



Extreme Food Risk Analytics

D5.3: Use-case Plan, Reports & Recommendations

(revised based on review recommendations)

Lead Author: Manos Karvounis (Agroknow)



**Funded by
the European Union**

Grant Agreement No.	101093026		
Project Acronym	EFRA		
Project Title	Extreme Food Risk Analytics		
Type of action	HORIZON Research and Innovation Actions		
Call Topic	HORIZON-CL4-2022-DATA-01-05		
Project Start Date	January 1 st , 2023	Project End Date	December 31 st , 2025
Project URL	efraproject.eu		
Work Package	WP5 Use-cases, Focus Groups, and Performance Experiments		
WP Lead Beneficiary	Agroknow		
Deliverable type ¹ Dissemination level ²	R PU		
Contractual due date	31 December 2023	Actual submission date	December 28 th , 2023 (re-submission on September 20 th , 2024)
Lead Author (s)	Manos Karvounis (Agroknow)		
Contributors	Marilena Dimitrakopoulou (Agroknow), Anne Richmond (MOY), Morgane Romeu (AGRIVI), Vassilis Kotsikoris (RAIN), Nicola Colombo (SGS), Franco Maria Nardini (CNR), Salvatore Trani (CNR)		
Internal reviewer(s)	Alessio Bosca (MAIZE)		

Disclaimer

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 the European Research Executive Agency (REA) (granting authority). Neither the European Union nor the granting authority can be held responsible for them.

Copyright message

This deliverable contains original unpublished work except where clearly indicated otherwise.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. Reproduction is authorised provided the source is acknowledged.

¹ R: Document, report; DEM: Demonstrator, pilot, prototype, plan designs; DEC: Websites, patents filing, press & media actions, videos, etc.; DATA: Data sets, microdata, etc.; DMP: Data management plan; ETHICS: Deliverables related to ethics issues; SECURITY: Deliverables related to security issues; OTHER: Software, technical diagram, algorithms, models, etc.

² PU – Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page); SEN – Sensitive, limited under the conditions of the Grant Agreement; Classified R-UE/EU-R – EU RESTRICTED under the Commission Decision No2015/444; Classified C-UE/EU-C – EU CONFIDENTIAL under the Commission Decision No2015/444; Classified S-UE/EU-S – EU SECRET under the Commission Decision No2015/444

Revision history (including peer reviewing & quality control)

Version	Issue Date	% Complete	Changes	Contributor(s)
V0.1	01/10/2023	10%	Initial Deliverable Structure	Manos (Agroknow) Karvounis
V0.2	20/10/2023	30%	First version of content in all sections	Manos (Agroknow) Karvounis
V0.3	25/10/2023	60%	Expanded content in all sections	Manos (Agroknow), Karvounis Dimitrakopoulou Marilena (Agroknow), Anne Richmond (MOY), Morgane Romeu (AGRIVI), Vassilis Kotsikoris (RAIN), Nicola Colombo (SGS), Franco Maria Nardini (CNR), Salvatore Trani (CNR)
V0.4	14/11/2023	90%	Added section updates	Manos (Agroknow), Karvounis Dimitrakopoulou Marilena (Agroknow), Anne Richmond (MOY), Morgane Romeu (AGRIVI), Vassilis Kotsikoris (RAIN), Nicola Colombo (SGS), Franco Maria Nardini (CNR), Salvatore Trani (CNR)
V0.9	23/11/2023	99%	Finalised delivered and prepared for review	Manos (Agroknow) Karvounis Babis Thanopoulos (Agroknow)
V1.0	13/12/2023	100%	Conducted review and finalised document	Alessio Bosca (MAIZE), Manos Karvounis (Agroknow)
V1.1	16/09/2024	90%	Updated content to address mid-term review revision requests. Specifically, we did the following: <ul style="list-style-type: none"> Enhanced section 4 to better lay out our plans for each of the EFRA scenarios and use-cases 	Manos (Agroknow) Karvounis

			<ul style="list-style-type: none"> Added detailed sections (4.2.3, 4.3.3, 4.4.3) on the piloting plan for each of the scenarios and use-cases, including 		
V1.2	20/09/2024	100%	Conducted review and finalised document	Alessio Bosca (MAIZE), Manos Karvounis (Agroknow)	

EFRA Consortium			
#	Participant Organisation Name	Short name	Country
1	AGROKNOW IKE	AGROKNOW	EL
2	CONSIGLIO NAZIONALE DELLE RICERCHE	CNR	IT
3	STOCKHOLMS UNIVERSITET	SU	SE
4	STICHTING WAGENINGEN RESEARCH	WR	NL
5	MAIZE SRL	MAIZE SRL	IT
6	AGRIVI DOO ZA PROIZVODNJU TRGOVINUI USLUGE	AGRIVI DOO	HR
7	RAINNO IDIOTIKI KEFALAIIOUCHIKI ETAIREIA	RAINNO	EL
8	SGS ROMANIA SA	SGS ROMANIA	RO
9	MOY PARK LTD	MOY PARK	UK

Table of Contents

1	Executive Summary	8
2	Introduction	9
3	Piloting and Evaluation Methodology	10
3.1	Piloting Methodology	10
3.1.1	Scope	10
3.1.2	Key Activities & Timing	10
3.1.3	Gaant Chart	16
3.2	Evaluation Methodology	16
3.2.1	Scope	16
3.2.2	Technical Performance Evaluation	17
3.2.3	Human Engagement Evaluation	17
3.3	Integration of Piloting Outcomes to Project Activities	18
3.4	Outreach Multiplication Measures	19
4	Scenarios, Use-cases & Piloting Plan	21
4.1	Interconnection of Pilot Scenarios with EFRA Platform	21
4.2	Enhancing Safety and Efficiency in Poultry Farming	22
4.2.1	Overview	22
4.2.2	Piloted Use-cases	22
4.2.3	Piloting Plan	23
4.3	Enhanced Predictive Capabilities for Pest Alarms	29
4.3.1	Overview	29
4.3.2	Piloted Use-cases	30
4.3.3	Piloting Plan	31
4.4	Automated Regulatory Analysis & Summarization	38
4.4.1	Overview	38
4.4.2	Piloted Use-cases	38
4.4.3	Piloting Plan	39
5	Conclusions	46

List of Tables

Table 1: Connection of technical WP work with tools developed within the piloted scenarios	21
--	----

Table 2: Piloting plan for the use-cases of the scenario “Enhancing Safety and Efficiency in Poultry Farming” during the 1 st piloting cycle.....	23
Table 3: Piloting plan for the use-cases of the scenario “Enhancing Safety and Efficiency in Poultry Farming” during the 2 ^d piloting cycle.....	26
Table 4: Piloting plan for the use-cases of the scenario “Enhancing Predictive Capabilities for Pest Alarms” during the 1 st piloting cycle focusing on the Data-Informed Rule Extraction phase	31
Table 5: Piloting plan for the use-cases of the scenario “Enhancing Predictive Capabilities for Pest Alarms” during the 1 st piloting cycle focusing on field farm data collection for the Feedback Loop phase	33
Table 6: Piloting plan for the use-cases of the scenario “Enhancing Predictive Capabilities for Pest Alarms” during the 2 ^d piloting cycle	35
Table 7: Piloting plan for the use-cases of the scenario “Automated Regulatory Analysis & Summarisation” during the 1 st piloting cycle	39
Table 8: Piloting plan for the use-cases of the scenario “Automated Regulatory Analysis & Summarisation” during the 2 ^d piloting cycle	42

List of Figures

Figure 1: Key activities during first piloting cycle.....	11
Figure 2: Key activities during first piloting cycle.....	13
Figure 3: Gaant chart of piloting activities throughout the lifetime of the project.....	16

1 Executive Summary

This deliverable outlines the methodologies for piloting and evaluating the EFRA use-cases and also describes the specific scenarios that will be tested, along with the activities and Key Performance Indicators (KPIs) associated with them. The methodology comprises two distinct evaluation phases: formative and summative, each targeting different stages of the use-case development.

The formative evaluation focuses on the early development stages, involving user interaction with preliminary tool versions. This phase is pivotal for gauging initial performance through KPIs like accuracy, AI model efficacy, and usability. Feedback from this phase, particularly from decision-makers, is integral for refining the tools, paving the way for the second cycle. The summative evaluation then assesses the final tool versions, aiming to validate their compliance with targeted KPIs, such as enhanced accuracy, user satisfaction, and overall business impact. This phase is crucial for confirming the technical robustness and practical utility of the developed tools in real-world settings.

Parallel to these methodologies, the deliverable presents concrete use-cases that will undergo these piloting phases. These include scenarios like root cause analysis for salmonella in poultry farms, enhanced predictive capabilities for pest alarms, and automated regulatory analysis. Each use-case is accompanied by specific activities designed to test and refine the tools in question.

For instance, the poultry farm use-case involves leveraging historical data and AI algorithms to predict salmonella outbreaks, with KPIs focusing on model accuracy and user satisfaction. Similarly, the pest alarm enhancement scenario aims to improve prediction algorithms using diverse data, with real-life tests planned to evaluate the model's accuracy and efficiency. The automated regulatory analysis module targets reducing manual effort in data interpretation, with KPIs focusing on the accuracy of automated summaries and user time saved.

In summary, this deliverable lays out a structured approach to piloting and evaluating EFRA tools and details the specific use-cases that will be put to the test. It highlights the planned activities and KPIs for each scenario, ensuring a comprehensive and focused approach to validating and enhancing the project's outputs. Through this dual emphasis on methodology and practical application, EFRA aims to deliver solutions that are technically sound and directly applicable and beneficial in addressing real-world challenges in the food safety and risk prevention sectors.

2 Introduction

The goal of this section is to provide a brief outline of the **objectives** of the at-hand EFRA Deliverable (outlined in [Table 1](#)) and how the rest of the document is structured to achieve them.

Table 1: Adherence to EFRA GA Deliverable & Tasks Descriptions

EFRA GA Component Title	EFRA GA Component Outline
Task 5.2 Experimental Methodology, Use- case Plan, and Recommendations	This task will design use-case, evaluation and valuation activities, subsequently deployed in T5.4, that will measure how real-life decisions can be supported with EFRA tools and periodically process, analyse and distil user feedback from the use-case trials, in order to generate recommendations to help inform the various project tasks and activities.

This deliverable aligns with the objectives set out in Task 5.2, as outlined in Table 1. The primary aim is to design a comprehensive framework encompassing use-case planning, evaluation methodologies, and actionable recommendations, which will be instrumental in guiding the deployment of EFRA tools in T5.4 and beyond.

The document begins with a detailed description of the experimental methodologies, which includes a two-phase evaluation approach: formative and summative. This section describes the methodologies and explains how they will be applied to assess the effectiveness and utility of the EFRA tools in real-world scenarios.

Following this, the deliverable presents a comprehensive use-case plan. This section details specific scenarios that the EFRA tools will address, such as salmonella management in poultry farms and enhanced pest prediction algorithms. For each use-case, the document outlines the objectives, the activities planned for piloting, and the Key Performance Indicators (KPIs) that will track and measure success.

In summary, this deliverable offers a comprehensive roadmap for the implementation, testing, and refinement of EFRA's tools and methodologies. It ensures that the project's deliverables are technically robust and aligned with real-world needs and user experiences, ultimately contributing to the project's overarching goal of enhancing decision support in the realm of food safety and risk prevention.

3 Piloting and Evaluation Methodology

In the EFRA project, each scenario will feature at least two distinct, specific use-cases that will be piloted and evaluated by appropriate users and decision makers to validate their effectiveness and utility in real-world scenarios. The piloting and evaluation methodologies are designed to unfold in two distinct cycles: formative and summative evaluation. In this section we detail the scope, timing, and types of actions that will be undertaken in each piloting cycle, as well as the evaluation methodologies and how their results will provide useful recommendations for the future evolution of the EFRA activities.

3.1 Piloting Methodology

3.1.1 Scope

Formative Evaluation (1st Piloting Cycle)

During the formative evaluation, the focus will be on the early-stage development and fine-tuning of each use-case. The pilot tests will engage both internal and external users to interact with preliminary versions of the tools and systems. Key performance indicators (KPIs) for this stage will measure aspects such as initial accuracy and the performance of any early AI models or useability aspects of envisioned decision dashboards. However, one essential KPI will gauge how well the end product incorporates feedback received from decision-makers during this phase. After collecting the feedback, necessary adjustments and improvements will be made to the tools, leading the way to the second cycle.

Summative Evaluation (2nd Piloting Cycle)

In the summative evaluation, the pilot tests will be focused on assessing the finalized versions of the tools and systems. This second cycle aims to validate whether the AI models and tools now meet or exceed the targeted KPIs, such as improved accuracy, user satisfaction, and business impact. The data gathered will validate both the technical performance of each tool and also its practical utility and impact on the end-users.

This two-step evaluation process ensures a rigorous and comprehensive validation of each use-case, offering multiple points for adjustment and improvement. The formative evaluation allows for proactive problem-solving, while the summative evaluation confirms that the final product meets the project objectives and user needs.

3.1.2 Key Activities & Timing

1st Piloting Cycle

1st Piloting Cycle: Timing & Key Activities

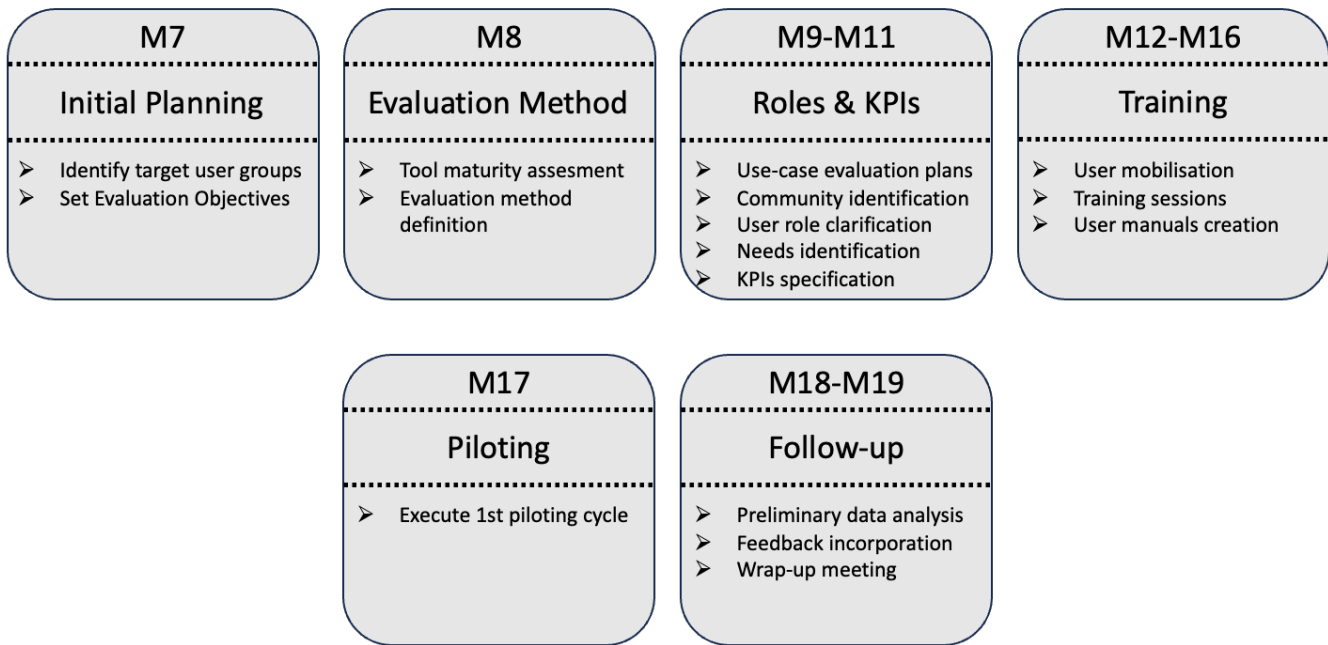


Figure 1: Key activities during first piloting cycle

M7 Initial Planning

During this month, the team will focus on the groundwork. Target user groups will be identified to determine who will be most impacted by the EFRA tools. Concurrently, key evaluation objectives will be set to outline what the pilot specifically aims to achieve. This could be anything from improving tool accuracy to incorporating user feedback effectively.

Activities during this period include:

- Identification of Target User Groups: List and categorize end-users, such as product compliance units and regulators.
- Setting Key Evaluation Objectives: Define what the pilot aims to achieve, such as feedback incorporation and initial accuracy.

M8 Tool Status and Evaluation Method

In this phase, an assessment of the EFRA tools involved in the pilot will be conducted to gauge their readiness level and steer their further development to better accommodate the real-world scenarios of the pilots. Alongside, the evaluation methods to be employed will be chosen. This could involve anything from expert reviews to user performance trials.

Activities during this period include:

- Tool Maturity Assessment: Evaluate the readiness level of the EFRA tools designated for the pilot.
- Evaluation Methods Definition: Decide on how the evaluation will be conducted, possibly employing methods like expert reviews and user trials.

M9-M11 Preparing Use Case-Specific Plans, User Roles, and KPIs

This three-month window will be utilized for the drafting of detailed evaluation plans for each use case. Specific communities of interest will be identified among both internal and external users to ensure that the most relevant groups are included in the pilot tests. By this time, the team will need to have a clear understanding of the roles within the target user groups that the tools will serve. For example, will it assist data engineers, or will it be more beneficial for regulators? Furthermore, the specific needs or challenges of these roles that the EFRA tools can address will be pinpointed. The focus in this period will be on defining KPIs specifically tailored to each case. These KPIs will directly stem from the needs and benefits identified in the previous phase, ensuring a coherent set of metrics that genuinely measure the effectiveness and impact of the tools.

Activities during this period include:

- Use-Case Evaluation Plans: Craft detailed plans specifying how each use-case will be evaluated.
- Community Identification: Determine which internal and external user communities the pilot will target.
- User Role Clarification: Identify the roles that the pilot functionality will serve, e.g., data engineer, regulator.
- Needs Identification: Enumerate the specific challenges and needs that the tool aims to address for the target group.
- Set Case-Specific KPIs: Define KPIs based on the benefits the tool is expected to bring to the end-users.
- Benefit Mapping: Correlate these KPIs to the identified needs of the target group.

M12-M16 Mobilization and Training

This phase will involve reaching out to and mobilizing the selected internal and external users for the evaluation. Training sessions will be organized to acquaint them with the EFRA tools. User manuals may also be developed and distributed to assist with the learning curve.

Activities during this period include:

- User Mobilization: Contact and enrol internal and external evaluators.
- Training Session Organization: Conduct training sessions for the pilot, possibly with the support of technical partners.
- User Manuals: Create and distribute user guides, if necessary.

M17 Actual Piloting

The actual piloting cycle will be conducted. The EFRA tools will be deployed to the target groups, and data will be collected based on the KPIs defined earlier. This is the critical stage where all the planning and preparations are put to test.

Activities during this period include:

- Execute 1st Piloting Cycle: Deploy the tool to the target user groups and collect data based on predefined KPIs.

M18-M19 Follow-Up

After the piloting, preliminary data analysis will be conducted to examine the effectiveness of the tools against the set objectives and KPIs. Feedback from users will also be integrated into subsequent iterations of the tool. The follow-up period will conclude with a wrap-up meeting to summarize the findings and discuss next steps.

Activities during this period include:

- Preliminary Data Analysis: Evaluate the data from the pilot test to assess how well objectives were met.
- Feedback Incorporation: Integrate the feedback into the tool for further refinement.
- Wrap-up Meeting: Conclude the 1st piloting cycle with a meeting summarizing findings and outlining next steps.

Each of these phases has been designed to build upon the previous, ensuring a systematic and thorough approach to piloting the EFRA tools.

2d Piloting Cycle

2d Piloting Cycle: Timing & Key Activities

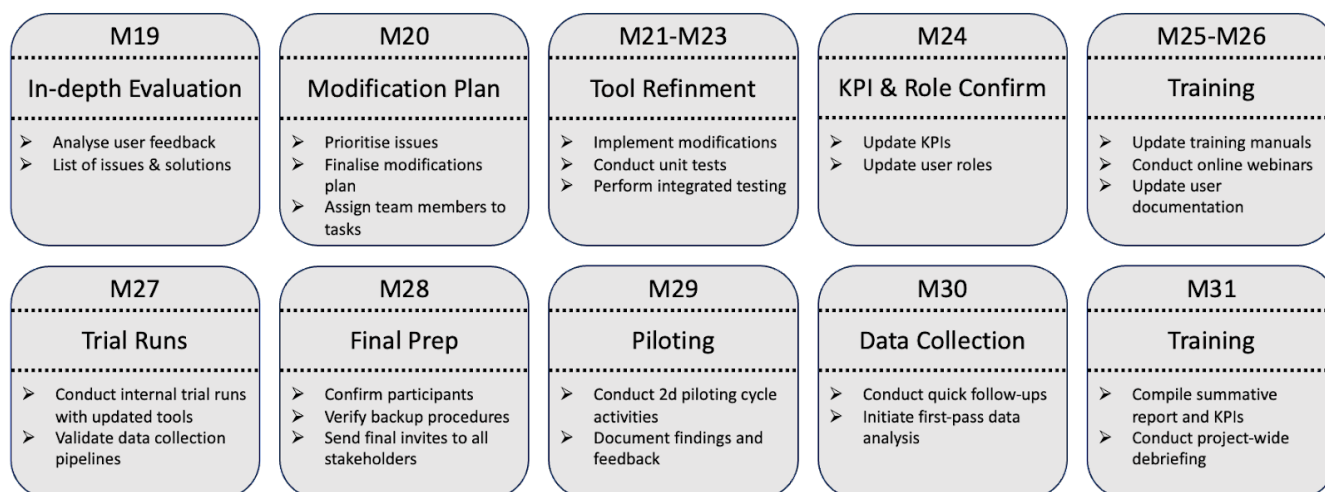


Figure 2: Key activities during first piloting cycle

M19 Evaluation of 1st Piloting Cycle

The initial phase will involve a thorough analysis of the results and feedback from the first piloting cycle. The objective is to understand what worked well and what needs refinement. Decisions will be made about tool enhancements, process modifications, or KPI adjustments for the second piloting cycle.

Activities during this period include:

- Conduct internal review meetings to discuss the 1st piloting cycle outcomes.
- Analyze user feedback and performance data.
- Create a list of identified issues and possible solutions.

M20 Finalizing Modifications

Based on the evaluation, the team will finalize the list of necessary modifications to the tools and pilot procedures. The finalized changes will be implemented during this month.

Activities during this period include:

- Prioritize issues and solutions identified in M19.
- Finalize a detailed modification plan.

- Assign team members to specific modification tasks.

M21-M23 Tool Refinement

This phase is dedicated to executing the tool modifications identified in the M20 planning. Rigorous internal testing will take place to ensure all changes are successfully implemented.

Activities during this period include:

- Implement code changes and modifications in the tools.
- Conduct unit tests to validate each modification.
- Perform integrated testing of the whole system.

M24 Reconfirmation of User Groups and KPIs

Reconfirming or potentially revising the user groups and KPIs will happen this month, taking into account lessons learned from the first cycle.

Activities during this period include:

- Review first-cycle data to validate user group relevance.
- Update KPIs based on first-cycle findings.
- Communicate any changes to stakeholders.

M25-M26 User Re-Training and Preparations

Training materials will be updated to reflect the tool modifications. Existing and potentially new end-users will be re-trained or trained respectively, setting the stage for the second cycle of piloting.

Activities during this period include:

- Update training manuals to include new features or modifications.
- Conduct online training webinars for end-users.
- Distribute updated user documentation.

M27 Trial Runs and Sanity Checks

A series of small-scale trial runs will be conducted internally to ensure everything is set up correctly for the main event. Sanity checks will involve ensuring that data collection mechanisms are operational, and the tools are performing as expected post-modification.

Activities during this period include:

- Conduct internal trial runs of the updated tools.
- Validate data collection pipelines.
- Perform a system-wide sanity check.

M28 Final Preparations and Communication

Last-minute checks and any required final adjustments will be made. A formal communication will be sent out to all stakeholders and participants to inform them about the upcoming second cycle, specifying dates, expectations, and any preparatory steps they need to take.

Activities during this period include:

- Confirm participant lists and schedules.
- Verify backup and rollback procedures.
- Send final pre-piloting communications to all stakeholders.

M29 Actual Piloting

The second piloting cycle will be executed, employing the revised tools and evaluation methods. This cycle will place more focus on validating the long-term utility and effectiveness of the tools.

Activities during this period include:

- Initiate the 2nd piloting cycle.
- Monitor system performance and user engagement in real-time.
- Document any immediate issues or failures.

M30 Initial Data Collection and Feedback

Immediate post-piloting data will be collected. Quick feedback loops will be established with end-users to gather their initial impressions and any urgent issues.

Activities during this period include:

- Collect initial data sets from the 2nd piloting.
- Conduct quick follow-up interviews with selected end-users.
- Initiate first-pass data analysis.

M31 Conclusion and Summative Analysis

Data will be analysed comprehensively to assess whether the KPIs were met and to what extent the tools fulfilled their intended objectives. Final evaluations will summarize the pilot's overall success and lay out the next steps for full-scale implementation or further refinement.

Activities during this period include:

- Complete the final round of data analysis.
- Compile a summative report comparing KPIs and objectives met.
- Conduct a project debriefing meeting to discuss next steps.

3.1.3 Gaant Chart

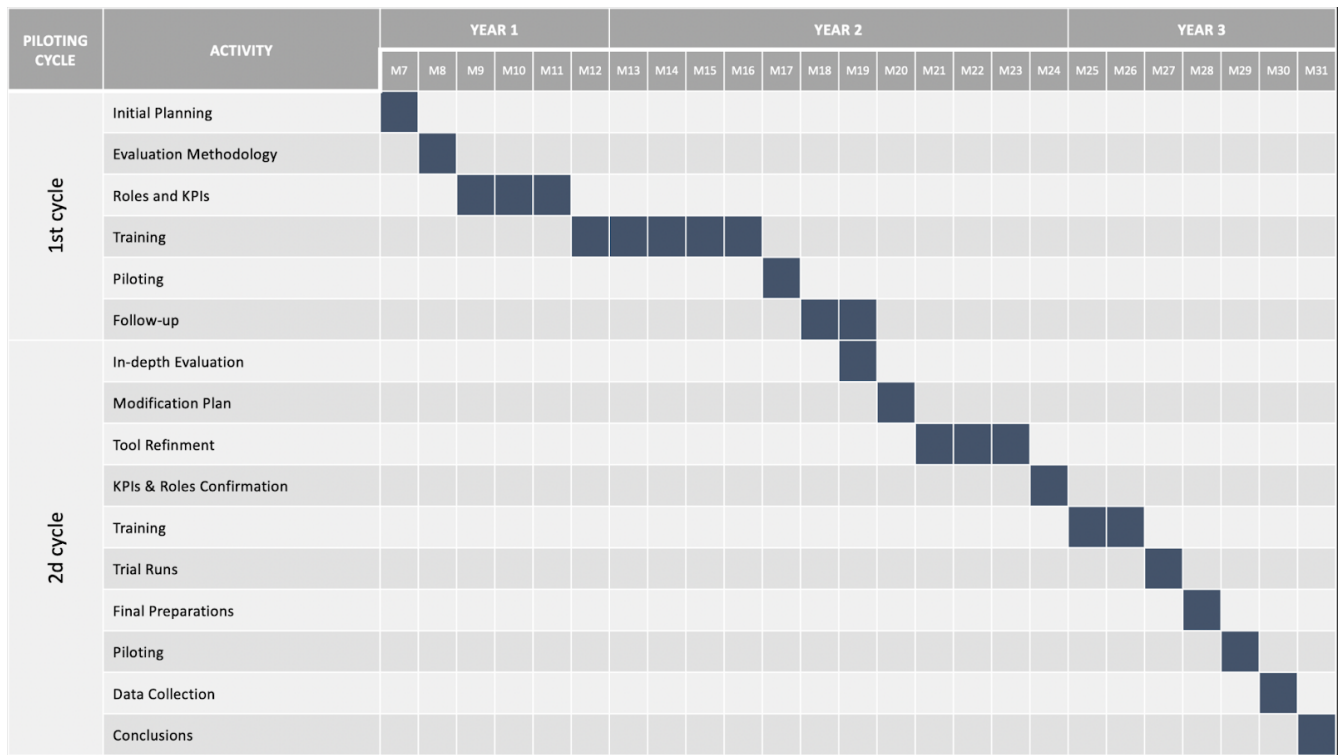


Figure 3: Gaant chart of piloting activities throughout the lifetime of the project

3.2 Evaluation Methodology

3.2.1 Scope

The primary focus of our evaluation process during the EFRA pilots is to collect direct user feedback from real users in their operational environments. This dual-pronged approach will involve presenting both design mock-ups and functional tools for evaluation. The evaluation strategy aims to be comprehensive, ensuring that both user needs and technical requirements are met, and providing a well-rounded view of the system's performance and impact.

Evaluation Methods

User Feedback: Users will interact with the system, providing qualitative data on usability, feature relevance, and overall experience. The users will be able to provide their feedback through appropriate online forms or, in some cases, directly inside the systems being used.

Technical Assessment: System logs, latency metrics, and other performance indicators will be recorded to assess the technical robustness of the tools and models.

Evaluation Outcomes

Quantitative Results: Various KPIs and evaluation measures will be benchmarked against initial requirements. These metrics will offer insights into the strengths and weaknesses of EFRA's components, informing both technical tasks and future business plans.

User Testimonials: Written testimonials from end users will be collected to add credibility and assist in the conversion of potential clients.

Lessons Learned: Key takeaways will be documented in annual reports. These will provide valuable insights for organizations planning to implement similar technologies in precision farming and food supply chain management.

Performance Metrics of AI Models: Objective measurements evaluating the speed and usability of the deployed AI models will be conducted. These metrics are critical for assessing the real-world applicability and reliability of the models in operational settings. They will be compared to predefined benchmarks to gauge effectiveness and areas for improvement.

3.2.2 Technical Performance Evaluation

Performance Metrics

To evaluate the answer speed and efficiency of the AI models, the following methods will be employed:

- **Latency Analysis:** This involves measuring the time taken from the point of request initiation to the receipt of a response, helping to identify latency issues that may affect user experience.
- **Resource Monitoring:** Here, metrics such as CPU usage, memory consumption, and network bandwidth are monitored during model operations to assess system load and overall efficiency.
- **Performance Profiling:** This allows us to locate bottlenecks and inefficiencies at the code level, pinpointing areas that could benefit from optimization.

These metrics aim to assess the performance of the AI models in terms of speed and usability as experienced by the user. The performance of the AI models in terms of predictive accuracy, recall, and similar metrics is in the scope of D5.2 and will be reported there.

System Robustness and Resilience

To assess how well the system can handle unexpected or adverse conditions, we will consider the following:

- **Error Analysis:** This involves cataloguing and analysing the various types of errors the models produce, such as false positives or false negatives, to understand their weaknesses and points of failure.
- **Chaos Engineering:** Here, we intentionally introduce faults into the system in a controlled manner to test its resilience and ability to recover from failures.

3.2.3 Human Engagement Evaluation

Evaluating how users interact with AI tools, models, and dashboards is crucial for refining usability and functionality. Below are comprehensive categories and methods that we will use for this evaluation. The relevant front-facing modules and dashboards will be developed within Task 5.4.

Usability and Functional Testing

This includes expert UX evaluation and pre-release testing to identify interface and functionality issues. Specific activities will include:

- **Expert Evaluation:** Mock-ups will be reviewed by expert decision makers to assess the interface and identify evident usability issues and possible implementation faults.
- **Thorough Testing:** Simple and complex scenarios will be tested to identify issues before platform release.

User Performance and Experience Metrics

These measurements occur in a controlled setting, capturing both quantitative and qualitative data about the user's experience and performance. Methods include:

- Task Completion Times: Measure the time users take to complete predefined tasks after brief training on the tool.
- Think Aloud Protocol: Users will be asked to verbalize their thoughts during task completion for evaluators to record.
- Post-Task Questionnaire: Users will fill out a questionnaire after completing tasks, further followed by a semi-structured interview focused on identifying both positive aspects and problems.

Longitudinal Studies

Users are asked to use the platform in their real work context for extended periods, with varying methods of feedback collection depending on the length of the study:

- Diaries: Daily or weekly logs maintained by users.
- Regular Interviews: Conducted at predefined intervals to collect feedback.

System Logging and Analytics

Capturing user interactions in log files provides supplementary data, such as:

- Time Analysis: Time taken by users to perform specific tasks.
- Activity Logging: Types of tasks performed, and duration spent using the tool.

Evaluator Feedback Methods

These methods aim to gather evaluative insights directly from those who interact with the system:

- Semi-Structured Interviews: Follow a guide but allow for additional follow-up questions to dig deeper into user experiences.
- User Experience Questionnaire (UEQ): Measures six dimensions of user experience such as efficiency, attractiveness, and novelty.
- Net Promoter Score (NPS): Assesses user loyalty by asking how likely they are to recommend the tool to others.

3.3 Integration of Piloting Outcomes to Project Activities

The outcomes of the pilot studies have a dual purpose, feeding directly into two key deliverables: D5.4 and D1.2. By conducting the pilot studies with the considerations reported below, we can ensure that their outcomes provide actionable insights for both D5.4 and D1.2, thereby fulfilling and enriching the broader objectives of the EFRA project.

Contribution to D5.4: Empowering Decision Support with EFRA Services

1. AI Explainable Dashboards: Feedback on the AI dashboards from the pilot tests will refine their user experience. This is particularly important for D5.4's focus on developing AI explainable dashboards.
2. HCI Insights: The pilot will include testing of hybrid AI explainability methods. These insights will directly inform the HCI perspective aspect of D5.4, particularly what experts may miss when AI results aren't easily interpretable.

3. Real-World Testing: The pilot's real-world scenarios align well with D5.4's goal to test novel solutions in real-world challenges. Feedback from the pilots will guide the iterative cycles for each use-case within D5.4.

Contribution to D1.2: Recommendations for EFRA Work

1. Data Source Challenges: Our pilot outcomes can directly inform the section on data mining and processing challenges, particularly if the pilot uncovers issues related to data heterogeneity or multilingual nature.
2. AI Approaches: Insights on the effectiveness, limitations, and efficiency of the AI models used in the pilot will contribute to D1.2's assessment of current AI methods in early warning systems.
3. User Needs on Decision Support: Feedback from the pilot studies on decision support tools, performance metrics, and security concerns will directly feed into D1.2's objectives to determine end-user needs in these areas.
4. Data Handling: Any challenges or insights related to data transfer and storage gained during the pilot will be valuable for D1.2's focus on Big/Extreme Data environments.

3.4 Outreach Multiplication Measures

Outreach multiplication measures play a pivotal role in the success of the pilots by enhancing participant engagement and fostering a diverse and robust involvement. These measures are critical because they extend the reach of the project, ensuring that it resonates with a broader audience. Implementing various outreach strategies, such as targeted social media campaigns, community engagement, and direct communication channels, helps create awareness and disseminate information effectively.

By multiplying outreach efforts, the project can tap into different market segments, encouraging diverse stakeholders' participation. This inclusivity not only enriches the diversity of perspectives but also increases the likelihood of attracting individuals who might otherwise be unaware of or hesitant to join the pilot initiative. Ultimately, a well-executed outreach multiplication strategy not only garners increased participant numbers but also contributes to the project's overall success by fostering a dynamic and engaged community.

To ensure the success of the pilots, EFRA's consortium employs Awareness, Interest, Desire, and Action (AIDA) measures, following a multi-actor, multi-step, and multi-channel approach. The specific user groups that these measures target for each scenario is the following:

- Targeted users for Scenario 1: *Food safety experts, Technical Support staff, Data Analysts, Veterinarians, Supply Chain Managers, Poultry Farm Managers, Process Managers.*
- Targeted users for Scenario 2: *Farmers, Farm Advisors, Decision Makers, Data Scientists, Agronomists.*
- Targeted users for Scenario 3: *Product Compliance Units, Regulators, Legal Advisors.*

And the KPIs been tracked in order to ensure pilot participation are:

- Scientific Publications: Publications in peer-reviewed journals, Publications in scientific conferences, Publications in scientific conferences.

- Technical/Thought Leadership Publications: white paper with the common conceptual model for EFRA and relevant standards to ensure data interoperability, discussion paper with contributions from key government, industry & standardization bodies, Blog contributions, discussion papers, best practice publications, sectoral data & prediction reports.
- Events/webinars organization: scientific / data science workshops and/or webinars, industry-oriented workshops and/or webinars, dedicated sessions at the annual GFSI/ fiin events, Working Group bringing together key stakeholders from the fiin & GFSI communities.
- Networking: booths at industry fairs, exhibitions and trade shows, participation in conferences, joint events with relevant EU funded projects and initiatives.
- Website & Branding materials: Website unique visitors, Blog posts, Bounce rate, visual identity and motto, banners, brochures, Printing and distribution of promotional materials.
- Social media platforms: Posts, Interactions, hashtags to use in social media.
- e-Newsletter & Email campaigns: e-newsletter to an existing database.

Multimedia & Multiplier campaigns: Published videos, Podcast series, leveraging print (newspapers, magazines) and online publishing platforms (e.g. Medium, WIRED), blogs related to food safety, food systems digitalization, IoT, AI, data management etc., to maximize the project's impact and also engage stakeholders outside the prioritized target groups.

4 Scenarios, Use-cases & Piloting Plan

4.1 Interconnection of Pilot Scenarios with EFRA Platform

The EFRA Platform's development is designed to be guided by real-world pilot scenarios, ensuring that the NLP and predictive AI models and related technical tools are applicable in actual real-world decision-support contexts. These tools are designed to leverage data from the EFRA Data Hub and the EFRA API Gateway, as outlined in D4.3, and contribute new datasets to the EFRA ecosystem. The aim is to create reusable, extensible AI and NLP models, and technical components, including dashboards, for integration into the EFRA Platform.

The EFRA Platform's development strategy places the pilot scenarios at the centre and allows them to guide the efforts of Work Packages 2, 3, and 5. These scenarios provide real-world contexts that anchor the theoretical aspects of data analysis, NLP, AI model development, and dashboard creation in tangible, applicable situations. This approach is instrumental in ensuring that the tools developed are viable solutions for real-world challenges in food safety.

Work Package 2 (WP2), focusing on Data and NLP, will utilize these scenarios to refine its methods for handling and interpreting vast datasets relevant to food safety. By engaging with data from actual field situations, WP2's NLP tools can be fine-tuned to understand and process industry-specific terminology and complex language constructs within food safety data, ensuring that the models developed are robust and capable of handling real-world data complexities.

Similarly, Work Package 3 (WP3) is tasked with AI Model Development. The pilot scenarios provide a critical testing ground for these models. By applying AI algorithms to real-world data and situations, WP3 ensures that the models are technically sound and effective in predicting and managing real-world food safety risks.

Work Package 5 (WP5) is responsible for creating Interactive Dashboards. The dashboards developed will be user-centric, with an emphasis on providing clear, actionable insights from complex data sets. The real-world scenarios will inform the design of these dashboards, ensuring that they are intuitive and meet the needs of various stakeholders, from farm operators to regulatory bodies.

Table 1: Connection of technical WP work with tools developed within the piloted scenarios

Scenario	Tool	Connected WPs
Enhancing Safety and Efficiency in Poultry Farming	AI model for outlier detection in poultry mortality and weight gain	WP2 (data analysis and integration), WP3 (AI model development)
	AI model for clustering of similar poultry flocks	WP2 (data analysis and integration), WP3 (AI model development)
	Outlier and Clustering Analysis Interactive Dashboards	WP5 (Interactive dashboard development)
Enhanced predictive capabilities for pest alarms	Pest Feedback Module	WP2 (data analysis and integration)
	AI Predictive Algorithm for Pest Occurrence	WP3 (AI model development)
	Interactive Dashboard	WP5 (Interactive dashboard development)
	Enhanced Farmer Scouting Module	WP2 (data analysis and integration), WP5 (Interactive dashboard development)
Automated regulatory analysis and summarisation	Automated Regulatory Analysis & Summarization Module	WP2 (data analysis and integration), WP3 (AI model development)

	Chatbot Interface	WP2 (data analysis and integration), WP5 (Interactive dashboard development)
--	-------------------	---

A crucial aspect of this strategy is the extension and generalization of the developed tools to make them applicable in contexts beyond the initial pilot scenarios. This adaptability is key to adding value to the EFRA Platform and ensuring its sustainability and relevance in the long term. For instance, a tool developed for outlier detection in early flock lifecycle in poultry farms might be extended to analyze similar risks in other livestock or agricultural settings.

This approach enhances the immediate utility of the EFRA Platform and lays a foundation for its continuous evolution. As new challenges emerge in the food safety domain, the platform can adapt and expand, incorporating new data, models, and dashboards to meet these challenges. This dynamic adaptability ensures that the EFRA Platform remains a valuable, sustainable resource in the ever-evolving field of food safety.

4.2 Enhancing Safety and Efficiency in Poultry Farming

4.2.1 Overview

In poultry farming, early identification of outliers in terms of mortality and weight gain rate, potentially caused by factors such as birds affected by diseases like Salmonella, pododermatitis (podo), hock burns, or other health issues—is crucial for maintaining both safety and operational efficiency. The timely detection of these anomalies enables faster intervention, which can prevent outbreaks and minimize economic losses.

To achieve this, the EFRA project is developing an advanced AI-driven monitoring system that (a) identifies potential outlier flocks in terms of mortality or weight gain rates and generates early warning signals, (b) clusters together similar flocks in terms of their behaviour regarding mortality and weight gain to identify flocks that behave differently than expected and allows closer analysis of the underlying flock characteristics. The underlying AI models are trained using historical data—including lab test results, Whole Genome Sequencing data, and specific production data, including mortality and weight gain rates during the first weeks of flock lifecycle.

The targeted users for the pilot phase include food safety experts, technical support staff, and process managers, all of whom play key roles in monitoring and responding to early warning signs. The success of the AI model will be measured by its accuracy, which is expected to surpass 60%, and user satisfaction ratings for decision-support dashboards, aiming for a score of above 7/10.

4.2.2 Piloted Use-cases

Use-Case 1: Outlier Detection in Early Poultry Flock Lifecycle

This use-case aims to employ appropriate statistical and AI methods to enable early outlier detection in a poultry flock lifecycle. The ultimate goal is to train an AI model that takes as input the characteristics of a given flock (e.g., current morality and weight gain rates, occurrence of disease, contextual and environmental farm factors) and identify the likelihood that this flock will behave abnormally in terms of mortality or weight gain rates. Provided that this is done early in the flock lifecycle, this can enable proactive measures to mitigate the underlying risk factors and ensure better animal welfare and farm productivity. Over this model, user-friendly decision support tools will be developed that will allow interaction with the underlying AI model and visualise its outputs.

Use-Case 2: Clustering and Analysis of Flock Characteristics

This use-case aims to leverage clustering algorithms and advanced data analytics to identify and group poultry flocks with similar characteristics. By analysing key parameters such as feed type, temperature, humidity, weight gain, mortality rates, and disease occurrence, the system will categorize flocks into distinct clusters. These clusters will help farm managers and veterinarians gain insights into which flocks are at higher risk of certain diseases or suboptimal growth performance.

The AI-based clustering tool will be integrated with a user-friendly dashboard that visualizes the clusters and allows the user to examine the different clusters. Decision-makers will be able to drill down into specific clusters to understand the unique characteristics of each group and make informed decisions regarding resource allocation, feed adjustments, or health interventions.

4.2.3 Piloting Plan

This section outlines specific plans for the 1st and 2^d piloting cycles for the scenario “Enhancing Safety and Efficiency in Poultry Farming” and its two use-cases “Outlier Detection in Early Poultry Flock Lifecycle” and “Clustering and Analysis of Flock Characteristics”. The piloting plan follows the methodology detailed in section 3 and provides details on the timeframe, participants, piloted EFRA tools, datasets, materials, and relevant KPIs.

Table 2: Piloting plan for the use-cases of the scenario “Enhancing Safety and Efficiency in Poultry Farming” during the 1st piloting cycle

Parameter	Description
Timeframe	Training: Will start on M12 by sending the training material (see “Materials” table row below) to an identified pool of participants (see “Participants” table row below). A follow up on month M15 with the participants will be scheduled to ensure they have understood the training material and that they will participate as expected in the piloting activity.
	Piloting: Will occur during M17 and will involve a live demo session with the identified participants that have successfully completed their training.
	Evaluation: Will happen immediately after the piloting session. Questionnaires will be sent to all piloting participants immediately after the live demo session. Selected participants will be asked for a follow-up interview on M18. All answers will be analysed on M19-M20, and the results will be reported in the next iteration of D5.3.
Participants	Targeted Groups: Our targeted groups include key stakeholders and domain experts from the private industry (poultry farming), academia, and public organisations (food safety authorities). Specifically, we are targeting to attract experts from the following organisations: Moy Park (private industry), Crème Global (private industry), Queens University Belfast (academia), WFSR (academia), Syreon Research Institute (academia), Brussels Federal Agency for the Safety of the Food Chain (public organisations) Food Safety Authority of Ireland (public organisation).

	Farm and Housing Information	Place W/C, FiscalTag, Growout No, Complex Entity Number, Farm No, Entity No, House No
	Product and Breed Details	Product No, Breed No
	Dates and Age-Related Data	FirstDatePlaced, Date When 21 days Old, Last Available Kill Date, Thin Date, Clear Date, Sample Receive Date, First Date Placed
	Hatchery and Placement Information	Hatchery No, Place / Transfer In Qty, Head Placed
	Growth and Weight Measurements	7 Day Weight, 14 Day Weight, 21 Day Weight, 28 Day Weight, Thin Weight, Clear Weight, Ave Weight
	Health and Quality Indicators	HealthStatus, SerotypeNo, Hock %, Podo %
	Swab and Sample Details	SwabDate, Swab Date Diff
	Feed and Nutrition	Grower, Finisher, CRUMB, WITHDRAWAL
	Culling and Mortality Statistics	Ave Age, D3_Cull_%, D7_Cull_%, D14_Cull_%, D21_Cull_%, D28_Cull_%, LOF_Cull_%, D3_Mort_%, D7_Mort_%, D14_Mort_%, D21_Mort_%, D28_Mort_%, LOF_Mort_%, Total ACM %
	Processing and Reference Information	Kill Plant, RefNo
Materials	<p>Training Materials: The training material will include digital files (in the form of a slide deck) that provide the participants with the following information:</p> <ul style="list-style-type: none"> (a) the problem statement and key information for the relevant use-cases, including our goals and approach (b) key metadata of the dataset that was used to train the AI models (c) figures and statistics that explain key correlations and observations from the dataset and which reinforce why the relevant AI model will be able to achieve high accuracy and recall in the outlier detection tasks 	
	<p>Evaluation Materials: Post-pilot questionnaires and semi-structured interviews with selected participants. The requested feedback will include, among others, the following questions:</p> <ul style="list-style-type: none"> (a) How helpful do you find the insights and AI models presented today? (b) How confident are you that the insights and AI models presented today are sound and trustworthy? (c) Were any of the findings surprising to you? If yes, please explain. (d) What other poultry management challenges would you like us to address next? 	
KPIs	<p>AI model accuracy: >70% in the outlier detection tasks</p> <p>AI model recall: >60% in the outlier detection tasks</p> <p>Users involved: >10 expert users covering both private industry, academia and public organisations</p>	

Table 3: Piloting plan for the use-cases of the scenario “Enhancing Safety and Efficiency in Poultry Farming” during the 2^d piloting cycle

Parameter	Description
Timeframe	Training: Will start on M25 by sending the training material (see “Materials” table row below) to an identified pool of participants (see “Participants” table row below) which will include as many of the 1 st cycle participants as possible, along with an extended list of participants with the added goal of driving exploitation of the presented EFRA assets. A follow up on month M26 with the participants will be scheduled to ensure they have understood the training material and that they will participate as expected in the piloting activity.
	Piloting: Will occur during M29 and will involve a live demo session with the identified participants that have successfully completed their training. After this the participants will receive accounts to the EFRA Platform and will be able to use the demonstrators on their own. Appropriate technical testing of the demonstrators will occur on M26 and M27 to ensure that everything will run smoothly during the live sessions.
	Evaluation: Will happen immediately after the piloting session. Questionnaires will be sent to all piloting participants immediately after the live demo session. Selected participants will be asked for a follow-up interview on M30. All answers will be analysed on M31, and the results will be reported in the next iteration of D5.3.
Participants	<p>Targeted Groups: Our targeted groups include key stakeholders and domain experts from the private industry (poultry farming), academia, and public organisations (food safety authorities).</p> <p>Specifically, we are targeting to attract several of the same experts from the 1st cycle, from the organisations: Moy Park (private industry), Crème Global (private industry), Queens University Belfast (academia), WFSR (academia), Syreon Research Institute (academia), Brussels Federal Agency for the Safety of the Food Chain (public organisations) Food Safety Authority of Ireland (public organisation).</p> <p>Additionally, to drive exploitation, we are targeting to attract additional organisations, especially from the poultry private industry. Given the scope of our AI models, we need to target vertically integrated industries which work directly with poultry farms for primary production, and then process the meat to manufacture end products. As a result, we will target large poultry companies, especially Pilgrim’s Pride and its subsidiaries in Europe (Pilgrim’s Europe). Pilgrim’s Pride is the largest poultry manufacturer globally and works in a similar fashion to Moy Park, making an ideal case for the generalisation of our AI models and interactive decision-making dashboards.</p>
	Targeted Expertise: The experts we wish to attract will have at least one of the following specialisations: (a) poultry farming, (b) microbiology, (c) food science, (d)

	<p>data analysis, (e) artificial intelligence. We will also target decision makers at the mid and senior management levels.</p> <p>Especially for this second pilot cycle, we will also target qualified leads in the private poultry companies to drive exploitation and uptake of the EFRA assets in a commercial setting.</p> <p>Targeted Countries: Ireland, United Kingdom, Netherlands, Belgium, Hungary, United States, France</p>
<p>Piloted EFRA Tools</p>	<p>Tool #1: EFRA Time series AI Engine / AI model for outlier detection in poultry mortality and weight gain Expected Status: The AI model will be trained using appropriate datasets (see “Datasets” table row below) that will allow it to successfully handle new cases. Interaction with the AI model will be done through appropriate interactive dashboards integrated within the EFRA Marketplace. The design and training methodology of the AI model is presented in deliverable D3.1. Demo: The pilot demonstration will include (a) an analysis of the dataset that was used to support the training of the AI model, including key characteristics and identified correlations (b) live outlier detection demo, where we will use the interactive dashboard and input some example cases of flocks and see the results produced by the AI model. After the live session, the participants will receive an account on the EFRA Marketplace to continue interacting with the interactive dashboard on their own time.</p> <p>Tool #2: EFRA Timeseries AI Engine / AI model for clustering of similar poultry flocks Expected Status: Appropriate clustering algorithm will be executed on the dataset (see “Datasets” below) and the clusters along with key insights will be identified. Interaction with the AI model will be done through appropriate interactive dashboards integrated within the EFRA Marketplace. The design and training methodology of the AI model is presented in deliverable D3.1. Demo: The pilot demonstration will include a live clustering demo, where we will execute use the interactive dashboards in the EFRA Marketplace and, in an interactive manner with the participants, examine selected clusters to delve deeper into their characteristics. After the live session, the participants will receive an account on the EFRA Marketplace to continue interacting with the interactive dashboard on their own time.</p> <p>Tool #3: EFRA Marketplace / Outlier and Clustering Analysis Interactive Dashboards Expected Status: Fully functioning, including interactive dashboards for Tools #1 and #2 above. The relevant demonstrators are presented in deliverable D4.3 (as part of the EFRA Marketplace). Demo: The demo will showcase the major functionalities of the interactive dashboards which include (a) Outlier Detection Dashboard: The dashboard will display detected outliers in flock performance, such as abnormal mortality or weight gain trends. The demo will walk</p>

	<p>through an example where participants explore the interactive dashboard to identify potential outliers flagged by the AI engine.</p> <p>(b) Clustering Dashboard: The dashboard will display clusters of flocks with similar characteristics, providing users with visual summaries and key insights from the clustering algorithm. The demo will show how users can interact with these clusters, delve into specific clusters, and understand underlying patterns.</p> <p>After the live session, the participants will receive an account on the EFRA Marketplace to continue interacting with the interactive dashboard on their own time.</p>
Datasets	<p>Detailed lifecycle reporting for 10327 flocks placed between 2022-2023, covering various parameters related to weight, mortality, and environmental conditions. The table below explains all the different parameters present in the dataset.</p> <p>For the second piloting cycle, we will include additional lifecycle reports from flocks placed in 2024. We expect this to add more than 5000 additional flock records to the dataset and increase the AI models accuracy and recall.</p>
Materials	<p>Training Materials: The training material will include digital files (in the form of a slide deck) that provide the participants with the following information:</p> <ul style="list-style-type: none"> (a) the problem statement and key information for the relevant use-cases, including our goals and approach (b) key metadata of the dataset that was used to train the AI models (c) figures and statistics that explain key correlations and observations from the dataset and which reinforce why the relevant AI model will be able to achieve high accuracy and recall in the outlier detection tasks. <p>Additionally, we will provide to all participants walkthrough videos that showcase the use of the interactive dashboards and the EFRA Marketplace. The participants will be guided on how to use create an account, log in, test the interactive dashboards with the sample datasets or even upload their own datasets to test the dashboards with their own data.</p> <p>Evaluation Materials: Post-pilot questionnaires and semi-structured interviews with selected participants. The requested feedback will include, among others, the following questions:</p> <ul style="list-style-type: none"> (a) How helpful do you find the insights and AI models presented today? (b) How confident are you that the insights and AI models presented today are sound and trustworthy? (c) Were any of the findings surprising to you? If yes, please explain. <p>Especially for the second pilot cycle, and for selected participants, we will focus also on exploitation related questions. The requested feedback will include, among others, the following questions:</p> <ul style="list-style-type: none"> (a) How interested are you to adopt AI-driven tools within your organization? (b) What potential barriers do you foresee in this endeavor? (c) In what ways could the EFRA assets be further developed to better meet the needs of your company?

	(d) Would you be interested in collaborating further for the deployment of the demonstrated EFRA tools in your organization?
KPIs	AI model accuracy: >75% in the outlier detection tasks AI model recall: >65% in the outlier detection tasks Satisfaction score with interactive dashboards: >7/10 Users involved: >25 expert users covering both private industry, academia and public organisations Follow-up actions: >5 demo requests from private organisations to drive commercial exploitation of the EFRA assets

4.3 Enhanced Predictive Capabilities for Pest Alarms

4.3.1 Overview

This scenario will improve AGRIVI's existing pest and disease algorithm, which currently relies on pest appearance criteria, weather forecast data, and location-specific inputs to provide timely alerts to farmers about potential pest invasions. The algorithm predicts the type of pest, the location, and the timing of potential outbreaks, aiming to help farmers take proactive protection measures. The updated algorithm, incorporating machine learning techniques and pest observations, aims to boost prediction accuracy and provide optimized recommendations for pest management. Real-life tests are scheduled with existing AGRIVI users, targeting farms with diverse crops and regional climates. This will help ensure the system's robustness and adaptability to various conditions.

Accurate pest prediction in agricultural production requires deep analysis of complex data, including environmental conditions (e.g., temperature, humidity, precipitation), crop types, and specific pest behaviors. This is critical because pests like the European corn borer, rootworms, and wireworms can cause significant damage during various crop life stages, from seedling to maturity. For example, pests can lead to 10-20% yield loss in crops such as corn, with specific pests like the European corn borer accounting for up to 15% yield loss. Additionally, diseases transmitted by pests can cause further damage, leading to financial and productivity losses for farmers.

The updated AGRIVI pest algorithm seeks to more effectively integrate historical on-field observations with climate and weather data to predict pest behaviors, especially in light of climate change, which is known to alter pest distribution patterns. Rising temperatures can expand the geographic range of certain pests, increase their overwintering survival, and disrupt the synchrony between host plants and natural predators. These shifts make it more difficult to rely on traditional pest management strategies, which are based on historical data and pest behavior models. In this context, the algorithm will support the implementation of Integrated Pest Management (IPM), a comprehensive approach that combines biological, cultural, and mechanical methods with selective pesticide use. By providing precise predictions, the updated algorithm aims to optimize pesticide use, reducing its environmental and health impacts. This is critical in minimizing risks such as pesticide poisoning, which accounts for an estimated 220,000 deaths annually, according to the World Health Organization.

The piloting of this enhanced algorithm is structured in two phases: data-informed rule extraction and predictive AI model development. Through extensive testing with AGRIVI's farmers, spanning different crops and regions, the algorithm will be validated for real-world application. This approach helps to optimize pest management strategies and ensure a reduction in unnecessary pesticide use, promoting sustainability and improved crop yield.

4.3.2 Piloted Use-cases

Use-case 1: AI-Enhanced Pest Prediction Alerts

This use case is designed to advance the understanding of how pests respond to changing climate conditions, enabling more precise and timely pest management. The approach is divided into two main phases: **Data-Informed Rule Extraction** and **Predictive AI Model Development**.

In the first phase, **Data-Informed Rule Extraction**, the focus is on identifying meaningful patterns in AGRIVI's extensive database. The data covers a wide range of factors, including direct pest observations, pesticide usage, and environmental variables such as weather conditions during pest treatments. The aim is to discover "soft rules"—subtle indicators that suggest when conditions may be ripe for pest outbreaks. Given the variability and sometimes incomplete nature of the data, machine learning algorithms like decision trees will be employed to handle inconsistencies and uncover these relationships. This phase is particularly important for understanding how climate change, which disrupts traditional weather patterns and pest behaviors, alters the risk landscape for farmers.

Once these foundational rules are established, the second phase, **Predictive AI Model Development**, will focus on building an AI model capable of forecasting pest invasions. The model will use a combination of weather data and real-time scouting inputs to predict pest behavior with greater accuracy. To ensure the model's reliability, AGRIVI will collaborate with farmers in different regions, who will provide ongoing feedback about pest activity in their fields. This regular, field-level input will be crucial in refining the model and validating its predictions against real-world conditions. The collected data will also help the AI system continuously adapt to new conditions, improving its long-term accuracy.

Ultimately, this use case aims to empower farmers with timely and actionable alerts about pest threats, allowing them to take preventative measures before infestations escalate. By providing reliable, data-driven insights, the system will help reduce the reliance on broad-spectrum pesticide applications, promoting more sustainable farming practices while safeguarding crop yields in an increasingly unpredictable climate.

Use-case 2: Real-Time Farmer Feedback Loop

This use case highlights the critical role that farmers' on-field feedback plays in refining and improving pest detection systems. Farmers serve as the first line of defense against pest invasions, and their real-time observations—whether confirming or refuting pest alarms—are essential for both timely interventions and the continuous enhancement of the AI model's accuracy.

By integrating feedback directly from those working in the fields, the system receives immediate validation or correction regarding the presence of pests. This feedback is crucial because it provides "ground truth" data, which strengthens the AI model's ability to adjust its predictions based on real-world conditions. Over time, this dynamic loop between human expertise and machine learning helps the model become more adept at detecting pests, even in complex or changing environments.

The feedback loop does not just rely on confirming the presence of pests; it also gathers valuable information when alarms are incorrect, allowing the system to learn from both positive and negative outcomes. This dual approach helps build a robust and well-rounded dataset, one that captures a wide array of environmental conditions and pest behaviors. As the AI model evolves, it becomes better at distinguishing between genuine threats and false alarms, providing more accurate alerts to farmers and reducing unnecessary interventions.

Ultimately, this real-time feedback loop creates a collaborative relationship between farmers and technology, where human insights continuously improve the system’s predictive capabilities, leading to more efficient pest management, reduced crop damage, and optimized pesticide use.

4.3.3 Piloting Plan

This section outlines specific plans for the 1st and 2^d piloting cycles for the scenario “Enhancing Predictive Capabilities for Pest Alarms” and its two use-cases “AI-Enhanced Pest Prediction Alerts” and “Real-time Farmer Feedback Loop”. The piloting plan follows the methodology detailed in section 3 and provides details on the timeframe, participants, piloted EFRA tools, datasets, materials, and relevant KPIs.

In the 1st piloting cycle, the aim for this scenario is two-fold and is analysed in two different tables below: (a) validate with agronomist experts the rules extracted during the Data-Informed Rule Extraction phase (Table 4), (b) execute a number of field trials, where a selected number of farmers will use the first version of the Feedback Loop tool to provide frequent positive/negative pest presence information (Table 6).

Table 4: Piloting plan for the use-cases of the scenario “Enhancing Predictive Capabilities for Pest Alarms” during the 1st piloting cycle focusing on the Data-Informed Rule Extraction phase

Parameter	Description
Timeframe	Training: Will start on M12 by sending the training material (see “Materials” table row below) to an identified pool of participants (see “Participants” table row below). A follow up on month M15 with the participants will be scheduled to ensure they have understood the training material and that they will participate as expected in the piloting activity.
	Piloting: Will occur during M17 and will involve a live demo session with the identified participants that have successfully completed their training.
	Evaluation: Will happen immediately after the piloting session. Questionnaires will be sent to all piloting participants immediately after the live demo session. Selected participants will be asked for a follow-up interview on M18. All answers will be analysed on M19-M20, and the results will be reported in the next iteration of D5.3.
Participants	Targeted Groups: Our targeted groups include expert agronomists with direct knowledge of pest behaviour and pesticide use. At this stage, this group will be composed of internal experts from the AGRIVI organization who have experience in managing crop health, pest detection, and the optimization of pesticide applications. These individuals will provide invaluable insights into how pest management practices can be improved and how the AI models can be fine-tuned for practical, on-field application.
	Targeted Expertise: The experts will have in-depth knowledge of several key areas, including <ul style="list-style-type: none"> • Pest lifecycle and behaviour: Understanding how various pests develop and the environmental conditions that promote their presence.

	<ul style="list-style-type: none"> • Pesticide usage and application strategies: Expertise in optimizing the type, timing, and quantity of pesticide use to achieve maximum efficacy with minimal environmental impact. • Crop-specific pest threats: Knowledge of the pests most likely to affect different types of crops and the specific countermeasures that are effective against these threats. • Integrated Pest Management (IPM) practices: Experience in combining biological, cultural, and mechanical pest control methods with selective pesticide application.
Piloted EFRA Tools	<p>Targeted Countries: Croatia, Bulgaria, Romania</p> <p>Tool: EFRA Analytics Powerhouse / Rule Extraction Module</p> <p>Expected Status: The rule extraction module will be totally operational and able to create a decision tree from the incomplete and heterogeneous dataset provided (see “Dataset” below). The dataset will be used to train machine learning models, specifically decision trees, which help in rule extraction. The goal is to create patterns or "rules" that can predict when pests are likely to appear based on weather conditions and historical data. The design and training methodology of the AI model is presented in deliverable D3.1.</p> <p>Demo: The pilot demonstration will include</p> <p>(a) an analysis of the dataset that was used to support the training of the decision tree, including key characteristics and identified correlations,</p> <p>(b) live decision tree creation demo, where we will showcase the creation of the decision tree and engage the participants in structured feedback.</p>
Datasets	<p>The dataset used in this pilot focuses primarily on observations from apple fields and weather data. More specifically:</p> <ol style="list-style-type: none"> 1. Pest Observations: <ul style="list-style-type: none"> ○ The dataset contains 22 direct pest observations collected between 2020 and 2022. ○ 927 pest alarms were recorded from 2020 to 2023. These alarms represent instances where the system predicted pest presence based on its current algorithm. 2. Weather Data: <ul style="list-style-type: none"> ○ Historical weather data is included for the same periods as the pest observations and alarms. Weather features consist of: <ul style="list-style-type: none"> ▪ Daily average, maximum, and minimum air temperature ▪ Wind speed (daily average, maximum, and minimum) ▪ Relative humidity (daily average, maximum, and minimum) ▪ Daily precipitation (accumulation in mm) ▪ Dewpoint (daily average)
Materials	<p>Training Materials: The training material will include digital files (in the form of a slide deck) that provide the participants with the following information:</p> <p>(a) the problem statement and key information for the relevant use-cases, including our approach to improving pest prediction accuracy under changing climate conditions</p>

	<p>(b) key metadata of the dataset used to train the AI models, including details about weather data, pest observations, and historical pest alarms</p> <p>(c) figures and statistics highlighting key patterns between weather conditions and pest outbreaks.</p>
	<p>Evaluation Materials: Post-pilot questionnaires and semi-structured interviews with selected participants. The requested feedback will include, among others, the following questions:</p> <p>(a) Are these rules deemed useful from a domain expert perspective?</p> <p>(b) Would the addition of more rules similar to these benefit stakeholders/domain experts?</p> <p>(c) What aspects do you find particularly useful, and conversely, what aspects do you deem less useful?</p> <p>(d) What important factors are missing from the rules, and conversely, what elements do you consider excessive or unnecessary?</p>
KPIs	<p>Signals used in the model: >6 parameters, including field observations and weather parameters.</p> <p>Usefulness of model: >7/10 approval of the data-extracted rules from experts</p> <p>Users involved: >5 expert users participating in the validation pilot</p>

Table 5: Piloting plan for the use-cases of the scenario “Enhancing Predictive Capabilities for Pest Alarms” during the 1st piloting cycle focusing on field farm data collection for the Feedback Loop phase

Parameter	Description
Timeframe	<p>Training: Will start on M12 by sending the training material (see “Materials” table row below) to an identified pool of participants (see “Participants” table row below). A follow up on month M15 with the participants will be scheduled to ensure they have understood the training material and that they will participate as expected in the piloting activity.</p>
	<p>Piloting: Will occur during M17 and will involve a live demo session with the identified participants that have successfully completed their training. The participants will then be asked to engage on field scouting on a regular basis and input their findings in the Feedback Loop module. This phase will conclude in August 2024.</p>
	<p>Evaluation: Will happen immediately after the piloting session. Questionnaires will be sent to all piloting participants immediately after the conclusion of the field scouting phase. Selected participants will be asked for a follow-up interview on M21. All answers will be analysed on M22-M23, and the results will be reported in the next iteration of D5.3.</p>
Participants	<p>Targeted Groups: Farmers actively engaged in crop production and with good understanding of digital technologies (e.g., through the use of the AGRIVI software in their daily operations).</p>

	<p>Targeted Expertise: The farmers should have the following expertise:</p> <p>Experience in crop management and pest scouting.</p> <ul style="list-style-type: none"> • Ability to monitor and record crop growth stages. • Skills in identifying and reporting the appearance or absence of pests. <p>Familiarity with documenting farming activities:</p> <ul style="list-style-type: none"> • Reporting natural disasters affecting crops. • Recording spraying activities and pesticide use. <p>Basic proficiency in using agricultural software tools for data input and communication.</p>
<p>Piloted EFRA Tools</p>	<p>Tool #1: EFRA Interaction Stack / Feedback Loop Application</p> <p>Expected Status: The Feedback Loop application will be operational for at least basic data input during the field scouting (e.g., presence and absence of pests, pest types).</p> <p>Demo: The pilot demonstration will include a live session where the Feedback Loop application will be showcased to the participants. Then the participants will be asked to use the application on their own during their field scouting, which should happen on a pre-agreed and frequent basis.</p>
<p>Datasets</p>	<p>The goal of this piloting phase is to create a more balanced and extensive dataset concerning field scouting observations for the presence or absence of pests across different regions and crop types. This will enable the creation of a robust dataset that combines real-time field data with historical weather patterns and pest behavior records. This dataset will play a crucial role in training and refining the AI models, improving their accuracy in predicting pest invasions based on local climate conditions and farmer feedback. This way, this scenario will effectively prepare the transition to the Predictive AI Model Development phase, which will be validated in the 2d piloting cycle.</p>
<p>Materials</p>	<p>Training Materials: The training material will include a live demo for the selected farmers to see the use of the new Feedback Loop application in action. The goal is to enable them to use the app independently during their field scouting to input the presence or absence of pests. The training will provide step-by-step guidance on how to:</p> <ul style="list-style-type: none"> • Log pest observations. • Report crop health status. • Input weather and environmental factors during field visits. <p>Evaluation Materials: Post-pilot questionnaires and semi-structured interviews with selected participants. The requested feedback will include, among others, the following questions:</p> <p>(a) How easy was the Feedback Loop application to use?</p> <p>(b) How confident are you in the accuracy of the pest alerts generated based on your input?</p> <p>(c) Did the system's predictions help you make better decisions on pest management in your field?</p>

	(d) What features or functionalities would you suggest improving the Feedback Loop application for future use?
KPIs	Useability of the interface: >70% satisfaction by the participating farmers Users involved: >3 farmers involved Frequency of scouting: Weekly for a period of 3 months

Table 6: Piloting plan for the use-cases of the scenario “Enhancing Predictive Capabilities for Pest Alarms” during the 2^d piloting cycle

Parameter	Description
Timeframe	Training: Will start on M25 by sending the training material (see “Materials” table row below) to an identified pool of participants (see “Participants” table row below) which will include as many of the 1 st cycle participants as possible, along with an extended list of participants with the added goal of driving exploitation of the presented EFRA assets. A follow up on month M26 with the participants will be scheduled to ensure they have understood the training material and that they will participate as expected in the piloting activity.
	Piloting: Will occur during M29 and will involve a live demo session with the identified participants that have successfully completed their training. After this the participants will receive accounts to the EFRA Platform and will be able to use the demonstrators on their own. Appropriate technical testing of the demonstrators will occur on M26 and M27 to ensure that everything will run smoothly during the live sessions.
	Evaluation: Will happen immediately after the piloting session. Questionnaires will be sent to all piloting participants immediately after the live demo session. Selected participants will be asked for a follow-up interview on M30. All answers will be analysed on M31, and the results will be reported in the next iteration of D5.3.
Participants	Targeted Groups: Farmers actively engaged in crop production and with good understanding of digital technologies (e.g., through the use of the AGRIVI software in their daily operations). We will endeavor to attract the same participants as in the first pilot cycle and also add additional ones taken from the current and active users of the AGRIVI software, which includes an extensive list of users across Europe.
	Targeted Expertise: The farmers should have the following expertise: Experience in crop management and pest scouting. <ul style="list-style-type: none"> • Ability to monitor and record crop growth stages. • Skills in identifying and reporting the appearance or absence of pests. Familiarity with documenting farming activities: <ul style="list-style-type: none"> • Reporting natural disasters affecting crops. • Recording spraying activities and pesticide use. Basic proficiency in using agricultural software tools for data input and communication.

	<p>Targeted Countries: Croatia, Bulgaria, Romania</p> <p>We will also aim to selectively test the applicability of the new pest alarm predictive algorithm in new countries for which the AI model has not been specifically trained with data. This will let us understand if transfer learning is a fruitful route.</p>
Piloted EFRA Tools	<p>Tool #1: EFRA Analytics Powerhouse / AI model for predicting pest presence Expected Status: The AI model will be trained using appropriate datasets (see “Datasets” table row below) that will allow it to successfully handle new cases. Interaction with the AI model will be done through appropriate interactive dashboards integrated within the EFRA Marketplace and/or the AGRIVI software. The design and training methodology of the AI model is presented in deliverable D3.1. Demo: The pilot demonstration will include (a) an analysis of the dataset that was used to support the training of the AI model, including key characteristics and identified correlations (b) live demo, where we will use the interactive dashboard to examine the predictions of the AI model for pest presence in various field conditions. After the live session, the participants will receive an account on the EFRA Marketplace to continue interacting with the interactive dashboard on their own time.</p> <p>Tool #2: EFRA Marketplace / Interactive dashboard for the AI model for predicting pest presence Expected Status: Fully functioning, including interactive dashboards for Tool #1 above. The relevant demonstrators are presented in deliverable D4.3 (as part of the EFRA Marketplace). Demo: The demo will showcase the major functionalities of the interactive dashboards which include examining the predictions of the underlying predictive AI model under different field conditions. After the live session, the participants will receive an account on the EFRA Marketplace to continue interacting with the interactive dashboard on their own time.</p>
Datasets	<p>The dataset used in this pilot will focus on newly generated data from the first piloting cycle, specifically incorporating field observations collected by farmers through the Feedback Loop module, combined with real-time and historical weather data. More specifically:</p> <ol style="list-style-type: none"> Pest Observations: <ul style="list-style-type: none"> The dataset will include new, real-time pest observations provided by farmers during their weekly scouting activities. These observations will cover the presence or <i>absence</i> of pests in their fields. Data will be continuously updated through the Feedback Loop module, where farmers record observations related to crop health, pest appearances, and any natural disasters affecting their crops. Weather Data: <ul style="list-style-type: none"> Real-time weather data will be collected in parallel with pest observations, consisting of: <ul style="list-style-type: none"> Daily average, maximum, and minimum air temperature Wind speed (daily average, maximum, and minimum) Relative humidity (daily average, maximum, and minimum) Daily precipitation (accumulation in mm)

	<ul style="list-style-type: none"> ▪ Dewpoint (daily average) <p>The new weather data will be integrated with the pest observations to refine the AI model's ability to predict pest presence under changing environmental conditions.</p>
Materials	<p>Training Materials: The training material will include digital files (in the form of a slide deck) that provide the participants with the following information:</p> <ul style="list-style-type: none"> (a) the problem statement and key information for the relevant use-cases, including our goals and approach (b) key metadata of the dataset that was used to train the AI models (c) figures and statistics that explain key correlations and observations from the dataset and which reinforce why the relevant AI model will be able to achieve high accuracy and recall in the pest prediction task. <p>Additionally, we will provide to all participants walkthrough videos that showcase the use of the interactive dashboards and the EFRA Marketplace. The participants will be guided on how to use create an account, log in, test the interactive dashboards with the sample datasets or even upload their own datasets to test the dashboards with their own data.</p>
	<p>Evaluation Materials: Post-pilot questionnaires and semi-structured interviews with selected participants. The requested feedback will include, among others, the following questions:</p> <ul style="list-style-type: none"> (a) How helpful do you find the insights and AI models presented today? (b) How confident are you that the insights and AI models presented today are sound and trustworthy? (c) Were any of the findings surprising to you? If yes, please explain. <p>Especially for the second pilot cycle, and for selected participants, we will focus also on exploitation related questions. The requested feedback will include, among others, the following questions:</p> <ul style="list-style-type: none"> (a) How interested are you to adopt AI-driven tools within your organization? (b) What potential barriers do you foresee in this endeavor? (c) In what ways could the EFRA assets be further developed to better meet the needs of your company? (d) Would you be interested in collaborating further for the deployment of the demonstrated EFRA tools in your organization?
KPIs	<p>AI model accuracy: >70% in predicting the presence of pests</p> <p>AI model recall: >60% in predicting the presence of pests</p> <p>Satisfaction score with interactive dashboards: >7/10</p> <p>Users involved: >25 users with commercial interest on this module</p> <p>Follow-up actions: >5 demo requests from farmers and private organisations to drive commercial exploitation of the EFRA assets.</p>

4.4 Automated Regulatory Analysis & Summarization

4.4.1 Overview

Food regulations are legal standards and guidelines established by governmental bodies and international organizations to ensure the safety, quality, and labeling of food products. These regulations encompass a broad range of requirements, including the use of additives, contamination limits, hygiene practices, and proper labeling for allergens or nutritional information. Relevant legal texts within this domain are often dense and complex, as they are designed to address every stage of food production and distribution, from farm to table. They are frequently updated to reflect new scientific findings, technological advances, and evolving public health priorities. Businesses, particularly those involved in food production and distribution, must continuously monitor and interpret these regulations to remain compliant and avoid legal or financial penalties. For professionals working in food safety and product compliance, timely and accurate interpretation of these regulatory changes is essential to ensure that products meet the required standards, and that public health is protected.

In the current state, SGS DIGICOMPLY provides users with access to vast amounts of regulatory data, but the burden of interpreting and summarizing this information remains largely manual. This process requires significant time and effort, as experts within the organisation must sift through complex, dense legal texts across multiple languages to extract relevant insights. The challenge is particularly pronounced in food safety regulations, where ensuring compliance in a timely manner is critical to business operations and public safety.

The future state envisions the integration of an **Automated Regulatory Analysis & Summarization Module**. This tool will use advanced language models to automatically generate concise, context-aware summaries of regulatory documents, thus drastically reducing the manual effort required from users. By offering key extracts and well-structured summaries, the module will help users quickly identify the most important compliance obligations and streamline decision-making processes. This will particularly benefit food safety professionals, product compliance teams, and regulatory authorities who regularly deal with the complexity of regulatory changes.

One of the core challenges is ensuring the module's **accuracy** and relevance, especially given the diverse, multi-lingual nature of regulatory texts, which may vary significantly in tone, structure, and specificity. Ensuring that the generated summaries maintain fidelity to the original legal nuances is essential. Finally, addressing linguistic variations and local regulatory specifics poses a considerable challenge.

4.4.2 Piloted Use-cases

Use-case 1: Automated Summary Generation for New Regulations

This use case focuses on automating the summarization of newly released or updated food safety regulations. Regulatory updates can be frequent, complex, and overwhelming, especially for companies operating in multiple jurisdictions. By leveraging the **Automated Regulatory Analysis & Summarization Module**, product compliance teams will be able to rapidly assess their compliance obligations and identify the specific actions required to adhere to new regulations.

The automated summaries will help users filter through the vast information, highlighting the key points that directly impact product compliance. This will allow compliance professionals to focus on implementing necessary changes without the burden of manual document reviews. In essence, this use case aims to provide users with faster, **more**

efficient access to relevant regulatory information, ensuring that they remain proactive in meeting compliance requirements.

Use-Case 2: Regulatory Summary Chatbot with Citation Capabilities

In this use case, the focus shifts to developing a **Regulatory Summary Chatbot**, which offers on-demand access to concise summaries of regulatory texts. The chatbot is designed to be **interactive**, allowing users to ask specific questions and receive tailored responses. Beyond just providing summaries, the chatbot will also cite exact sections of the legal texts, adding a layer of transparency and enabling users to verify the information against the original documents.

This feature will be particularly useful for **food safety managers and legal professionals**, who often need instant access to regulatory information but may lack the time to manually search through lengthy texts. The chatbot will save users considerable time by reducing the need for manual interpretation and searches while ensuring the accuracy of the responses provided. Additionally, by offering citations, it enhances **user trust** and **transparency**, allowing professionals to make informed decisions with confidence.

4.4.3 Piloting Plan

This section outlines specific plans for the 1st and 2^d piloting cycles for the scenario “Automated Regulatory Analysis & Summarisation” and its two use-cases “Automated Summary Generation for New Regulations” and “Regulatory Summary Chatbot with Citation Capabilities”. The piloting plan follows the methodology detailed in section 3 and provides details on the timeframe, participants, piloted EFRA tools, datasets, materials, and relevant KPIs.

Table 7: Piloting plan for the use-cases of the scenario “Automated Regulatory Analysis & Summarisation” during the 1st piloting cycle

Parameter	Description
Timeframe	Training: Will start on M12 by sending the training material (see “Materials” table row below) to an identified pool of participants (see “Participants” table row below). A follow up on month M15 with the participants will be scheduled to ensure they have understood the training material and that they will participate as expected in the piloting activity.
	Piloting: Will occur during M17 and will involve a live demo session with the identified participants that have successfully completed their training.
	Evaluation: Will happen immediately after the piloting session. Questionnaires will be sent to all piloting participants immediately after the live demo session. Selected participants will be asked for a follow-up interview on M18. All answers will be analysed on M19-M20, and the results will be reported in the next iteration of D5.3.
Participants	Targeted Groups: Our targeted groups include legal and compliance experts with deep knowledge of regulatory frameworks and food safety legislation. This group will be composed of internal experts from the SGS organization, as well as potentially external legal professionals and food compliance officers. These individuals will provide essential insights into how legal texts can be summarized effectively while maintaining regulatory accuracy and compliance integrity. Their expertise will help

	<p>refine the summarization tools and ensure they meet the practical needs of regulatory professionals.</p> <p>Targeted Expertise: The experts will have in-depth knowledge of several key areas, including</p> <ul style="list-style-type: none"> • Regulatory interpretation and compliance: Expertise in understanding and applying food safety regulations, ensuring that products comply with legal standards across various regions. • Legal text analysis: Skills in analyzing complex legal texts, identifying key obligations, and summarizing lengthy documents without losing critical legal meaning or nuance. • Regulatory change management: Experience in monitoring, interpreting, and applying updates to regulations, ensuring that businesses remain compliant as legal frameworks evolve. <p>Targeted Countries: Germany, Romania, Switzerland</p>
<p>Piloted EFRA Tools</p>	<p>Tool: EFRA Analytics Powerhouse / Text Summarisation Module</p> <p>Expected Status: The text summarization module will be fully operational and fine-tuned for the task of summarizing regulatory texts, using the provided SGS dataset (see “Dataset” below). The dataset will be used to train machine learning models capable of generating concise and accurate summaries from complex legal texts. This module aims to automate the process of regulatory text summarization, significantly reducing the time and effort required by users to interpret and act on regulatory updates. The design and training methodology of the AI model is presented in deliverable D3.1.</p> <p>Demo: The pilot demonstration will include:</p> <p>(a) An analysis of the SGS dataset, detailing the types of documents used for model training, key characteristics of the regulatory texts, and the various summary lengths and languages covered in the dataset. This will also highlight how the model was fine-tuned for the summarization task.</p> <p>(b) Live text summarization demo, where we will showcase the model’s ability to generate summaries for a new regulatory document. The demonstration will involve real-time interaction with participants, allowing them to evaluate the generated summaries and provide structured feedback on accuracy, clarity, and relevance.</p>
<p>Datasets</p>	<p>The datasets used in this use-case for regulatory and legal document summarization focus primarily on the SGS Dataset, a large-scale collection of regulatory texts and their corresponding summaries:</p> <ol style="list-style-type: none"> 1. SGS Dataset v1.0: <ul style="list-style-type: none"> ○ Total Documents: 14,308 documents. ○ Languages: The dataset spans 28 languages, with each document translated into English for broader accessibility and consistency. ○ Documents with Summaries: Out of the total, 9,995 documents have associated summaries. <ul style="list-style-type: none"> ▪ Manual Summaries: A small subset of 44 summaries was created manually by legal experts.

	<ul style="list-style-type: none"> ▪ Scraped Summaries: The remaining 9,951 summaries were automatically scraped from various sources. ○ Usage: This version of the dataset will be primarily used for finetuning models in early experiments. <p>2. SGS Dataset v2.0:</p> <ul style="list-style-type: none"> ○ Total Documents: This updated version includes 137,128 documents. ○ Languages: Documents are available in 45 languages, with translations into English. ○ Documents with Summaries: 124,235 documents have associated summaries. <ul style="list-style-type: none"> ▪ Manual Summaries: 2,124 summaries are manually created by experts. ▪ Scraped Summaries: 122,111 summaries were automatically scraped. ○ Usage: This version represents a more comprehensive dataset, used for more advanced experiments and finetuning of summarization models.
Materials	<p>Training Materials: The training material will include digital files (in the form of a slide deck) that provide participants with the following information:</p> <p>(a) The problem statement and key information for the relevant use-cases, including our approach to automating the summarization of complex regulatory texts to improve efficiency and accuracy in compliance management.</p> <p>(b) Key metadata of the SGS dataset used to train the AI models, including details about the number of documents, the languages covered, and the summary types (manual and scraped).</p> <p>(c) Figures and statistics illustrating the lengths of documents and summaries, accuracy of the model-generated summaries, and examples of successfully summarized regulatory texts.</p> <p>(d) An overview of how the model is finetuned using machine learning techniques to handle multi-lingual regulatory texts and ensure that summaries maintain critical legal details and nuances.</p>
	<p>Evaluation Materials: Post-pilot evaluation will involve both questionnaires and semi-structured interviews with legal experts who participated in the pilot. The feedback will focus on assessing the quality and utility of the generated summaries, based on the following key areas:</p> <ul style="list-style-type: none"> • Manual Evaluation of Generated Summaries: Participants will be asked to review the summaries produced by the summarization models and compare them to the original regulatory documents. The evaluation will focus on: <ul style="list-style-type: none"> ○ Main Parts for Summary: How effectively the summaries identify and capture the most important sections of the regulatory texts. ○ Key Points: Whether the summaries highlight critical legal concepts and compliance obligations that need to be addressed. ○ Summary Quality: Overall clarity, conciseness, and accuracy of the generated summaries in terms of reflecting the essential content. • Feedback Loop for Improvement: A key element of the evaluation will involve integrating participant feedback into the model improvement process. The following questions will guide the feedback collection:

	<ul style="list-style-type: none"> ○ Did the generated summaries cover the most relevant sections of the documents? ○ Were there any important aspects of the regulatory text that were missing or inaccurately summarized? ○ How did the generated summaries compare to manual summaries in terms of accuracy and usefulness? ○ What improvements would you suggest to enhance the summarization tool's ability to capture critical legal information? <ul style="list-style-type: none"> • Scoring and Ranking: Participants will score the generated summaries based on pre-defined criteria, with scores ranging from 1 to 5 for aspects such as: <ul style="list-style-type: none"> ○ Brevity and conciseness: How well the summary distills the regulatory text into a shorter format without losing critical meaning. ○ Accuracy: The ability of the summary to capture the key legal obligations and details. ○ Relevance: Whether the summary prioritizes the most relevant portions of the text for compliance professionals. • Summaries Comparison: Legal experts will compare different models used in the summarization process, providing feedback on which models performed best. Each summary will be annotated by legal experts to ensure the summaries correctly represent the content of the original documents. This feedback will be used to guide further fine-tuning of the models.
KPIs	<p>Summary Accuracy: ROUGE-1 score > 75% in generated summaries compared to the manual summaries provided by legal experts</p> <p>Relevance of Key Points: >80% of the generated summaries must include the most critical regulatory points identified by legal experts</p> <p>Summarization Length: Summaries generated must reduce document length by >85% on average</p> <p>User Satisfaction: >70% user satisfaction rate, based on feedback from legal and compliance professionals involved in the pilot</p> <p>Users involved: >5 expert users participating in the pilot.</p>

Table 8: Piloting plan for the use-cases of the scenario “Automated Regulatory Analysis & Summarisation” during the 2^d piloting cycle

Parameter	Description
Timeframe	<p>Training: Will start on M25 by sending the training material (see “Materials” table row below) to an identified pool of participants (see “Participants” table row below). A follow up on month M26 with the participants will be scheduled to ensure they have understood the training material and that they will participate as expected in the piloting activity.</p>
	<p>Piloting: Will occur during M29 and will involve a live demo session with the identified participants that have successfully completed their training. After this the participants will receive accounts to the EFRA Platform and will be able to use the demonstrators on their own. Appropriate technical testing of the demonstrators will</p>

	occur on M26 and M27 to ensure that everything will run smoothly during the live sessions.
	Evaluation: Will happen immediately after the piloting session. Questionnaires will be sent to all piloting participants immediately after the live demo session. Selected participants will be asked for a follow-up interview on M30. All answers will be analysed on M31, and the results will be reported in the next iteration of D5.3.
Participants	<p>Targeted Groups: The target groups in this piloting cycle will focus on engaging potential commercial users to drive exploitation of the EFRA assets. More specifically:</p> <ul style="list-style-type: none"> • Compliance Managers within food companies: These professionals are responsible for ensuring that products comply with constantly evolving regulatory standards, making them the primary users of the summarization and chatbot tools. Their input will help refine the tools to meet the practical needs of compliance and legal departments. • Senior Managers and Decision Makers: This group will focus on driving exploitation and adoption of the EFRA assets in a commercial setting. By engaging senior leadership, we aim to showcase the business benefits of the tools, including improved compliance efficiency and reduced manual workload.
	<p>Targeted Expertise: The expertise we will focus to attract is as follows:</p> <ul style="list-style-type: none"> • Compliance Management: Experience in ensuring compliance with food safety regulations across multiple regions and markets, with a focus on keeping up with regulatory updates. • Legal and Regulatory Analysis: Expertise in interpreting and applying food-related laws and regulations, particularly within the EU and international contexts. • Business Strategy and Decision Making: Senior management experience in implementing and exploiting digital tools to improve regulatory compliance processes, efficiency, and risk management. • Automation in Compliance: Knowledge of how automation tools like the summarization module and chatbot can streamline regulatory analysis and help ensure timely responses to new or changing regulations.
	Targeted Countries: Germany, Netherlands, United Kingdom, Switzerland, United States
Piloted EFRA Tools	<p>Tool: EFRA Interactions Stack / Interactive chatbot for Regulatory Texts</p> <p>Expected Status: The chatbot will be fully functional, capable of providing real-time, concise summaries of regulatory texts on demand. It will also include features for citing specific sections of legal texts, ensuring transparency and trust in the information provided. The chatbot will be integrated within the EFRA Interactions Stack (EFRA Marketplace and/or directly in the SGS Digicomply product), enabling users to access it as part of their compliance management workflows. The relevant</p>

	<p>demonstrators are presented in deliverable D4.3 as part of the EFRA Interactions Stack.</p> <p>Demo: The demo will highlight the core functionalities of the interactive chatbot, demonstrating how users can query the system for regulatory summaries in real-time. The chatbot will provide quick responses, summarizing key compliance obligations and citing specific portions of the regulatory text for validation. The participants will interact with the chatbot, inputting various questions related to food safety regulations, and receive tailored summaries. This will showcase how the chatbot reduces the need for manual text searches, streamlining compliance workflows. After the live session, participants will receive accounts on the EFRA Marketplace and/or SGS Digicomply product, allowing them to continue exploring and using the chatbot on their own time.</p>
Datasets	<p>We expect to use an update version of the same dataset used in the first pilot iteration, incorporating any recent changes in food regulations and relevant legal texts.</p>
Materials	<p>Training Materials: The training material will include digital files (in the form of a slide deck) that provide participants with the following information:</p> <ul style="list-style-type: none"> (a) The problem statement and key information for the relevant use-cases, including our approach to automating the summarization of complex regulatory texts to improve efficiency and accuracy in compliance management. (b) Key metadata of the SGS dataset used to train the AI models, including details about the number of documents, the languages covered, and the summary types (manual and scraped). (c) Figures and statistics illustrating the lengths of documents and summaries, accuracy of the model-generated summaries, and examples of successfully summarized regulatory texts. (d) An overview of how the model is finetuned using machine learning techniques to handle multi-lingual regulatory texts and ensure that summaries maintain critical legal details and nuances. <p>Additionally, we will provide to all participants walkthrough videos that showcase the use of the interactive chatbot and the EFRA Marketplace / SGS Digicomply product. The participants will be guided on how to use create an account, log in, test the interactive chatbot.</p> <p>Evaluation Materials: Participants will complete a structured questionnaire to evaluate their experience with the chatbot. This will include both quantitative and qualitative questions to assess various aspects of the tool:</p> <ul style="list-style-type: none"> • Ease of Use: <i>How easy was it to navigate and interact with the chatbot?</i> <ul style="list-style-type: none"> ○ Rating scale (1-5): Very difficult – Very easy ○ Open-ended question: <i>What specific aspects of the interface did you find challenging or easy to use?</i> • Quality of Summaries: <i>How would you rate the quality and accuracy of the summaries generated by the chatbot?</i> <ul style="list-style-type: none"> ○ Rating scale (1-5): Poor – Excellent ○ Open-ended question: <i>Were there any key points in the regulatory text that the summary missed or misinterpreted?</i>

	<ul style="list-style-type: none"> • Citations and References: <i>How useful did you find the chatbot's ability to cite specific sections of the regulatory text?</i> <ul style="list-style-type: none"> ○ Rating scale (1-5): Not useful – Very useful ○ Open-ended question: <i>Were the citations accurate and helpful in understanding the context of the regulatory obligations?</i> • Relevance of Responses: <i>How relevant were the chatbot's responses to your queries?</i> <ul style="list-style-type: none"> ○ Rating scale (1-5): Not relevant – Highly relevant ○ Open-ended question: <i>Did the chatbot provide relevant summaries and actionable insights based on your questions?</i> • Time Saved: <i>Did the chatbot help you save time in reviewing and interpreting regulatory texts compared to your previous methods?</i> <ul style="list-style-type: none"> ○ Rating scale (1-5): No time saved – A significant amount of time saved ○ Open-ended question: <i>How much time do you estimate you saved on average when using the chatbot?</i> • Overall Satisfaction: <i>How satisfied are you overall with the chatbot's ability to assist in regulatory compliance tasks?</i> <ul style="list-style-type: none"> ○ Rating scale (1-5): Very unsatisfied – Very satisfied ○ Open-ended question: <i>What improvements would you suggest for the chatbot to enhance its functionality?</i> <p>Especially for the second pilot cycle, and for selected participants, we will focus also on exploitation related questions. The requested feedback will include, among others, the following questions:</p> <p>(a) How interested are you to adopt AI-driven tools within your organization?</p> <p>(b) What potential barriers do you foresee in this endeavor?</p> <p>(c) In what ways could the EFRA assets be further developed to better meet the needs of your company?</p> <p>(d) Would you be interested in collaborating further for the deployment of the demonstrated EFRA tools in your organization?</p>
KPIs	<p>Repeat Usage: >50% of users return to use the chatbot at least 3 times during the pilot period</p> <p>User Satisfaction Rate: > 75% satisfaction rate based on post-pilot feedback from participants</p> <p>Chatbot Usability and Accessibility: >70% of users rate the chatbot's interface and interaction flow as easy to navigate and use</p> <p>Users involved: >25 users with commercial interest on this module</p> <p>Follow-up actions: >5 demo requests from private organisations to drive commercial exploitation of the EFRA assets.</p>

5 Conclusions

In conclusion, this EFRA project deliverable provides a detailed and structured approach to piloting and evaluating a variety of use-cases, ensuring that the tools developed are not only technically proficient but also practically applicable in real-world scenarios. By methodically outlining the two-phase evaluation process, detailing the specific activities and timing, and integrating the outcomes into the broader project objectives, this document serves as a comprehensive guide. It charts the course for the successful implementation and refinement of EFRA tools and sets a precedent for future initiatives aimed at enhancing decision support systems in the realm of food safety and risk prevention.