate impacts, inspection of time-chunk contributions in the ReWTS models, and the development of tailored metrics and visualisations. The latter were iteratively refined in close collaboration with Veas domain experts during bi-weekly validation meetings. As a result, model behaviour could be related to known process conditions, seasonal effects, and operational events, enabling meaningful human-centred validation.

Third, the initial pilot enabled user and stakeholder involvement as a core part of the AI development and validation process. Stakeholders , engineers, and operators at Veas have been involved through stakeholder meetings, validation meetings, and evaluation workshops – the latter seeing the broadest involvement with both strategic and operational orientations. User and stakeholder involvement helps model advances to be aligned with operational needs and constraints, foster shared understanding of AI capabilities and limitations, and support needed organisational maturity for the transition from AI model validation to trials of AI-based predictive control.

Fourth, the pilot provided a basis for early operational trials of AI-based predictive control. Following from positive results in the initial pilot, we have been able to move more rapidly than originally planned toward preparations for MPC trials. A secure and segregated technical setup for predictive control has been designed, simulation environments for validating MPC behaviour are near completion, and preliminary work on objective functions, safeguards, and evaluation

methods has been carried out. This readiness enables the replication pilot to begin exploratory MPC trials early, creating room for iterative refinement and potentially multiple trial iterations.

Fifth, the initial pilot successfully applied the FAITH AI TAF and associated tools, TrustGuard, TrustSense, and the AI Model Hub. TrustGuard supported systematic identification of assets, users, threats, vulnerabilities, risks, and controls, allowing for a particular focus in LSP5 on validity and reliability, explainability and interpretability, and accountability and transparency. TrustSense facilitated reflection on team maturity for trustworthy AI, among both AI developers and domain experts, motivating along with the TrustGuard outputs, concrete plans for training, procedures, and safeguards. The AI Model Hub enabled documentation of models and datasets through passports. This strengthens transparency, traceability, and accountability of models. Together, the FAITH framework and tools supported trustworthiness assessments for LSP5. Furthermore, our use of the framework and tools in LSP5, generated feedback to guide their further development.

In summary, the initial pilot suggests the feasibility of developing accurate, robust, and interpretable AI prediction models for wastewater treatment. Furthermore we have validated these models with uses and stakeholders, thereby making the process human-centred to strengthen trustworthiness of models and – eventually – well-funded user and stakeholder trust in the outcomes. This provides a good basis for the replication pilot, where the focus will shift from validation toward live trials of AI-based predictive control.

6.2 Plan for replication pilot

The replication pilot runs in the period M20-M51. In the replication pilot, the aim is to advance AI prediction models for denitrification to a point where we can run live trials of model predictive control.

Due to promising advancement and validation of prediction models in the initial pilot, we are currently planning for a first trial of MPC already in the period M32-M39. Following this, provided the first trial is successful, we may also run a second trial of MPC in the period M40-M51.

The planned MPC trial in the replication pilot consists of the following activities and statuses, listed in bullet points for brevity. Following the listed activities, a Gantt chart overview of the plan for the MPC trial of the replication pilot is provided.

- **Initial preparations for MPC trial.** Following successful validation of prediction models, initial preparations for the MPC trial was onset in M20.
- **Establish test infrastructure.** A test infrastructure for the MPC trial, including communication of control signal from the MPC through an OPC-UA server to the wastewater treatment control system (DCS/SCADA), has been developed. Only some adaptations to DCS/SCADA to read AI model control signals remain.
- **Objective function for MPC.** The selection of the MPC objective function is currently being done in close collaboration between the AI model researchers at SINTEF and

domain experts at Veas. As part of this, a dedicated workshop on MPC development was conducted in M27.

- **Verification methods for MPC.** Methods for verification of the effect of MPC have been identified, but not yet detailed. These are foreseen to include assessments of effectiveness, efficiency, and responsiveness – in addition to robustness concerning changing weather conditions and events in the treatment facility – such as data fallout. Verification methods is planned to be completed by M29.
- **Validate MPC in simulated environment.** An environment for simulating the effect of MPC is currently being finalized by AI researchers at SINTEF. This environment will be used to validate MPC output prior to the trials. Initial validation activities have been conducted, and we plan for these to be completed by M32.
- **Safeguards for MPC trial.** Veas and SINTEF in collaboration will establish safeguards to ensure the model does not lead to excessively low or high methanol consumption will be established by Veas. This will be conducted in parallel with validation of MPC in simulated environment and also through exploratory use of MPC until M34.
- **Exploratory use of MPC.** Exploratory use of the MPC in the form of several short trials (duration of days) will be conducted **t**o identify and mitigate initial issues and gather data from live process application. This is planned for the fall of 2026 – M31-M34.
- **MPC trial.** Provided simulations and explorations of MPC give positive results, a live trial of model-based regulation of methanol (duration of weeks) is planned for the period M36-M38.
- **MPC trial evaluation.** Immediately following the MPC trial, an evaluation of the completed initial trial will be conducted to guide the subsequent trial of MPC v2.
- **MPC v2.** Following the initial trial, a subsequent trial of an MPC version 2 may be conducted provided successful testing of version 1. Version 2 may be an iteration of the current chunk-based time-segmented models, or may include transformer-based approaches, depending on experiences from ongoing model advances.

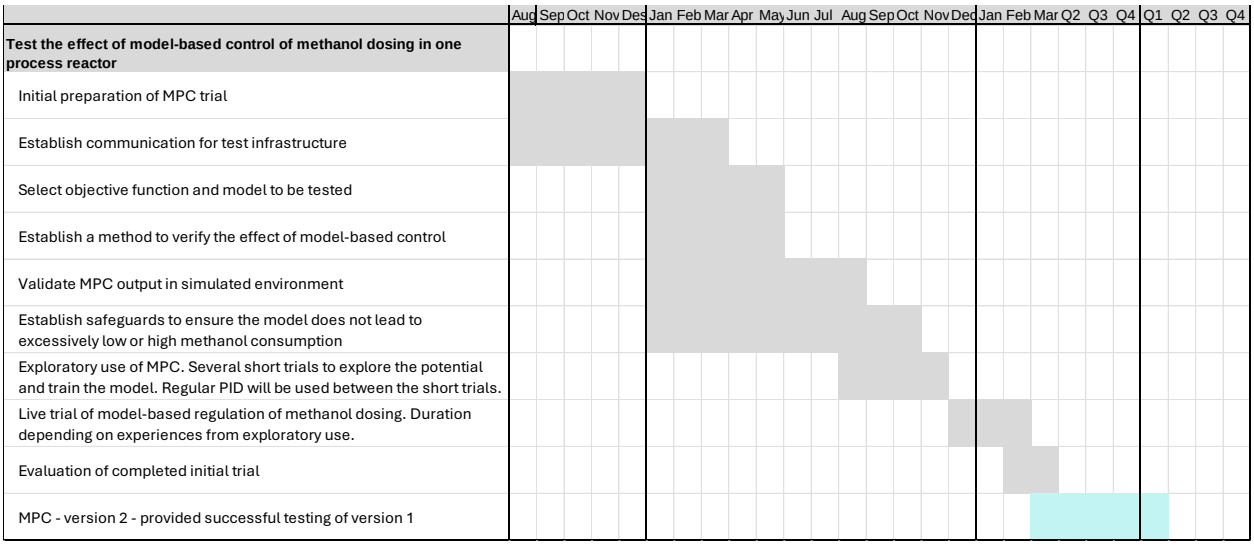


Figure 16. Gantt chart of MPC trial in LSP5 replication pilot.

6.3 Identified risks for the replication pilot

For the replication pilot, the risks identified in the FAITH risk registry continue to be relevant. Specifically, risks concerning stakeholder engagement and resistance to change (Risks #6, #11). Furthermore, the identified risk concerning not meeting user requirements remains particularly relevant (Risk #7).

Stakeholder engagement and resistance to change: As we move from validation of AI model predictions to trials of AI models for predictive control in Veas wastewater treatment, user and stakeholder engagement is critical. The successful implementation of trials depends on the close collaboration between AI researchers and SINTEF and process owners, engineers and operators at Veas. Furthermore, the implementation must be aligned with Veas operational requirements and practices. Without sufficient stakeholder engagement, the MPC trial will be challenging to implement and may also result in resistance to change with regards to uptake of AI.

To control this threat, we have focused strongly on the human-centred development process in both the initial pilot and the initial steps of the replication pilot. We will continue this human-centred approach with substantial involvement of process owners, engineers, and operators. Our established structure of involvement with stakeholder meetings, validation meetings, and evaluation workshops has worked well in the initial pilot, and has also substantially enhanced to AI models advanced in the pilot. We are therefore confident that we will be able to have the needed user and stakeholder engagement also in the replication pilot but will continue to consider it a relevant risk.

Meeting user requirements: As we move to trials of AI models in the replication pilot, it will also be critical to provide results that are in line with the expectations and requirements of users and stakeholders at Veas. The aim of taking up AI to support wastewater processing is to strengthen process effectiveness and efficiency, and the MPC trials in the replication pilot will be the ultimate test in the project for whether these aims can be reached.

To control the threat of not meeting user requirements, we have a detailed plan for conducting the MPC trial as described in Section 6.2. Furthermore, our sequencing of the replication pilot into phases of exploration, trial, and – if relevant – a repeated trial of an MPC version 2, allows more room for tuning and advancing predictive control to a point where it meets user and stakeholder expectations. Furthermore, both users, stakeholders, and AI developers within the LSP5 team understand the challenges involved in the MPC trials, and that reaching the stated aims may require some repeated trials.

An overview of the identified risks for LSP5 is provided in Figure 17. Risks particularly discussed above are highlighted.

Risk ID	Description	Probability	Impact	Mitigation measures	Contengency plans, additional actions, effects on project
3	Data leaks and/or security breaches in LSP or in FAITH infrastructure	Low	High	Detailed data management and security processes in place at all LSPs, including data minimization and anonymization. Any breaches to be promptly reported to coordinator DPO and any other necessary instances. Infrastructure performance monitored regularly.	Data sharing with research partners only following confidentiality agreement
4	Insufficient stakeholder engagement in initial requirements' gathering	Low	Medium	Stakeholder involvement from the beginning in all LSPs. Clear communication of benefit to stakeholders. All LSPs selected to be of core partner interest, to enhance stakeholder engagement.	Regular biweekly status meetings involving key project resources at LSP partner, and additional meetings on technical issues
5	Users / stakeholders reluctant to participate in LSPs due to security and/or data protection concerns	Low	Medium	Clear communication on confidentiality and privacy of data, ethical AI, and benefits of participation will be provided for all LSPs. Detailed informed consent procedures will be detailed for participating users.	Security handled through agreements for data sharing. No personal data in data for AI modelling
6	Insufficient stakeholder engagement in validation activities	Low	Medium	Stakeholder engagement in validation ensured through business relevance of initial pilot and replication, Engagement monitored in project as part of LSP KPIs.	Plans for stakeholder involvement in process
7	Identified user requirements cannot be met within the scope of the project	Medium	Medium	WPs designed to ensure that user requirements collection and end user involvement is adequately linked to the technical specs delivery and architecture design in the LSPs. Close follow-up of progress in WP1 and WP2.	Close and continuous collaboration between LSP partner and research partner
8	Infrastructure computational power insufficient to deal with data volume or complexity of calculations.	Low	Low	Technical partners will collaborate in WP2 to properly foresee the required computational resources needed to run FAITH tools. This will be done during system specification and architectural design, as well as the setup of the technological infrastructure across LSPs.	Computational requirements not of a kind likely to create issues.
9	Integration issues between FAITH approach and existing systems	Medium	Medium	The FAITH approach is developed iteratively in close collaboration with LSP partners. Integration issues will be addressed in WP2 for benefit for all LSPs.	Interative development and close collaboration with regular status meetings (bi-weekly)
10	Inadequate stakeholder communication	Medium	Medium	Stakeholder communication in the specific LSPs (WP4-10) will be supported by specific communication support developed by communication expertise in WP3.	Stakeholder communication for LSP have been sufficeint until now. Will be followed up
11	Resistance to change within the organizations of the LSP partners	Medium	Medium	A strong focus on communicating the benefits of the FAITH approach to LSP stakeholders. Stakeholder engagement throughout the process, monitored and supported in WP3, will minimize resistance to change.	Close and continuous collaboration between LSP partner and research partner

Figure 17. Known risks for LSP5, with risks of particular relevance for replication pilot highlighted (grey)

7 References

Böhn, E., Eidnes, S., & Jonassen, K. R. (2026). Machine learning for forecasting and simulation of a denitrification pilot reactor. In review for the 67th International Conference of Scandinavian Simulation Society, SIMS 2026.

EU (2022). NIS 2 Directive. https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng

EU (2024). Artificial Intelligence Act. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Johnsen, P. V., Bøhn, E., Eidnes, S., Remonato, F., & Riemer-Sørensen, S. (2025). Recency-Weighted Temporally-Segmented Ensemble for Time Series Modeling. Journal of Artificial Intelligence Research 84, article No. 12. https://doi.org/10.1613/jair.1.17572

Uddin, M. Z., Eidnes, S., Jonassen, K. R. (2026). A Transformer-based Approach for Nitrate Concentration Estimation in a Pilot-scale Denitrification Reactor. Accepted for the 21st Annual System of Systems Engineering (SoSE) Conference 2026.