

Responsible and Inclusive Artificial Intelligence LabS

Triple E: Ethics = Explicability + Equity



Contents

Executive summary	03
01 About AI LabS	04
02 Current context – European framework	07
03 Explainability and equity: introduction	09
04 Explainability: understanding algorithms’ decision-making processes.	
4.1 The need for explainable models	
4.2 Explainability is important at all levels of the organization	
4.3 Relationship between the complexity of the model and the degree of explainability	
	11
05 Equity: ensure fair outcomes for all	
5.1 Know the types of biases in AI models	
	18
06 How do we apply explainability and equity in AI projects?	23
Conclusions	25
About SERES FOUNDATION & NTT DATA	27

Executive summary

Artificial intelligence (AI) is a technology that has the potential to transform many aspects of our lives. However, there is also a risk that AI will be used in a discriminatory or biased manner or that errors will occur, causing it to malfunction. To avoid this risk, AI systems must be understandable and fair, and companies and organizations must adopt ethical AI principles in their operations.

In the 2023 AI LabS, the SERES Foundation and NTT DATA have delved into the principles of explainability and equity, the strong connection of these principles with the ethical plane, and how they directly affect companies.

Explainability is the capability to explain how an algorithm works. It enables us to trust the algorithm and adopt a continuous improvement process. It has also been proven

that explainability is proportional to the complexity of the algorithm.

Equity is an algorithm's capacity to be fair and impartial and guarantee the inclusion and diversity of individuals and social groups. We have gone deeply into the concept of bias and shown how it can occur in all phases of AI-model lifecycles.

Benchmarks such as the *CDO Journey*, developed by NTT DATA, enable companies to develop and implement ethical AI systems that benefit all stakeholders, comply with existing regulations and are prepared for future regulations.

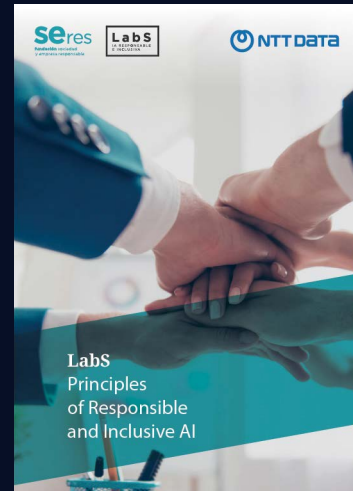
In conclusion, the principles of explainability and equity are essential if we are to develop and use AI ethically.

“ To ensure they use this technology responsibly, companies must incorporate these principles into their culture and processes from the first stages of their AI projects.

1. About AI LabS

Since 2020, NTT DATA and the SERES Foundation have collaborated to create themed laboratories on artificial intelligence that aim to help companies and organizations face the challenges currently posed by AI.

The first edition of AI LabS held in 2020 sought to promote companies' role in developing responsible and inclusive AI that would avoid harming society by drawing up strategic objectives and creating training programmes. To this end, this first laboratory drew up a set of 12 statements on how to help companies promote the ethical AI paradigm from various perspectives. Taking the seven ethical principles drawn up by the European Union for building trustworthy AI as the basis of the session, we drafted a final report entitled the "Common Principles for Responsible and Inclusive AI" ¹



¹AI LabS' Common Principles for Responsible and Inclusive AI



Illustration 1

Twelve statements of AI LabS' Common Principles for Responsible and Inclusive AI

In 2021 we held the second edition of AI LabS that aimed to **design people-centric AI services**. That is, a model that focuses on people’s needs and encompasses clients, users and collaborators. To understand the end-to-end design of people-centric AI services, we applied our AI Service Design methodology to this concept in a workshop. As in the first LabS, this workshop also culminated in a deliverable ².



² NTT DATA (2022) AI LabS' People-centric AI Services Design

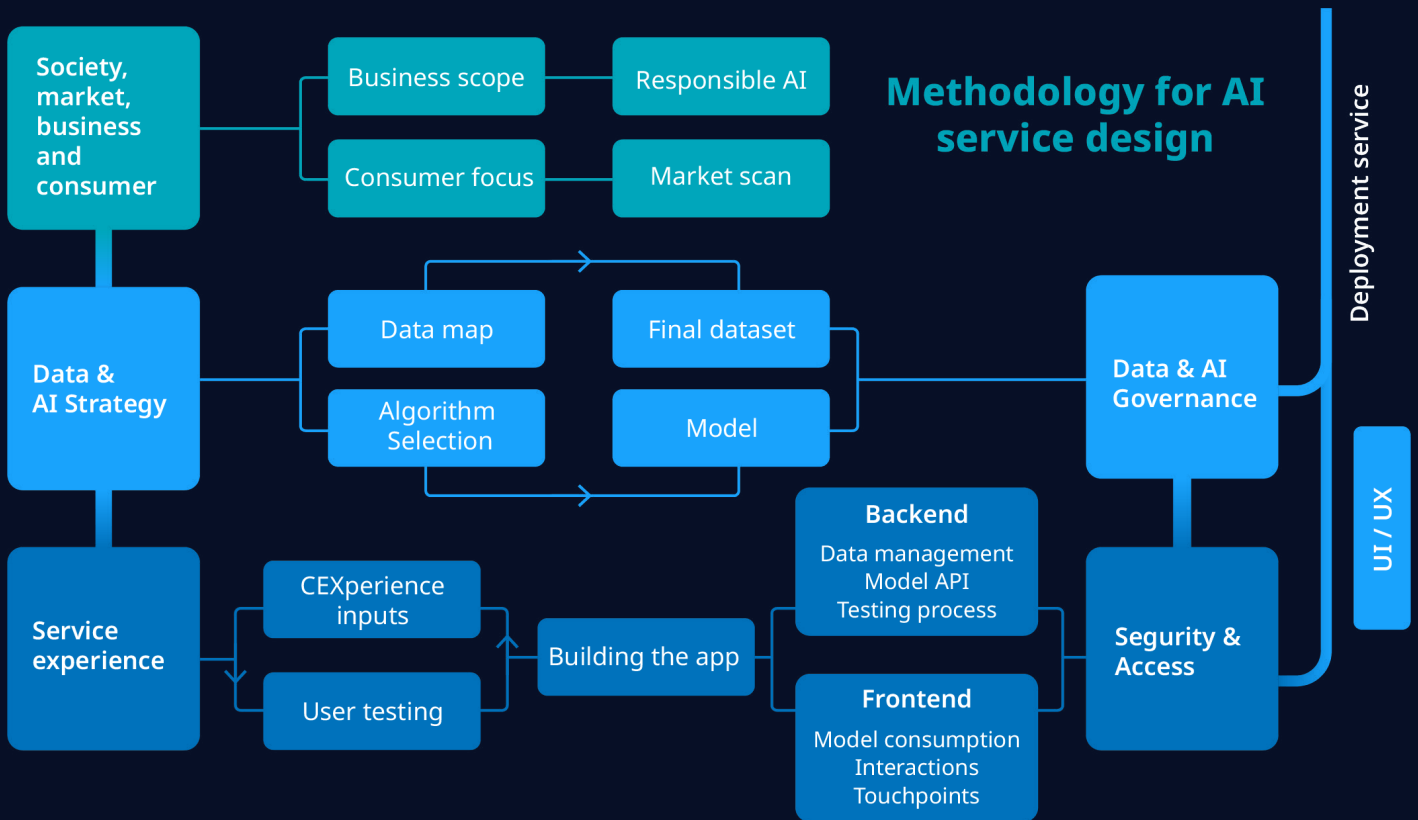


Illustration 2: Methodology for designing an AI service

Thus, we drew on the experience and knowledge acquired in previous laboratories to create this new edition entitled AI LabS “Triple E: Ethics = Explainability + Equity”. The 2022/2023 laboratory delves deeply into the principles of **explainability** and **equity**. To this end, we take as a benchmark one of the statements of the Common Principles from the first edition on diversity and promoting the inclusion of individuals in communities: the principle of equity.

This principle aims to reduce the negative effects of technology and the vulnerabilities that AI may bring about through biases, for example. We also took the principles of explainability and transparency from the second LabS; these are the principles that make AI trustworthy. **This third edition of the LabS aims to break down the principles of explainability and equity and study the ethical implications of developing and deploying AI and of its lifecycle.**



Artificial intelligence is part of everyday life, but its responsible management and global regulation are essential, as leaders such as Brad Smith and Sam Altman have explained during the 2024 Davos World Economic Forum.

Ensuring workforce preparedness and maintaining an ongoing ethical debate, as emphasized by Nick Clegg, are crucial aspects to prevent unintended harm and move towards ethical technological development that benefits society. In this context, organizations should build a relationship of trust with people based on the transparency and explainability of technology.

“ The LabS’ objective is to develop action frameworks that consider the social perspective necessary to address the ethical issues associated with AI from its conception to product design, thus promoting Responsible and Inclusive Artificial Intelligence.

2. Current context – European framework

The principles of explainability and equity are not two isolated ideas; they stem from the European Commission’s commitment to ethical AI expressed in many publications and regulatory proposals dating back as far as 2018.

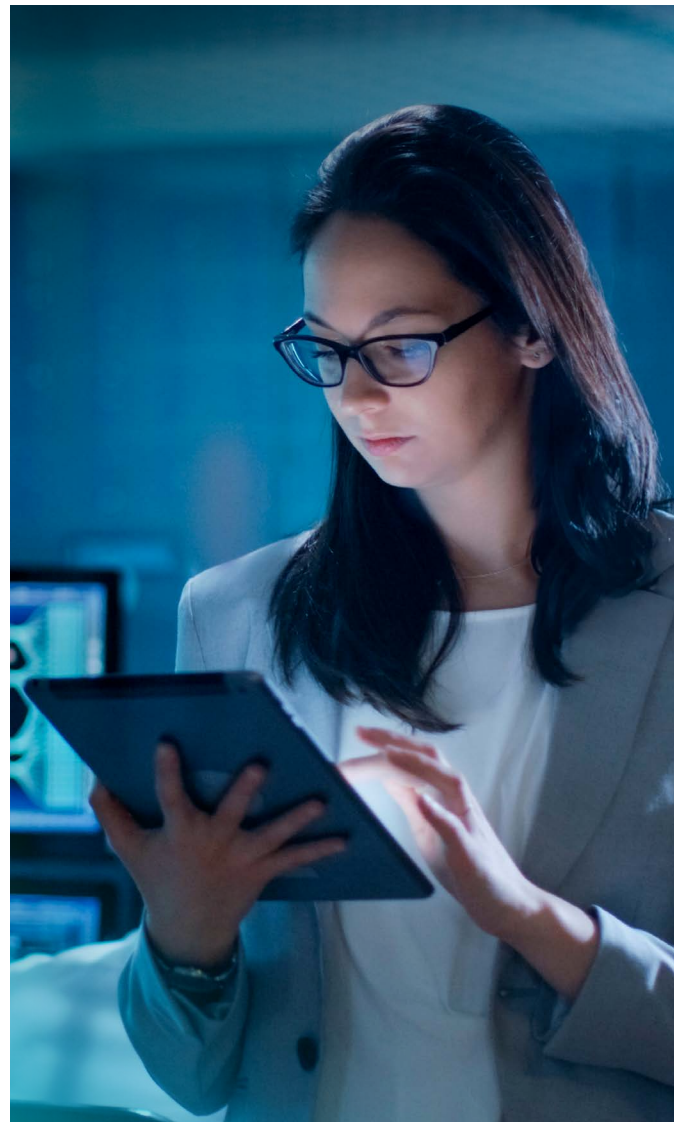
These proposals are intended to provide guidance to businesses and organizations and lay the groundwork for how AI is used.

Among them, the **Coordinated Plan on Artificial Intelligence** published in 2018 is particularly significant ³. It is a joint effort of the Member States and the European Commission to align matters of strategy, policies, regulation, investment, and others in which AI is the engine of the European economy.

Then, in 2019, the **Ethics Guidelines for Reliable AI were published** ⁴. This was the first document to draw a roadmap towards ethical AI based on making the technology trustworthy. That is, for AI to be trusted, it must be robust, secure and promote people’s rights. This same document details the four fundamental principles of trustworthy AI: damage prevention, human autonomy, equity and explainability.

Finally, the **AI Act** ⁵ (2021) of the European Commission is one of the most important regulatory proposals at European level, since it sets out the underlying obligations of service providers and users, taking into account the risk levels that the use of artificial intelligence may pose. The categories are:

- Unacceptable risk, prohibiting the use of AI for autonomous weapons or mass surveillance systems, for example.
- High risk, encompassing AI systems used for access to employment, education or public services.
- Limited risk, which establishes transparency obligations for such uses as chatbots or biometric categorization.
- Minimal risk, for systems that have no obligations or limitations of any kind.



³ [European Commission \(December 7 2018\) Coordinated Plan on Artificial Intelligence](#)

⁴ [European Commission \(2019\) Ethics Guidelines for Trustworthy AI](#)

⁵ [European Commission \(April 21, 2021\) Artificial Intelligence Act](#)

Moreover, during 2023, the European Parliament incorporated new clauses into the AI Act that affect generative AI service providers, who will be required to comply with additional transparency requirements (risk assessment and mitigation, design, information and environmental requirements and registration in the EU database). In any case, the text is still under negotiation and is expected to be formally adopted in the first quarter of 2024 and to enter into force in 2026.

In this way, through co-creation spaces such as AI LabS and especially this new edition, NTT DATA and SERES Foundation join forces to help companies and organizations prepare for all the implications and adjustments they require to align with this new regulation that will structurally affect all companies that develop or use AI.

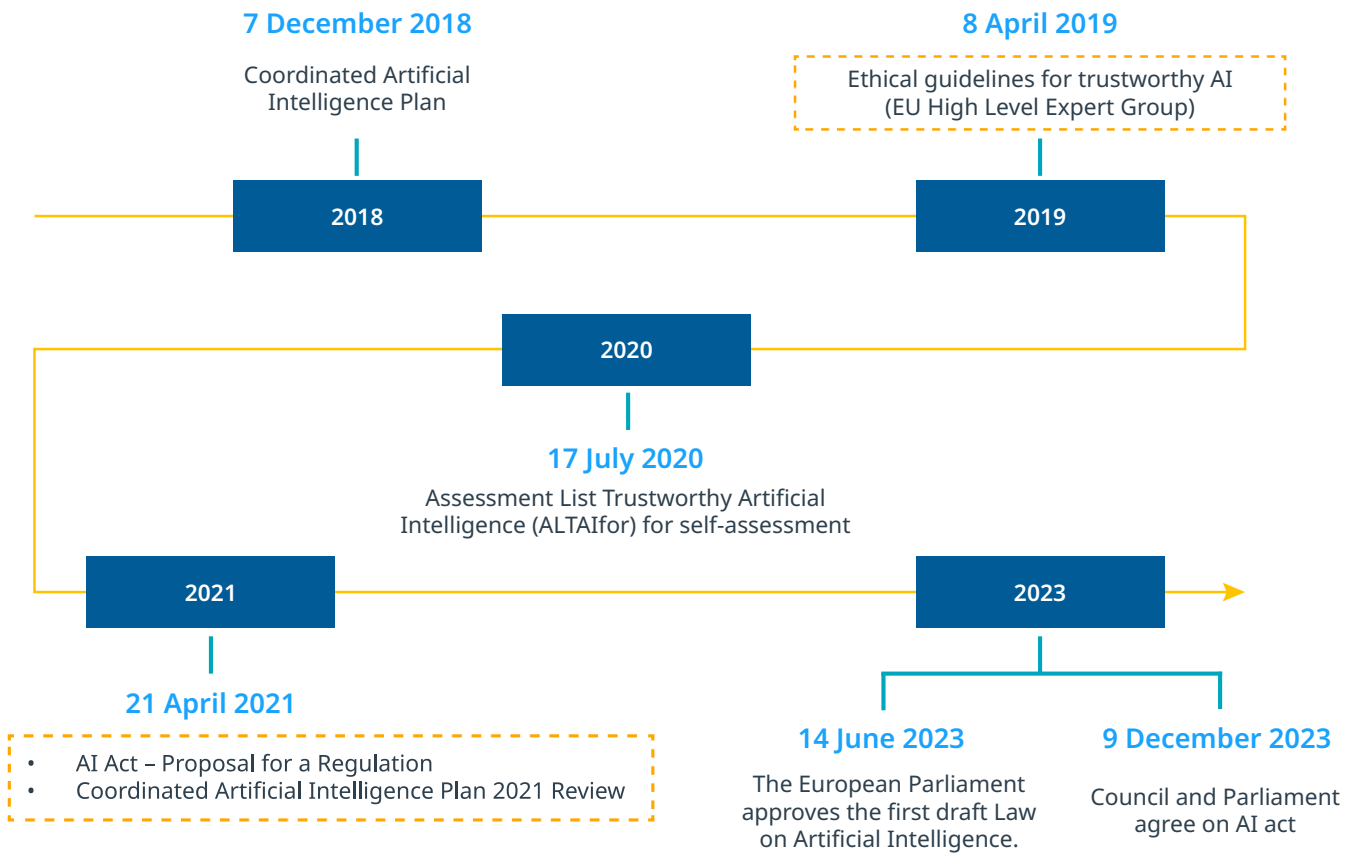


Illustration 3:
Timeline of the most important milestones in AI regulation in the European Union

3. Explainability and equity: introduction

Explainability is the principle that deals with the interpretability of AI models. In the field of AI, it is also known as explainable AI (xAI) and is often associated with transparency. This principle aims to understand and explain the decisions of AI models and ensure that they are safe, impartial, explainable and respect privacy.

We can define **equity** as the field of research that deals with promoting equality and mitigating biases in AI models. The aim of this principle is to deploy measures to address biases, errors and inaccuracies, improve data quality and empower people.

Gartner (2022) ⁶ analyses people-centric AI and defines it as an essential innovation in the AI area and techniques in the coming years. This people-centric AI is built on including business and social value, transparency, reducing bias, fairness, safety, and regulatory compliance, among others. According to this study, **responsible AI will have been widely adopted in five to ten years, and will have a profoundly transformational effect on companies**, so it is important to structurally integrate ethics into AI strategies to encourage their adaptation.

To explore people's knowledge of these topics, we asked the AI LabS attendees whether they were familiar with the concepts of explainability and equity. **48%** responded that they were not; **20%** answered that they knew about them but did not yet apply them. The remaining **32%** said they were familiar with the principles of explainability and equity and applied them in their organization.



⁶New in Artificial Intelligence from the 2022 Gartner Hype Cycle



“ Although many companies have begun to adopt basic measures to understand the responses given by AI models, getting the most out explainability requires a comprehensive strategy at all levels of the company.

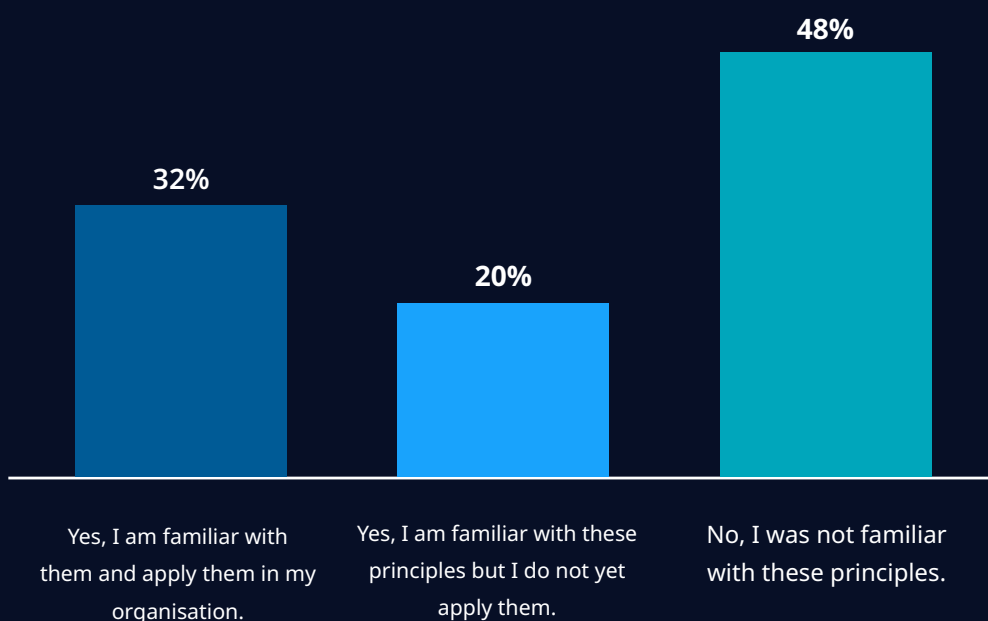


Illustration 4:

Results to the question launched in the AI LabS edition on familiarity with the concepts of equity and explainability

4. Explainability: understanding algorithms' decision-making processes

4.1 The need for explainable models

If we can interpret a model and explain its decisions, we can create a continuous improvement process. However, AI models are so complex it is often impossible for a human being to understand and comprehend their decisions. This is known as the black box effect. In its article "**Why businesses need explainable AI-and how to deliver it**", McKinsey ⁷ reveals that companies increasingly rely on AI systems to make decisions that can significantly affect individual rights, human safety and critical business operations. But how do these models reach their conclusions? What data do they use? Can we trust the results? Addressing these questions is the essence of "explainability" and getting it right is becoming crucial.

In the following diagram (page 12) we can see how including -or excluding- interfaces and tools that favour explainability and mitigate the black box effect can affect an AI project.

We apply an unexplainable AI model for the task we have to perform and use a series of training data that are processed through machine learning and generate a function. This process ends up offering us a decision or recommendation in the form of a conclusion. However, without the appropriate measures, it is highly likely that we will not be able to explain how the algorithm has reached its result, what may be considered a correct answer, what may be considered an incorrect answer, how to correct an error and most importantly, whether the model can be trusted.



⁷ Grennan, L., Kremer, A., Singla, A., & Zipparo, P. (2022, 29 de septiembre) Why businesses need explainable AI—and how to deliver it. McKinsey

On the other hand, in an explainable AI model incorporating an explainability interface we can see how a feedback relationship is established between the task and the results of the algorithmic function, so we can trace the route that the algorithm follows to give us one answer and not another. We can correct errors more precisely and, ultimately, we can trust the model.

Moreover, organizations that establish digital trust among consumers through practices such as making AI explainable are more likely to see their annual revenue and EBIT by 10% or more.

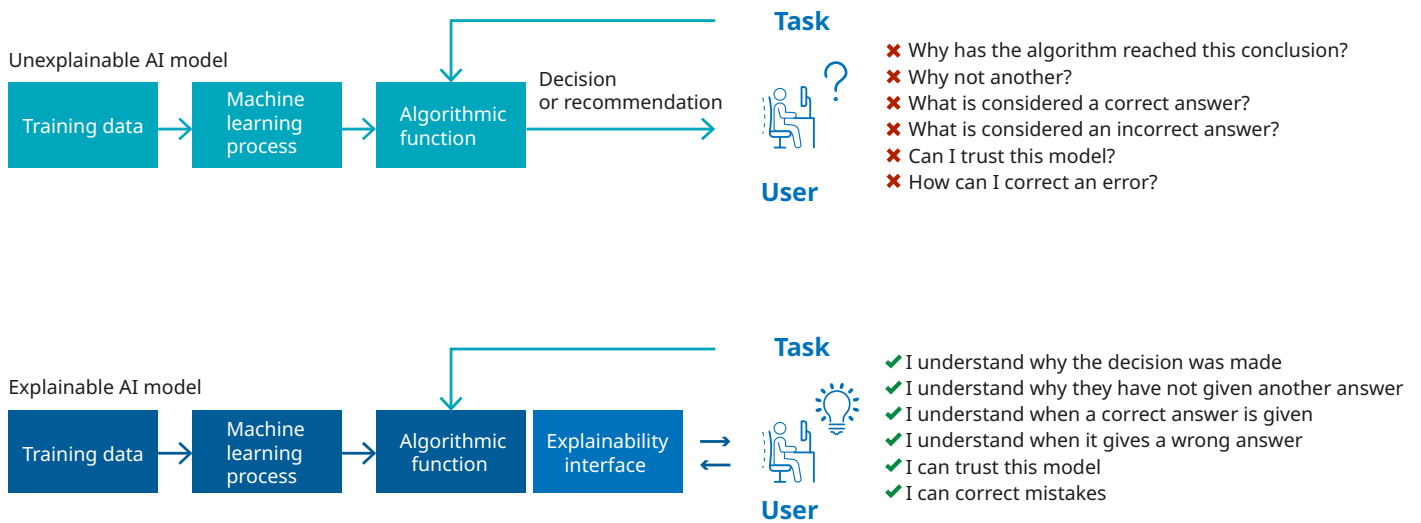


Illustration 5:

Comparison of an unexplainable AI model with an explainable one

“ According to McKinsey, companies that achieve the greatest economic benefits from using AI are those that allocate at least 20% of their EBIT⁸ to its utilization. These companies are more likely to adhere to best practices in making AI more explainable.

⁸ Earnings before interest and taxes: indicator of a company's profitability calculated as income minus expenses excluding taxes and interest.

4.2 Explainability is important at all levels of organization

The ethical AI paradigm affects AI projects transversally, not only in technical processes. All teams involved in the life cycle can benefit from applying measures that favour explainability.

Explainability can help **executive profiles** improve the value proposition, facilitate compliance with legislation (the future AI Act regulation requires AI service providers to establish transparency standards for AI systems intended to interact with natural persons). Explainability can also help us increase clients' trust, attract new clients and, ultimately, improve the company's image.





But it does not just affect executive levels. It is important we incorporate explainability measures in our **technical teams**, since it will favour the production of new algorithmic models, improve knowledge of algorithms, and increase transparency and performance. We will also understand more clearly our projects' effects more clearly, increase safety by creating reports and analytics, and spot any biases more efficiently.

It provides benefits such as increased trust and loyalty in our clients. Explainability can help us win new clients, increase sales and, ultimately, improve service quality.



4.3 Relationship between the complexity of the model and the degree of explainability

In its publication, “**Four Principles of Explainable Artificial Intelligence**” (2020)⁹, the National Institute of Standards and Technology (NIST) presents four fundamental principles for explainable AI systems. These are:

-  **Explainability:** forces artificial intelligence systems to provide evidence, support or reasoning for each result.
-  **Meaningful:** systems must provide understandable explanations for each individual user.
-  **Accuracy of explanation:** the explanation must correctly reflect the system’s process for generating the result.
-  **Knowledge limits:** the system should only operate under the conditions for which it was designed and when it achieves sufficient confidence in its outcome.



However, it is important to note that explainability is not a binary factor, but a spectrum, and that the degree of explainability is closely related to the accuracy of the AI model.

As we have said throughout this report, explainability is the capability to understand and explain how the AI model makes decisions or predictions using its input data. On the other hand, the accuracy of an AI model is its capability to predict accurately. Thus, the complexity of the model is related to the algorithm’s levels of explainability and accuracy.

There is a balance between the factors of accuracy and explainability. We can observe it in the following chart, which contains some of the most common algorithms used in AI systems:

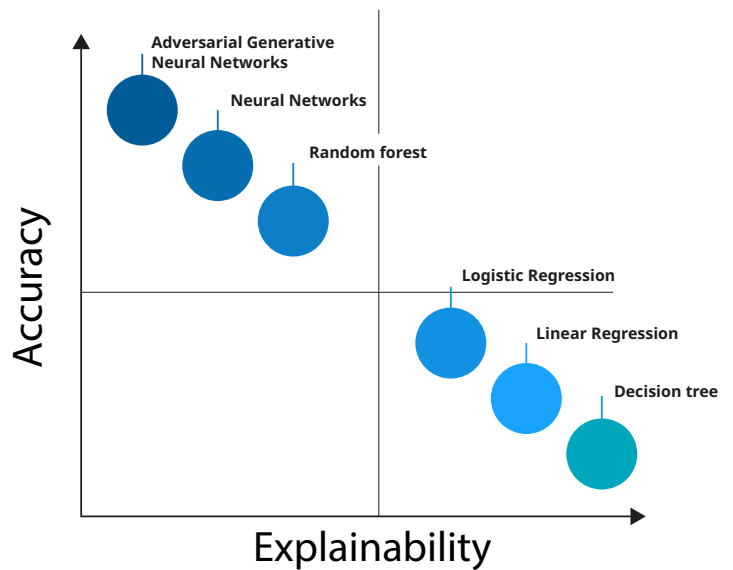


Illustration 6:

Graph that depicts accuracy versus explainability of real AI algorithm models

⁹Phillips, P. J., Hahn, C. A., Fontana, P. C., Broniatowski, D. A., & Przybocki, M. A. (2020). [Four principles of explainable artificial intelligence](#). Gaithersburg, Maryland, 18



In the chart above we can see some of the main types of AI algorithms (adversarial generative neural networks, neural networks, linear regression, among others), divided into two axes: accuracy and explainability. We observe that more explainable algorithms (such as decision trees) are often less accurate than more complex algorithms, for example, neural network systems, which are highly accurate, but less explainable. In short, the more complex the AI system, the more accurate it is, but the less explainable its answers are. However, we must note that we cannot say it is better to use one type of algorithm or another, since we need to develop an algorithmic model that can solve the task in each case.

Let us look at this relationship in a few examples:

Suppose we have a hypothetical algorithm that helps us predict the probability of a customer’s defaulting on a loan. In this case, the bank determines the types of customers and generates an AI model (such as the one below) to calculate their solvency considering their age, whether they study, and their income.

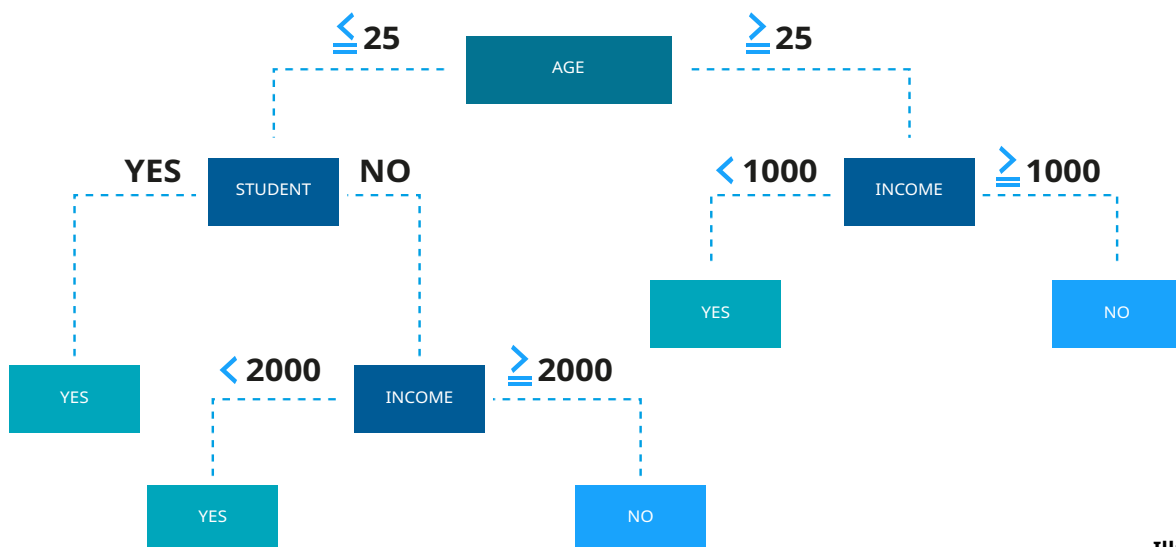


Illustration 7:
Example of how a decision tree works

Let’s see how this relates to a series of examples:

Here, we can easily determine whether Marta (who is 30 years old and has an income of over €3,000) is going to default on her loan. For this specific example, we are talking about a decision tree model in which the decision-making path can be easily followed, and we can see why it has given us this answer.

On the other hand, if we take another example and imagine a hypothetical system for diagnosing pneumonia from x-rays, we will see how explainability changes radically. This classification system allows us to analyse and detect patterns and abnormalities through images that may indicate the presence of lung diseases.

When we analyse the x-rays of our patient Lola, who is 60, we realize that the results are not accurate; they contain errors, and this complicates early detection. In this case, we have a neural network. As we can see in the illustration, the system is very accurate. However, the lack of transparency and the complexity of the network make it exceedingly difficult to identify at what point in the AI system lifecycle the error is occurring.

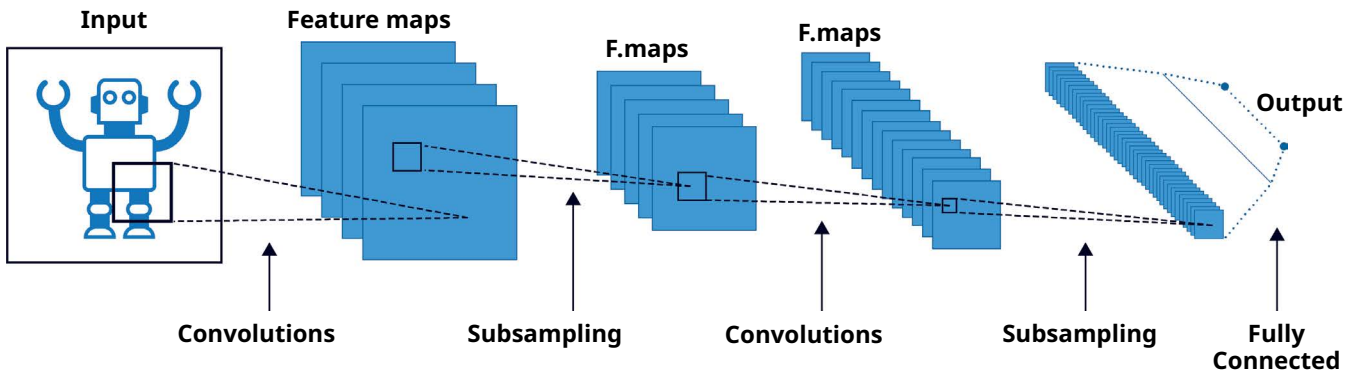


Illustration 8:

Example of a Convolutional Neural Network. We can see its several layers and complexity

These two examples highlight the relationship between the degree of accuracy and explainability. They also help us understand that artificial intelligence models move within the spectrum of interpretability. Within this spectrum we can classify the models into:



Interpretable models: An interpretable model can be understood by a human without the need for special techniques or tools. For example, a decision tree.



Explainable models: Explainable models are too complex for humans to grasp, and additional techniques are required to understand them. Models of this type are also known as black boxes (see reference on page 11). In the above case, the neural network.

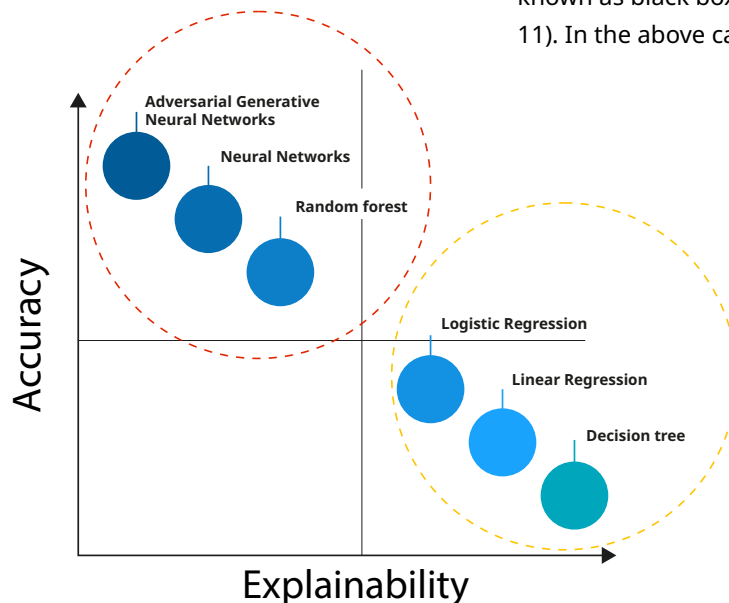
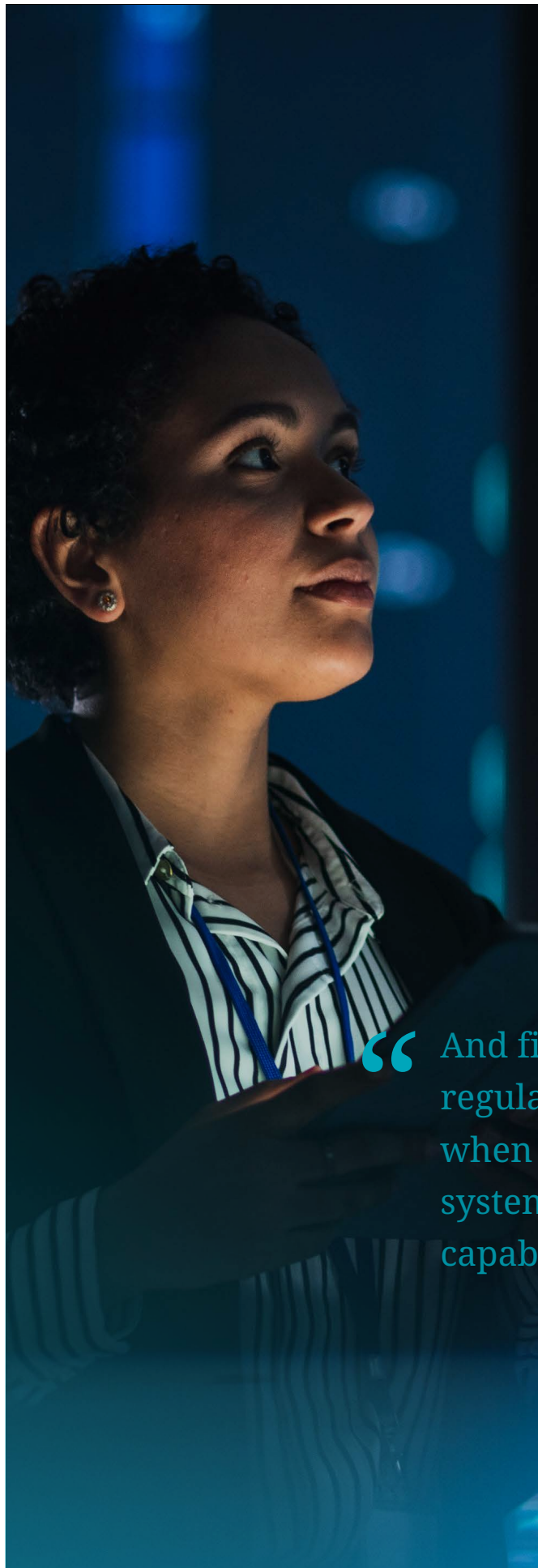


Illustration 9:

Chart containing different actual AI algorithms that contrasts the accuracy and explainability of the model. In this case, the algorithms are grouped into Interpretable Models and Explainable Models



We should also note that we need an audit methodology and training programmes that equip our teams to explain the system's decisions or recommendations.

IBM¹⁰ points out that explainable AI mitigates the legal, compliance, safety, and reputational risks of AI in production. Therefore, it is crucial that an organization fully understands AI decision-making processes, applies AI model oversight and accountability, and does not blindly rely on the processes. IBM also stresses that explanatory AI can make an important contribution by helping people understand and explain machine learning, deep learning and neural network algorithms.

As pointed out in a report by Bain & Company and the World Economic Forum (2021), documentation processes are essential for enabling traceability and auditability and for increasing transparency¹¹. The report stresses that digital traceability enables companies to meet their sustainability objectives and achieve a broader set of business objectives, including efficiency, resilience and responsiveness.

“ And finally, in line with the future AI Act regulations, people have the right to know when they are interacting with an AI system and should be informed about its capabilities and limitations.

¹⁰ [What is explainable AI. IBM](#)

¹¹ [Bain & Company, World Economic Forum \(2021, 16th of September\) The traceability transformation: How transparent value chains can help companies achieve their sustainability goals](#)

5. Equity: ensure fair outcomes for all

5.1 Know the types of biases in AI models

To understand how the principle of equity intervenes in AI projects, we must first know an AI model's lifecycle and understand what bias means.

According to the Merriam-Webster dictionary, one definition of bias is a "systematic error introduced into sampling or testing by selecting or encouraging one outcome or answer over others". Biases can influence the way we perceive and process information and can affect our decisions and judgements. Thus, a biased or unfair algorithm is one whose decisions are influenced by certain unethical attributes, resulting decisions that discriminate against a certain group of people.

AI models can encounter biases throughout their lifecycle since they are not agnostic to the person or team that develops them or to the data and patterns that feed their learning. Below, we can see the life cycle diagram of an AI model:

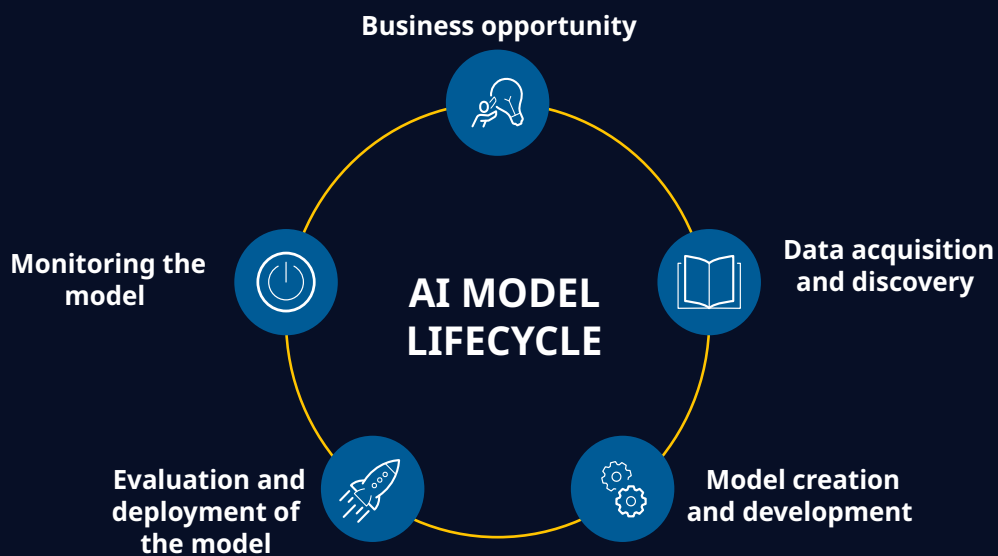


Illustration 10:
Phases of an AI model lifecycle

By having a clear understanding of the process and the potential sources of bias that can arise at each stage of the model lifecycle, we can take steps to ensure that our models are fair and equitable. Generally, biases are classified according to the phase of the cycle in which they are generated. Below are some examples of biases that we may come across:



1. Business opportunity:

This step involves conceptualizing, planning, and compiling business information. In this phase, we determine the objectives and requirements of the machine learning model and plan the development process. The biases that may occur when drawing up the business proposal are:

- **Confirmation bias:** Data that support the business opportunity or the previous hypothesis are selected or searched for, while those that contradict it are ignored or discarded.
- **Misalignment of the initiative:** Risk of misalignment with the purpose for which the model was created. This misalignment may result in unethical results
- **Effects on stakeholders:** No measures are taken to assess the effects of the initiatives on other stakeholders.



2. Data acquisition and discovery:

In this step, the data used to train the model is compiled and processed. This may include cleaning and processing the data to eliminate noise and ensure its quality. The biases inherent in data acquisition may be seen in:

- **Data quality bias:** Using outdated and inaccurate databases may cause the algorithm to ingest incorrect information and therefore make incorrect predictions.
- **Historical bias:** This occurs when the data set used to train an algorithm reflects society's historical inequalities, which may lead to unfair and discriminatory predictions in the future.

- **Sampling bias:** this bias occurs when the sample represented in the data used to train an algorithm is not representative of the population for which a prediction is being made.
- **Multiple labels within the Data Base:** problems arising from the large number of values in the database that affect quality and accuracy, and which may not be monitored by the database administrator.



3. Creation and development of the model:

During this phase, a training model is created, the data are transformed, the predictive model is refined, and its performance is evaluated. The data are aggregated, manipulated and formatted to identify patterns. The biases we may observe are:

- **Attribute bias:** this occurs when attributes are used in the model that are irrelevant to the problem to be solved.
- **Programmer bias:** unconscious bias introduced by the programmer or team working with known parameters without realizing who they are leaving out. It occurs especially in non-diverse teams.
- **Measurement error bias:** this bias occurs when the data used to train the model is subject to measurement errors or uncertainty.





4. Evaluation and deployment of the model:

Once the model has achieved satisfactory performance, it is implemented and goes into production. Some of the biases and risks of this phase are:

- **Feedback bias:** This bias arises when the model receives biased or erroneous information in the retraining data it receives as feedback, which may perpetuate the model's biases.
- **Popularity bias:** The AI model favours certain aspects or outcomes simply because they are popular or occur frequently in training data.
- **Characteristic bias:** The characteristics used to train the model may become irrelevant or insufficient as the data evolves. That is, characteristics that were previously informative may become redundant or lose their predictive capacity.
- **Data drift bias:** This occurs when the underlying concepts that describe the data change over time. For example, in social media sentiment analysis, user opinions and trends may change over time, affecting how data are interpreted and how models perform.



5. Model monitoring

finally, we monitor the performance of the model in production, which enables us to adjust and update it to maintain its performance and accuracy. The biases and risks of this phase are:

- **Interpretability problems:** Difficulties understanding the relationship between inputs and outputs and predicting the model's response to changes in the inputs.
- **Changes in the environment:** The model does not function suitably for the intended objectives because biases or unwanted factors were introduced during the training process, or there were changes in the environment or in the way the model is used.
- **Deterioration of performance:** Drop in the capacity to perform the task (either due to not updating training data, or to changes in data distribution, or other causes).



“ The World Economic Forum (2021)¹² discusses how biases are quite often transferred to AI models in its article “Research shows AI is often biased. Here’s how to make algorithms work for all of us”.

It refers to the need for greater transparency and accountability to contend with this situation and proposes strategies to identify and mitigate the risks of equity and non-discrimination.

But how can the appearance of biases affect AI development? One of the most high-profile real cases of bias in AI is the **COMPAS Case**, an algorithm used in the US to predict the risk of recidivism in defendants. The system was designed to capture the details of a defendant’s profile (age, gender, education level, income level, place of residence, social environment, family criminal background, among others) and use them to estimate the probability of recidivism.

In the results the system produced, we see how African American people received more convictions and harsher sentences than white people from similar backgrounds (risk 10 versus 3); that is, they were twice as likely as white people to be mistakenly classified as “high risk”. Machine learning algorithms use statistics to find patterns in the data; in this case, the system was trained using historical data that was biased against the African American population, which has been disproportionately persecuted by law enforcement over the years, especially in low-income communities. The result is that the algorithm falsely labels African American defendants as future offenders, mislabelling them at nearly twice the rate of white defendants, meaning they are at risk of receiving higher recidivism scores. As a result, the algorithm can amplify and perpetuate existing biases and generate even more biased data that feeds a vicious cycle.



Illustration 11:

Mage of the COMPAS Case showing the degree of recidivism estimated by the system

Source: ProPublica in Psychology Today

In conclusion, we can say that to mitigate biases in our organizations and steer our projects towards ethical and equitable AI we must ensure that the data we use takes into account the multitude of cultural and social expressions, guarantees inclusion and reflects the diversity of individuals and social groups. It is also vital that the results given by the models focus on preserving and protecting the individuals’ cultural characteristics. We must establish measures and procedures to identify and mitigate the occurrence of biases in the model throughout its lifecycle. And as we pointed out in the chapter on explainability, we must train teams on diversity and inclusion.

¹²[World Economic Forum \(2021, 19th of July\) Research shows AI is often biased. Here’s how to make algorithms work for all of us. Artificial Intelligence](#)



Finally, including all types of stakeholders is critical if we are to prevent and mitigate bias. Including different points of view makes for a fuller, holistic evaluation of the possible biases that may arise. Moreover, engaging actors from a diversity of cultures and contexts can help ensure that the model is responsive and equitable for all communities represented in the data.

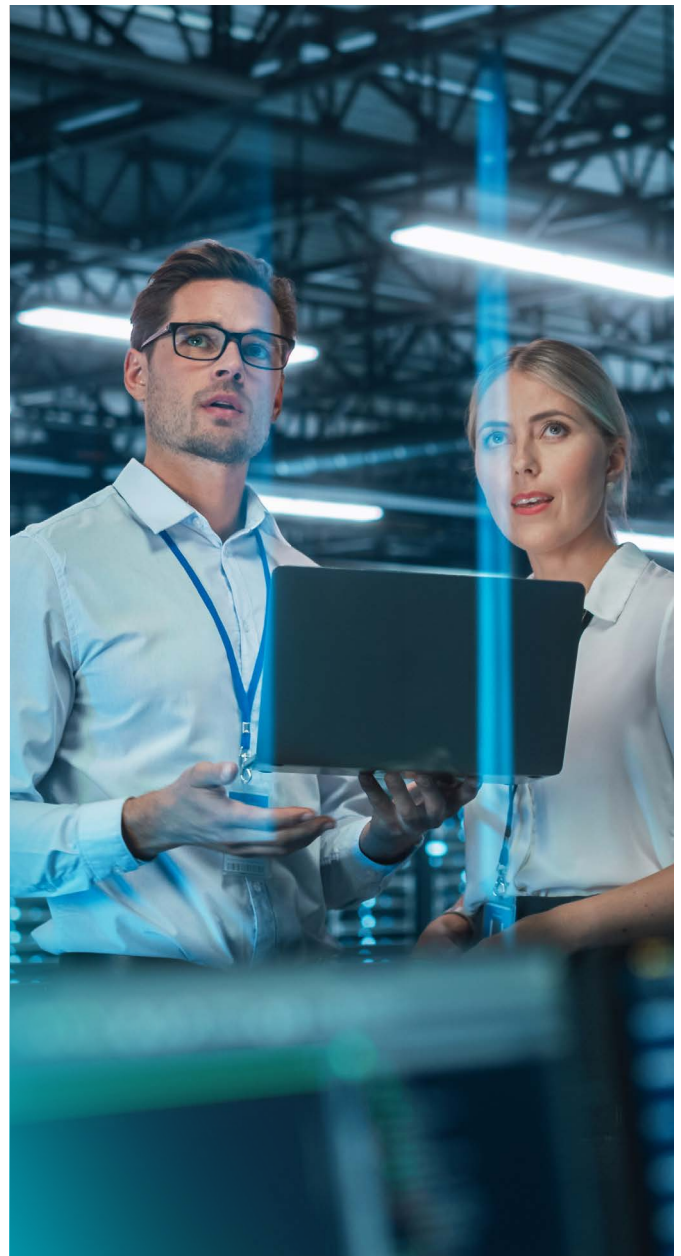
Including diverse stakeholders with knowledge of ethics, laws and regulations, and social and political matters can help ensure that we make equitable decisions and address ethical concerns. Even end-user participation can play a crucial role in how AI systems are adopted and accepted. If users perceive the model to be biased or unfair, they are less likely to trust it and use it appropriately. Involving users in the development process increases our understanding of their needs and concerns, which encourages users to trust and adopt the model.

6. How do we apply explainability and equity in AI projects?

At NTT DATA we ensure that these principles of equity and explainability do not remain on the theoretical plane as a statement of good intentions, but that we assist companies to apply them to their daily operations.

To this end, NTT DATA has created the CDO Journey, a framework that gives clients an overview of the Data & Intelligence (D&I) practice and identifies key areas for improvement, shortening the learning curve. Thus, the paradigms of explainability and fairness, together with the other ethical requirements, guide the actions in each business phase: Business Value, Responsible Governance, Core Tech & Next Gen Operations, Ecosystem & Innovation and Culture & Change Management.

In the Business Value area, we lay the foundations for the responsible use of AI in organizations, by promoting a culture of responsible D&I that spreads ethical principles throughout the organization. To this end, we provide a deep understanding of Data Ethics and AI to guide organizations by drawing up a **Data Ethics and AI Guide**. Likewise, we provide audit tools such as the **AI Audit Model** that helps clients understand how well their AI models comply with the new European regulatory regulations of the AI Act.



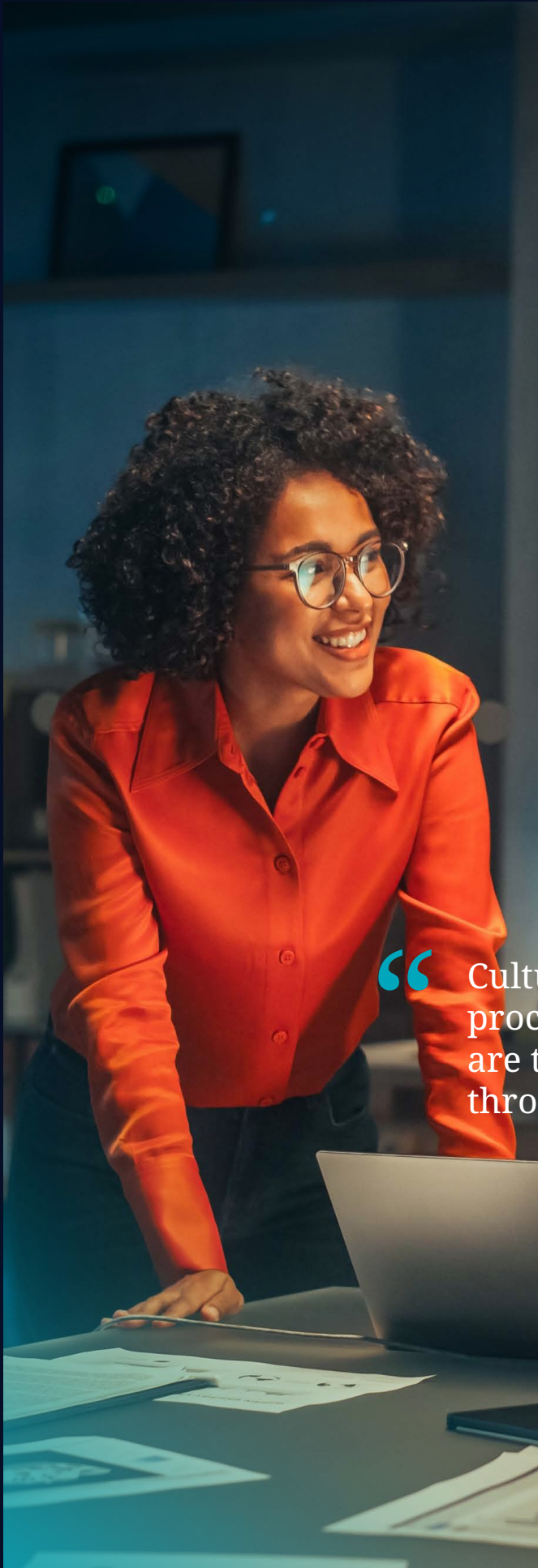
The way the lifecycle of the models is covered is also particularly important. Our Responsible Governance proposal helps companies understand the processes involved in each phase of the AI algorithm and establishes standards that guide the responsible and ethical governance of D&I solutions. It provides a **standardized model for supervising and controlling the biases and risks** involved in putting AI solutions into production and ensures their scalability.

NTT DATA assists organizations design a training plan to help augment analytical and responsible knowledge of AI by creating personalized paths for each role and profile. As a result, in the Culture & Change Management section we provide training programmes in AI ethics in workshops aimed at the entire organization. These are awareness sessions on ethical AI and specific workshops for technical staff that address explainability in machine learning models or data equity and algorithms.



Illustration 12:
Outline of the CDO Journey developed by NTT DATA

“ As a result, employees with advanced D&I-trained skills will contribute to fostering and spreading a culture of D&I in all areas, developing the organization’s D&I maturity, awareness, and analytical skills.



Conclusion

During the AI LabS session, we delved into ethical AI to understand the principles of explainability and equity and how they affect AI projects.

Regarding explainability, we stressed that understanding how an algorithm works encourages us to trust it more, increases service quality, and improves the value proposition and the response to regulatory requirements. The session's practical examples described in this report show that the degree of explainability is closely related to the complexity of the algorithm.

In the equity section, we demonstrate the importance of using algorithms that take into account the many cultural and social expressions required to guarantee inclusion and reflect the diversity of individuals and social groups. We also look at how biases can occur at all stages of AI model life cycles.

“ Cultural change and awareness processes are fundamental if we are to prioritize the ethical use of AI throughout the organization.

This is why the principles of equity and explainability must be present from the very first stages of designing an AI project and why they must be part of our companies' culture. Benchmarks such as the CDO Journey enable companies to initiate a transformation path or identify innovation processes in which applying ethical AI benefits all business areas and helps prepare companies for future regulatory requirements.

Companies participating in the Responsible and Inclusive AI LabS since its launch:

BBVA

 **CaixaBank**


CUATRECASAS

El Corte Inglés

ferrovial

FUJITSU

FFP FUNDACIÓN FERNANDO POMBO

gsk

 **IESE**
Business School
University of Navarra

 **ILUNION**

MELIÀ
HOTELS & RESORTS

 **NTT DATA**

Pérez-Llorca

 **randstad**

randstad
fundación.

^BSabadell

 **Santander**

TENDAM
GLOBAL FASHION RETAIL

URÍA
MENÉNDEZ

 **vodafone**


Willis Towers Watson

About SERES FOUNDATION & NTT DATA

Founded over 15 years ago, SERES Foundation is a non-profit organization that assists companies in their transformation and drives their leadership in the face of social challenges. Its objective is to encourage organizations to position social matters as strategically indispensable. As a pioneering movement with around 150 member companies representing 30% of GDP and 75% of the Ibex 35, it addresses the companies' social commitment through a strategic, practical and innovation-focused approach.

From the Foundation, we have worked with companies to address major corporate challenges in social matters, combining purpose and strategy. In the field of Artificial Intelligence, we encourage organizations to manage technology responsibly and contribute to a model of inclusive progress that leaves no one behind.

NTT DATA, a Japanese company and one of the TOP 10 largest IT services companies in the world, employs over 140,000 professionals and operates in more than 50 countries. In NTT DATA we accompany our clients in their digital development through a wide range of strategic consulting and advisory services, innovative technologies, applications, infrastructure, IT-service modernisation and BPO. We contribute our experience in all sectors of economic activity and our vast knowledge of the geographies where we are present.

We strive to build a single yet open community of people, led by shared values, which continues to grow into an even larger network of collective talent capable of multiplying our capabilities and knowledge to respond swiftly to our clients' changing needs and intelligently anticipate the future. In the Data & Intelligence field, we accelerate our clients' business transformation through innovation and a full-service portfolio.

Contact



David Pereira Paz

Head of Data & Intelligence Europe,
NTT DATA



Cristina Aliaga Ibañez

Company Director, SERES Foundation



Beatriz Zamora
Project Manager,
Fundación SERES

Authors



Alicia de Manuel Lozano

Expert Analyst in AI Ethics



Alberto Martinez Caballero

Tech Advisory Consultant

