Development
Chapter

SAP AI Ethics Handbook

Handbook for applying SAP's Global AI Ethics Policy across the AI Factory Process

External

SAP.com

Authors
Camila Lombana Diaz,
Saskia Welsch,
Sebastian Wieczorek

Special Thanks to:
Claus Holzknecht, Detlef Plümper,
Jean-Christophe Pazzaglia,
Julia Schöndienst, Nicole Sanchez Cordero
Pavithra GK, Philip John, Rainer Kerth,
Vikram Nagendra.

© 2023 SAP SE or an SAP affiliate company.

Updated version: August 2023
All rights reserved
https://www.sap.com/about/legal/copyright.html

Content.

SAP Definition of AI — 0.6
Ethical AI Guiding Principles — 0.8
The AI Ethics Policy — 0.10
Red Lines — 0.12
AI Factory Process — 0.18
A Human Centered Approach — 0.20
01. Ideation — 0.24
02. Validation — 0.28
03. Realization — 0.34
04. Productization — 0.38
05. Operations — 0.42

Disclaimer

In addition to this handbook, employees should regularly check on and must adhere to other existing policies such as Data Protection and Privacy (DPP) policy as well as applicable product development standards.

This document is intended to serve as a general guide for applying the SAP Global AI Ethics Policy in everyday work and is not meant to advise, implement, or replace either existing or new processes, policies, or procedures. This guide is not intended to be comprehensive and is not a substitute for legal or data protection and privacy advice.

To the extent that a 3rd Party AI system (e.g., 'TensorFlow') is embedded in SAP solutions, the requirements laid down in this handbook shall apply to the overall SAP software solution.

Foreword.

We at SAP believe that Artificial Intelligence (AI) has great potential to create opportunities for businesses, governments, and societies. But for people to trust in AI solutions the development, deployment, use, and sale of AI systems must be governed by clear moral rules.

To ensure adherence and account-ability, SAP has made 'Trustworthy AI' a key pillar of its strategy and has committed to guaranteeing the sustainable and safe development of all AI systems. This promise SAP keeps by adhering to a governance structure, which ensures the ethical and transparent development of AI solutions.

The foundation of SAP's approach to AI Ethics is our pledge to uphold and support the Universal Declaration of Human Rights. An essential part of this is prohibiting discrimination and harassment of humans based on personal factors, such as race, ethnicity, religion, age, and gender. 'Trustworthy AI' at SAP is a pledge to safeguard our customers and users from harm, treating all individuals fair and just.

The SAP Global AI Ethics Policy ensures the development and deployment of SAP's AI systems in line with the established guiding principles and core organizational values. This AI Ethics Handbook will guide SAP employees in implementing the AI Ethics Policy during all stages of the AI development lifecycle. All policy requirements are mapped to the respective stages of SAP's AI Factory processes.

My sincere gratitude goes to all people who have helped shape Trustworthy AI at SAP and set up the structures and instruments for implementing it. This handbook contributes to these efforts and enables and supports employees in developing ethical, trustworthy AI at SAP.

I would like to encourage each development colleague to get familiar with the handbook and address any questions you might have related to the processes to ai.ethics@sap.com. Together we can ensure compliance with ethical development standards and establish SAP's position as a thought leader in this field.

Dr. Feiyu Xu
Global Head of Artificial Intelligence, SAP
Senior Vice President

SAP's Definition of AI

AI is typically defined as the ability of a machine to perform cognitive functions we associate with human minds, such as perceiving, reasoning, learning, and problem-solving.

It requires a system to correctly interpret external data, learn from such data, and use those learnings to achieve specific goals through flexible adaptation.

SAP differentiates between two types of AI systems:

Rule-Based AI

- Rule-based AI systems are characterized by the fact that their behavior is fully defined by rules created by human experts. These systems are often described as symbolic or expert systems.

Learning-Based AI

- Learning-based AI systems are differentiating themselves by the fact that humans define the problem and the goal, but the behavior, rules and relationships required for the system are learned in an automatize way. With the help of data, they train how to solve a problem and continuously adapt their function in this process.

At SAP, Risk Assessments shall be performed on such Learning-based AI systems during the development phase and subsequent phases to ensure that there are no unintended biases.

SAP Ethical AI Guiding Principles.

01.
We are driven by our values

02.
We design for people

03.
We enable businesses beyond bias

04.
We strive for transparency and integrity

05.
We uphold quality and safety standards

06.
We place data protection and privacy at our core

07.
We engage with the wider societal challenges of AI

General Guidelines.

How this handbook is structured

SAP's AI Ethics Guiding Principles, and AI Ethics Policy are presented in the first two chapters.

The "Red Lines" chapter clarifies which AI use-cases are prohibited within SAP. It includes the definition of high-risk use-cases pointing out the necessary processes for approval.

SAP's AI Factory Process chapter allows for an overview of its different stages, including the people involved in the process. Before each stage is introduced in more detail on how the AI Ethics Policy is applicable, some general guidelines are described. At the end of the handbook, you can find helpful links and sources.

SAP personnel shall approach ethical dilemmas and trade-offs related to their use via reasoned, context-relevant, and evidence-based decision making rather than intuition or random discretion.

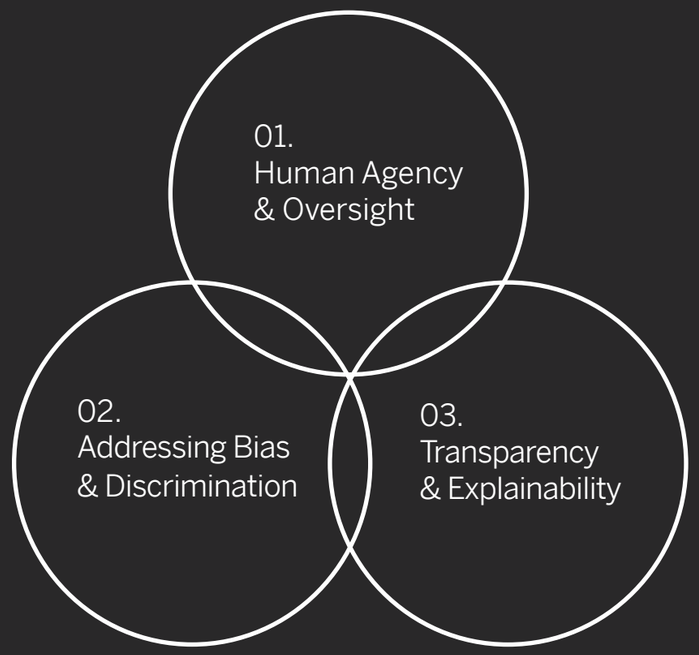
SAP employees should raise any issues related to possible policy breaches for evaluation by their immediate L1 unit, even if employees only have doubts or concerns.

If questions or concerns remain, or a decision is unable to be made by the appropriate L1 unit, then employees should inform the AI Ethics Office via an e-mail to ai.ethics@sap.com describing the use case.

The AI Ethics Policy.

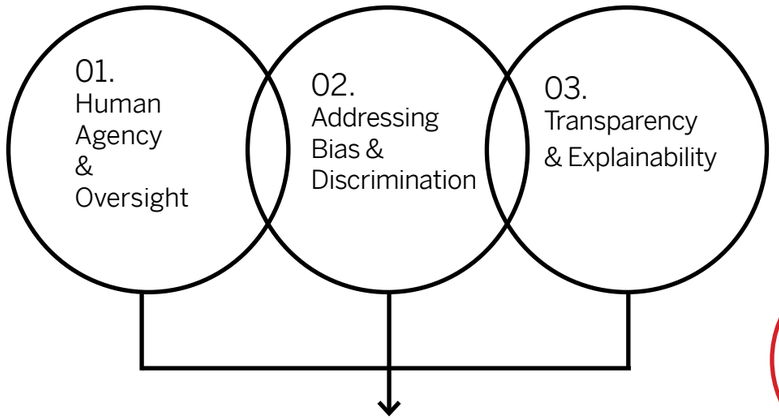
The AI Ethics Policy clarifies how SAP's Guiding Principles for Artificial Intelligence relate to AI use cases. The policy applies to SAP and all its employees worldwide. It defines intent, expectations and obligations for employees involved in development, deployment and sale of AI systems.

The Pillars of the Policy.



Product Standards

The pillars of the policy have been linked to the existing product standards to provide a transparent guidance. Please beware that any use case falling into the "Red-Line" category should be immediately dismissed.



DPP-350 (Product standard)

Human Intervention
Allow human intervention for all automated decision processes.

Transparency
Provide transparency when using AI during processing of personal data.

Bias
Ensure used AI methods do not result in inaccuracies detrimental to data subjects' rights.

Safeguards
Implement safeguards to protect personal data used to train AI models.

RED LINE
Deanony-
mization

DPP-361 (Product standard)

Apply adequate anonymization techniques

Red Lines

SAP Red Line Cases

No SAP AI use cases should fall under the so-called Red Lines defined in the SAP AI Ethics Policy. These are defined as highly unethical under the purposes for AI.

If your use case is built for these purposes, you are required to immediately stop developing, deploying, and selling that use case.

SAP's Red Lines are outlined on the right.

01. Personal Freedom

Human Surveillance
Human surveillance that is utilized for targeting individuals or groups, either by biometrics, facial recognition, or other identifiable features, with the purpose of disregarding or abusing the rights of the individuals or groups.

Discrimination
Purposes which cause individuals or groups to be discriminated against or excluded from equal access to AI's benefits and opportunities to the wider population.

Deanonymization
Deanonymization of already anonymized data which may result in the identification of individuals or groups.

02. Society

Manipulation
for deception or unfair manipulation of individuals or groups via public forums, media, or moderation of other similar uses.

Undermine Debate
systems which undermine human debate or democratic electoral systems.

Intentionally harmful impacts on users and/or those directly and indirectly affected by the system.

03. Environment

Environmental Harm
AI system development or deployment shall be conducted with minimum to no explicit damage to the environment.

High Risk

SAP High Risk Cases

SAP classifies AI use cases as high-risk under certain circumstances. We base these criteria on what kind of AI has led to negative consequences for individuals or whole populations in the past (see, for example, Amazon's case of a discriminating AI recruiting tool).

High-risk use cases are not prohibited within SAP; however, they must first go through an assessment process of the AI Ethics Steering Committee before they can be further developed, deployed, and sold.

Please be mindful that further changes could be necessary after the assessment process of the AI Ethics Steering Committee.

Determining High Risk Cases

Processing personal data:
Does the use case process any information relating to an identified or identifiable natural person for training purposes or during productive usage? – *Use Cases with anonymized data sets or only the process of anonymizing Personal data does not qualify as High Risk Case.*

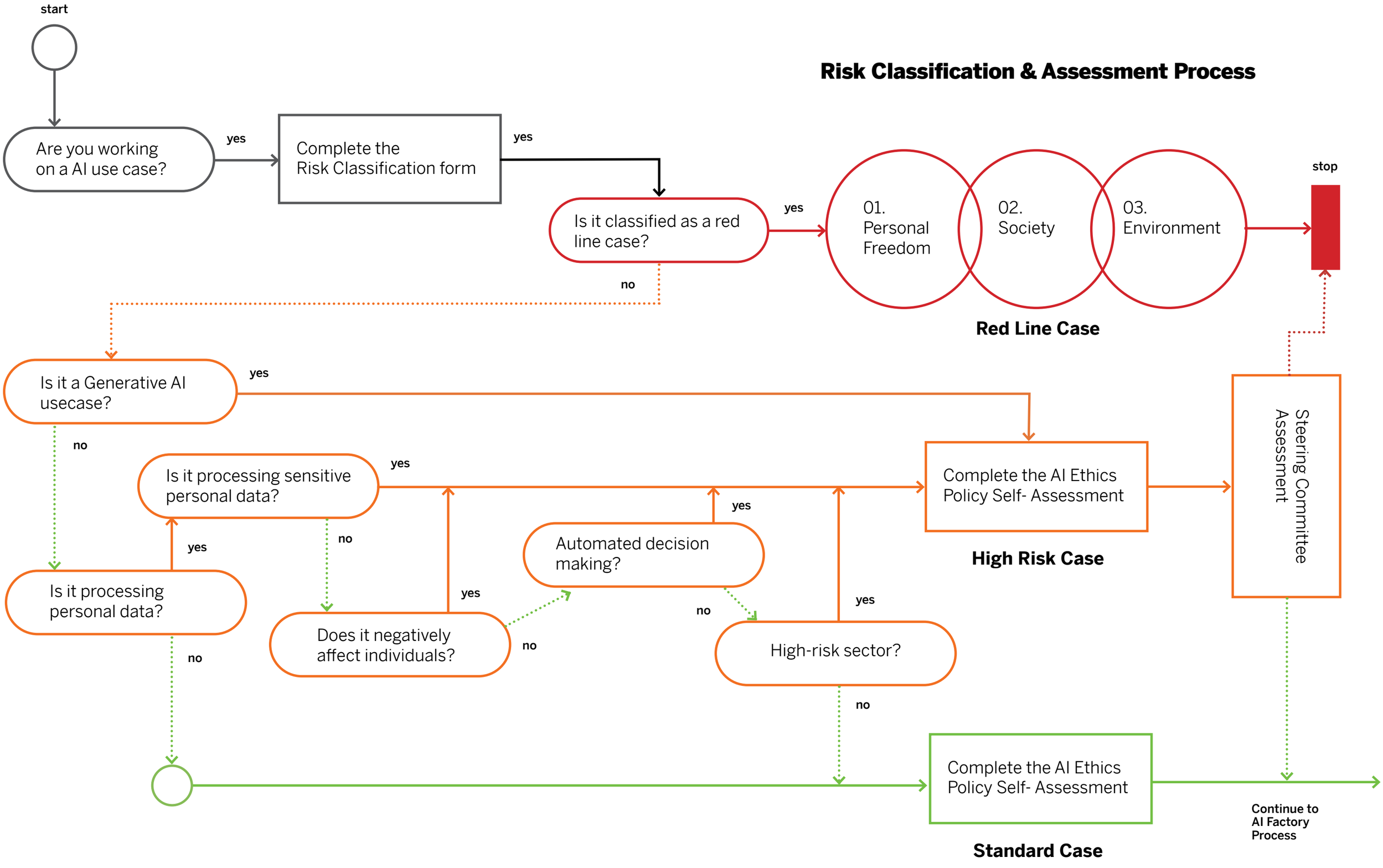
Processing sensitive personal data:
Is the use case including the processing of sensitive personal data like information on sexual orientation, religion, biometric data (including face imaging &/or voice recognition)?

Automated decision making:
Is the use case exhibiting fully or partially automated decision making? Does it include cases where no human intervention or human supervision takes place? – *This excludes any recommender system*

Negatively affected individuals:
Could the use case negatively affect the social well-being of individuals (health and safety), or intrude/restrict an individual's fundamental rights and freedom?

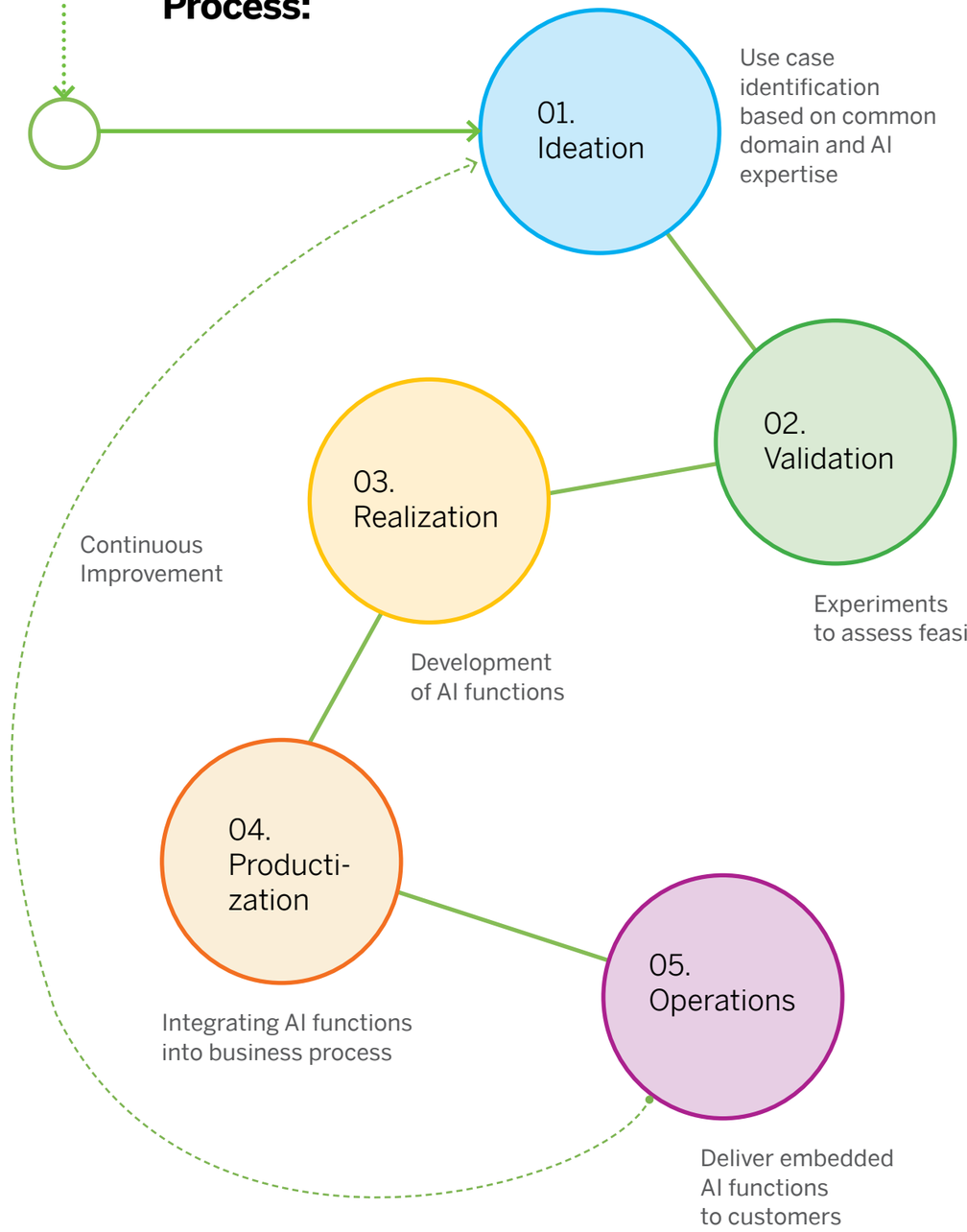
High-risk application:
Does the use case belong to one of the following domains:
E.g. Categorisation of natural persons, Management and operation of critical infrastructure, Employment/HR, Healthcare, Private services and Public services and benefits, Law Enforcement, Migration, Democratic processes?

Risk Classification & Assessment Process



SAP

AI Factory Process



A Human Centered Approach

People Involved

In SAP we believe that a greater emphasis and understanding on “Human-Centered AI” will strengthen the benefits of users and society in business, by reducing the fears and/or existential threats our users may encounter.

Thus, our AI processes must be conducted through the understanding of our user and stakeholder needs. Methods of user research, design thinking, and correctly defining user stories are mandatory aspects of our development.

This enables that our products are as close as possible to the real needs of our target groups avoiding unconscious bias, or possible unintended harms.

AI Use Case Owner

Responsible for the coordination of the delivery of an AI enabled product or feature end to end across all phases of the AI factory process.

Factory Process Activity:

AI Data Scientist

Develops, implements or applies AI methods to derive solutions to business problems that can be translated into AI functions.

Factory Process Activity:

AI Engineer

Responsible for the design, implementation and maintenance of AI functions of a product.
Works according to specifications and project plans.

Factory Process Activity:

Product Developer

Develops the business logic of a product or at least parts of it. This task includes the integration of the provided AI functions into the product context.

Factory Process Activity:

Designer

Provides the Human-Centered understanding of the product by developing a user need strategy including stakeholder, end users, and product roadmap.
Conducts: user research, flows, usability evaluation, including voice response and conversational user interaction.

Factory Process Activity:

AI OPS Engineer

Operates the AI enabled products or individual AI functions and manages life-cycle aspects like version updates.

Factory Process Activity:

User Assistance Developer

Responsible for designing, developing, reviewing, and maintaining content for product documentation, user interface messages, and conversational applications.
Collaborates with internal stakeholders to design customer-facing content and implements embedded help delivery mechanisms that improve how technical information is delivered to customers.

Factory Process Activity:

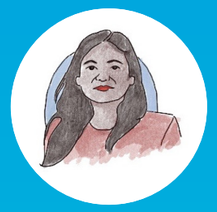
Customer

Represents the user of the AI enabled product.
As user he must be active in the ideation, validation, and operations phases.

Factory Process Activity:

01 Ideation

AI Use Case Owner



AI Data Scientist

Designer

Customer

Example:

A product owner within SuccessFactors has a deep understanding of his customers' pain points. To help them run efficiently, he identifies a business process that can be automated. Before the team builds a prototype of the AI use case, they together collect all the relevant information about the use case and discuss whether any of SAP's red lines are touched (e.g., AI used for mass surveillance) and identify risks of unethical behaviour (E.g., automated decision-making).

They fill out the use case risk assessment template and send it to the AI Ethics Office for due diligence.

We will follow & document these requirements:

A. Human Agency & Oversight.

- Before implementation, the degree of freedom of the AI system must be defined.

The decision-making degrees of freedom of the AI system must be defined.

- The target definition of the AI system must be given by a human.

- Decisions by an AI system may always be overruled by a human.

AI systems shall be subject to appropriate human oversight, and the rights and freedoms of a human shall exceed that of AI systems.

- An appropriate governance mechanism has to be chosen

Human oversight shall be achieved through an appropriate governance mechanism. This could include but not be exclusive to human-in-the-loop, human-on-the-loop, or human-in-command, and shall be decided on a case-by-case basis.

- Human oversight must be introduced where humans are directly impacted.

In situations where humans may be directly impacted by a decision made by SAP's AI system, human oversight shall be introduced to safeguard that AI system does not undermine human autonomy or introduce unintended consequences.

B. Addressing Bias & Discrimination

- Affected users should be involved in the development process

Wherever possible, developers shall seek to involve impacted/affected users to evaluate and check that outputs are diverse and discrimination free.

- The AI system should address the widest possible range of end-users.

AI software shall be user-centric, addressing the widest possible range of applicable end-users, and following relevant accessibility standards, regardless of users' age, gender, abilities, or characteristics.

Transparency & Explainability

- AI systems interacting directly with humans must be identifiable as such.

Where applicable, when interacting directly with humans (including via Conversational AI or 'Chat-bots'): AI systems shall be made identifiable as such to appropriate end users. AI systems shall be developed such that it does not encourage humans to develop attachment and/or empathy of users towards the AI system. AI systems shall clearly signal to end users that its social interaction is simulated.

- Prevent misuse of the AI system in production.

AI system development shall take into account the context and environment in which the system will operate such that, even with good intentions, no harm or misuse is likely to occur to humans when AI systems are deployed.

02 Validation

AI Use Case Owner

AI Data Scientist

Designer

Customer

Example:

The SuccessFactors team moves forward with the use case considering the instructions of the AI Ethics Steering. They are interacting with potential customers and user groups to discuss expectations in regards to the functional scope as well ethical considerations like transparency and human agency.

Based on their findings they carry out experiments and build prototypes, iteratively aligning the results with their stakeholders. The team makes sure that the data used for exploration and experimentation is balanced and representative and that the used AI methods are suitable for the desired level of explainability.

We will follow & document these requirements:

A. Human Agency & Oversight

- Before implementation, the degree of freedom of the AI system must be defined.

Before implementation, the decision-making degrees of freedom of the AI system must be defined.

- Decisions by an AI system may always be overruled by a human

AI systems shall be subject to appropriate human oversight, and the rights and freedoms of a human shall exceed that of AI systems.

- An appropriate governance mechanism has to be chosen.

Human oversight shall be achieved through an appropriate governance mechanism. This could include but not be exclusive to human-in-the-loop, human-on-the-loop, or human-in-command, and shall be decided on a case-by-case basis.

- Human oversight must be introduced where humans are directly impacted.

In situations where humans may be directly impacted by a decision made by SAP's AI system, human oversight shall be introduced to safeguard that AI system does not undermine human autonomy or introduce unintended consequences.

- Clear and simple explanations have to be provided for automated decisions.

As far as is practical, a clear and simple explanation shall be provided as to how decisions were made by an AI system used in automated decision processes.

02 Validation

B. Addressing Bias & Discrimination

- Unfair bias must be avoided.

SAP shall endeavour to achieve fairness; AI systems shall not intentionally generate unfairly biased outputs.

- Inclusive data must be used for training.

Where relevant, the data used to train AI systems shall be as inclusive as possible, representing as diverse a cross-section of the population or past situations as possible, and as free as possible from (or accounted and mitigated for) any historic or socially constructed biases, inaccuracies, errors, and mistakes.

- Measures to detect bias must be realized.

SAP shall endeavour to detect unfairly biased outputs and shall implement technical and/or organizational measures to prevent direct or indirect prejudice, discrimination, or marginalization of groups or individuals, e.g. by reducing bias in training data.

- Affected users should be involved in the development process.

Wherever possible, developers shall seek to involve impacted/affected users to evaluate and check that outputs are diverse and discrimination free.

- Data used for training and testing has to be representative and generalizable.

It shall be trained and tested on as expansive as is feasible, representative, relevant, accurate, and generalizable datasets.

- Model architectures must not include ethically objectionable features or structures.

The model architectures shall not include target variables, features, processes, or analytical structures which are unreasonable, ethically objectionable, or unable to be validated according to the principles laid out in this document.

- The AI system should address the widest possible range of end-users.

AI software shall be user-centric, addressing the widest possible range of applicable end-users, and following relevant accessibility standards, regardless of users' age, gender, abilities, or characteristics.

- The use of data for testing must comply with DPP Policy.

The use of data for the testing of AI systems shall comply with applicable data protection, and privacy laws.

- The data sets and the development processes must be documented.

The data sets and the processes that produce an AI system's decisions, including those of data gathering and data labelling as well as the algorithms used by the developed AI system, shall be documented to allow for traceability and transparency.

- AI systems interacting directly with humans must be identifiable as such.

Where applicable, when interacting directly with humans (including via Conversational AI or 'Chat-bots'): AI systems shall be made identifiable as such to appropriate end users. AI systems shall be developed such that it does not encourage humans to develop attachment and/or empathy of users towards the AI system. AI systems shall clearly signal to end users that its social interaction is simulated.

- Prevent misuse of the AI system in production.

AI system development shall take into account the context and environment in which the system will operate such that, even with good intentions, no harm or misuse is likely to occur to humans when AI systems are deployed.

02
Validation

C. Transparency and Explainability

- AI systems interacting directly with humans must be identifiable as such.

Where applicable, when interacting directly with humans (including via Conversational AI or 'Chat-bots'): AI systems shall be made identifiable as such to appropriate end users. AI systems shall be developed such that it does not encourage humans to develop attachment and/or empathy of users towards the AI system. AI systems shall clearly signal to end users that its social interaction is simulated.

AI Use Case Owner

AI Data Scientist

AI Engineer

Designer

User Assistant Developer

Example:

The SuccessFactors team has successfully built a prototype and now enters the realization stage. After the model has been trained, the data scientists and engineers explore a technical solution to provide local explainability. This way, they can provide the factors behind a particular decision by the AI system for the user.

03 Realization

We will follow & document these requirements:

A. Human Agency & Oversight

- Additional testing is necessary in case of composed AI systems.

When two or more AI systems are connected to each other or embedded within each other, additional testing and control measures should be performed for the individual AI systems as well as for the overall system.

- Clear and simple explanations have to be provided for automated decisions.

As far as is practical, a clear and simple explanation shall be provided as to how decisions were made by an AI system used in automated decision processes.

- AI systems automating decisions must be tested extensively to avoid unintended behaviour.

When human-on-the-loop models are used, appropriate extensive testing and governance shall be conducted during development and deployment to ensure the system behaves as intended by the developers and does not have any unintended behaviour, outputs, or usage.

03 Realization

B. Addressing Bias & Discrimination

- Unfair bias must be avoided.

SAP shall endeavour to achieve fairness; AI systems shall not intentionally generate unfairly biased outputs.

- Inclusive data must be used for training.

Where relevant, the data used to train AI systems shall be as inclusive as possible, representing as diverse a cross-section of the population or past situations as possible, and as free as possible from (or accounted and mitigated for) any historic or socially constructed biases, inaccuracies, errors, and mistakes.

- Measures to detect bias must be realized.

SAP shall endeavour to detect unfairly biased outputs and shall implement technical and/or organizational measures to prevent direct or indirect prejudice, discrimination, or marginalization of groups or individuals, e.g. by reducing bias in training data.

- Affected users should be involved in the development process.

Wherever possible, developers shall seek to involve impacted/affected users to evaluate and check that outputs are diverse and discrimination free.

- Data used for training and testing has to be representative and generalizable.

It shall be trained and tested on as expansive as is feasible, representative, relevant, accurate, and generalizable datasets.

- Model architectures must not include ethically objectionable features or structures.

The model architectures shall not include target variables, features, processes, or analytical structures which are unreasonable, ethically objectionable, or unable to be validated according to the principles laid out in this document.

- Where feasible, a fairness function shall be used.

Where feasible, a fairness function shall be applied to test AI systems for unbiased output.

- The AI system should addresses the widest possible range of end-users.

AI software shall be user-centric, addressing the widest possible range of applicable end-users, and following relevant accessibility standards, regardless of users' age, gender, abilities, or characteristics.

- The use of data for testing must comply with DPP Policy.

The use of data for the testing of AI systems shall comply with applicable data protection, and privacy laws.

C. Transparency & Explainability

- The data sets and the development processes must be documented.

The data sets and the processes that produce an AI system's decisions, including those of data gathering and data labelling as well as the algorithms used by the developed AI system, shall be documented to allow for traceability and transparency.

- The AI system's capabilities and limitations must be documented.

The capabilities and limitations shall be documented as part of the development process in a manner appropriate to the use case at hand. This shall include information regarding the AI system's level of accuracy (performance metric), as well as its limitations and capabilities.

- Transparency on how personal data is processed must be provided.

In alignment and compliance with applicable data protection, and privacy laws, products that use AI systems in the processing of personal data must provide transparency to the extent possible as to how the AI system was used in clear and simple language if requested by the data subject.

- For decisions about affecting humans explanations have to be provided to the data subject.

In alignment and compliance with applicable data protection and privacy law, AI systems that engage in profiling or automated decision-making must be able to provide explanations to the extent possible to data subjects upon request, describing the data segment the subject was placed into and the reasons they were placed there.

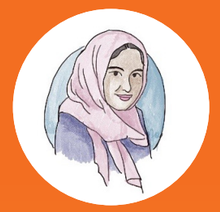
In addition, the reasons as to why the decision was made shall be provided if requested by the data subject. The explanation must be such as to provide the data subject grounds to challenge the decision.

- The methods used for development, testing and validation must be documented.

The methods used for developing, testing and validating, and the outcomes of or decisions made by the AI system shall be fully documented as part of the development process according to SAP's Global Development Policy and Product Development Standards.

AI Use Case Owner

AI Engineer



Developer

Designer

AI OPS Engineer

User Assistant Developer

Example:

The developed AI functions are now ready to be embedded into the product. Since the data sets in production will be located somewhere else than the training data sets, the corresponding documentation is refined and extended by the AI engineer. Developers and UX designers work on a user-friendly way to display local explanations for the predictions made by the AI directly in the application. Furthermore, they include a feedback control in the product, which allows users to share feedback on the quality of the AI recommendations. This feedback could be used to re-train the algorithm, if needed.

04 Productization

We will follow & document these requirements:

A. Human Agency & Oversight.

- Additional testing is necessary in case of composed AI systems.

When two or more AI systems are connected to each other or embedded within each other, additional testing and control measures should be performed for the individual AI systems as well as for the overall system.

- Clear and simple explanations have to be provided for automated decisions.

As far as is practical, a clear and simple explanation shall be provided as to how decisions were made by an AI system used in automated decision processes.

- AI systems automating decisions must be tested extensively to avoid unintended behaviour.

When human-on-the-loop models are used, appropriate extensive testing and governance shall be conducted during development and deployment to ensure the system behaves as intended by the developers and does not have any unintended behaviour, outputs, or usage.

04 Producti-zation

B. Addressing Bias & Discrimination

- Unfair bias must be avoided.

SAP shall endeavour to achieve fairness; AI systems shall not intentionally generate unfairly biased outputs.

- Inclusive data must be used for training.

Where relevant, the data used to train AI systems shall be as inclusive as possible, representing as diverse a cross-section of the population or past situations as possible, and as free as possible from (or accounted and mitigated for) any historic or socially constructed biases, inaccuracies, errors, and mistakes.

- Measures to detect bias must be realized.

SAP shall endeavour to detect unfairly biased outputs and shall implement technical and/or organizational measures to prevent direct or indirect prejudice, discrimination, or marginalization of groups or individuals, e.g. by reducing bias in training data.

- Affected users should be involved in the development process.

Wherever possible, developers shall seek to involve impacted/affected users to evaluate and check that outputs are diverse and discrimination free.

- Data used for training and testing has to be representative and generalizable.

It shall be trained and tested on as expansive as is feasible, representative, relevant, accurate, and generalizable datasets.

- Model architectures must not include ethically objectionable features or structures.

The model architectures shall not include target variables, features, processes, or analytical structures which are unreasonable, ethically objectionable, or unable to be validated according to the principles laid out in this document.

- Where feasible, a fairness function shall be used.

Where feasible, a fairness function shall be applied to test AI systems for unbiased output.

- The AI system should addresses the widest possible range of end-users.

AI software shall be user-centric, addressing the widest possible range of applicable end-users, and following relevant accessibility standards, regardless of users' age, gender, abilities, or characteristics.

- The use of data for testing must comply with DPP Policy.

The use of data for the testing of AI systems shall comply with applicable data protection, and privacy laws.

C. Transparency & Explainability

- The data sets and the development processes must be documented.

The data sets and the processes that produce an AI system's decisions, including those of data gathering and data labelling as well as the algorithms used by the developed AI system, shall be documented to allow for traceability and transparency.

- The AI system's capabilities and limitations must be documented.

The capabilities and limitations shall be documented as part of the development process in a manner appropriate to the use case at hand. This shall include information regarding the AI system's level of accuracy (performance metric), as well as its limitations and capabilities.

- Transparency on how personal data is processed must be provided.

In alignment and compliance with applicable data protection, and privacy laws, products that use AI systems in the processing of personal data must provide transparency to the extent possible as to how the AI system was used in clear and simple language if requested by the data subject.

- For decisions about affecting humans explanations have to be provided to the data subject.

In alignment and compliance with applicable data protection and privacy law, AI systems that engage in profiling or automated decision-making must be able to provide explanations to the extent possible to data subjects upon request, describing the data segment the subject was placed into and the reasons they were placed there.

In addition, the reasons as to why the decision was made shall be provided if requested by the data subject. The explanation must be such as to provide the data subject grounds to challenge the decision.

- The methods used for development, testing and validation must be documented.

The methods used for developing, testing and validating, and the outcomes of or decisions made by the AI system shall be fully documented as part of the development process according to SAP's Global Development Policy and Product Development Standards.

AI Use Case Owner

AI OPS Engineer

User Assistant Developer

Customer

Example:

When the embedded AI function is deployed the AI Ops Engineer makes sure that the data used for on-boarding the customer is representative and regularly checks for undesired bias during execution. Further the AI Ops Engineer provides potential user requests in regards to transparency and explainability to the Use Case Owner.

Because of the need for continues improvements by the AI Factory Process, one year after the successful product launch, the feedback control shows that the accuracy of the predictions provided by the AI by including two new parameters needs to improve. The AI OPS Engineer advises the development team to re-train the model including the new parameters. Before the optimized model is again built into the product, the team ensures that the model is free from any biases.

05 Operations

We will follow & document these requirements:

A. Human Agency & Oversight

- Additional testing is necessary in case of composed AI systems.

When two or more AI systems are connected to each other or embedded within each other, additional testing and control measures should be performed for the individual AI systems as well as for the overall system.

- Clear and simple explanations have to be provided for automated decisions.

As far as is practical, a clear and simple explanation shall be provided as to how decisions were made by an AI system used in automated decision processes.

- AI systems automating decisions must be tested extensively to avoid unintended behaviour.

When human-on-the-loop models are used, appropriate extensive testing and governance shall be conducted during development and deployment to ensure the system behaves as intended by the developers and does not have any unintended behaviour, outputs, or usage.

05 Operations

B. Addressing Bias & Discrimination

- Unfair bias must be avoided.

SAP shall endeavour to achieve fairness; AI systems shall not intentionally generate unfairly biased outputs.

- Inclusive data must be used for training.

Where relevant, the data used to train AI systems shall be as inclusive as possible, representing as diverse a cross-section of the population or past situations as possible, and as free as possible from (or accounted and mitigated for) any historic or socially constructed biases, inaccuracies, errors, and mistakes.

- Measures to detect bias must be realized.

SAP shall endeavour to detect unfairly biased outputs and shall implement technical and/or organizational measures to prevent direct or indirect prejudice, discrimination, or marginalization of groups or individuals, e.g. by reducing bias in training data.

- Affected users should be involved in the development process.

Wherever possible, developers shall seek to involve impacted/affected users to evaluate and check that outputs are diverse and discrimination free.

- Data used for training and testing has to be representative and generalizable.

It shall be trained and tested on as expansive as is feasible, representative, relevant, accurate, and generalizable datasets.

- Model architectures must not include ethically objectionable features or structures.

The model architectures shall not include target variables, features, processes, or analytical structures which are unreasonable, ethically objectionable, or unable to be validated according to the principles laid out in this document.

- Where feasible, a fairness function shall be used.

Where feasible, a fairness function shall be applied to test AI systems for unbiased output.

- The AI system should address the widest possible range of end-users.

AI software shall be user-centric, addressing the widest possible range of applicable end-users, and following relevant accessibility standards, regardless of users' age, gender, abilities, or characteristics.

- The use of data for testing must comply with DPP Policy.

The use of data for the testing of AI systems shall comply with applicable data protection and privacy laws.

C. Transparency & Explainability

- The data sets and the development processes must be documented.

The data sets and the processes that produce an AI system's decisions, including those of data gathering and data labelling as well as the algorithms used by the developed AI system, shall be documented to allow for traceability and transparency.

- The AI system's capabilities and limitations must be documented.

The capabilities and limitations shall be documented as part of the development process in a manner appropriate to the use case at hand. This shall include information regarding the AI system's level of accuracy (performance metric), as well as its limitations and capabilities.

- Transparency on how personal data is processed must be provided.

In alignment and compliance with applicable data protection, and privacy laws, products that use AI systems in the processing of personal data must provide transparency to the extent possible as to how the AI system was used in clear and simple language if requested by the data subject.

- For decisions about affecting humans explanations have to be provided to the data subject.

In alignment and compliance with applicable data protection and privacy law, AI systems that engage in profiling or automated decision-making must be able to provide explanations to the extent possible to data subjects upon request, describing the data segment the subject was placed into and the reasons they were placed there.

In addition, the reasons as to why the decision was made shall be provided if requested by the data subject. The explanation must be such as to provide the data subject grounds to challenge the decision.

- The methods used for development, testing and validation must be documented.

The methods used for developing, testing and validating, and the outcomes of or decisions made by the AI system shall be fully documented as part of the development process according to SAP's Global Development Policy and Product Development Standards.

2023

SAP AI ETHICS COMMUNITY