Data sovereignty AI algorithms classification for data spaces

DOCUMENT DATA

Client	**Stand.ICT Call 08**		
Titte	**Data sovereignty AI algorithms classification for data spaces**		
Date	**24-04-23**		
Author	**Alberto Abella**		
Contact	**alberto.abella@meloda.org**		
License	**Creative Commons 4.0**		
Confidentiality	**Public**		
Language	**English**		

CHANGE LOG

Version	Date	Changes	Autor
0.8	6-4-23	Draft to be reviewed	AAG
0.9	23-4-23	Draft to be reviewed	AAG
1.0	24-4-23	First release	AAG

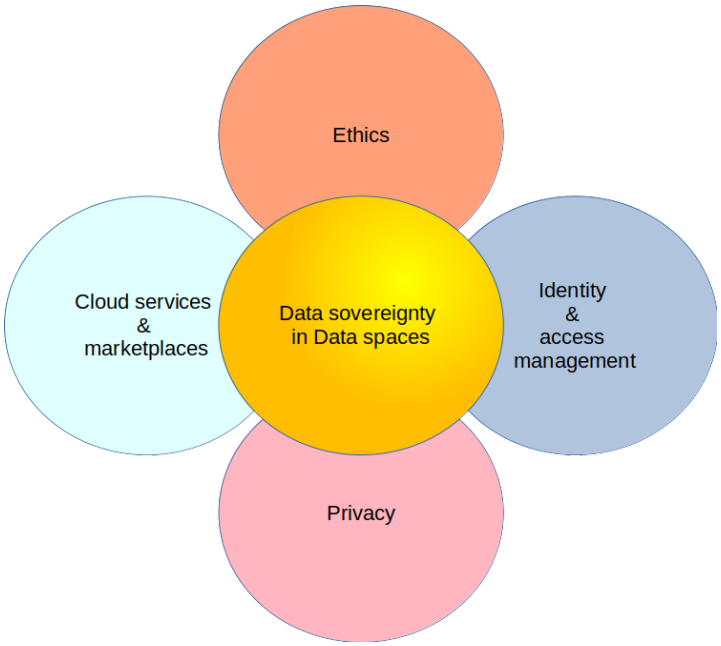
INDEX

1 Executive summary

The following report presents an analysis of the feasibility of standardizing algorithms provided as a service within data spaces. The study investigates potential sources of trust for users and proposes potential mechanisms that could be established as standard practices to instill trust. Establishing trust in algorithmic services is a critical factor in facilitating the growth of the data economy. This work has received support from StandICT under the 8th call for projects, and the licensing of this work will be assigned by StandICT.
The activities around this work include:

- A survey run between members of AI interest groups
- Interviews with experts on AI
- Connection with AI groups at different organizations
- Research on the context situation
- A proposal for a possible solution based on an independent certification, with an agile standardization approach and only after running a specific pilot for testing sustainability and market adaptation.
- Conclusions with the most remarkable points

The main findings of this study indicate that the topic under investigation is of significant interest within a dynamic market environment with multiple concurrent initiatives. It is noteworthy that the concept of data sovereignty has received limited attention, primarily intersecting with other dimensions such as privacy, access & identity management in data spaces, marketplaces, cloud services, and ethics. The topic is of considerable interest to all involved agents, but the complexity of the challenge presents organizational and technical difficulties. The study includes a couple of data models to collect critical information necessary for certification, classified into three distinct categories. A practical implementation of these proposals would help validate or evolve the findings further. It is evident that traditional standardization approaches may not suffice to address this challenge, and a complementary approach such as agile standardization is necessary.



1.1 Other data

Thanks to the experts collaborating on the survey and special thanks to Juan Tomás García, Nerea Luis, Sonia Jiménez, Mariano Blaya, Tania Marcos, Chus Garcia, Martin Bauer, and Juan Jose Hierro)

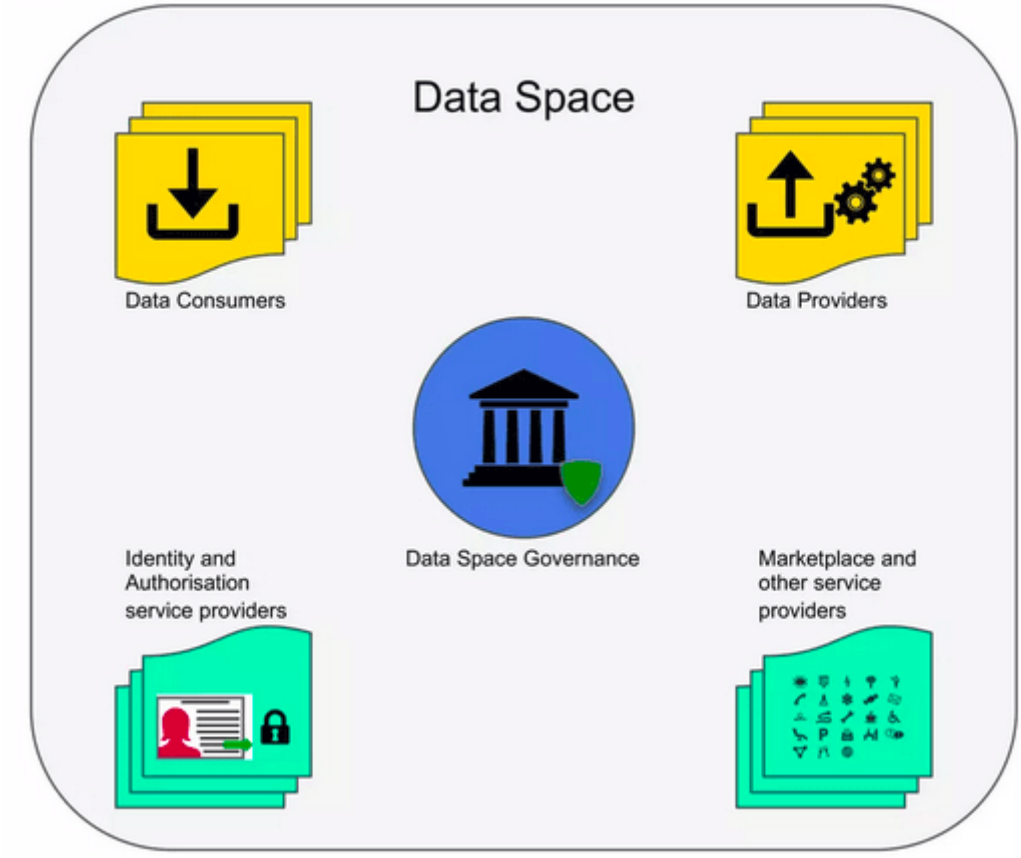
1.2 Acronyms

Acronym	**Explanation**
AWS	Amazon Web Services
DDBB	Database
DMA	Digital Market Act
DSA	Digital Services act
EAIS	European AI Strategy

2 Introduction

2.1 Scope

This study focuses on the potential standardization of AI algorithms within data spaces in the context of data sovereignty. A data space is a secure and trusted platform that enables the sharing of data among authorized data providers and consumers. In addition, it facilitates the growth of associated service marketplaces. To realize the full potential of data spaces, a robust technological framework is necessary to facilitate secure data sharing and economic transactions. A marketplace of services would further enhance the appeal of the infrastructure to providers and consumers and promote its sustainability.

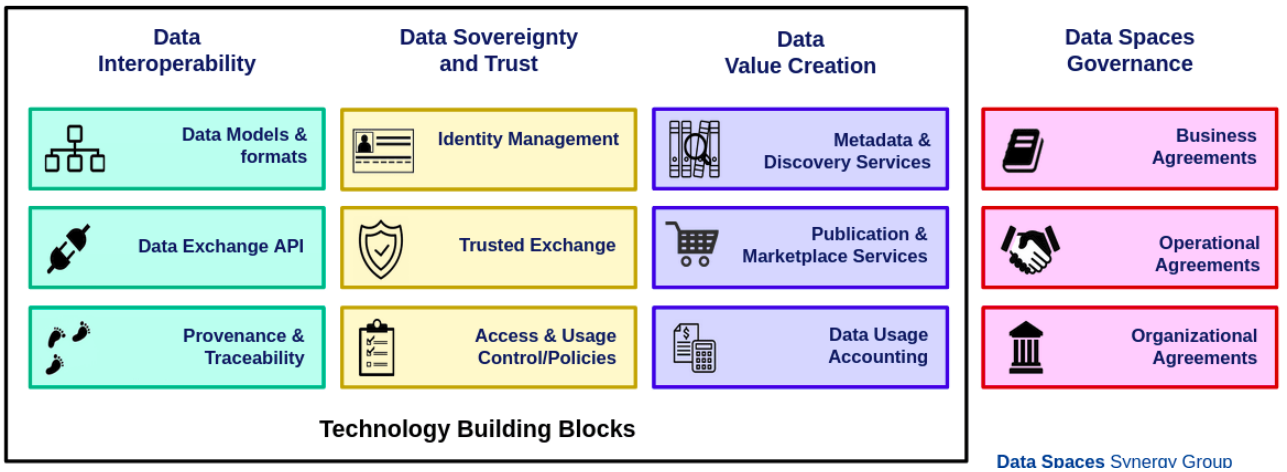


Source: European project I4trust[1]

In order for services to participate in data spaces, they must adhere to the governance rules of the data space and possess the necessary technical mechanisms for accessing, authenticating, and authorizing the data space. The next diagram includes the pink blocks on the right, which pertain to governance, as well as the yellow blocks concerning identity management. Services must

[1] https://i4trust.org/

demonstrate their identity and trustworthiness in a reliable manner, as indicated by the second yellow block, and comply with the access and usage policies outlined in the third block.



Source: Based on the basic blocks described in the EU project Open DEI

It is worth noting the potential issues addressed by the data spaces regarding the associated services. These potential issues include, for instance, the criteria to accept, reject and monitor the services in the data space.

It has benchmarked the approach in Google services[2] and Amazon services and also the AI governance framework of Singapur[3].
Besides these sources, it was also reviewed the Ethics guidelines for trustworthy AI[4](2019) from the EU thanks to the European AI Alliance[5]. Out of the 7 principles listed in these guidelines, the most relevant for the topics related to the aim of this study are:

Privacy and data governance.
Transparency.
Accountability

These guidelines are aimed at promoting a trustworthy AI and therefore exceed the scope of this analysis. For them, trust should be based on 3 elements, to be lawful, ethical, and robust. Data sovereignty, the goal of this study, is somehow transversal to these elements.

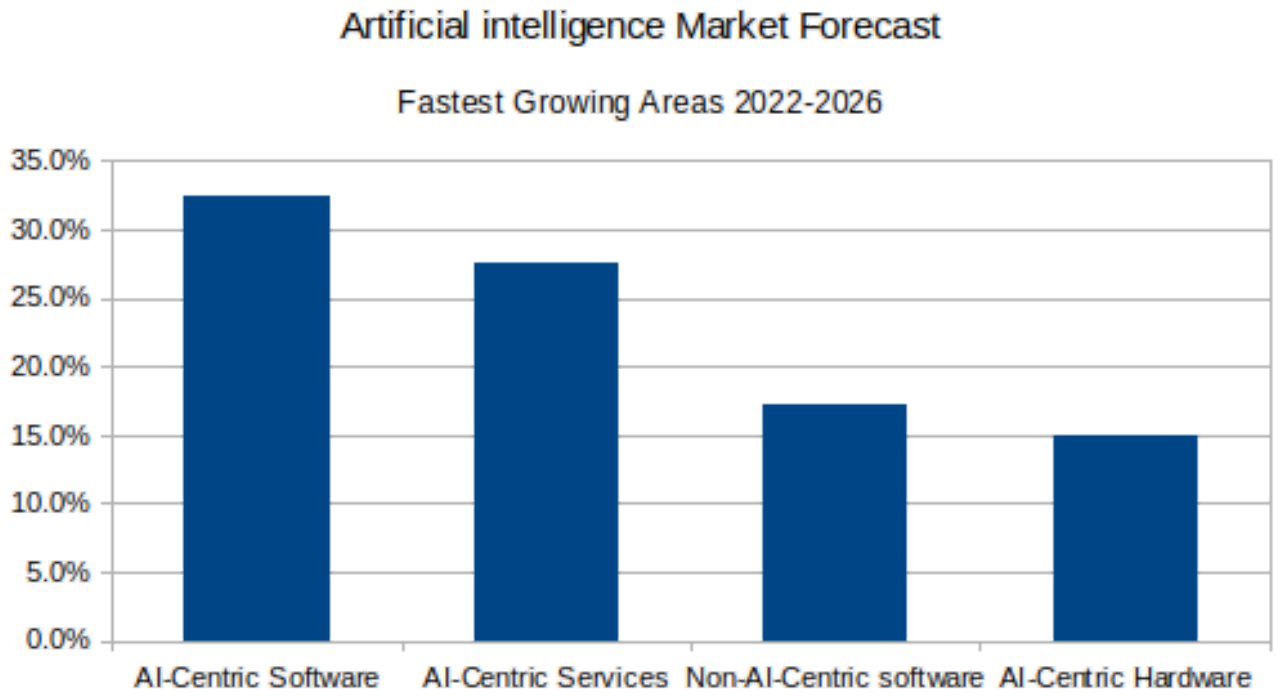
[2] Google AI principles. https://ai.google/static/documents/ai-principles-2022-progress-update.pdf

[3] https://www.pdpc.gov.sg/Help-and-Resources/2020/01/Model-AI-Governance-Framework

[4] https://ec.europa.eu/newsroom/dae/redirection/document/60419
[5] https://futurium.ec.europa.eu/en/european-ai-alliance/pages/about

2.2 AI services market expected evolution

The adoption of AI in the market is boosting rapidly. And one of the main uses is the processing of data in order to provide different services, from extracting patterns, images, trends, etc.
What will be the share of the AI services running in data spaces is part of these predictions, but, inevitably, it will be growing at a high rate like the rest of the market.



Source: Based on IDC worldwide Semiannual Artificial Intelligence tracks 2022 H1. Annual growth

2.3 List of initiatives related to data sovereignty in AI services

The study conducted an analysis of the state-of-the-art by examining several working groups and initiatives related to the topic. The input from these sources was compiled to inform the potential standardization efforts, and the most notable initiatives are summarized in the table provided. Some of these initiatives have provided valuable information to the report, as noted in the relevant sections, while others are too recent to provide information or have not yet released their findings to the public.

Institution	Initiative
StandICT	Working groups on Artificial intelligence[6]
ISO	ISO/IEC JTC 1/SC 42 Artificial Intelligence[7]
EU	European AI Alliance[8]

[6] https://www.standict.eu/discussion-groups/artificial-intelligence/267
[7] https://www.iec.ch/dyn/www/f?p=103:22:217276730919533::::FSP_ORG_ID,FSP_LANG_ID:21538,25
[8] https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance

ETSI	OCG, subgroup on AI[9]
AI	The center for governance of AI[10]
Gaia-X	Service characteristics work group is defining how to describe a cloud service like the ones provided by AI services in the marketplaces of data spaces.
AI Act	A site on the follow up of the EU AI act[11]

2.4 The questions addressed in this report

The classification of algorithms is necessary to provide trust for the users of these AI services in the marketplaces of the data spaces. The approach was simplified to answer these 3 questions:

1.-How could we trust those algorithms applied to our data?
If we can trust somehow
2.- What are the major concerns regarding using AI services applied to your data?
and then in order to deal with these concerns, what are the potential solutions by
3.- What would be a possible solution to trust in AI algorithms?
After these questions, and based on the results of the survey and the interviews, a draft on how this could be implemented and what would be the agents involved?
Although the analysis is focused on those marketplaces associated with data spaces, most of their conclusions could be applicable in very different scenarios where AI services are provided.

2.5 Alignment with existing policies

Based on the Coordinated Plan on Artificial Intelligence of EU[12] these are the current policies for AI development

- Set enabling conditions for AI development and uptake in the EU
- Make the EU the right place for excellence from lab to the market
- Ensure AI technologies work for people
- Build strategic leadership in the sectors

the main regulations applicable to these algorithms are based on the AI act[13], Digital Services Act[14](DSA) and the Digital Market Act[15].
DSA focuses on online platforms and what they should do to provide greater transparency about the way algorithms work (their internal algorithms, not the algorithms offered as services) and how they are used to recommend content or target users with advertising. This includes disclosing information about the data used to train algorithms, the parameters used to make decisions, and the impact of the algorithm on user behavior.

[9] Private link for ETSI members. https://portal.etsi.org/TB-SiteMap/OCG/OCG-AI-ToR
[10] https://www.governance.ai
[11] https://artificialintelligenceact.eu/
[12] https://digital-strategy.ec.europa.eu/en/policies/plan-ai
[13] https://data.consilium.europa.eu/doc/document/ST-8115-2021-INIT/en/pdf in progress to be voted at EU parliament is expected to be voted during 2023
[14] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2065&qid=1666857835014
[15] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R1925

2.6 Some practices around data sovereignty in cloud providers and standards related

This section compiles remarkable approaches (AWS and Google) regarding the closest terms to data sovereignty. It is included here as illustrative examples and to benchmark some of their practices in order to create the standardization. it is true that the aims of these examples are initially far from those in a standardization.

2.6.1 AWS for AI services

AWS for AI services establishes these 3 conditions for users to trust your services.:

2.6.1.1 ISO 27017 for cloud security.

ISO 27017 is a standard that provides guidelines for information security controls within cloud computing environments. Specifically, it focuses on the security of cloud services, addressing cloud-specific threats, risks, and vulnerabilities.
ISO 27017 provides guidelines for cloud service providers and cloud customers to ensure the confidentiality, integrity, and availability of information stored in the cloud. It covers a range of topics, including

1. Virtualization and multi-tenancy
2. The separation between the customer and cloud service provider
3. Access control and identity management
4. Compliance with legal and regulatory requirements
5. Incident management and response
6. Data classification and handling
7. Business continuity management
8. Monitoring and logging

2.6.1.2 ISO 27701 for privacy information management

ISO 27701 is a standard that provides guidelines for implementing and managing a Privacy Information Management System (PIMS). This standard builds upon the framework established in ISO 27001 and ISO 27002, which provide guidelines for information security management systems.
ISO 27701 specifically addresses privacy concerns and provides guidance on how organizations can effectively manage the privacy of personal information. It provides a framework for implementing and maintaining a PIMS that takes into account the requirements of various privacy regulations and laws, such as the General Data Protection Regulation (GDPR)
Some of the key areas addressed by ISO 27701 include:

1. Privacy policy and objectives: Establishing privacy policies and objectives that align with the organization's overall goals and objectives.
2. Risk management: Identifying, assessing, and managing privacy risks to ensure that personal information is protected.
3. Controls and measures: Implementing appropriate controls and measures to ensure the confidentiality, integrity, and availability of personal information.

4. Training and awareness: Providing training and raising awareness among employees, contractors, and other stakeholders on the importance of privacy and their responsibilities in protecting personal information.
5. Incident management: Establishing procedures for managing and responding to privacy breaches and incidents.

2.6.1.3 ISO 27018 for cloud privacy.

ISO 27018 is a standard that provides guidelines for protecting the privacy of personal data in public cloud computing environments. It focuses specifically on the protection of Personally Identifiable Information (PII) in cloud computing.
ISO 27018 establishes a set of controls that cloud service providers should implement to protect personal data in accordance with privacy laws and regulations. Some of the key areas addressed by the standard include:

1. Information security policies: Establishing policies for the handling of personal data in the cloud, including data processing, retention, and deletion.
2. Consent and disclosure: Ensuring that cloud customers provide informed consent for the processing of personal data and that any disclosures of personal data are made in accordance with applicable laws and regulations.
3. Data retention: Specifying the minimum and maximum periods for which personal data can be retained in the cloud.
4. Data handling: Describing the procedures for handling personal data, including collection, storage, and sharing.
5. Access controls: Establishing access controls to prevent unauthorized access to personal data in the cloud.
6. Data transfer: Specifying the conditions under which personal data can be transferred to third parties, including cross-border transfers.
7. Incident management: Describing the procedures for managing and reporting data breaches and incidents involving personal data in the cloud.

2.6.2 Google AI principles

Here there are those principles with some kind of relation to data sovereignty.

1. Be socially beneficial: With the likely benefit to people and society substantially exceeding the foreseeable risks and downsides
2. Avoid creating or reinforcing unfair bias: Avoiding unjust impacts on people, particularly those related to sensitive characteristics such as race, ethnicity, gender, nationality, income, sexual orientation, ability, and political or religious belief.
3. Be built and tested for safety: Designed to be appropriately cautious and in accordance with best practices in AI safety research, including testing in constrained environments and monitoring as appropriate.
4. Be accountable to people: Providing appropriate opportunities for feedback, relevant explanations, and appeal, and subject to appropriate human direction and control.
5. Incorporate privacy design principles: Encouraging architectures with privacy safeguards, and providing appropriate transparency and control over the use of data.
6. Uphold high standards of scientific excellence: Technology innovation is rooted in the scientific method and a commitment to open inquiry, intellectual rigor, integrity, and collaboration.

7. Be made available for uses that accord with these principles: We will work to limit potentially harmful or abusive applications.

Explanation of potential relationship between these principles and data sovereignty
1.- Be socially beneficial. It is supposed that data leakage or data misuse should be opposite to this principle.
3.- Be built and tested for safety. It is supposed that the services should be tested in terms of data sovereignty.
4.- Being accountable to people. It is supposed that we could ask for our concerns in terms of data sovereignty
5.- Incorporate privacy design principles. The control of the use of data together with the transparency should allow users to understand the actual use of this data.
7.- It is made explicit the limitation to harmful applications (like those that work differently with our data in terms of data sovereignty)

2.7 ISO activity for the creation of standards on AI

Although there is more AI related standardization activity at ISO, here are the closest standards to data sovereignty created by the group 2 of subcommittee 42.

2.7.1 Group 2 at ISO-IEC JTC-1 SC 42

Regarding the standards that it has been reviewed for the creation of the report. It was explored the works of the SC 42, WG-2 in the JTC 1[16] but none of the 4 standards planned or published cover specifically the data sovereignty of AI services.

- ISO/IEC 24668:2022(Main) Information technology - Artificial intelligence - Process management framework for big data analytics
- ISO/IEC 20547-3:2020(Main) Information technology - Big data reference architecture - Part 3: Reference architecture
- ISO/IEC TR 20547-1:2020(Main) Information technology - Big data reference architecture - Part 1: Framework and application process
- ISO/IEC 20546:2019(Main) Information technology - Big data - Overview and vocabulary

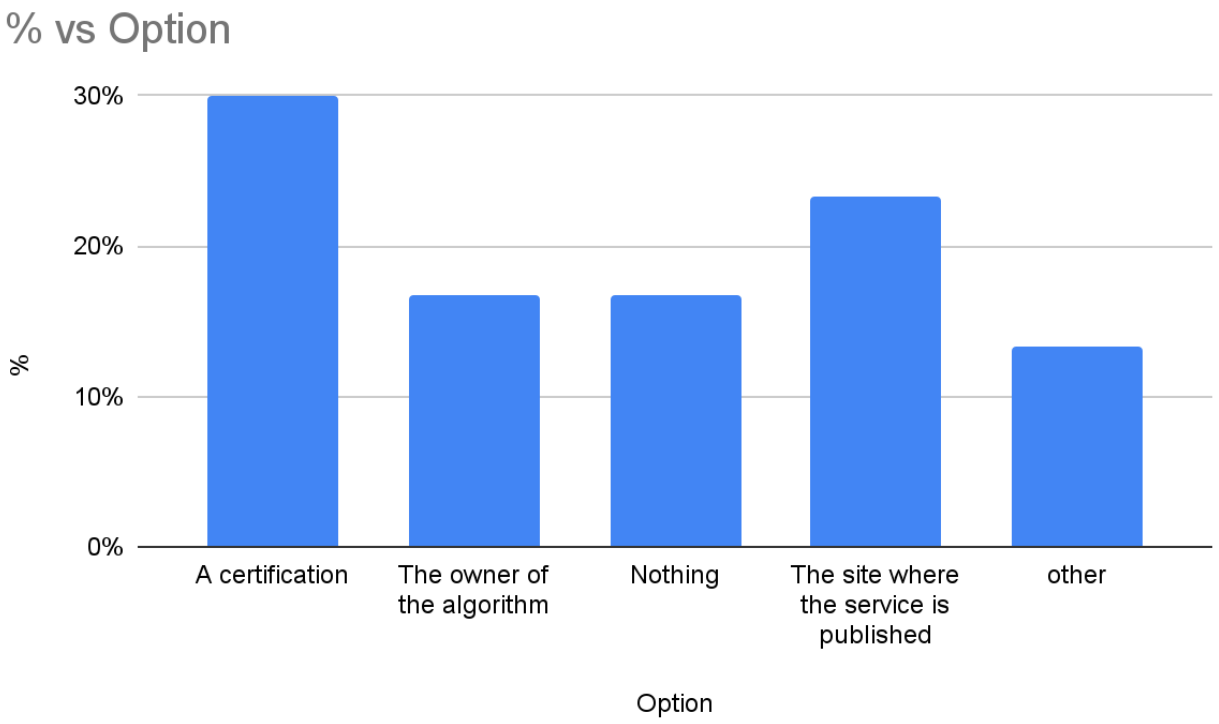
16 https://standards.iteh.ai/catalog/tc/iso/5dc4630c-b8e5-434f-88b5-d3fbe904cb31/iso-iec-jtc-1-sc-42-wg-2

3 Is it possible to standardize the algorithms in terms of data sovereignty in data spaces?

The study is centered around three primary questions. The first question pertains to the mechanisms for establishing trust in AI services with regard to data sovereignty, excluding other factors such as result reliability, bias, energy consumption, etc. The second question explores the main concerns regarding AI services in data spaces. Finally, assuming a viable solution exists, the study investigates potential solutions. It is worth noting that if no solution were found, which is not currently the case, the practical utility of this work would be limited. The study allowed for multiple answers to these questions.

3.1 The trust mechanism

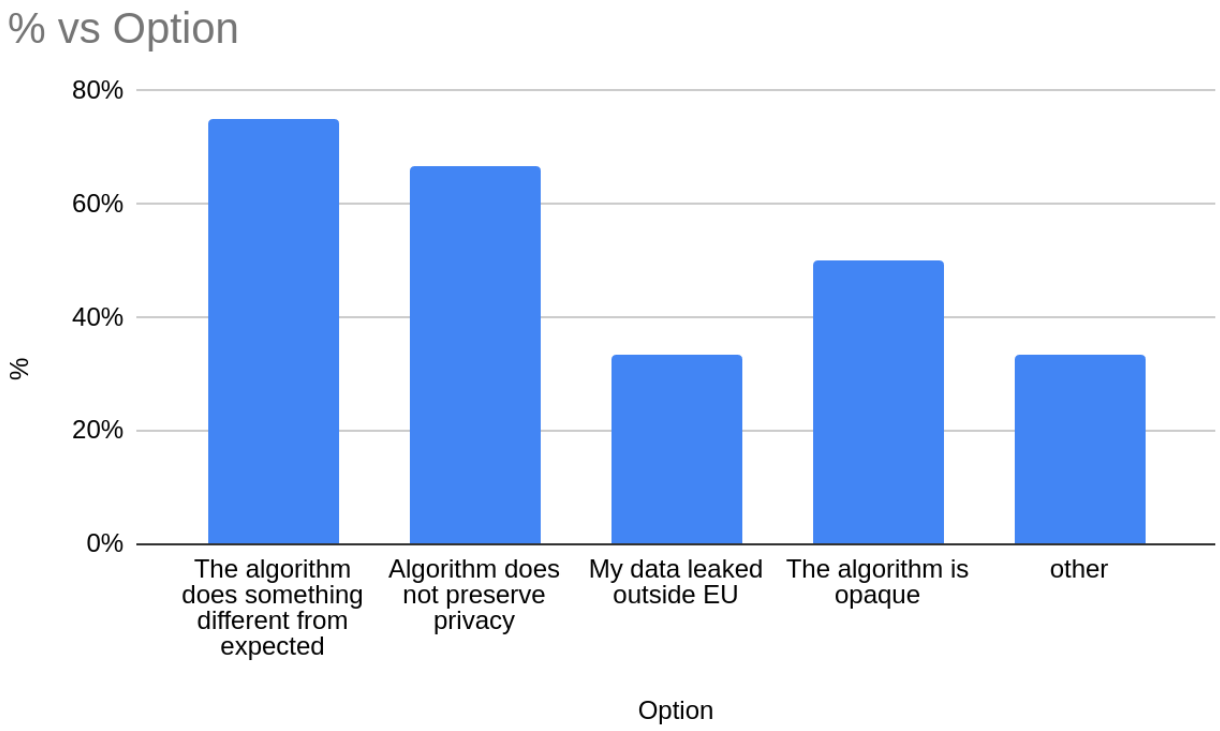
The first question of the study focused on the trust mechanisms for AI services in data spaces concerning data sovereignty. Initially, an open-ended question was posed to AI experts, but the results were not as expected in terms of the responses received. Therefore, a new survey was created with a limited set of options to simplify the answering process. The survey also included an "other" option, which allowed respondents to provide additional input in a separate text box. Participants were allowed to provide multiple answers to this question. It is important to note that the results must be interpreted qualitatively due to the lack of a statistically fair sampling approach. Nonetheless, the report includes actual percentages of the responses received.



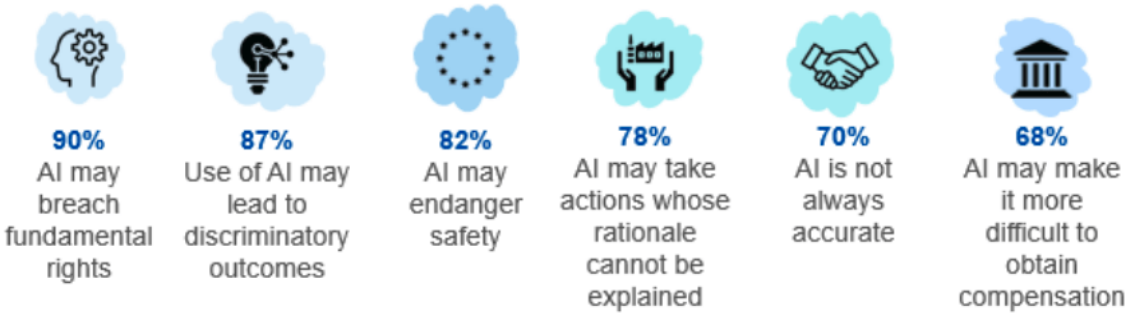
The certification was the first option followed by the site where the service is published.
Although option *other* was selected by several participants there was not any content in the ad hoc box deployed.

3.2 Main concern about algorithms

The methodology for the second question was also modified from an initial open-ended survey to a list of possible concerns with an additional option for "other" concerns. The list of possible concerns included: (1) data leakage outside the EU, (2) lack of privacy preservation by the algorithm, (3) algorithm behaving unexpectedly, and (4) algorithm being opaque. Multiple answers were allowed for this question as well. The results are subject to qualitative interpretation due to the sampling methodology and lack of statistical approach, although actual percentages are provided.



The two solutions with more votes are that the algorithm does something different from expected and that the algorithm does not preserve privacy. Again the option other was chosen but no additional information has been provided in the deployed box.



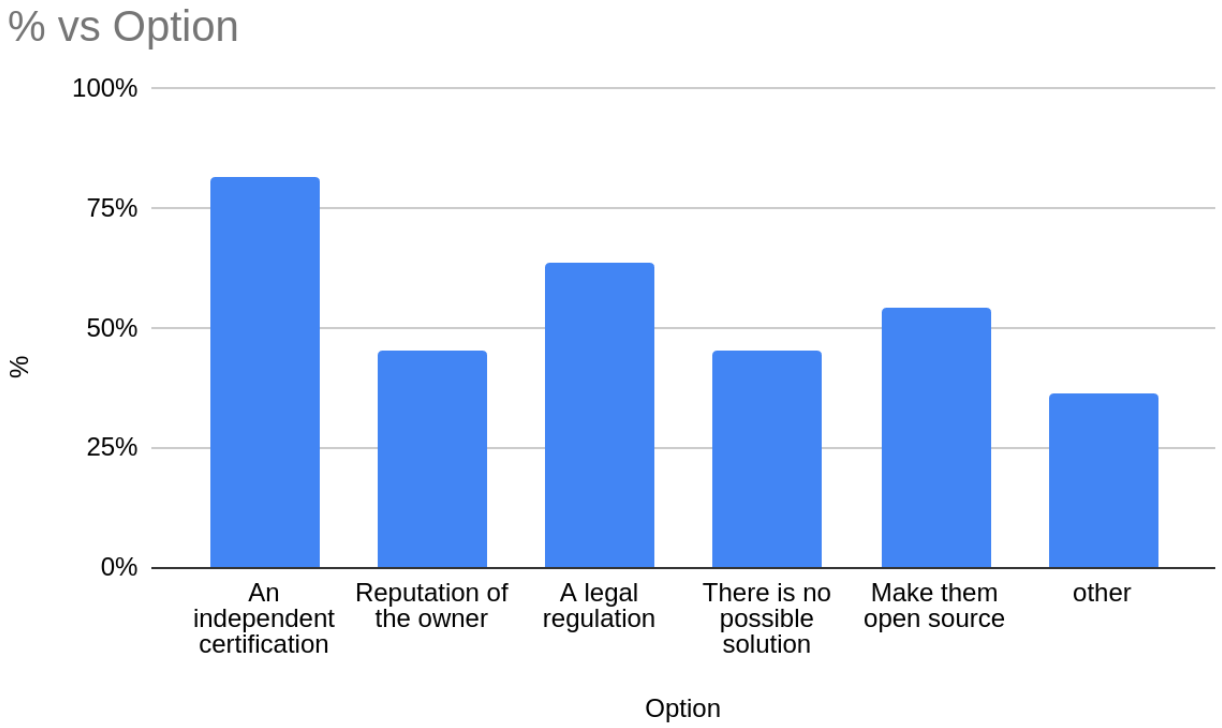
Concerns of the survey on the trust dimension of AI in the white paper on AI[17] (2020) look somehow aligned

3.3 Potential solutions

The last question was regarding the potential solutions. Like the previous questions, it was also redone after the initial survey, and the following possible answers were added after an initial discussion with some experts.

- An independent certification
- The reputation of the algorithm's owner
- A legal regulation
- Make the algorithm open source
- There is no possible solution
- Other

Like previous questions multiple answers were possible for the participants.



The results show the most voted is an independent certification followed by a legal regulation.

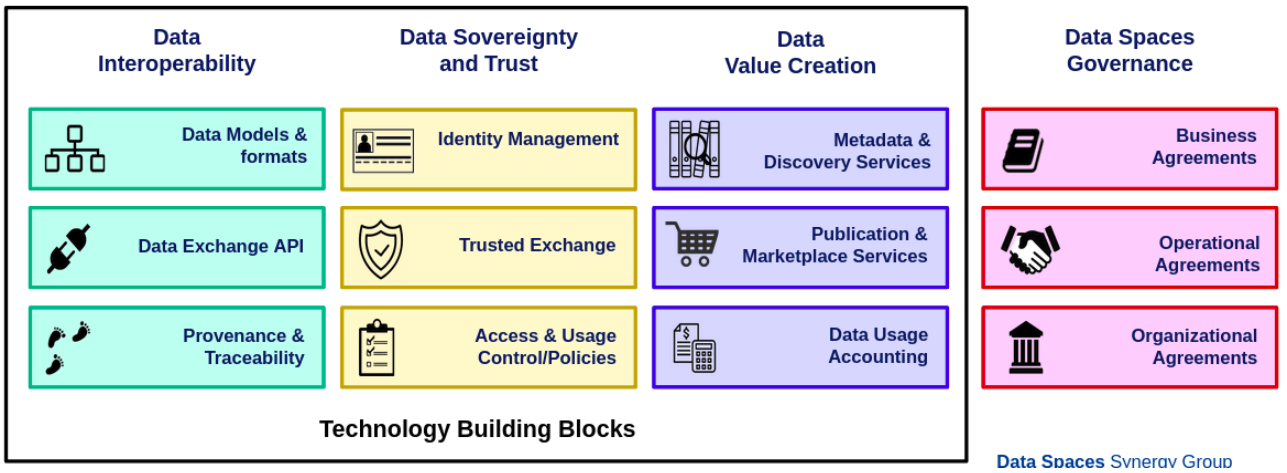
3.4 Results of the personal interviews

Seven personal interviews were run during the period of the work. Some considerations:

1) **Relevance of the topic**. All participants were people deeply involved in either data spaces, AI algorithms, or data sovereignty. All of them agree that the topic deserves to be considered from several points of view, and it is of remarkable importance.

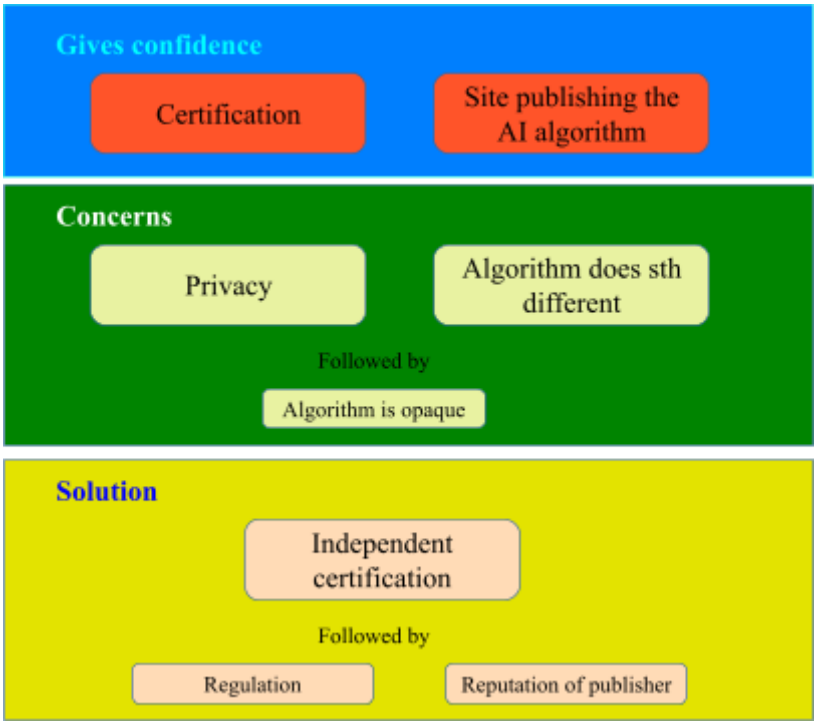
[17] https://ec.europa.eu/newsroom/dae/redirection/document/68462

2) **Legal / governance**. Running a data space (allowing the interchange of data between the participants in a sustainable way, and involving economic transactions), requires governance. Data sovereignty should be part of this governance. See the OpenDEI blocks diagram in the next figure. The pink blocks column are blocks related to the governance of the data space.
3) **Technical complexity**. If there is a technical solution for ensuring that the services operating associated with a data space comply with some requirements of the data space governance requires some kind of technical tests to be run on them. Explainability of the AI algorithms is not always possible and therefore those tests should take this limitation into account (or exclude those which are not somehow explainable). Additionally not always the services run completely in the limits of the data space imposing an additional complexity level.
4) **Limited experience**. The interviewed experts commented that one limitation of the creation of such standardization is the limited experience accumulated in the operation of the existing data spaces. According to the schema defined in the OpenDEI project, which includes governance components, it also includes strong mechanisms of identification, complex authorization & data use policies, and economic transactions, the oldest data spaces hardly exceed two years of operation. It is true that other data-sharing ecosystems existed many years ago and some experience could be extracted. But it is true that few, if any, of them have AI services as services.



Source: Original from OpenDEI EU project. Synergy Group

5) **Organizational complexity of certification**. A running certification implies the creation of a complex organization with strong technical capabilities and agreed procedures in the case of a certification that could be in operation.
6) **Uncertain sustainability**. The sustainability of the data spaces is currently an open question. A certification associated with them would share the same sustainability problems. It is also true that some of the interviewed people remarked that such certification could help to provide trust more in the data spaces and therefore help to populate them.



Qualitative results of the survey on AI algorithms in terms of data sovereignty for data spaces

4 Proposal for drafting a standardization

The scope, especially in the time span, of this work, does not allow a deep study and creation of the potential standardization from the standpoint of the classical standardization which includes an independent group of experts, a consensus procedure, a public review and a final version before publication.
Usually, this process could take more than 24 months. Especially for emerging subjects like artificial intelligence and data spaces a complementary approach, agile standardization[18], is recommended to be used. Otherwise, the standardization will be ready when the market will have made the standardization obsolete and the divergence between the solutions created while the standard was in progress, will not have a turning back.

Due to the time constraints of the project, the traditional approach to standardization - which involves an independent group of experts, consensus procedure, public review, and final version before publication - is not feasible. This process can take up to 24 months, which is not suitable for emerging topics like artificial intelligence and data spaces. Therefore, the complementary approach of agile standardization is recommended, which allows for a more flexible and adaptable process that can keep up with the rapidly changing market environment. This section also notes that if the standardization process is not agile (so it is approached in the classical way), its results may become obsolete by the time it is completed, and the divergence between the solutions created during the standardization process will be irreversible.

4.1 Process of acceptance

Here there is a drafted process of acceptance of an AI service to be in the data spaces marketplace. It is an adaptation of other certification processes with the possible adaptations required for data spaces.

1. Develop your service: The service should comply with the recommendation for development (in terms of data sovereignty). A document with the ***guidelines*** has to be published by ***the certification entity***
2. Prepare your service for testing: It will be required that you ***pack your service according to the instructions*** of data space, and include some ***metadata*** describing the main purpose and characteristics of the application, especially where the data is stored, how it is processed, and how the service is deployed in the marketplace.
3. You'll need to prepare your service to meet the ***identification, and authorization requirements*** of the data space. Besides this, the access policies will be applicable to the service.
4. Access to the ***interface for validation*** of services: You'll need to access this service once an identity would be registered and accepted in the data space.
5. Request for publication. Once completed the application will be reviewed in the ***sandbox*** (provisionally accepted or rejected) and if there are economic transactions involved further information will be required from the owner of the service.
6. Wait for approval: Once registered an ***internal process*** will be carried out to test that the instructions and limitations claimed in the metadata are actually enforced. The ***certification***

[18] The principles of agile standardization are more extensively described in this manifesto.

entity of the data space (it could be the governance board of the data space) should inform the service owner about the potential delays in this process.

7. Receive feedback and make necessary changes: If your service is rejected, you'll receive feedback from the data space certification entity explaining the reasons for rejection. You can make the necessary changes and resubmit your service for review.
8. Service is published in the data space marketplace: Once your service is approved, it will be published in the catalog of available services for data space participants. Eventually, a badge regarding the level of data sovereignty would inform users about their characteristics in terms of data sovereignty.

4.2 Artifacts to be created for a certification

According to the proposed procedure these are the artifacts to be created to implement such certification. The procedure is quite generic and therefore can be adapted to cover not only data sovereignty but other elements as well.

4.2.1 Guidelines for the development of services.

These guidelines will help the potential participants of the data space marketplace with the main best practices to be implemented.

4.2.2 Recommendations for packaging the service, metadata

These recommendations will provide information to the service owners about the specific elements to be covered in the metadata of the service (the information shared in the catalog) and about the internal use of data according to the levels defined in the specification

4.2.3 Interface for the publication request

The data space manager entity has to publish an interface for gathering all the information of the services to be published in the data space marketplace. It will include all technical information, metadata, economic details for transactions, etc.

4.2.4 Sandbox for testing the services and its programming testing services

The sandbox will carry out the tasks to check the compliance with the claimed data management of the service and with the real use. The results will be sent as feedback to the service owner. One important dilemma is where the services will be running. It could be either in the data space resources or in their own. Consequently, the control over the executed service will be very different and the requirements for acceptance should vary accordingly.

4.2.5 Testing instructions and procedures

The operation of the sandbox will involve a group of procedures that has to be known to the certification entity workforce and partially by the owners of the service. The location of the operation of the service will make clear differences here.

4.2.6 Standard description with different levels in terms of data sovereignty

Obviously some type of text has to be created describing the possible levels of data sovereignty. These levels have to describe the use of the owners' data, their final storage and treatment. It has to be clear to the user of the service what will happen with their data. This is critical to allow trust for the users of the service inside a data space where non-public data can be shared. Although eventually it would be possible that specific requirements would be created by every data space, it is much more likely that a global certification would be created and the data space would adhere to it.

4.2.7 Governance of the entity/entities managing the certification

The data space has to be managed by an entity or a group of entities. They could be the responsible ones for the assessment of the data sovereignty, however, it is more likely that this task will be delegated to a specific entity because of the specialization economy. Therefore the governance of the data space will need to state what will be the requirements for the certification.

4.2.8 Certificate badges linked to the approval registry

It is true that the marketplace catalog of services would contain an accurate description of the service in terms of data sovereignty. But it is also true that a certification can be identified by a badge/logo by the different levels identified in the standard.

4.2.9 DDBB of valid and revoked / obsolete certificates of services

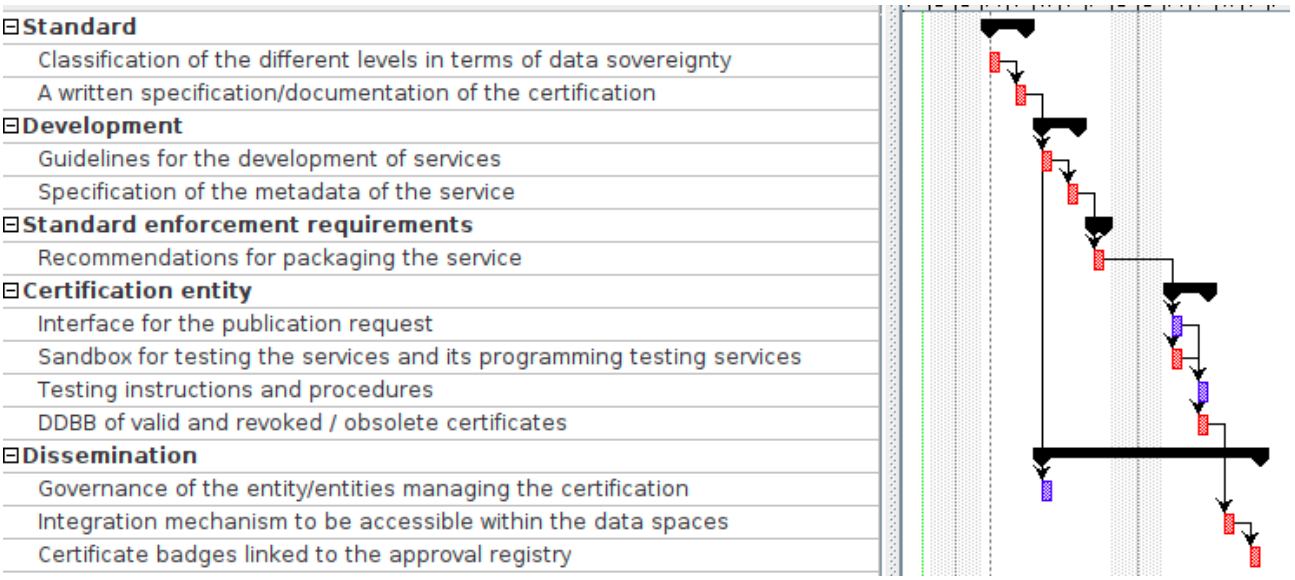
The users of the cloud service in the data space, the marketplace, need to know when a service is being successfully tested and when the service is out of certification for whatever reason. the interface for the users of the marketplace has to provide easy ways of recognizing this fact (badges?).

4.2.10 Integration mechanism to be accessible within the data spaces

The certification has to have some mechanisms to get seamlessly integrated with the different architectures of data spaces. Not only for the certification badges but possibly for the integration of instances of the test beds for the different data spaces.

4.3 Planification of artifacts in order to launch the certification

The following diagram provides a simplified approach for planning the dependencies among different artifacts required to initiate the independent certification process. The diagram only depicts the main dependencies and excludes bidirectional relationships and multiple dependencies that may be present in reality.



5 Proposal of a data model for algorithm service classification in terms of data sovereignty

The data models proposed here are based on agile standardization and its seven principles

1. Don't just standardize, be agile and standardize
2. Do not reinvent the wheel
3. Normalize real cases
4. Be open
5. Don't be overly specific
6. Flat not Deep
7. Sustainability is key

This proposal aligns with these seven principles. The first principle, "*Don't just standardize, be agile and standardize*", emphasizes the need for a quick response to market demands. Traditional standardization bodies often take too long to draft specifications, resulting in obsolete standards that struggle to regulate a market already diverged into different approaches.
The second principle, "*Do not reinvent the wheel*", is followed by linking the proposed service description to the Gaia-x service description defined in the service characteristics group[19]. Although the cloud service description is incomplete, this initiative will eventually define it.
However, the proposal fails to comply with the third principle, "*Normalize real cases*", as it lacks an actual use case in a data space and relies on expert opinions and surveys rather than in a working data space. Therefore, it is recommended to launch a pilot to test the proposed approach.
The fourth principle, "*Be Open*", is upheld by publicly sharing the data model in the Smart Data Models initiative's incubated repository[20] with an open license.
The fifth principle, "*Don't be overly specific*" is also met as the proposed models have no dependence on a specific data space implementation.
The sixth principle, "*Flat not Deep*", is observed by designing a simple structure without additional complexity.
The seventh principle, "*Sustainability is key*", is proposed to be enforced through an independent certification organization with a sustainable funding mechanism beyond the initial pilot project.

Compliance with Agile standardization principles.	
Principle	Compliance
Don't just standardize, be agile and standardize	
Do not reinvent the wheel	
Normalize real cases	
Be open	
Don't be overly specific	
Flat not Deep	

[19] https://gitlab.com/gaia-x/technical-committee/service-characteristics
[20] https://github.com/smart-data-models/incubated/tree/master/CROSSSECTOR/DataSovereignty

Sustainability is key	

Legend. Green: compliant, Yellow: partially compliant, Red: Not compliant

Thus, a JSON schema is constructed as a single-source-of-truth for the proposed standard, and all related documentation can be automatically generated based on this technical file. There are two schemas, one for the submission of the service owner (request for certification) and another for the response from the certification (answer to assessment). It is possible that additional data models may be required, and this need should be identified during the pilot execution.

5.1 Schema for requesting the certification of an AI-based service on data sovereignty

The schema includes not only the attributes for the request on data sovereignty but also assesses the potential bias of the algorithm.
Every attribute or subattribute includes a tag description with explanations about the meaning of the attribute. The standard is in json schema format.

```
{
 "$schema": "http://json-schema.org/schema#",
 "$schemaVersion": "0.0.1",
 "$id":
"https://smart-data-models.github.io/datamodel.MachineLearning/DataSovereigntyMe
tadata/schema.json,",
 "title": "Algorithm data sovereignty request of certification",
 "description": "Algorithm data sovereignty request of certification. The data
has to be provided by the owner of the service",
 "modelTags": "StandICT",
 "derivedFrom": "",
 "required": [
   "id",
   "type"
 ],
 "license":
"https://smart-data-models.github.io/datamodel.MachineLearning/AssessedAlgorithm
/LICENSE.md",
 "type": "object",
 "allOf": [
   {
     "$ref":
"https://smart-data-models.github.io/data-models/common-schema.json#/definitions
/GSMA-Commons"
   },
   {
     "$ref":
"https://smart-data-models.github.io/data-models/common-schema.json#/definitions
/Contact-Commons"
   },
   {
     "properties": {
       "type": {
         "type": "string",
         "enum": [
           "DataSovereigntyMetadata"
```

```
          ],
          "description": "Property. It has to be DataSovereigntyMetadata"
        },
        "request": {
          "type": "object",
          "description": "Property. identifier of the request for certification",
          "properties": {
            "requestId": {
              "type": "string",
              "description": "Property. Unique identifier of the request"
            },
            "requestDate": {
              "type": "string",
              "format": "date-time",
              "description": "Property. Date of the request according to ISO 8601
format"
            },
            "serviceId": {
              "type": "string",
              "description": "Property. Unique identifier of the service provided
by the data space management system"
            },
            "serviceContact": {
              "$ref":
"https://smart-data-models.github.io/data-models/common-schema.json#/definitions
/Contact-Commons"
            }
          }
        },
        "serviceCharacteristics": {
          "type": "string",
          "format": "uri",
          "description": "Relationship. Pointer to where the description of the
service is available compliant with the service description specification of
Gaia-X. https://gitlab.com/gaia-x/technical-committee/service-characteristics"
        },
        "execution": {
          "type": "string",
          "description": "Property. Whether the service is executed in the data
space's resources or in their own resources",
          "enum": [
            "remote",
            "dataspace"
          ]
        },
        "executionUrl": {
          "type": "string",
          "format": "uri",
          "description": "Property. Pointer to where the service is executed
(usually for remotely executed services)"
        },
        "allowedPersonalData": {
          "type": "boolean",
          "description": "Property. if the service is capable to deal with
personal data affected by the GDPR regulation"
        },
        "storage": {
          "type": "string",
          "description": "Property. What is the service doing with the data
```

```
storage. noStorage means that the service is not storing the data at all.
aggregatedStorage is that some data are stored but in an aggregated form so it
avoids individual access to the source registers. The aggregationThreshold
attribute complement with the minimum size of an aggregation. individualStorage
means that the service can store individual registries",
          "enum": [
            "noStorage",
            "aggregatedStorage",
            "individualStorage"
          ]
        },
        "storageJurisdiction": {
          "type": "string",
          "description": "Property. The jurisdiction where the data are stored in
the case they are. Registration coded in ISO 3166-1 alpha2, alpha-3 or numeric
format."
        },
        "aggregationThreshold": {
          "type": "number",
          "minimum": 1,
          "description": "Property. Minimum size of aggregated data stored by the
service."
        },
        "processingIntegrity": {
          "type": "string",
          "description": "Property. Description of the mechanisms for ensuring
the integrity of the transmission chain of data for processing, obtaining
results and delivering to the user"
        },
        "resultsTrainAlgorithm": {
          "type": "boolean",
          "description": "Property. it indicates if the processed information is
incorporated into the AI algorithm for improving its performance, accuracy,
etc."
        },
        "qualificationRequested": {
          "type": "string",
          "description": "Property. Global qualification of the algorithm
requested for the certification. asIs means that the algorithm does not provide
evidence of compliance with data sovereignty regulations or standards.
documented means that the algorithm provides self-certification of some data
sovereignty standards compliance. sandbox means that the algorithm has been
tested in an independent sandbox to check data sovereignty ",
          "enum": [
            "No storage",
            "storage of aggregated data",
            "storage of individual data"
          ]
        },
        "socialImpact": {
          "type": "string",
          "description": "Property. if the algorithm has any kind of social
impact assessment"
        },
        "checkBias": {
          "type": "object",
          "description": "Property. If the algorithm has been checked against
these potential biases. Race, ethnicity, gender, nationality, income, sexual
orientation, ability and political or religious belief.",
```

```
          "properties": {
            "race": {
              "type": "boolean",
              "description": "Property. If the algorithm has been assessed in
terms of race bias "
            },
            "ethnicity": {
              "type": "boolean",
              "description": "Property. If the algorithm has been assessed in
terms of ethnicity bias "
            },
            "gender": {
              "type": "boolean",
              "description": "Property. If the algorithm has been assessed in
terms of gender bias "
            },
            "nationality": {
              "type": "boolean",
              "description": "Property. If the algorithm has been assessed in
terms of nationality bias "
            },
            "income": {
              "type": "boolean",
              "description": "Property. If the algorithm has been assessed in
terms of income bias "
            },
            "sexualOrientation": {
              "type": "boolean",
              "description": "Property. If the algorithm has been assessed in
terms of sexual orientation bias "
            },
            "politicalBelief": {
              "type": "boolean",
              "description": "Property. If the algorithm has been assessed in
terms of political beliefs bias "
            },
            "religiousBelief": {
              "type": "boolean",
              "description": "Property. If the algorithm has been assessed in
terms of religious beliefs bias "
            }
          }

        },
        "termsOfUse": {
          "type": "string",
          "format": "uri",
          "description": "Property. Link to the terms of use of the algorithm,
especially for the data sovereignty section"
        },
        "integrity": {
          "type": "object",
          "description": "Property. Preventions took by the owner of the AI
service to guarantee the integrity of the evidence of data",
          "properties": {
            "secureStorage": {
              "type": "string",
              "description": "Property. The mechanism for preventing unauthorized
access to data"
```

```
            },
            "chainOfCustody": {
              "type": "string",
              "description": "Property. Mechanism to ensure that the evidence has
not been altered"
            },
            "encryption": {
              "type": "string",
              "description": "Property. Mechanism to protect the evidence during
communications and storage"
            },
            "digitalSignatures": {
              "type": "string",
              "description": "Property. Description of the mechanism for digital
signing of the evidence"
            }
          }
        },
        "trustee": {
          "type": "string",
          "format": "uri",
          "description": "Property. URI pointing to the id of the organization
providing the trust. Next version could provide a VC trust framework for the
identification of the certification"
        },
        "opensource":  {
          "type": "boolean",
          "description": "Property. if the algorithm is open source and can be
publicly checked"
        },
        "sourceCode": {
          "type": "string",
          "format": "uri",
          "description": "Property. Link to the source code available to be
examined by the certification experts and processes”
        }
      }
    }
 ]
}
```

5.2 Schema for answering a request for certification of an AI-based service on data sovereignty

The schema includes not only the attributes for the request on data sovereignty but also assesses the potential bias of the algorithm.
Every attribute or subattribute includes a ta description with the explanations for the meaning

```
{
 "$schema": "http://json-schema.org/schema#",
 "$schemaVersion": "0.0.1",
```

```
 "$id":
"https://smart-data-models.github.io/datamodel.MachineLearning/AssessedAlgorithm
/schema.json,",
 "title": "Algorithm assess in terms of data sovereignty",
 "description": "This data model represents the assessment of an AI algorithm,
initially offered as a service",
 "modelTags": "StandICT",
 "derivedFrom": "",
 "required": [
   "id",
   "type"
 ],
 "license":
"https://smart-data-models.github.io/datamodel.MachineLearning/AssessedAlgorithm
/LICENSE.md",
 "type": "object",
 "allOf": [
   {
     "$ref":
"https://smart-data-models.github.io/data-models/common-schema.json#/definitions
/GSMA-Commons"
   },
   {
     "$ref":
"https://smart-data-models.github.io/data-models/common-schema.json#/definitions
/Contact-Commons"
   },
   {
     "properties": {
       "type": {
         "type": "string",
         "enum": [
           "AlgorithmAssessed"
         ],
         "description": "Property. It has to be AlgorithmAssessed"
       },
       "request": {
         "type": "object",
         "description": "Property. identifier of the request for certification",
         "properties": {
           "requestId": {
             "type": "string",
             "description": "Property. Unique identifier of the request"
           },
           "requestDate": {
             "type": "string",
             "format": "date-time",
             "description": "Property. Date of the request according to ISO 8601
format"
           },
           "serviceId": {
             "type": "string",
             "description": "Property. Unique identifier of the service provided
by the data space management system"
           },
           "serviceContact": {
             "$ref":
"https://smart-data-models.github.io/data-models/common-schema.json#/definitions
/Contact-Commons"
```

```
            }
          }
        },
        "dateResponse": {
          "type": "string",
          "format": "date-time",
          "description": "Property. Date of the answer to the request according
to ISO 8601 format"
        },
        "certification": {
          "certificationGranted": {
            "type": "boolean",
            "description": "Property. If the request has generated some kind of
certification "
          },
          "typeOfCertification": {
            "description": "Property. What is the service doing with the data
storage. noStorage means that the service is not storing the data at all.
aggregatedStorage is that some data are stored but in an aggregated form so it
avoids the individual access to the source registers. The aggregationThreshold
attribute complement with the minimum size of an aggregation. individualStorage
means that the service can store individual registries",
            "enum": [
              "noStorage",
              "aggregatedStorage",
              "individualStorage"
            ]
          },
          "validFrom": {
            "description": "Property. Defines the date and time, when the
certification becomes valid",
            "type": "string",
            "format": "date-time"
          },
          "validUntil": {
            "description": "Property. Defines the date and time, when the
certification expires",
            "type": "string",
            "format": "date-time"
          }
        }
      },
      "biasResults": {
        "type": "object",
        "description": "Property. ",
        "properties": {
          "race": {
            "type": "string",
            "description": "Property. Results of the algorithm test about race
bias"
          },
          "ethnicity": {
            "type": "string",
            "description": "Property. Results of the algorithm test about
ethnicity bias"
          },
          "gender": {
            "type": "string",
            "description": "Property. Results of the algorithm test about gender
```

```
bias"
          },
          "nationality": {
            "type": "string",
            "description": "Property. Results of the algorithm test about
nationality bias"
          },
          "income": {
            "type": "string",
            "description": "Property. Results of the algorithm test about income
bias"
          },
          "sexualOrientation": {
            "type": "string",
            "description": "Property. Results of the algorithm test about sexual
orientation bias"
          },
          "politicalBelief": {
            "type": "string",
            "description": "Property. Results of the algorithm test about
political beliefs bias"
          },
          "religiousBelief": {
            "type": "string",
            "description": "Property. Results of the algorithm test about
religious beliefs bias"
          }
        }
      },
      "termsOfUse": {
        "type": "string",
        "format": "uri",
        "description": "Property. Link to the terms of use of the algorithm,
especially for the data sovereignty section"
      },
      "resultsIntegrity": {
        "type": "object",
        "description": "Property. Assessment of the preventions taken by the
owner of the AI service to guarantee the integrity of the evidence of data",
        "properties": {
          "secureStorage": {
            "type": "string",
            "description": "Property. Results on the mechanisms for preventing
unauthorized access to data"
          },
          "chainOfCustody": {
            "type": "string",
            "description": "Property. Results on the mechanisms to ensure that
the evidence has not been altered"
          },
          "encryption": {
            "type": "string",
            "description": "Property. Results on the mechanisms to protect the
evidence during communications and storage"
          },
          "digitalSignatures": {
            "type": "string",
            "description": "Property. Results on the mechanisms for digital
signing of the evidence"
```

```
          }
        }
      },
      "trustee": {
        "type": "string",
        "format": "uri",
        "description": "Property. URI points to the id of the organization
providing the trust. Next version could provide a VC trust framework for the
identification of the certification"
      }
    }
  ]
}
```

6 Future lines of work

6.1 Pilot on data sovereignty for data spaces for AI services

The rapid pace of the data market necessitates alternative approaches to classical standardization and its associated consensus procedures. This report suggests that agile standardization is a complementary alternative approach that can be used. However, in order to effectively implement agile standardization, it is necessary to have actual use cases as examples. It is therefore strongly recommended to run a pilot in an actual data space ecosystem. This pilot should validate the proposed standardization outlined in the previous chapter and assess the organizational and economic feasibility of such certification. Collaborating with existing data spaces or EU projects on data spaces would support the validation and evolution of the proposed certification, and, possibly, it is the simplest way to launch this pilot.

6.2 Extend the study to more stakeholders

Due to the short period of the study some interviews were not achieved, the participation in the working groups (Gaia-X, ETSI, etc) was limited and other main stakeholders cannot share their position (Microsoft Azure). The extension of the working period could facilitate the gathering of input from these stakeholders.

6.3 Validate the standard out of the data spaces marketplaces

The proposed standard is especially suitable for the data spaces' marketplaces but it is also true that the proposed certification mechanisms do not need to be in data space to operate, therefore other marketplaces, not associated with data spaces, could apply the same proposed standard to assess the data sovereignty of their services.

6.4 Apply the certification schema for other purposes

As mentioned in the previous chapter there are a set of attributes to assess the possible biases of the algorithm, which are out of the scope of this work but they could share the same structure for the validation.

- Race
- Ethnicity
- Gender
- Nationality
- Income
- Sexual orientation
- Political beliefs
- Religious beliefs

The reason for including this possibility is that it could share most of the elements and in some of the sources for information they were already identified (on the contrary to data sovereignty).

6.5 Further developments on the data sovereignty tests

Defining the aspects to be assessed in terms of data sovereignty doesn't mean that such a test can be performed successfully. There are many open questions, here are some that were collected from the interviews with the experts.

- Technical viability of some of the tests.
    - How to check the integrity of the data transmission outside of the data space boundaries
    - How to ensure that the aggregation of data claimed by a service really meets the declared limits
    - How to deal with the constant evolution of AI services. What will be the cost of re-certification of new versions of the services
    - How to check that the data are not being transferred outside the jurisdiction limits on a daily basis?
    - How to access the technical details of the services when they are not open-sourced?
    - Even if there is the possibility of access, will there be enough information to really grant that some requirements of the certification are met?
    - Integration with data spaces. will the services assume the identification and authorization of data spaces' technical architectures?
- Sustainability
    - What will be the cost of the certification for the AI services?
    - The data spaces would pay for having certified services? Will the users pay for the difference between certified and not certified AI services?
    - How can an independent organization run the certification?
- Legal framework and regulation
    - Will the public regulators launch a regulation on the use of non-certified services (in terms of data sovereignty, the bias of the algorithms, etc)?
    - How could be translated into the governance of the data spaces the requirement of being compliant with a data sovereignty certification?

7 Conclusions

AI services is an extremely dynamic market

The AI market is quite dynamic and in constant evolution to a speed rate well above the usual capability of regulation and standardization. For instance, ChatGPT has exceeded more than 100M users in less than 2 months with a significant percentage in the UE.

This fact urges complementary approaches to get the same results as classical standardization to solve the main issues like unique markets, prevention of problems with citizens' rights, and keeping a competitive industry.

Agile standardization for approaching certification in this market

The use of agile standardization, and its seven principles, could cope with the AI services market speed and produce a pre-certification mechanism that prevents the divergence of solutions in the market.

Based on actual experiences complemented with:

- An open survey shared in AI groups
- Personal interviews with AI experts
- Participation in standardization groups
- Research on the existing initiatives and working groups

Most remarkable conclusions

- There is a lack of control regarding data sovereignty in AI services
- One possible solution would be a certification of AI services

Practice is challenging

However, all individuals interviewed acknowledge the technical and organizational challenges that come with such a certification. Technical hurdles arise from the fact that data spaces are technical environments that lack uniform definition and standardization. The DSBA's efforts aim to align various initiatives (such as Gaia-X, IDSA, FIWARE, and BDVA) in a competitive environment with the main providers (such as Amazon, Microsoft, Google, etc.). Furthermore, technical difficulties stem from the complexity of certifying AI systems that may not always be explainable or may involve the handling of sensitive information. Organizational challenges arise in determining how the certification can be launched, who the responsible entities are, and how sustainability can be ensured. Finally, although there is a consensus on the need for a solution, it remains unclear how much the market will differentiate between certified and uncertified services.

Drafted schemas

The technical schemas proposed in Chapter 5 provide a solid foundation for discussing the advancement of data sovereignty certification and other AI-related matters, such as diverse biases. The schemas propose a basic classification system with three categories: no storage of data, storage of aggregated data, and storage of individual data. Additionally, the transmission and processing of information are taken into account. However, a more intricate classification system with specific levels for transmission and processing may be necessary, and this requires validation in the pilot or other future projects.

Doing nothing (let the market flow) is, possibly, a bad course of action

Taking no action (i.e., allowing the market to operate freely) is potentially a detrimental course of action. If left unchecked, dominant market players may expand their operations at the cost of innovation and a fair market, potentially compromising the rights of certain individuals.

8 Annexes

Some additional information is included here for clarification purposes or just as background information.

8.1 Dimensions of classification of AI algorithms

These dimensions are potential dimensions involving or related to data sovereignty.

8.1.1 By location of the algorithm

8.1.1.1 On premises

These are algorithms that are installed and run on a user's own computer or server. Since the data and algorithms are both located in the same jurisdiction, there is little risk of violating data sovereignty laws. However, this approach may limit the ability to scale and collaborate across different jurisdictions.

8.1.1.2 Cloud Based

Cloud-Based Algorithms: These are algorithms that are hosted on cloud infrastructure, which may be located in a different jurisdiction than the data. In this case, data sovereignty can be a more significant issue. Organizations using cloud-based algorithms may need to ensure that they comply with local data protection regulations, such as the General Data Protection Regulation (GDPR) in the European Union.

8.1.1.3 Federated Learning

Federated Learning Algorithms: These are algorithms that allow multiple parties to collaborate on a machine learning model without sharing their data. Instead, each party trains the model on their own data and shares only the model parameters. Federated learning can be a useful approach for preserving data sovereignty, as each party retains control over their own data. However, it may also require additional technical and legal measures to ensure that data privacy is maintained.

8.1.2 By access to individual registries

8.1.2.1 Centralized Algorithms

These are algorithms that require access to a centralized registry or database in order to function. In this case, the algorithm may require access to individual registries in order to train on or analyze the data. Centralized algorithms may pose a challenge in terms of data sovereignty, as the registry may be located in a different jurisdiction than the organization using the algorithm.

8.1.2.2 Decentralized Algorithms

These are algorithms that do not require access to a centralized registry or database in order to function. In this case, the algorithm may be designed to operate on data that is distributed across different locations or that is owned by different parties. Decentralized algorithms may offer a

greater degree of data sovereignty, as they can be designed to operate without requiring access to individual registries.

8.1.3 Transparency and explainability

Algorithms that are transparent and explainable may be more conducive to ensuring data sovereignty, as they allow organizations to understand how the algorithm operates and how it is making decisions based on the data. This can be particularly important in cases where the algorithm is being used to make decisions that have significant legal or social implications.

8.1.3.1 Black box algorithms

Black box algorithms: These are algorithms in which the internal workings are not transparent or explainable. Black box algorithms are difficult to audit or evaluate, as it is unclear how they arrive at their decisions. This can be problematic, particularly in cases where the algorithm is used to make decisions that have significant legal or social implications, as it can be difficult to understand or challenge the algorithm's decision-making process.

8.1.3.2 Glass box algorithms

Glass box algorithms: These are algorithms that are more transparent and explainable than black box algorithms. Glass box algorithms provide some degree of visibility into the decision-making process, and can be audited or evaluated to some extent. However, there may still be aspects of the algorithm that are difficult to understand or that are not fully transparent.

8.1.3.3 Explainable algorithms

Explainable algorithms: These are algorithms that are specifically designed to be explainable. Explainable algorithms are designed to provide a clear and understandable explanation of how they arrive at their decisions and may use techniques such as natural language generation or interactive visualizations to make the decision-making process more transparent. Explainable algorithms can be particularly useful in contexts where decisions are being made based on sensitive or personal data.

8.1.3.4 Transparent algorithms

Transparent algorithms: These are algorithms that are fully transparent and auditable. Transparent algorithms provide complete visibility into the decision-making process and can be audited or evaluated in a rigorous and systematic way. Transparent algorithms can be particularly useful in contexts where there is a need for high levels of trust and accountability, such as in the case of safety-critical systems or in the regulation of financial markets.

8.1.4 Data ownership

Algorithms may be designed to operate on data that is owned by different parties, such as customers or partners, especially in data spaces. In this case, organizations running the algorithms must ensure that they have the proper legal and regulatory agreements in place to access and use the data in compliance with local laws and regulations.
Therefore this should be a requirement for those providers of services to operate on clients' data.
According to this dimension the algorithms could be classified as

- single owner treatment
- multi owner treatment

8.1.5 Security and privacy

High-security data: This refers to data that is particularly sensitive and requires the highest level of security protection. Examples of high-security data may include government secrets, military intelligence, or financial transactions.

1. Medium-security data: This refers to data that is not necessarily high-security, but still requires a moderate level of security protection. Examples of medium-security data may include trade secrets, confidential business information, or sensitive personal information.
2. Low-security data: This refers to data that is not particularly sensitive and does not require a high level of security protection. Examples of low-security data may include public records, non-sensitive business information, or non-sensitive personal information.
3. Regulated data: This refers to data that is subject to specific regulations, such as HIPAA or GDPR. Regulated data requires a high level of security protection to ensure compliance with relevant regulations.
4. Proprietary data: This refers to data that is owned by an individual or organization and is not intended for public consumption. Proprietary data requires a high level of security protection to prevent unauthorized access or use.
5. Open data: This refers to data that is intended for public consumption and does not require a high level of security protection. However, open data still requires some level of security protection to prevent unauthorized access or use.

8.1.6 By Cross-border data flows

Some algorithms will demand the transfer of data between different legislation areas. Suggested classification would include

Cross-border
Domestic

8.2 Existing standards on AI potentially affected

8.2.1 SC38 PWI Data space

The SC38 just had a resolution to start a standard about this.

1. The SC38 PWI Data Space is a standard that defines a common data model for the exchange of data between programming languages. It is designed to enable interoperability between different programming languages and systems, allowing data to be exchanged between them without loss of information or functionality.
2. Purpose: The purpose of the standard is to provide a common data model that can be used by different programming languages and systems to communicate with each other. By providing a common data model, the standard aims to reduce the complexity of data exchange between different systems and to facilitate interoperability.
3. Scope: The standard defines a set of abstract data types, such as strings, numbers, and arrays, as well as operations that can be performed on them. It also includes rules for

mapping these data types to specific programming languages and for encoding them in a standard way to facilitate data exchange.

4. Data Types: The standard defines a set of data types that can be used in the data model, including:
   Integer
   Real
   String
   Boolean
   Array
   Structure
5. Operations: The standard defines a set of operations that can be performed on the data types, including:
   Arithmetic operations (e.g. addition, subtraction, multiplication, and division)
   Comparison operations (e.g. equal, not equal, greater than, and less than)
   Logical operations (e.g. AND, OR, and NOT)
   Conversion operations (e.g. converting between different data types)
6. Encoding: The standard defines a standard way of encoding the data types to facilitate data exchange between different systems. The encoding scheme is designed to be platform-independent and language-independent.
7. Mapping: The standard also includes rules for mapping the abstract data types to specific programming languages. These rules ensure that the data types are represented consistently across different systems, regardless of the programming language used.

Overall, the SC38 PWI Data Space is an important standard for enabling interoperability between different programming languages and systems, and for facilitating the exchange of data between them. By providing a common data model and encoding scheme, the standard aims to reduce the complexity of data exchange and promote interoperability.

9 Regulations of the EU

Article 23, point d)

2. The officials and other accompanying persons authorized by the Commission to conduct an inspection are empowered to:
(d) require the undertaking or association of undertakings to provide access to and explanations on its organisation, functioning, IT system, algorithms, data-handling and business practices and to record or document the explanations given by any technical means;

Article 30 . Point 3

The Commission may adopt a decision, imposing on undertakings, including gatekeepers where applicable, and associations of undertakings, fines not exceeding 1 % of their total worldwide turnover in the preceding financial year where they intentionally or negligently:
(e) fail to provide access to data, algorithms or information about testing in response to a request made pursuant to Article 21(3);

Article 31. Periodic penalty payments. Point e

1.The Commission may adopt a decision imposing on undertakings, including gatekeepers where applicable, and associations of undertakings periodic penalty payments not exceeding 5% of the average daily worldwide turnover in the preceding financial year per day, calculated from the date set by that decision, in order to compel them:
(d) to ensure access to data, algorithms and information about testing in response to a request made pursuant to Article 21(3) and to supply explanations on those as required by a decision pursuant to Article 21;9.1.2 Ethics guidelines for trustworthy AI
The seven principles

Human agency and oversight: AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights. At the same time, proper oversight mechanisms need to be ensured, which can be achieved through human-in-the-loop, human-on-the-loop, and human-in-command approaches

Technical Robustness and safety: AI systems need to be resilient and secure. They need to be safe, ensuring a fallback plan in case something goes wrong, as well as being accurate, reliable, and reproducible. That is the only way to ensure that also unintentional harm can be minimized and prevented.

Privacy and data governance: besides ensuring full respect for privacy and data protection, adequate data governance mechanisms must also be ensured, taking into account the quality and integrity of the data, and ensuring legitimized access to data.

Transparency: the data, system, and AI business models should be transparent. Traceability mechanisms can help achieve this. Moreover, AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned. Humans need to be aware that they are interacting with an AI system, and must be informed of the system's capabilities and limitations.

Diversity, non-discrimination, and fairness: Unfair bias must be avoided, as it could have multiple negative implications, from the marginalization of vulnerable groups to the exacerbation of prejudice and discrimination. Foster diversity, AI systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life circle.

Societal and environmental well-being: AI systems should benefit all human beings, including future generations. It must hence be ensured that they are sustainable and environmentally friendly. Moreover, they should take into account the environment, including other living beings, and their social and societal impact should be carefully considered.
Accountability: Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes. Auditability, which enables the assessment of algorithms, data, and design processes plays a key role therein, especially in critical applications. Moreover, adequate and accessible redress should be ensured.

10 Other References reviewed to generate the contents of study

1. ISO/IEC 29100:2011. Information technology — Security techniques — Privacy framework.
2. ISO/IEC TS 27570:2020.
3. European AI Alliance. https://futurium.ec.europa.eu/en/european-ai-alliance
4. White Paper on Artificial Intelligence – a European approach to excellence and trust. (Jul 2020)
5. ETSI White paper #52. ETSI Activities in the field of Artificial Intelligence Preparing the implementation of the European AI Act (Dec 2022)
6. WP5 – Digital Platforms and Marketplace. D5.2 Data Flow Management. European project Interconnect.
7. Manifesto for agile standardization.