AI·PROFICIENT

Artificial intelligence
for improved production efficiency,
quality and maintenance

Deliverable 6.1

D6.1: Validation methodology, ethical and acceptance criteria

WP6: Use case evaluation and ethical considerations

T6.1: Validation methodology, ethical and acceptance criteria.

Version: 2.0 (revised version)

Dissemination Level: PU

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957391.

Table of Contents

List of Tables

Disclaimer

This document contains a description of the AI-PROFICIENT project work and findings.

The authors of this document have taken any available measure for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any responsibility for actions that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of the AI-PROFICIENT consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 28 Member States of the Union. It is based on the European Communities and the Member States cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors (http://europa.eu/).

AI-PROFICIENT has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 957391.

Title: D6.1: Validation methodology, ethical and acceptance criteria

Lead Beneficiary:	IBER
Due Date:	31/10/2022 (revised version; initial one was submitted M18)
Submission Date	31/10/2022 (revised version; initial one was submitted M18)
Status	**Final** ~~Preliminary~~ ~~Draft~~
Description	Definition of the requirements and the Key Performance Indicators for each 8 use cases.
Authors	Pedro de la Peña (IBER), Alexandre Voisin (UL), Kerman Lopez de Calle (TEK), Julien Hintenoch (CONTI), Sirpa Kallio (VTT), Christophe Van Loock (INEOS), Katarina Stanković (IMP), Dea Pujić (IMP), Vasillis Spais (INOS), Karen Fort (UL); Marc Anderson (UL)
Type	Report
Review Status	~~Draft WP Leader accepted~~ PC + TL accepted
Action Requested	~~To be revised by partners~~ ~~For approval by the WP leader~~ ~~For approval by the Project Coordinator & Technical Leaders~~ **For acknowledgement by partners**

VERSION	ACTION	OWNER	DATE
0.1	First Draft	IBER	10/02/2022
0.2	First Complete version	IBER	12/04/2022
0.2	Update & Final version for revision	IBER	22/04/2022
1.0	Final version	IBER	25/04/2022
2.0	Revised version	IBER	31/10/2022

Executive Summary

The content of this revised version of D6.1. (requested in the 2nd review report) is similar to the initial D6.1. version (except executive summary). No changes have been made.

Indeed, some weeks after receiving the 2nd review report, it was planned, at M24, to submit final version of D6.1. in terms of D6.6. The main scope of D6.6. was of course to improve the content of D6.1. mainly by adding material related to the two major reasons for which D6.1. was not accepted: No KPIs identified, Not sufficient information is provided about the validation methodology and its connection to WP1 requirements.

So, it was approved by EU that details of the responses to the D6.1. issues highlighted in the 2nd review report must be integrated in the D6.6. (D6.1. remains the same). This decision makes it possible to develop a content of D6.6. which will therefore be a real improvement of D6.1. (which would not really have been the case, if D6.1. had been deeply modified).

Initial Executive Summary

Deliverable D6.1 is a public document of the AI-PROFICIENT project delivered in the context of WP6 (Use case evaluation and ethical considerations.), and more specifically T6.1 (Validation methodology, ethical and acceptance criteria), relating to the creation of a validation methodology. This methodology intends to allow the establishment of objective measurement criteria for the results obtained in AI-Proficient. These criteria will make it possible to measure in complementary ways the results obtained in the different use cases that have been developed. To meet this objective, the task has worked based on the information gathered mainly in WP1 ("Pilot site characterization, requirements and system architecture") and more specifically in D1.4 (“Project requirements and performance assessment KPIs") which incorporates the list and description of the different user requirements related to the eight use cases.

The work of the current task, which has generated this deliverable, is to propose a generic methodology to measure the degree of compliance with different AI modules deployed in an industrial facility.

This deliverable will serve as a basis for the rest of the WP6 activities. In the following activities, the proposed methodology will be applied to the different tasks of the work package. It will be applied at the use case level, which will be carried out in T6.2 “Use case validation analysis and reporting”. The measurement of the interaction of applications with the user will be detailed in Task 6.3: “Qualitative assessment of user experience and feedback”. The ethical aspects will be developed in Task 6.4“Instantiation of HLEG guidelines and ethical recommendations”.

This deliverable will be followed by a second and more detailed version of it, in which the progress of the work on this task will be compiled.

1 Introduction

This deliverable aims to create the AI-PROFICIENT project validation methodology. Although it will only be used within this project, it is intended to serve as a basis for future projects where it is necessary to objectify the impact of AI in productive industrial environments.

The impacts to be measured will be divided in turn into the following 3 main typologies.

- **Production level.** Outcomes such as production performance and product quality will be evaluated at this level. This evaluation can be carried out at the different stages of a product. This will involve being able to cover from the product engineering phase, through the introduction of new products/testing on production lines, as well as the planning and operation of production lines. This level can also be measured at two levels.
    - The first level is the measurement of the impact of a use case at a **specific location** in the production line/facility. It will be a use case by **use case** approach.

      Within this typology, two sub-levels will be measured
        - **Qualitative** criteria that will be measured subjectively.
        - **Quantitative** criteria that will be measured objectively.
    - The second level will be related to the collective impact of several use cases on the same production line/ facility. This level will measure the impact **as a whole.**
- **User-level experience.** It will define the data to be collected, the means to collect it, the roles from which the data should be collected, e.g., operators, plant personnel, etc., and when it should be collected.
- **Ethical approach leve**l This section will cover both the ethical impact of the results in the workplace and the compliance with the ethical advice given to the different use cases.

Different formulas will also be designed to allow a comparative evaluation of the numerical results of the three levels of evaluation to homogeneously assess the real impact of each use case, both level by level and as a whole.

As mentioned above, the outcome of this task will be a harmonized validation methodology capable of providing an impact assessment procedure at different levels of the application of

an AI-based solution such as AI-PROFICIENT to both the project plants and possible future industrial cases.

2 Measuring the impact on a use cases basis

This section of the methodology aims to establish the groundwork for measuring the performance of AI-Proficient at the use case level. This evaluation will be measured at different levels. Qualitative evaluation of the result, quantitative evaluation of the result, and measurement of the IA models must be measured numerically. In addition, a calculation of the result obtained at the global level of each use case will be established.

2.1 General information

For each of the use cases, the following information will be collected

- **Description** of the use case identifying the current and target situation.
- **Maturity to be achieved**: Initial development validation, experimental (non-functional) pilot, limited pilot, final roll-out.
- **Goal of the use case**: Concise and, if possible, numerical description of the purpose of the use case

2.2 Qualitative evaluation of the result

The qualitative results show the perception that the different users of an application have of it. Although they are subjective assessment criteria, they will show how users perceive the impact of the developments on their daily work.

Qualitative QID	*ID*	*Meaning*
Qualitative	QUA	Qualitative
Identification of the industrial site	C	Continental
	I-C	Ineos Cologne
	I-G	Ineos Geel
Identifying ID	ID X	Number

Table 1 Identifiers for qualitative results

An instance of the following table will be filled in with information about the qualitative results at different times: first deployment, successive deployments, after training of workers. This will make it possible to measure the impact that changes in the use case have on the users'

perception concerning the use case. It will be collected when and by whom the information is collected.

The rating of the results will range from

- Strongly disagree -1
- Disagree 2
- Neutral **3**
- Agree **4**
- Strongly agree **5**

ID	Description	Result
	Do you think that AI displays helpful information for your job?	
	Do you think the AI provides information at an optimal periodicity?	
	Do you think AI provides enough information in advance to be useful?	
	Does the solution capture the feedback from the user?	
	Does the solution display information to the relevant user in an understandable way?	
	Do you believe that AI provides accurate information?	
	Does the proposed solution reduce the number of corrective actions to be taken by the operators?	
	Do you think that the information provided by AI improves the process in any significant way?	
	Do you think that the recommendations provided by AI could induce a negative impact on another segment of the production process?	

	Do you think that the information provided by the AI helps to understand the functioning of the line better?	
	Does the system offer explanations of the recommendations so that trustworthiness in the AI system is ensured?	

Table 2 Qualitative Evaluation Result

2.2.1 Qualitative formula

Since the final objective of the methodology is to establish an objective criterion of the success obtained, the use of different formulas is proposed. They allow measuring the progress achieved in a certain area, as well as in its whole.

For the qualitative compliance percentage calculation, the following method shall be used:

Qualitative Result Value =

(QUA 1 Result *1/NQ)+ (QUA 2 Result */NQ)+…+ (QUA 10 Result */NQ)

*NQ number of questions

2.3 Quantitative evaluation of the result

2.3.1 General information

This section will collect the result of the progress achieved by the use case about the KPIs of industrial parameters to be improved (Speed, quality, ...). In the case of AIProficient, these KPIs were detailed in Deliverable D1.4.

Quantitative ID -QE	*ID*	*Meaning*
Quantitative	QUT	Quantitative
Identification of the industrial site	C	Continental
	I-C	Ineos Cologne
	I-G	Ineos Geel
Use Case	UC X	Number
Identifying ID	ID X	Number

Table 3 Identifiers for Quantitative Evaluation

Additional Information should also be collected to provide context for the collection process itself. This **context informatio**n is

- **When has this variable been collected**? The same use case can be deployed multiple times, so it is necessary to keep track of the results to be able to see the evolution of the use case.
- **Have there been any changes in the plant that may affect the KPI since the Baseline was measured?** If the answer is yes, a new baseline will need to be collected, and a new historical series of results will need to be started.

The weight variable in the following table will be a figure that allows adjusting the importance of compliance of a certain KPI against the rest.

ID	Description	Reference From D1.4	Target	The current value of the KPI /Unit	Final measurement	% of achievement	Weight

Table 4 Quantitative results information

2.3.2 Quantitative formula

The calculation of the percentage of improvement for the current state of the process affected by the use case will be carried out according to the following method:

- Each QUTID X is assigned a weight out of the total 100 in percentage
- Each QUTID X is assigned a value as an (% of achievement.
- The Weights and Values for each subKPI are integrated into the following formula:

Quantitative Result Value=

(QUTID1 % of achievement Result x WEIGHT)+ (QUTD1% of achievement Result x WEIGHT)+ (QUTID3% of achievement Result x WEIGHT)+ (QUTIDN% of achievement Result x WEIGHT)

2.3.3 Evaluation of the AI Models

Because AI-PROFICIENT is a project mainly based on the use of AI for problem-solving, it is necessary to consider measuring the outcome of the different AI models performed.

Quantitative ID -QE	*ID*	*Meaning*
Quantitative AI Models	AI	Quantitative measurement at AI Model Level
Identification of the industrial site	C	Continental
	I-C	Ineos Cologne
	I-G	Ineos Geel
	DIA	Diagnostic and anomaly detection
	HEA	Health state evaluation
	PRO	Component prognostics
	HYB	Hybrid models of production processes and digital twins
	PRE	Predictive Production quality assurance
	ROO	Root-cause identification
	EAR	Early anomaly detection
	OPP	Opportunistic maintenance decision-making
	GEN	Generative holistic optimization
	LSL	Future scenario-based Lifelong self-learning system
	ETD	Explainable and transparent decision making
Use Case	UC X	Number
Identifying ID	ID X	Number

Table 5 Identifiers for AI – Models results

Additional Information should also be collected to provide context on the collection process. This **context informatio**n is

- **Date** When it is carried out: This procedure can be applied at different stages in the life of the process, from engineering to series production... and it can be deployed at different levels of maturity.
- **Evaluation criteria** Standard ML classification metrics (e.g., accuracy, F1-score, MSE, etc.), that will be calculated upon available data from the production line

The last of the instances in the table will be considered as the final result.

ID	Use Case	Final Or in a pipeline	Evaluation criteria

Table 6 AI models results information

2.4 Evaluation of User Experience

2.4.1 General information

The objective of this part of the methodology is to repeatedly and continuously measure the perceptions of the users of the application and their interaction with it. The criteria to be measured will be usability, usefulness, learning curve, etc.

To carry out these measurements, a variety of surveys will be carried out among the different operators that interact with the developments. These questions will focus on the impact of the developments on day-to-day work, the user-friendliness of the developments, the usefulness of the developments in solving day-to-day problems, etc. Some of these questions could be:

- Does the proposed AI-based solution provide information on the estimated cause of a specific product feature?
- Are these recommendations and the estimation of the potential cause presented in an understandable and user-friendly way?

Additional Information should also be collected to provide context on the collection process. This **context informatio**n is:

- **Frequency of collection**. A priori these questions will be asked the first time once the solution is deployed in an industrial process. Once the first survey has been carried out, other surveys will be carried out periodically (monthly, bimonthly, quarterly). These repeat surveys will determine whether the implementation of corrective actions

improves user perception in case of negative feedback after a survey and whether the changes improve user use of the solution.

- **Which Tools are used to display the survey**? Mail, web.
- **TimeLine** it is necessary to define when the different phases of the survey will be carried out. These phases are
    - Invitation
    - Survey Phase
    - Evaluation Phase
    - Results Sharing At this stage, it is necessary to define the template of actions and the information to share the results.
- **Actions**. The results of the surveys will be communicated to selected roles within a set timeframe. Where the survey reveals user dissatisfaction, the results will be complemented with an action plan to improve the user experience with the solution.
- **Date** when it was measured
- **Role** Although it is not necessary to know who has filled in the survey, it is advisable to know the role of that worker within the plant (operator, maintenance, quality, etc.).

Role-ID	Rol Description	Description of the interaction

Table 7 Roles involved

The following table shows which codes will be used to construct the identifiers.

UX ID	*ID*	*Meaning*
User Experience	UX	User experience
identification of the industrial site	C	Continental
	I-C	Ineos Cologne
	I-G	Ineos Geel
Identifying ID	ID X	Number

Table 8 Table of identifiers for User Experience

The following table will present the information on the evaluation of each iteration used in the use case.

ID	The stage at which the test was carried out :	Summary of the result	Numeric value of the result (0-10)	Weight
	~~Preliminary test, engineering phase test, real execution test~~			

Table 9 Evaluation of user experience

2.4.2 User Experience Formula

The calculation of the degree of usability of use will be carried out according to the following method:

- Each interaction element is assigned a weight out of the total of 100 as a percentage
- Each interaction element is assigned a numeric value as a result (0 to 10).
- The weights and results are integrated into the following formula

Human Feedback Value =

(UXID1 % of achievement Result x WEIGHT)+ (UXID1% of achievement Result x WEIGHT)+ (UXID3% of achievement Result x WEIGHT)+ (UXIDN% of achievement Result x WEIGHT)

2.5 Ethical approach

2.5.1 General information

This chapter of the methodology aims to determine the degree of ethical impact that AI solutions have on operators and their working practices. The approach to be adopted in Deliverable D6.1 will be designed to be used by industrial partners.

On the other hand, the development of Deliverable 6.4 will not be limited to completing this questionnaire. Still, it will include a more detailed approach to the ethical impact of AI in real-life scenarios. It will aim to support ethicists who may be involved in future projects.

The following table shows which encodings will be used to construct the identifiers, which will allow distinguishing the different elements of the tables.

Eth ID	*ID*	*Meaning*
Ethics	ETH	Ethics
identification of the industrial site	C	Continental
	I-C	Ineos Cologne
	I-G	Ineos Geel
Group	GAI	General AI/Operator(s) interaction
	ERRH	Error Handling
	WkL	Identification and minimization of (additional) workload
	IN	Facilitate interaction/engagement with AI system
	EtbD	Ethics by Design Developer and Industrial Partner Engagement
Identifying ID	ID X	Number

Table 10 Table of identifiers for ethical outcomes

The Ethical questionnaire will be divided into four main groups, which will make it possible to assess the different ethical impacts that the deployment of the project will have on

If necessary, this information can be collected on more than one occasion for the same technological development, reflecting the impact of possible deployments of the technical developments in the different phases of a specific industrial process, such as design, engineering, production, etc.

Not all questions apply to all situations in an industrial process. Therefore, during T6.4 they will be adapted to the casuistry of each development by filling in those that are deemed necessary.

Additional Information should also be collected to provide context on the collection process.

This **context informatio**n is

- **Date** When it is carried out: This procedure can be applied at different stages in the life of the process, from engineering to series production and it can be deployed at different levels of maturity.
- **Responsible Partner(s)** who are the partners involved in the development

2.5.2 General AI/Operator(s) interaction

Total WEIGHT of this category 1: X%

ID	Description	Result (Y/N)	Method Used
	Are the limits of the AI and the operators' actions clear?		
	Is there more than one human role involved? If so, has the chain of responsibility been clearly defined?		
	Is it defined who/when/how receives the information from the AI system?		
	Is it defined what is the degree of flexibility of the operator?		
	If not, is the operator expected to always follow the AI approach?		
	Are there situations where the default control is only human, e.g., alerts?		
	Have the risks of the user giving a default acceptance been assessed?		
	Has it been considered who (e.g., operator, process engineer, maintenance) is best placed to undertake the new AI-related task based on the time and space considerations of the work context?		
	Where used, have the targets of control concepts been specified, e.g., if HITL who is the human?		

	When the AI use is exploratory and engaging processes that are beyond operator/engineer human capacity, are the related limitations to responsibility formally clarified?		

Table 11 4.1General AI/Operator(s) interaction

2.5.3 AI Errors handling

Total WEIGHT of this category 2: X%

ID	Description	Result	Method Used +
	Is there an AI error-handling protocol?		
	If so, has it been defined which role should manage each step of the process?		
	Is it defined which are the guidelines to continue the production process in case of an AI error?		
	Is it defined what feedback the human should give to the system in case of an error?		
	Is it defined how to handle incorrect feedback from the operator to the system as the origin of the failure?		

Table 12 General AI/Operator(s) interaction

2.5.4 Identification and minimization of (additional) workload

Total WEIGHT of this category 3: %

ID	Description	Result (Y/N)	Method Used +
	Has a maximum reaction time been defined for the testing period / normal working period?		
	Has it been determined which role is the end-user of the tool?		

	Has the additional workload of using the tool been estimated?		
	Has resistance to the use of AI been assessed and measured?		
	Have measures been considered to minimize this resistance?		
	Is the user forced to accept the outcome of the tool?		
	Have the operator's previous workload/task expectations been formally adjusted in view of additional time required for new tasks		
	Has it been clarified who will undertake processing tasks for AI training and feedback: e.g., marking images?		

Table 13 Identification and minimization of (additional) workload

2.5.5 Facilitate interaction/engagement with AI system

Total WEIGHT of this category 4: X%

ID	Description	Result (Y/N)	Method Used +
	Has a phased deployment approach been considered?		
	Will operator involvement be gradual and phased?		
	Will there be a specific training period?		
	If the system involves the use of specific hardware, is there an assessment of the ergonomic impact of its use?		
	Are the interfaces redundant?		
	Has fatigue in the use of automatic decision-making tools been assessed? Have any mitigation measures been considered?		

	Have choices of XAI been tailored to the primary user?		

Table 14 Facilitate interaction/engagement with the AI system

2.5.6 Ethics by Design Developer and Industrial Partner Engagement

Total WEIGHT of this category 5: X%

ID	Description	Result (Y/N)	Method Used +
	Have tech developers worked directly with operators from the prototype stages to understand their needs in terms of HMIs and XAI? (Y/N)		
	Do diagrams and figures in Deliverables specify which people are carrying out which tasks? (Y/N)		
	Has written content avoided anthropomorphizing the AI?		
	Has the work team cohesion been monitored after AI integration?		
	Have operators and process engineers been formally notified about how their roles will change after AI integration?		

Table 15 Ethics by Design Developer and Industrial Partner Engagement

2.5.7 Ethical approach formula

The calculation of the percentage of compliance with the ethical recommendations will be carried out according to the following method:

Ethical approach value=
(Category1 [(ETH_IDX Result) …..(ETH_IDX Result)]* (Total WEIGHT of this category 1) +
(Category2 [(ETH_IDX Result) …..(ETH_IDX Result)]* (Total WEIGHT of this category 2) +
(Category3 [(ETH_IDX Result) …..(ETH_IDX Result)]* (Total WEIGHT of this category 3) +
(Category4 [(ETH_IDX Result) …..(ETH_IDX Result)]* (Total WEIGHT of this category 4) +
(Category5 [(ETH_IDX Result) …..(ETH_IDX Result)]* (Total WEIGHT of this category 5)]

The sum of all total Weight must be 100%

2.6 Use Case formula

The formula for the global estimation of the impact of the use case on the production process is detailed hereafter. Each use case shall adjust the formula to its specific needs. The general formula will include the result of the 4 big partial modules of the methodology, which are the following:

- Qualitative Results
- Quantitative Results
- Human Feedback
- Ethical approach

All these results, which will have a value between (0-100), will additionally have a coefficient to adjust the impact of this partial result on the final result.

The final result will also be in the range of 0-100.

Overall Estimation = [[Qualitative Result Value * Qualitative result Coefficient] +

[Quantitative Result Value * Quantitative result Coefficient] +

[Human Feedback Value * Human Feedback Coefficient] +

[Ethical approach Value * Ethical approach Coefficient]]

3 Production Level Impact of AI-PROFICIENT at site level

This section of the methodology aims to measure the impact of the application of AI-PROFICIENT results at the plant level. This approach allows measuring positive and negative impacts, which are the result of various interventions implemented.

1. The first step will be to determine what production improvement targets have been set for each site.

   The objectives that have been set for the Continental Sarreguemines site are:

   - Objective 1 System breakage: reduction by 50%
   - Objective 2 Production of scrap: reduction by 0.05%
   - Objective 3 Low-quality products: reduction by 50%
   - Objective 4 Extrusion line speed: improvement by 2%
   - Objective 5 Number of trial loops before release: improvement by 12,5%

   The target for the plant of INEOS Cologne:

- Off-spec production: 1M€ saving per annum

Targets for the plant of INEOS Geel:

- Reduction of human errors: reduction by at least 50%
- Plant availability: improvement by at least 0,5%

2. **Baseline**. To measure the combined impact of the developments made in AI-PROFICIENT when they are deployed simultaneously in an industrial plant, it is necessary to capture a baseline that reflects the current values of the variables to be optimized. For this purpose, the values corresponding to the last year of these variables will be collected and homogenized.
3. **Metering after the installation of the developments** Subsequently, the different use cases will be applied jointly. As detailed in the previous chapters of the methodology, the quantitative measurement of the impact of the different use cases on the variables to be optimized will be carried out
4. **Coefficient of adjustment**. Given that the application of the use cases will not lead to real production over a long period, it is important to count on the production technicians of the plants to extrapolate the impact over long periods of application based on the quantitative impact of the different models over time. They will determine a coefficient to adjust the impact measured during the tests to long-term periods.

Ethical ID	*ID*	*Meaning*
Site Level	SL	Site-level impact
identification of the industrial site	C	Continental
	I-C	Ineos Cologne
	I-G	Ineos Geel
Identifying ID	ID X	Number

Table 16 Table of identifiers for production-level impact

ID	Description	Target	The current value of the KPI /Unit	Final measurement	Coefficient of adjustment	% of achievement

Table 17 User Experience

4 Conclusion

In this document we have detailed a methodology that aims to measure as objectively as possible the results of the application of AI-based developments in production lines. In those cases where objectivity is not feasible, it has been decided to perform surveys that will be executed multiple times on different users, to objectify the results. In the following versions of this document, we will work on further detailing the different sections of the document as well as working statistically on the results in order to minimize the biases of the users' assessment in the final result.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 957391.